SYMMETRICAL COMPLEMENTATION DESIGNS

by

William H.armann Beyer, B.S., M.S.

Thesis submitted to the Graduate Faculty of the
Virginia Polytechnic Institute
in candidacy for the degree of
DOCTOR OF PHILOSOPHY
in
Statistics

July 1961
Blacksburg, Virginia

TABLE OF CONTENTS

I. INTRODUCTION

1.1 General Description

The "Symmetrical Complementation Design" which is discussed in this dissertation is intended for those experimental situations where the levels of three factors always sum to the same constant. Thus, when we have three factors, the level of the third will be strictly determined in terms of the levels of the two first factors. For example, if we are dealing with a nutrition experiment in which a constant caloric intake is required, changing the composition of the diet will imply, for example, an increase in carbohydrates and fat and, at the same time, a reduction in the protein intake. The diagram below illustrates how such a design could be set up. It has been assumed that we have five different levels of each of three components.

	1	2	3	4	5	= Level of B
1		5	4	3	2	
2	5	4	3	2	1	
3	4	3	2	1		
4	3	2	1			
5	2	1				
Level of A						

= Level of C

Thus, if level 1 of Factor A and level 2 of Factor B are applied to one or more experimental units, the diagram shows that Factor C has to be applied at level 5. As Factor A is increased to level 3 and Factor B to level 4, Factor C should be applied at the lowest level, again so as to make the total sum of the levels a constant. Finally Factor A at the third level and Factor B at the fifth level is an inadmissible combination under this design, as complete interchangeability of the three factors is desired. By complete interchangeability, we mean that any one of the factors can be chosen as the diagonal factor, and complete symmetry in the design would be assured. We have chosen Factor C as the diagonal factor, and Factors A and B as the row and column factors, respectively. This notation will be used throughout this dissertation. We merely want to point out that Factors A and C could be used as the row and column factors, as could be Factors B and C, without changing the scheme of the design. Thus, for the case of five levels, eighteen different treatment combinations are available corresponding to the cells in the diagram. It should be noted that, if the levels of the factors are referred to a common unit of measurement, e.g., weight per cent, calory equivalents, the levels of these factors must be equally spaced. Each of these combinations may be applied to one or more experimental units, and the subsequent chapters deal with cases of n (n=1, 2, ...) experimental units per cell. The reason for leaving the corner cells empty will be stated in more detail later. At this time, we merely wish to point out that a design obtained from the present one by extending it to six levels and filling the corner cells would confound the effects in the corner cells with the general effect.

Examples calling for such types of design are quite numerous. In education research, we may be interested in the best type of curriculum to be offered to a group of high school students. Whereas the total time devoted per day has to be a constant, the amount attributed to each subject may be of some importance in the general development of educational abilities. As another example, we may consider the composition of chemical mixtures, for instance, solid propellants. In addition to the relatively large amount of nitroglycerin, we need several plasticizers. These plasticizers may be equivalent in their effects on the quality of the resulting propellant. Thus, if we have nitroglycerin plus two plasticizers, we may consider both increasing the relative amount of the plasticizers and reducing the amount of nitroglycerin. As an example of such a design, consider the case where the factors are given in percent of the total amount. Thus the diagram may take the following form:

	0	2	4	6	8
17		81	79	77	75
19	81	79	77	75	73
21	79	77	75	73	
23	77	75	73		
25	75	73			

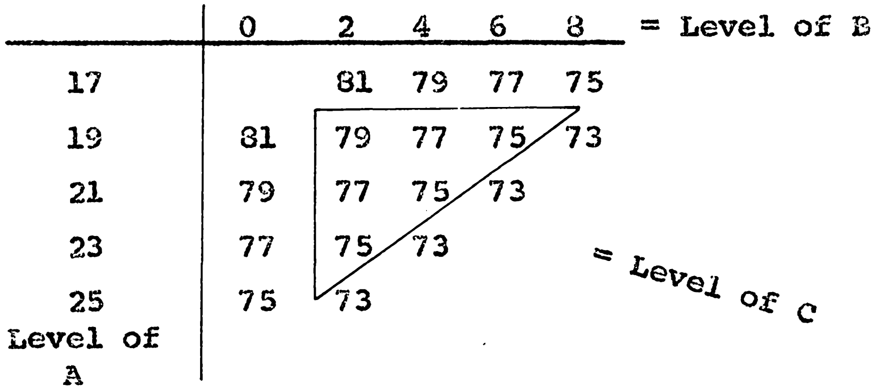
= % Plasticizer B

% Plasticizer A

= % Nitroglycerin

Here, we might study the effects on specific impulse due to changes in composition of the solid propellant.

The most conspicious experimental situation which calls for a design for this type would be met in economics. If one has a fixed budget available, and wants to apportion ones inventory in, say, three different types of merchandise, such a design might be helpful to investigate the type of combination which will produce maximum sales. As a matter of fact, this dissertation will discuss the great importance of the "hypothesis of substitution". If this kind of hypothesis is accepted, it will say that applying one factor at a low level and another one at a high level does not produce results which are different when the levels of the two factors are exchanged. Thus, if this situation holds, we need not worry about the composition of two of the three components. Further, we might test for this exchangeability of the levels of two factors for a fixed level of the third. To illustrate, let us consider the previous example on the solid propellant. Acceptance of the hypothesis of substitution for the two plasticizers, say, where the level of nitroglycerin is the 75% level, would imply that the response surface has reached a plateau along the indicated diagonal. Suppose, however, that the hypothesis had been rejected, and further that hypotheses for the 77% and 79% levels are also rejected. We might then consider testing the hypothesis of substitution where, say, Plasticizer B is applied at a higher level than the first. Thus, we may test hypotheses of substitution for those combinations as outlined by the triangle in the diagram below.



	0	2	4	6	8	= Level of B
17		81	79	77	75	
19	81	79	77	75	73	
21	79	77	75	73		
23	77	75	73			
25	75	73				
Level of A						

= Level of C

Acceptance of these hypotheses of substitution would further indicate the attainment of a plateau in the response function, where it would be immaterial what combination of levels is administered.

This design differs from the usual types of designs, in that the number of contrasts that can be estimated is limited. For example, simple comparisons of two levels which would combine into the hypothesis of equality of effects of all levels of one factor, are not testable in this particular design. The restriction is also intuitively obvious, since increase of the level of one factor necessitates a decrease in the combination of the other two. On the other hand, the contrasts which compare one level with the mean or weighted mean of others are estimable for each factor separately. "Quadratic" contrasts* are a special case of these. Contrasts of higher order are also estimable; they are, however, somewhat difficult to interpret. We will combine them in the different hypotheses of substitution. A detailed study about the types of functions which are estimable will be presented in the chapter on linear estimation. Estimable functions in one factor only and in two factors are presented for the general case of p levels.

Estimation of parameters is rather simple. There are several methods which can be employed in order to obtain estimates of the treatment effects under various constraints. It must be noted, however, that these estimates of treatment effects are always rather meaningless quantities. It is only when these are combined in estimable functions that unique results will be obtained. Thus, if the so-called "high-low"

---

* What is called the "linear contrast" is the first degree Tchebycheff polynomial , given, e.g., by [-3, -1, 1, 3] in the case of four levels. "Quadratic" and higher order contrasts are also names given to the higher Tchebycheff polynomials. Actual quadratic and higher order contrasts with reference to this design will be described in Section 4.5.

method (see Section 3.2) is applied for the estimation of treatment effects, the treatment effects will look completely different from the ones obtained by employing a different set of constraints in an inversion procedure. However, the estimable functions will always come out the same. If only estimation is required, the simpler high-low method should be used. If however, both estimation and tests of hypotheses are required, the "modified high-low" method (see Section 3.3) should be used throughout, since the formulas producing estimates will be a by-product of the matrix inversion required. These methods are described in complete detail in the chapter on estimation of parameters, with the complete inverse matrix, or a method of obtaining this patterned inverse, being presented.

The testing of hypotheses is a little more difficult than the estimation of parameters. In this case, the inversion method seems to be the most effective one. There is a general technique, based upon the inversion of the matrix in the modified high-low method, which can be used for the testing of any testable hypothesis. Sums of squares and test statistics are presented for the various hypotheses formulated. The high-low method itself is not suitable for the computation of sums of squares and quadratic forms necessary for the test statistics which test the various hypotheses. Sections are also included, in this chapter on testing of hypotheses, which indicate how one might approximate the response due to any combination of levels and obtain response functions for single factors.

After we consider the various degrees of freedom available for each factor and factor combinations, there will be a certain residual which we have referred to as "lack-of-fit". If we assume interactions to be present among the factors, the present design corresponds to an incomplete fractional factorial of order 1/p. In other words, if all the cells were filled we would have a complete fractional factorial design. Since, however, the limitation of a total level sum makes it impossible to take certain combinations, the present design is not even a complete fractional factorial. Even if it were, main effects in a three-factor factorial would always be aliased with two-way interactions. It seems thus that this type of design should not be used in those situations where a significant interaction may be present.

The mean square due to lack-of-fit may serve to assess the magnitude of the error, if only one experimental unit is available per cell, in the same way that the mean square due to interaction is used in a two-way classification with one replication. In the preferable situation where we have n (n=2, 3, ...) replications per cell, i.e., n experimental units per cell, we would use, as is usually done, the mean square within cells as an estimate of the experimental error. In this connection we would have to distinguish between three different cases:

(a) the replications are strictly repetitions of the experiment under otherwise identical conditions. In this case, the analysis proceeds in the customary two-way or three-way analysis with n replicates per treatment combination.

(b) the experiments within a cell represent repetitions over a period of time, during which some kind of a trend may be present. In this case, the analysis is readily extended into an analysis of covariance with one concomitant variable if only linear trend is postulated and more than one if trend of a higher degree is expected. This case would prevail particularly in nutrition studies with growing children. If experiments are repeated after a certain period of time, a growth effect has taken place and must be eliminated before analysis of the effect of the different compositions can be performed.

(c) a case may arise, which is somewhat similar to case (b), where the replications constitute several experiments or responses with the same experimental units, so that the observations within a cell are dependent. On the assumption that the covariance matrix of observations in a cell is the same from one cell to another, we may extend this method readily to a multivariate situation. This will not interfere with the estimation problem, but presents some additional problems to be discussed for purposes of analysis. These methods are described in detail in the chapter on extensions.

If the levels of the three factors are referred to a common unit of measurement, but are not equally spaced, symmetry will be lost in that the third factor will require more levels. Although this design does not actually belong to the class of designs discussed in this dissertation, we feel it is important enough to warrant special discussion, and this discussion is presented in the chapter on extensions. Several demonstration studies of these extensions are given in the chapter on applications.

In conclusion it should be stated that this type of design requires a rather careful consideration of the type of functions that can be estimated and the type of hypotheses that can be tested. The user of such a design should have a clear concept of what is estimable and testable. Thus, for example, if the user is merely interested in studying the difference between the levels in the same factor, he must be cautioned that such a function cannot be estimated in this design. He cannot simply test the usual hypotheses of equality of effects for each factor. He may, however, compare a particular level with several others in the same factor and obtain a perfectly estimable function and consistent conclusion. For the case of three, four, and five levels the formulas have been completely worked out. All that is required of the user of such a design is to substitute numbers in the expressions developed. Recommendations for interpretation and statement of limitations are also made in detail.

1.2 Review of Literature

The Symmetrical Complementation Design can be regarded as a Latin Square design with missing observations. It is so incomplete, however, that application of usual missing value techniques would lead to questionable results, for even the estimability conditions are more restrictive in this design than the usual Latin Square design.

Allan and Wishart (1930) were among the first writers to develop a method of analysis when a single observation was missing. They obtained a formula for estimating the missing observation by minimizing the error sum of squares. The augmented data was then analyzed in the usual way, with the number of degrees of freedom for the error and for the total sum of squares each decreased by unity. Yates (1933) showed that this method of analysis resulted in an approximate F-test for the treatment mean square, and he furnished a method to estimate the bias.

Much of the writing on missing observations in Latin Square designs has been done by Yates (1936) and Yates and Hale (1939). These men have written articles which present the method of analysis when a single observation is missing; when one or more rows, columns, or treatments are missing; and when one column and one or more single observations are missing. They furnish a method of solving the normal equations by successive approximation. DeLury (1946) brought a number of these methods together.

Yates (1933) has shown how the method for dealing with one missing observation may be extended to the case where several observations are missing. However, an iterative procedure is necessary when the method is used to estimate more than one missing value. Kramer and Glass (1961) have also dealt with the estimation of several missing values in the Latin Square design, by minimizing the error sum of squares. They present a direct method of analysis of variance, not requiring a correction for bias in the treatment sum of squares. Wilkinson (1958a), following Yates' procedure of minimizing the error sum of squares, has shown that correct estimates of the treatment effects and other parameters for incomplete data can be given by standard formulae for various designs. The estimates depend upon the solution of matrix equations. When the data are completed with these estimates, a standard analysis of variance yields the correct error sum of squares, with degrees of freedom reduced. However, other components of variance in the analysis are only approximations. Wilkinson (1958b) has dealt with correcting the standard analysis of variance, deriving general formulae for the necessary corrections.

These various writers have produced elegant solutions for these special cases of non-orthogonality. Stevens (1948) has shown that it is still possible to construct exact analyses of variance and to make valid tests of significance when orthogonality restrictions have completely broken down. He has furnished methods for analysis by means of arithmetical computations which can be reduced to a routine which follows a simple and recognizable pattern.

Ditchburne (1955) gives the method of analysis for a double classification arranged in a triangular table, on the assumptions that interactions do not exist. Here, the sum of squares for each factor, freed from the effects of the other factor, are obtained from a set of orthogonal comparisons. This method of analysis is only applicable under certain restrictions: either the number of replicates in each subclass is equal, or the number of replicates in each subclass within a single level of one of the factors is equal.

Since, in the present design, missing observation techniques are not easily applied, the analysis of the general linear model is used, see, e.g., Kempthorne (1952) or Graybill

(1961). The concept of non-estimable parametric functions was introduced by Bose (1944). Rao (1945a) further generalized these results and derived tests of significance connected with linear hypotheses. Rao (1945b) also extended the problem of linear estimation to cases where the parameters are subject to linear restrictions.

The multivariate extensions are based upon results given by Roy (1957). Here the test statistic for testing the general linear hypothesis is the largest root of a certain determinantal equation. The hypothesis postulated that each of a specified set of linear combinations of the regression parameters be equal to zero. It is possible, however, to make this hypothesis more general, by hypothesizing that the linear combinations of regression parameters are equal to constants not necessarily zero. This has been done and presented in the 1959 multivariate analysis lectures at Virginia Polytechnic Institute by Rolf E. Bargmann.

CHAPTER II

LINEAR ESTIMATION

In most experimental design situations, the problem of estimability needs no specific discussion. From intuitive considerations it is usually quite obvious which functions are estimable and which are not. Thus, for example, in a simple one-way classification design it is clear that the treatment effect in each category is non-estimable, for its magnitude depends upon the size of the arbitrarily fixed general effect. However, all differences between treatment effects and, for that matter, all contrasts in the treatment effects, are estimable. A similar result holds for a two-way cross-classification without interaction. In the present design, however, no such intuitive argument will be obvious, and we are thus compelled to present a detailed mathematical description in order to decide which function can be estimated.

2.1 General Theory

The concept of non-estimable parametric functions was introduced by R. C. Bose (1944). C. R. Rao (1945a) further generalized these results and derived tests of significance connected with linear hypotheses. Rao (1945b) also extended the problem of linear estimation to cases where the parameters are subject to linear constraints.

To indicate what functions of the parameters can be estimated, we must consider the general-linear-hypothesis model of less than full rank. The linear matrix model is

$$\underline{y} = A\underline{\xi} + \underline{e} , \qquad (2.1.1)$$

where

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1M} \\ a_{21} & a_{22} & \cdots & a_{2M} \\ & \cdots & & \\ & & & \\ a_{N1} & a_{N2} & \cdots & a_{NM} \end{bmatrix}, \quad \underline{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \cdot \\ \cdot \\ \cdot \\ \xi_M \end{bmatrix}, \quad \underline{e} = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_N \end{bmatrix}.$$

The ξ's are fixed unknown parameters, and the e's are independently distributed with mean zero and variance σ^2. We assume that the design matrix A, of order $(N \times M)$, is less than full rank, i.e., the rank of $A = r < \min (N, M)$.

If we wish to find an unbiased linear estimate of a function $\underline{\lambda}'\underline{\xi}$ of the parameters, where $\underline{\lambda}$ is a known $(M \times 1)$ vector of constants, we must have, that for a vector $\underline{b}$

$$E(\underline{b}'\underline{y}) \equiv \underline{\lambda}'\underline{\xi} \quad , \qquad (2.1.2)$$

where $\underline{b}'\underline{y}$ is the desired estimate.

Since the matrix A is less than full rank, let us partition A into a basis, say A_1, of order $(N \times r)$, and an extension, say A_2, of order $(N \times M-r)$, such that

$$A = [A_1, A_2] = [A_1, A_1Q] \quad , \qquad (2.1.3)$$

where Q will be called the transformation matrix. Equation (2.1.3) indicates that the columns in the extension A_2 are linear combinations of the columns in the basis A_1 (in line with the fact that A have rank r).

Then

$$\begin{aligned} E(y) &= A\underline{\xi} \\ &= A_1\underline{\xi}_1 + A_2\underline{\xi}_2 \\ &= A_1\underline{\xi}_1 + A_1Q\underline{\xi}_2 \\ &= A_1[\underline{\xi}_1 + Q\underline{\xi}_2] \quad , \end{aligned} \tag{2.1.4}$$

where $\underline{\xi}' = [\underline{\xi}_1', \underline{\xi}_2']$ is a partitioning of the row vector of parameters analogous to that of the design matrix so that $\underline{\xi}_1$ has length r and $\underline{\xi}_2$ has length M-r. Equation (2.1.4) indicates that the elements of the vector $\underline{\xi}$ cannot be estimated separately, but merely in the combination $\underline{\xi}_1 + Q\underline{\xi}_2$. Thus, for a linear function to be estimable, it must be represented by

$$\begin{aligned} \underline{\lambda}'\underline{\xi} &= \underline{\lambda}_1'\underline{\xi}_1 + \underline{\lambda}_2'\underline{\xi}_2 \\ &= \underline{\lambda}_1'[\underline{\xi}_1 + Q\underline{\xi}_2] \quad . \end{aligned}$$

Hence

$$\underline{\lambda}_2' = \underline{\lambda}_1'Q \quad . \tag{2.1.5}$$

Equation (2.1.5) is the fundamental relation for estimability; it states that the extension of a vector of coefficients must be related to its basis in the same way that the extension of the design matrix is related to its basis. If we attempt to estimate a linear function which does not satisfy the relation (2.1.5) we may produce any arbitrary value for the estimate from the same observations. Estimability of parametric functions in the Symmetrical Complementation design will be investigated by the application of this relation. We will first set up, in detail, the normal equations.

2.2 Derivation of Estimates and Estimable Functions

2.2.1 Model and Notation

The yield of the ij(k)th observation will be expressed by the model*

$$y_{ij(k)} = \mu + \alpha_i + \beta_j + \gamma_{(k)} + e'_{ij(k)} , \qquad (2.2.1)$$

where

$i = 1, 2, \ldots, p$ = number of rows,
$j = 1, 2, \ldots, p$ = number of columns,
$(k) = 1, 2, \ldots, p$ = number of diagonals.

It is to be understood that all cells where $i+j > p+2$ will be empty, as will the corner cells where $i+j$, $i+(k)$, and $j+(k) = 2$.

In the model, μ, α_i, β_j, and $\gamma_{(k)}$ are the general effect, row effects, column effects, and diagonal effects, respectively. The $e'_{ij(k)}$, representing random effect, are assumed to be independently distributed with mean zero and variance σ^2. In hypotheses testing, they will be assumed normally distributed. In the case of one observation per cell, this random effect will be referred to as the contribution due to "lack-of-fit", in the same way that the contribution due to interaction is considered in the two-way classification with one observation per cell. In subsequent sections, where we consider n (n=2,3,...) observations per cell, we will consider the error within cells as an estimate of the random effect;

---

*The subscript in parentheses is not an independent subscript. Once values for i and j are chosen, the value for (k) is determined so that $i+j+(k) = p+3$.

and we will there consider the contribution due to lack-of-fit as both a fixed and a random variable. We shall, however, in the remainder of this chapter, consider the case of one observation per cell, as the derivation of estimable functions is somewhat simplified. The design with only one observation per cell, however, suffers to some extent from the same disadvantage as the Latin Square design, in that only a few degrees of freedom are provided for estimation of error when the number of levels of the factors is small. For this reason, it is advantageous to use a design with several observations per cell, and to thus use the error within cells as the estimate of the random effect.

We shall denote row totals, column totals, diagonal totals, and the grand total, for the case of one observation per cell, by $y_{i..}$, $y_{.j.}$, $y_{..(k)}$, and $y_{...}$, respectively.

2.2.2 Normal Equations for the Model

The least squares estimates of the effects for the above linear model are obtained by minimizing the residual sum of squares (2.2.2)

$$\sum_{i,j,(k)} e'^{2}_{ij(k)} = \sum_{i,j,(k)} (y_{ij(k)} - \mu - \alpha_i - \beta_j - \gamma_{(k)})^2.$$

Equating to zero each of the partial derivatives of the above residual sum of squares with respect to the parameters μ, α_i, β_j, and $\gamma_{(k)}$, results in the following normal equations, with $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\beta}_j$, and $\hat{\gamma}_{(k)}$ being used for the estimators of the parameters in equation (2.2.1):

Normal equations in matrix notation:

$$\begin{bmatrix} \frac{(p-1)(p+4)}{2} & \underline{z}' & \underline{z}' & \underline{z}' \\ \underline{z} & D & E & E \\ \underline{z} & E & D & E \\ \underline{z} & E & E & D \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \underline{\hat{\alpha}}_i \\ \underline{\hat{\beta}}_j \\ \underline{\hat{\gamma}}_{(k)} \end{bmatrix} = \begin{bmatrix} Y_{\cdots} \\ Y_{i\cdot\cdot} \\ Y_{\cdot j\cdot} \\ Y_{\cdot\cdot(k)} \end{bmatrix}, \quad (2.2.3)$$

where

$i = j = (k) = 1, 2, 3, \ldots, p$;

$\underline{z}'$ is a $(1 \times p)$ vector of the form

$$\underline{z}' = [\ (p-1) \quad p \quad (p-1) \quad (p-2) \quad \ldots \quad 4 \quad 3 \quad 2\ ] \quad ;$$

D is a $(p \times p)$ diagonal matrix of the form

$$D = \begin{bmatrix} (p-1) & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & p & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & (p-1) & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & (p-2) & \ldots & 0 & 0 & 0 \\ & & \cdot & \cdot & \cdot & & & \\ 0 & 0 & 0 & 0 & \ldots & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 2 \end{bmatrix} ;$$

and E is a (p x p) matrix of the form

$$E = \begin{bmatrix} 0 & 1 & 1 & 1 & \cdots & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & \cdots & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & \cdots & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & \cdots & 1 & 0 & 0 \\ \\ 1 & 1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix} .$$

$\hat{\underline{\alpha}}_i$ denotes a (p x 1) vector of unknown row effects, $\hat{\underline{\beta}}_j$ a (p x 1) vector of unknown column effects, and $\hat{\underline{\gamma}}_{(k)}$ a (p x 1) vector of unknown diagonal effects.

2.2.3 Estimable Functions for the Symmetrical Complementation Design

The design matrix A, of rank (3p-3), for the design under consideration, can be written in such a way that the columns corresponding to the first (p-1) levels of each factor are stated first, i.e., fall into the basis A_1. Then the columns, corresponding to the pth level of each factor and the mean, respectively, will be stated next, i.e., fall into the extension A_2. Then A is of the following form:

$$
A = \begin{array}{l|cccc|cccc|}
\text{effect} \diagdown \text{cell } [i,j,(k)] & (\alpha_1 \ldots \alpha_{p-1}) & (\beta_1 \ldots \beta_{p-1}) & (\gamma_{(1)} \ldots \gamma_{(p-1)}) & (\alpha_p & \beta_p & \gamma_{(p)} & \mu) \\
\hline
[1,2,(p)] & 1\,0\,0\,0\,\ldots\,0\,0 & 0\,1\,0\,0\,\ldots\,0\,0 & 0\,0\,0\,0\,\ldots\,0\,0 & 0 & 0 & 1 & 1 \\
[1,3,(p-1)] & 1\,0\,0\,0\,\ldots\,0\,0 & 0\,0\,1\,0\,\ldots\,0\,0 & 0\,0\,0\,0\,\ldots\,0\,1 & 0 & 0 & 0 & 1 \\
[1,4,(p-2)] & 1\,0\,0\,0\,\ldots\,0\,0 & 0\,0\,0\,1\,\ldots\,0\,0 & 0\,0\,0\,0\,\ldots\,1\,0 & 0 & 0 & 0 & 1 \\
 & \cdots & \cdots & \cdots & & & \cdots & \\
[1,p-2,(4)] & 1\,0\,0\,0\,\ldots\,0\,0 & 0\,0\,0\,0\,\ldots\,1\,0 & 0\,0\,0\,1\,\ldots\,0\,0 & 0 & 0 & 0 & 1 \\
[1,p-1,(3)] & 1\,0\,0\,0\,\ldots\,0\,0 & 0\,0\,0\,0\,\ldots\,0\,1 & 0\,0\,1\,0\,\ldots\,0\,0 & 0 & 0 & 0 & 1 \\
[1,p,(2)] & 1\,0\,0\,0\,\ldots\,0\,0 & 0\,0\,0\,0\,\ldots\,0\,0 & 0\,1\,0\,0\,\ldots\,0\,0 & 0 & 1 & 0 & 1 \\
[2,1,(p)] & 0\,1\,0\,0\,\ldots\,0\,0 & 1\,0\,0\,0\,\ldots\,0\,0 & 0\,0\,0\,0\,\ldots\,0\,0 & 0 & 0 & 1 & 1 \\
[2,2,(p-1)] & 0\,1\,0\,0\,\ldots\,0\,0 & 0\,1\,0\,0\,\ldots\,0\,0 & 0\,0\,0\,0\,\ldots\,0\,1 & 0 & 0 & 0 & 1 \\
[2,3,(p-2)] & 0\,1\,0\,0\,\ldots\,0\,0 & 0\,0\,1\,0\,\ldots\,0\,0 & 0\,0\,0\,0\,\ldots\,1\,0 & 0 & 0 & 0 & 1 \\
 & \cdots & \cdots & \cdots & & & \cdots & \\
[2,p-2,(3)] & 0\,1\,0\,0\,\ldots\,0\,0 & 0\,0\,0\,0\,\ldots\,1\,0 & 0\,0\,1\,0\,\ldots\,0\,0 & 0 & 0 & 0 & 1 \\
[2,p-1,(2)] & 0\,1\,0\,0\,\ldots\,0\,0 & 0\,0\,0\,0\,\ldots\,0\,1 & 0\,1\,0\,0\,\ldots\,0\,0 & 0 & 0 & 0 & 1 \\
[2,p,(1)] & 0\,1\,0\,0\,\ldots\,0\,0 & 0\,0\,0\,0\,\ldots\,0\,0 & 1\,0\,0\,0\,\ldots\,0\,0 & 0 & 1 & 0 & 1 \\
 & \cdots & \cdots & \cdots & & & \cdots & \\
[p-2,1,(4)] & 0\,0\,0\,0\,\ldots\,1\,0 & 1\,0\,0\,0\,\ldots\,0\,0 & 0\,0\,0\,1\,\ldots\,0\,0 & 0 & 0 & 0 & 1 \\
[p-2,2,(3)] & 0\,0\,0\,0\,\ldots\,1\,0 & 0\,1\,0\,0\,\ldots\,0\,0 & 0\,0\,1\,0\,\ldots\,0\,0 & 0 & 0 & 0 & 1 \\
[p-2,3,(2)] & 0\,0\,0\,0\,\ldots\,1\,0 & 0\,0\,1\,0\,\ldots\,0\,0 & 0\,1\,0\,0\,\ldots\,0\,0 & 0 & 0 & 0 & 1 \\
[p-2,4,(1)] & 0\,0\,0\,0\,\ldots\,1\,0 & 0\,0\,0\,1\,\ldots\,0\,0 & 1\,0\,0\,0\,\ldots\,0\,0 & 0 & 0 & 0 & 1 \\
[p-1,1,(3)] & 0\,0\,0\,0\,\ldots\,0\,1 & 1\,0\,0\,0\,\ldots\,0\,0 & 0\,0\,1\,0\,\ldots\,0\,0 & 0 & 0 & 0 & 1 \\
[p-1,2,(2)] & 0\,0\,0\,0\,\ldots\,0\,1 & 0\,1\,0\,0\,\ldots\,0\,0 & 0\,1\,0\,0\,\ldots\,0\,0 & 0 & 0 & 0 & 1 \\
[p-1,3,(1)] & 0\,0\,0\,0\,\ldots\,0\,1 & 0\,0\,1\,0\,\ldots\,0\,0 & 1\,0\,0\,0\,\ldots\,0\,0 & 0 & 0 & 0 & 1 \\
[p,1,(2)] & 0\,0\,0\,0\,\ldots\,0\,0 & 1\,0\,0\,0\,\ldots\,0\,0 & 0\,1\,0\,0\,\ldots\,0\,0 & 1 & 0 & 0 & 1 \\
[p,2,(1)] & 0\,0\,0\,0\,\ldots\,0\,0 & 0\,1\,0\,0\,\ldots\,0\,0 & 1\,0\,0\,0\,\ldots\,0\,0 & 1 & 0 & 0 & 1
\end{array}
\qquad (2.2.4)
$$

Then, with A partitioned, as indicated in equation (2.2.4), into $[A_1, A_2] = [A_1, A_1Q]$, with $A_2 = A_1Q$, we have

$$Q = (A_1'A_1)^{-1} A_1'A_2 \quad . \tag{2.2.5}$$

Thus the last four columns (extension) of equation (2.2.4) are linear functions of the basis A_1 with a transformation matrix Q. In general Q, as derived later in this section, will be of the following form:

$$Q = \frac{1}{2p-3} \begin{array}{c|cccc|}
 & (\alpha_p & \beta_p & \gamma_{(p)} & \mu) \\
\alpha_1 & -(p-2) & p-1 & p-1 & p-1 \\
\alpha_2 & -(p-1) & p-2 & p-2 & p-2 \\
\alpha_3 & -p & p-3 & p-3 & p-3 \\
\alpha_4 & -(p+1) & p-4 & p-4 & p-4 \\
\cdot & \cdot & \cdot & \cdot & \cdot \\
\alpha_{p-1} & -(2p-4) & 1 & 1 & 1 \\
\hdashline
\beta_1 & p-1 & -(p-2) & p-1 & p-1 \\
\beta_2 & p-2 & -(p-1) & p-2 & p-2 \\
\beta_3 & p-3 & -p & p-3 & p-3 \\
\beta_4 & p-4 & -(p+1) & p-4 & p-4 \\
\cdot & \cdot & \cdot & \cdot & \cdot \\
\beta_{p-1} & 1 & -(2p-4) & 1 & 1 \\
\hdashline
\gamma_{(1)} & p-1 & p-1 & -(p-2) & p-1 \\
\gamma_{(2)} & p-2 & p-2 & -(p-1) & p-2 \\
\gamma_{(3)} & p-3 & p-3 & -p & p-3 \\
\gamma_{(4)} & p-4 & p-4 & -(p+1) & p-4 \\
\cdot & \cdot & \cdot & \cdot & \cdot \\
\gamma_{(p-1)} & 1 & 1 & -(2p-4) & 1
\end{array} \quad . \tag{2.2.6}$$

The fact that $(A_1'A_1)^{-1}$ exists and that A_2 is, in fact, equal to A_1Q, proves that the rank of the design matrix is four less than its order.

Now, equation (2.1.5) states that the estimable functions must be such that

$$\underline{\lambda}_2' = \underline{\lambda}_1'Q \quad .$$

Also, for the estimation of contributions due to the effect, these estimable functions should not involve the arbitrary mean μ, so that $\underline{\lambda}_1'$ should be chosen in such a way that

$$(p-1)(\lambda_{\alpha_1} + \lambda_{\beta_1} + \lambda_{\gamma(1)}) + (p-2)(\lambda_{\alpha_2} + \lambda_{\beta_2} + \lambda_{\gamma(2)})$$
$$+ \ldots + (\lambda_{\alpha_{p-1}} + \lambda_{\beta_{p-1}} + \lambda_{\gamma(p-1)}) = 0 ,$$

where λ_{α_i}, λ_{β_j}, and $\lambda_{\gamma(k)}$ denotes the coefficient of α_i, β_j, and $\gamma_{(k)}$ in the estimable function. Suppose we wish to estimate the contrast given by

$$\underline{\lambda}_1' = [1, -1, 0, \ldots, 0 \;\; 0, 0, 0, \ldots, 0 \;\; 0, 0, 0, \ldots, 0].$$

Then

$$\underline{\lambda}_2' = \underline{\lambda}_1'Q = [\,\frac{1}{2p-3}, \;\; \frac{1}{2p-3}, \; \frac{1}{2p-3}, \; \frac{1}{2p-3}\,] \quad .$$

The estimable function is thus given by

$$\alpha_1 - \alpha_2 + \frac{1}{2p-3}(\alpha_p + \beta_p + \gamma_{(p)} + \mu) \; ,$$

a rather useless quantity. On the other hand, the contrast given by

$$\underline{\lambda}_1' = [1, -1, 0, \ldots, 0 \;\; -1, 1, 0, \ldots, 0 \;\; 0, 0, 0, \ldots, 0]$$

leads to

$$\underline{\lambda}_2' = \underline{\lambda}_1'Q = [0,\ 0,\ 0,\ 0] \quad .$$

The estimable function in this case is thus

$$\alpha_1 - \alpha_2 - \beta_1 + \beta_2 = (\alpha_1 - \beta_1) - (\alpha_2 - \beta_2) \quad .$$

In a similar way, it can be shown that the following contrasts are estimable:

$$\begin{aligned} &(\alpha_c - \beta_c\ ) - (\alpha_d - \beta_d\ ) \\ &(\alpha_c - \gamma_{(c)}) - (\alpha_d - \gamma_{(d)}) \qquad (c \neq d) \qquad (2.2.7) \\ &(\beta_c - \gamma_{(c)}) - (\beta_d - \gamma_{(d)}) \ . \end{aligned}$$

Suppose now we wish to estimate the contrast given by

$$\underline{\lambda}_1' = [-1,\ 2,\ -1,\ 0,\ \ldots,\ 0\quad 0,0,0,\ldots,0\quad 0,\ 0,\ 0,\ \ldots,\ 0].$$

Then

$$\underline{\lambda}_2' = \underline{\lambda}_1'Q = [0,\ 0,\ 0,\ 0] \quad .$$

The estimable function is given by

$$-\alpha_1 + 2\alpha_2 - \alpha_3\ ,\ \text{or}\quad \alpha_2 - \frac{\alpha_1 + \alpha_3}{2} \quad .$$

In a similar way, it can be shown that the following contrasts are estimable:

$$\begin{aligned} &\alpha_i - \frac{\alpha_{i-1} + \alpha_{i+1}}{2} \\ &\beta_j - \frac{\beta_{j-1} + \beta_{j+1}}{2} \qquad i,j,(k) = 2,3,\ldots,(p-1) \\ &\gamma(k) - \frac{\gamma_{(k-1)} + \gamma_{(k+1)}}{2} \quad . \qquad (2.2.8) \end{aligned}$$

This holds for any three adjacent observations, since in Q, the middle element of three adjacent ones in the same column is always the mean of the two bordering ones. It should be noted that the estimable functions in (2.2.8) involve only one factor each.

The above can be generalized to the following estimable functions:

$$\begin{aligned} &\alpha_i - \frac{c\alpha_{i-n} + d\alpha_{i+m}}{c + d} \\ &\beta_j - \frac{c\beta_{j-n} + d\beta_{j+m}}{c + d} \\ &\gamma_{(k)} - \frac{c\gamma_{(k-n)} + d\gamma_{(k+m)}}{c + d} \quad , \end{aligned} \qquad (2.2.9)$$

where

$$\begin{aligned} c &= 1,\ 2,\ \ldots,\ (p-i) \\ d &= 1,\ 2,\ \ldots,\ (i-1) \\ i, j, (k) &= (d+1,\ \ldots,\ p-c). \end{aligned}$$

Hence, simple differences in one factor only are not estimable, but means (weighted or unweighted) can be estimated in each factor separately.

2.2.31 Derivation of the Transformation Matrix Q

We now proceed to show the evaluation of the transformation matrix Q. We will consider two Symmetrical Complementation designs; one where each factor appears at three levels ($p = 3$), and one where each factor appears at four levels ($p = 4$). Finally we will conjecture the general form of the matrix, and prove that it is correct.

(1) Three Level Case

For the case where each factor appears at three levels, the design is set out diagrammatically as

$$\begin{array}{l|cccl} & 1 & 2 & 3 & = \text{Level of B} \\ \hline 1 & & 3 & 2 & \\ 2 & 3 & 2 & 1 & \\ 3 & 2 & 1 & & \\ \text{Level of A} & & & & = \text{Level of C} \end{array}$$

The design matrix A is given by

$$A = \begin{array}{r} \text{cell } [i,j,(k)] \backslash \text{effect} \\ [1,2,(3)] \\ [1,3,(2)] \\ [2,1,(3)] \\ [2,2,(2)] \\ [2,3,(1)] \\ [3,1,(2)] \\ [3,2,(1)] \end{array} \begin{array}{c} (\mu \ \alpha_1 \ \alpha_2 \ \alpha_3 \ \beta_1 \ \beta_2 \ \beta_3 \ \gamma_{(1)} \ \gamma_{(2)} \ \gamma_{(3)})' \\ \left[\begin{array}{cccccccccc} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{array}\right] \end{array} .$$

The design matrix can now be partitioned into a basis A_1 and an extension A_2, as given by equation (2.2.4), where

$$A_1 = \begin{array}{c} (\alpha_1 \ \alpha_2 \ \beta_1 \ \beta_2 \ \gamma_{(1)} \ \gamma_{(2)})' \\ \left[\begin{array}{cccccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right] \end{array} ,$$

and

$$A_2 = \left[\begin{array}{cccc} (\alpha_3 & \beta_3 & \gamma_{(3)} & \mu) \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{array}\right] .$$

Then, using equation (2.2.5), we obtain the Q matrix as

$$Q = \frac{1}{3} \left[\begin{array}{cccc} (\alpha_3 & \beta_3 & \gamma_{(3)} & \mu) \\ -1 & 2 & 2 & 2 \\ -2 & 1 & 1 & 1 \\ 2 & -1 & 2 & 2 \\ 1 & -2 & 1 & 1 \\ 2 & 2 & -1 & 2 \\ 1 & 1 & -2 & 1 \end{array}\right] . \qquad (2.2.10)$$

(2) Four Level Case

For the case where each factor appears at four levels, the design is set out diagrammatically as

	1	2	3	4	= Level of B
1		4	3	2	
2	4	3	2	1	
3	3	2	1		
4	2	1			
Level of A					

= Level of C

The design matrix A is then given by

$$
A = \begin{array}{c|ccccccccccccc|}
\text{effect / cell } [i,j,(k)] & (\mu & \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \beta_1 & \beta_2 & \beta_3 & \beta_4 & \gamma_{(1)} & \gamma_{(2)} & \gamma_{(3)} & \gamma_{(4)}) \\
\hline
[1,2,(4)] & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
[1,3,(3)] & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
[1,4,(2)] & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
[2,1,(4)] & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
[2,2,(3)] & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
[2,3,(2)] & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
[2,4,(1)] & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
[3,1,(3)] & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
[3,2,(2)] & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
[3,3,(1)] & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
[4,1,(2)] & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
[4,2,(1)] & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0
\end{array} .
$$

The design matrix can be partitioned into a basis A_1 and an extension A_2, as given by equation (2.2.4), where

$$
A_1 = \begin{array}{ccccccccc}
(\alpha_1 & \alpha_2 & \alpha_3 & \beta_1 & \beta_2 & \beta_3 & \gamma_{(1)} & \gamma_{(2)} & \gamma_{(3)}) \\
\left[\begin{matrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{matrix}\right. & \begin{matrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{matrix} & \begin{matrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{matrix} & \begin{matrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{matrix} & \left.\begin{matrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{matrix}\right]
\end{array} ,
$$

and

$$A_2 = \begin{array}{c} \begin{array}{cccc} (\alpha_4 & \beta_4 & \gamma_{(4)} & \mu) \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \end{array} .$$

Then, using equation (2.2.5), we obtain the matrix Q as

$$Q = \frac{1}{5} \begin{array}{c} \begin{array}{cccc} (\alpha_4 & \beta_4 & \gamma_{(4)} & \mu) \end{array} \\ \begin{bmatrix} -2 & 3 & 3 & 3 \\ -3 & 2 & 2 & 2 \\ -4 & 1 & 1 & 1 \\ 3 & -2 & 3 & 3 \\ 2 & -3 & 2 & 2 \\ 1 & -4 & 1 & 1 \\ 3 & 3 & -2 & 3 \\ 2 & 2 & -3 & 2 \\ 1 & 1 & -4 & 1 \end{bmatrix} \end{array} . \qquad (2.2.11)$$

The general expression (2.2.6) was conjectured on the basis of these results and can be verified to be accurate as follows:

The $A_1'A_1$ matrix can be obtained from the matrix in equation (2.2.3) by eliminating the very first column and row, and the last column and row in each of the nine submatrices, due to rows, columns, and diagonals, respectively. Thus, let

$$\begin{bmatrix} p-1 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 1 & 1 & \cdots & 1 & 1 & 0 & 1 & 1 & 1 & \cdots & 1 & 1 \\ 0 & p & 0 & 0 & \cdots & 0 & 1 & 1 & 1 & 1 & \cdots & 1 & 1 & 1 & 1 & 1 & 1 & \cdots & 1 & 1 \\ 0 & 0 & p-1 & 0 & \cdots & 0 & 1 & 1 & 1 & 1 & \cdots & 1 & 1 & 1 & 1 & 1 & 1 & \cdots & 1 & 1 \\ 0 & 0 & 0 & p-2 & \cdots & 0 & 1 & 1 & 1 & 1 & \cdots & 1 & 0 & 1 & 1 & 1 & 1 & \cdots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 3 & 1 & 1 & 1 & 0 & \cdots & 0 & 0 & 1 & 1 & 1 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

$$= [\ D\ E\ E\ ]\ , \text{ say.}$$

Then

$$A_1'A_1 = \begin{array}{ccc} \left[\begin{array}{ccc} D & E & E \\ E & D & E \\ E & E & D \end{array}\right] & \begin{array}{c} (p-1) \\ (p-1) \\ (p-1) \end{array} \\ \begin{array}{ccc} (p-1) & (p-1) & (p-1) \end{array} & \end{array} \quad ,$$

where the numbers in parenthesis indicate the order of the submatrices D and E.

Also, Q is the form

$$Q = \frac{1}{2p-3} \begin{array}{cc} \left[\begin{array}{cccc} \underline{u} & \underline{v} & \underline{v} & \underline{v} \\ \underline{v} & \underline{u} & \underline{v} & \underline{v} \\ \underline{v} & \underline{v} & \underline{u} & \underline{v} \end{array}\right] & \begin{array}{c} (p-1) \\ (p-1) \\ (p-1) \end{array} \\ \begin{array}{cccc} (1) & (1) & (1) & (1) \end{array} & \end{array} ,$$

where

$$\underline{u}' = \left[-(p-2),\ -(p-1),\ -p,\ -(p+1),\ \ldots,\ -(2p-4) \right] ,$$

and

$$\underline{v}' = \left[(p-1),\ (p-2),\ (p-3),\ (p-4),\ \ldots,\ 1 \right] .$$

Then

$$A_1'A_1Q = \frac{1}{2p-3} \left[\begin{array}{cccc} D\underline{u} + 2E\underline{v} & (D+E)\underline{v} + E\underline{u} & (D+E)\underline{v} + E\underline{u} & (D+2E)\underline{v} \\ (D+E)\underline{v} + E\underline{u} & D\underline{u} + 2E\underline{v} & (D+E)\underline{v} + E\underline{u} & (D+2E)\underline{v} \\ (D+E)\underline{v} + E\underline{u} & (D+E)\underline{v} + E\underline{u} & D\underline{u} + 2E\underline{v} & (D+2E)\underline{v} \end{array}\right] .$$

Now

$$\underline{u}'D' = \underline{u}'D = \left[-(p-2)(p-1),\ -(p-1)p,\ -p(p-1),\ -(p+1)(p-2),\ \ldots, -(2p-4)(3) \right] ,$$

and

$$2\underline{v}'E' = 2\underline{v}'E = \left[(p-2)(p-1),\ (p-1)p,\ p(p-1),\ (p+1)(p-2),\ \ldots,\ (2p-4)(3) \right] .$$

Hence, summing these two row vectors, we obtain

$$\underline{u}'D + 2\underline{v}'E = \underline{0} ,$$

or

$$D\underline{u} + 2E\underline{v} = \underline{0} .$$

Also

$$\underline{v}D' = \underline{v}'D = \left[\ (p-1)^2,\ (p-2)p,\ (p-3)(p-1),\ (p-4)(p-2),\ \ldots,\ (1)(3)\ \right] .$$

Then

$$\begin{aligned} \underline{v}'\ (D+2E) &= \underline{v}'D + 2\underline{v}'E \\ &= \left[\ (p-1)(2p-3),\ p(2p-3),\ (p-1)(2p-3),\ (p-2)(2p-3),\ \ldots,\ 3(2p-3)\right] \\ &= (2p-3)\ \left[\ p-1,\ p,\ p-1,\ p-2,\ \ldots,\ 3\ \right] . \end{aligned}$$

Also

$$\begin{aligned} \underline{v}'E &= \frac{1}{2}\left[\ (p-2)(p-1),\ (p-1)p,\ p(p-1),\ (p+1)(p-2),\ \ldots,\ (2p-4)\ (3)\ \right] , \\ \underline{v}'D &= \left[\ (p-1)^2,\ (p-2)p,\ (p-3)(p-1),\ (p-4)(p-2),\ \ldots,\ (1)(3)\ \right] , \\ \underline{u}'E &= -\frac{1}{2}\left[\ (3p-5)(p-2),\ (3p-6)(p-1),\ (3p-6)(p-1),\ (3p-7)(p-2),\ \ldots,\ (2p-2)(3)\ \right] . \end{aligned}$$

Then, summing these three row vectors, we obtain

$$\begin{aligned} (D+E)\underline{v} + E\underline{u} &= \underline{v}'(D+E) + \underline{u}'E \\ &= \underline{v}'D + \underline{v}'E + \underline{u}'E \\ &= \left[\ 2p-3,\ 2p-3,\ 0,\ 0,\ \ldots,\ 0\ \right] . \end{aligned}$$

Hence

$$A_1'A_1Q = \begin{bmatrix} 0 & 1 & 1 & p-1 \\ 0 & 1 & 1 & p \\ 0 & 0 & 0 & p-1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 3 \\ 1 & 0 & 1 & p-1 \\ 1 & 0 & 1 & p \\ 0 & 0 & 0 & p-1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 3 \\ 1 & 1 & 0 & p-1 \\ 1 & 1 & 0 & p \\ 0 & 0 & 0 & p-1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 3 \end{bmatrix} . \qquad (2.2.12)$$

The last column in equation (2.2.12) is clearly recognized as the first column in equation (2.2.3) after elimination of the rows corresponding to A_2; the first three columns will be recognized as the corresponding last columns in each set of equations due to rows, columns, and diagonals, respectively. Thus, $A_1'A_1Q = A_1'A_2$, and the correctness of the general transformation matrix Q is verified.

2.2.4 Generally Estimable Functions in One and Two Factors

In Section 2.2.3, we indicated that certain contrasts in one and two factors were estimable. The results are easily verified for the general case of p levels, as will be shown in this section.

2.2.4.1 Estimable Functions in One Factor Only

Let us consider an estimable function in one factor only, say α . Due to the complete symmetry of the design, estimable functions verified for the factor α also hold for the other two factors, β and γ . If the factor appears at p levels, the estimable function will be of the form

$$\alpha_1 + a_2\alpha_2 + a_3\alpha_3 + \dots + a_{p-1}\,\alpha_{p-1} + a_p\alpha_p \qquad , \qquad (2.2.13)$$

where $a_2, a_3, \dots, a_{p-1}, a_p$ are unknown constants which are to be found. The transformation matrix Q was given by equation (2.2.6), and we can write the transpose of Q as

$$Q' =$$

$$\frac{1}{2p-3}\left[\begin{matrix} -(p-2) & -(p-1) & -p & \cdots & -(2p-4) & p-1 & p-2 & p-3 & \cdots & 1 \\ p-1 & p-2 & p-3 & \cdots & 1 & -(p-2) & -(p-1) & -p & \cdots & -(2p-4) \\ p-1 & p-2 & p-3 & \cdots & 1 & p-1 & p-2 & p-3 & \cdots & 1 \\ p-1 & p-2 & p-3 & \cdots & 1 & p-1 & p-2 & p-3 & \cdots & 1 \end{matrix}\right.$$

$$\left.\begin{matrix} p-1 & p-2 & p-3 & \cdots & 1 \\ p-1 & p-2 & p-3 & \cdots & 1 \\ -(p-2) & -(p-1) & -p & \cdots & -(2p-4) \\ p-1 & p-2 & p-3 & \cdots & 1 \end{matrix}\right]^{*} .$$

Now, to find the unknown constants in equation (2.2.13), we must solve equation (2.1.5), which states that

$$\underline{\lambda}_2' = \underline{\lambda}_1' Q \quad .$$

For the case under consideration

$$\underline{\lambda}_1' = [\, 1,\ a_2,\ a_3,\ \ldots,\ a_{p-1} \quad 0,\ 0,\ 0,\ \ldots,\ 0,\ 0,\ \ 0,\ 0,\ \ldots,\ 0 \,],$$

and

$$\underline{\lambda}_2' = [\ 1,\ 0,\ 0,\ 0\ ] \quad .$$

---

* Large matrices are presented in two parts.

Thus, we obtain a system of two equations in (p-1) unknowns, given by

$$-(p-2) - (p-1)a_2 - pa_3 - \dots - (2p-4)a_{p-1} = a_p \qquad (2.2.14)$$

$$(p-1) + (p-1)a_2 + (p-3)a_3 + \dots + a_{p-1} = 0 \quad . \qquad (2.2.15)$$

To solve this system, we may arbitrarily specify (p-3) of the unknowns and solve for the remaining two unknowns in terms of the specified ones. Let us solve equation (2.2.15) for a_{p-1}, obtaining

$$a_{p-1} = -\left[(p-1) + (p-2)a_2 + (p-3)a_3 + \dots + 2a_{p-2} \right] \quad .$$

If we now substitute this result into equation (2.2.14), we find the value of a_p as

$$a_p = (p-2) + (p-3)a_2 + (p-4)a_3 + \dots + a_{p-2} \quad .$$

Thus the estimable function, given by equation (2.2.13), is

$$\begin{aligned} &\alpha_1 + a_2\alpha_2 + a_3\alpha_3 + \dots + a_{p-2}\alpha_{p-2} \\ &\quad - \left[(p-1) + (p-2)a_2 + (p-3)a_3 + \dots + 2a_{p-2} \right] \alpha_{p-1} \quad . \\ &\quad\quad + \left[(p-2) + (p-3)a_2 + (p-4)a_3 + \dots + a_{p-2} \right] \alpha_p \end{aligned}$$

$$(2.2.16)$$

Analogously, and because of symmetry, we obtain the estimable function in β only, and in γ only.

It is easily seen that the usual linear contrast is not estimable, but that all contrasts of a higher order are estimable. As an example, let us consider the case of four levels (p=4). The estimable function in α is then of the form

$$\alpha_1 + a_2\alpha_2 - [3 + 2a_2]\,\alpha_3 + [2 + a_2]\,\alpha_4 \quad .$$

Suppose now, we let $a_2 = -1$. Then the estimable function becomes

$$\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 \quad ,$$

which is recognized as the contrast usually called "quadratic".[*] By letting $a_2 = -3$, we obtain the function

$$\alpha_1 - 3\alpha_2 + 3\alpha_3 - \alpha_4 \quad ,$$

which is the contrast, usually called "cubic".

Let us now consider estimable functions involving three levels only. Here we may let $2 + a_2 = 0$, or $a_2 = -2$. We thus obtain

$$\alpha_1 -2\,\alpha_2 + \alpha_3 \quad ,$$

the contrast which compares one level with the mean of two other levels. In a similar manner, we might let $a_2 = 0$.

---

[*] The actual "quadratic" contrast will be described in Section 4.5. In this section we only try to find some reference contrasts from which all others can be constructed, and we arbitrarily choose the usual "orthogonal polynomials".

Here we would obtain

$$\alpha_1 - 3\alpha_3 + 2\alpha_4 \quad ,$$

which is a contrast comparing one level with the weighted mean of two other levels. Other estimable functions can easily be found by arbitrarily choosing values of a_2 .

We have shown in the special case of four levels how one can find various estimable functions. This procedure can be extended to the general case by arbitrarily choosing the (p-3) unknown constants in equation (2.2.16), or in analogous equations involving the β only and γ only.

2.2.4.2 Estimable Functions in Two Factors

Let us now consider estimable functions involving two factors, say α and β , with no more than (p-1) levels in each. Once again, estimable function in α and γ , and β and γ , will follow immediately due to the symmetry of the design. The estimable function will be of the form

$$\begin{aligned} &\alpha_1 + a_2\alpha_2 + a_3\alpha_3 + \ldots + a_{p-1}\alpha_{p-1} \qquad (2.2.17)\\ &\quad +b_1\beta_1 + b_2\beta_2 + b_3\beta_3 + \ldots + b_{p-1}\beta_{p-1} \quad , \end{aligned}$$

where the a's and the b's are unknown constants. Employing the same method as used in Section 2.2.4.1, with

$\underline{\lambda}_1' =$

$$[1, a_2, a_3, \ldots, a_{p-1}, b_1, b_2, b_3, \ldots, b_{p-1}, 0, 0, 0, \ldots, 0] ,$$

and

$$\underline{\lambda}_2' = [0, 0, 0, 0,] ,$$

we obtain a system of three equations in (2p-3) unknowns:

$$\begin{aligned}
&-(p-2) - (p-1)a_2 - pa_3 - \ldots - (2p-4)a_{p-1} \\
&\qquad + (p-1)b_1 + (p-2)b_2 + (p-3)b_3 + \ldots + b_{p-1} = 0 \\
&(p-1) + (p-2)a_2 + (p-3)a_3 + \ldots + a_{p-1} - (p-2)b_1 \\
&\qquad - (p-1)b_2 - pb_3 - \ldots - (2p-4)b_{p-1} = 0 \\
&(p-1) + (p-2)a_2 + (p-3)a_3 + \ldots + a_{p-1} \\
&\qquad + (p-1)b_1 + (p-2)b_2 + (p-3)b_3 + \ldots + b_{p-1} = 0 .
\end{aligned}$$

To solve this system, we may arbitrarily specify (2p-6) of the unknowns and solve for the remaining three unknowns in terms of the specified ones. If we specify all but a_{p-1}, b_{p-2}, and b_{p-1} as arbitrary, we may write the above system, after transposing the arbitrary terms to the right-hand sides of the equations, in matrix notation as

$$\begin{bmatrix} -(2p-4) & 2 & 1 \\ 1 & -(2p-5) & -(2p-4) \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} a_{p-1} \\ b_{p-2} \\ b_{p-1} \end{bmatrix} =$$

$$\begin{bmatrix} (p-2) & (p-1) & p & \cdots & (2p-5) & -(p-1) & -(p-2) & -(p-3) & \cdots & -3 \\ -(p-1) & -(p-2) & -(p-3) & \cdots & -2 & (p-2) & (p-1) & p & \cdots & (2p-6) \\ -(p-1) & -(p-2) & -(p-3) & \cdots & -2 & -(p-1) & -(p-2) & -(p-3) & \cdots & -3 \end{bmatrix}$$

$$\times \begin{bmatrix} 1 \\ a_2 \\ a_3 \\ \vdots \\ a_{p-2} \\ b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_{p-3} \end{bmatrix}$$

The solution of this system for the three unknowns a_{p-1}, b_{p-2}, and b_{p-1} is given by

$$\begin{bmatrix} a_{p-1} \\ b_{p-2} \\ b_{p-1} \end{bmatrix} = \frac{1}{2p-3} \begin{bmatrix} -1 & 0 & 1 \\ 1 & 1 & (2p-5) \\ -1 & -2 & -(2p-6) \end{bmatrix}$$

$$\times \begin{bmatrix} (p-2) & (p-1) & p & \cdots & (2p-5) & -(p-1) & -(p-2) & -(p-3) & \cdots & -3 \\ -(p-1) & -(p-2) & -(p-3) & \cdots & -2 & (p-2) & (p-1) & p & \cdots & (2p-6) \\ -(p-1) & -(p-2) & -(p-3) & \cdots & -2 & -(p-1) & -(p-2) & -(p-3) & \cdots & -3 \end{bmatrix}$$

$$\times \begin{bmatrix} 1 \\ a_2 \\ a_3 \\ \vdots \\ a_{p-2} \\ b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_{p-3} \end{bmatrix}$$

$$= \begin{bmatrix} -1 & -1 & -1 & \dots & -1 & 0 & 0 & 0 & \dots & 0 \\ -(p-2) & -(p-3) & -(p-4) & \dots & -1 & -(p-2) & -(p-3) & -(p-4) & \dots & -2 \\ (p-2) & (p-3) & (p-4) & \dots & 1 & (p-3) & (p-4) & (p-5) & \dots & 1 \end{bmatrix}$$

$$x \begin{bmatrix} 1 \\ a_2 \\ a_3 \\ \vdots \\ a_{p-2} \\ b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_{p-3} \end{bmatrix} .$$

Hence we have

$$a_{p-1} = -[1 + a_2 + a_3 \dots + a_{p-2}] ,$$

$$b_{p-2} = -[(p-2) + (p-3)a_2 + (p-4)a_3 + \dots + a_{p-2} + (p-2)b_1 + (p-3)b_2 + \dots + 2b_{p-3}] ,$$

$$b_{p-1} = [(p-2) + (p-3)a_2 + (p-4)a_3 + \dots + a_{p-2} + (p-3)b_1 + (p-4)b_2 + \dots + b_{p-3}] .$$

With these results the estimable function in α and β is given by

$$\alpha_1 + a_2\alpha_2 + a_3\alpha_3 + \ldots + a_{p-2}\alpha_{p-2}$$
$$-[1 + a_2 + a_3 + \ldots + a_{p-2}] \quad \alpha_{p-1}$$

$$+ b_1\beta_1 + b_2\beta_2 + \ldots + b_{p-3}\beta_{p-3}$$

$$-[(p-2) + (p-3)a_2 + \ldots + a_{p-2} + (p-2)b_1 + \ldots$$
$$+ 2b_{p-3}] \beta_{p-2}$$
$$+[(p-2) + (p-3)a_2 + \ldots + a_{p-2} + (p-3)b_1$$
$$+ \ldots + b_{p-3}] \beta_{p-1} \quad .$$

(2.2.18)

Analogously, and because of symmetry, we obtain the estimable function in α and γ, and β and γ, respectively. As an example, let us consider the case of four levels ($p=4$). The estimable function in α and β, say, is then of the form

$$\alpha_1 + a_2\alpha_2 - (1 + a_2)\alpha_3 + b_1\beta_1 - (2 + a_2 + 2b_1) \beta_2$$
$$+ (2 + a_2 + b_1)\beta_3 = 0 \quad .$$

Suppose now, we let $a_2 = 0$ and $b_1 = -1$. Then the estimable function becomes

$$\alpha_1 - \alpha_3 - \beta_1 + \beta_3 \quad .$$

This is an example of an estimable function that leads to the "hypothesis of substitution", to be discussed in the next chapter, viz.

$$H_o: \quad \alpha_1 + \beta_3 = \alpha_3 + \beta_1 \quad .$$

By letting $a_2 = -1$ and $b_1 = -1$, we obtain the estimable function

$$\alpha_1 - \alpha_2 - \beta_1 + \beta_2 \quad .$$

By letting $a_2 = -1$ and $b_1 = 0$, we obtain

$$\alpha_1 - \alpha_2 - \beta_2 + \beta_3$$

as the estimable function.
Hence we have shown that we can obtain the estimable functions

$$\alpha_1 - \alpha_3 - \beta_1 + \beta_3 \quad ,$$

and

$$\alpha_1 - \alpha_2 - \beta_2 + \beta_3 \quad .$$

From these we can formulate the hypothesis

$$H_o: \quad \alpha_1 + \beta_3 = \alpha_3 + \beta_1$$

and, simultaneously, the hypothesis

$$H_o: \quad \alpha_1 + \beta_3 = \alpha_2 + \beta_2 \; .$$

But this is the hypothesis

$$H_o: \qquad \alpha_1 + \beta_3 = \alpha_2 + \beta_2 = \alpha_3 + \beta_1 \quad ,$$

which if accepted, says that applying one factor at a low level and another one at a high level does not produce results which are different when the levels of the two factors are exchanged, nor does this result differ when the factors are both applied at the second level.

The above mentioned hypothesis of substitution can be tested in the general case of p levels, and it is one of the most important to be investigated in the present design. The test statistics which test these various hypotheses will be discussed in Chapter IV.

2.3 Reason for Leaving Corner Cells Empty

Suppose we have one observation per cell. Then, with the corner cells filled, we would have the model

$$y_{ij(k)} = \mu + \alpha_i + \beta_j + \gamma_{(k)} + e'_{ij(k)} \quad ,$$

where i, j, (k) = 1, 2, ..., p+1 . Let us consider the value of the observation in cell [1, 1, (p+1)], which would be

$$y_{11(p+1)} = \hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_1 + \hat{\gamma}_{(p+1)} + e'_{11(p+1)} \quad ,$$

where $e'_{11(p+1)}$ is (supposedly) the contribution due to lack-of-fit. However, since $\gamma_{(p+1)}$ occurs only in $y_{11(p+1)}$, it will be estimated in such a way that $e'_{11(p+1)} = 0$.

Therefore

$$y_{11(p+1)} = \hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_1 + \hat{\gamma}_{(p+1)} \quad .$$

For the same reason

$$y_{1(p+1)(1)} = \hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_{p+1} + \hat{\gamma}_{(1)} \quad .$$

and

$$y_{(p+1)1(1)} = \hat{\mu} + \hat{\alpha}_{p+1} + \hat{\beta}_1 + \hat{\gamma}_{(1)} \quad .$$

Now suppose these corner cells are empty. Then, all the parameters, save $\hat{\alpha}_{p+1}$, $\hat{\beta}_{p+1}$, and $\hat{\gamma}_{(p+1)}$ can be estimated from the remaining cells.

Retaining the corner cells would imply no lack-of-fit in the corner cells but presence of lack-of-fit in the others. To call this discrepancy between y in a corner cell and parameters estimated from other cells a (p+1)st effect, or a (p+1)st effect plus a contribution due to lack-of-fit, or a contribution due to lack-of-fit, is semantic. Regardless of what is put into the corner cells, the estimable functions involving the non-corner parameters will remain unchanged; the corner parameters, on the other hand, are artifacts, since they are not shared with any other treatment combination.

The design could be viewed as a p^3 factorial. A usual fractional replication would immediately alias main effects with two-way interaction. The present design is, as can be shown, an incomplete fractional replication. That main effects, for any p, will be aliased with two-way interaction effects can be proved more directly.

For p levels, there are $\binom{p+2}{2}$ -3 experiments, and there are (3p-2) main effects (including μ). Hence there will be $\binom{p+2}{2} -3 -(3p-2) = \frac{p(p-3)}{2}$ degrees of freedom due to lack-of-fit. In a p^3 experiment, there will be $3(p-1)^2$ degrees of freedom due to two-factor interaction. If none of these are to be aliased with main effects, they must all be contained in the degrees of freedom due to lack-of-fit. Let us see whether this can be true:

$$\begin{aligned}
\frac{p(p-3)}{2} &\geqslant 3(p-1)^2 \\
p^2 - 3p &\geqslant 6p^2 - 12p + 6 \\
0 &\geqslant 5p^2 - 9p + 6 \\
0 &\geqslant p^2 - 1.8p + 1.2 \\
-1.20 + 0.81 &\geqslant p^2 - 1.8p + 0.81 \\
-0.39 &\geqslant (p - 0.9)^2
\end{aligned}$$

Thus it is impossible to obtain main effect contrasts free of two-way interaction contrasts.

Let us now find a lower bound (which may not be attained) for p such that there is a possibility that no quadratic terms of interaction (linear by linear) will be aliased with main effects:

$$\begin{aligned}
\frac{p(p-3)}{2} &\geqslant 3 \\
p(p-3) &\geqslant 6 \\
p^2 - 3p + 2.25 &\geqslant 8.25 \\
(p - 1.5)^2 &\geqslant 8.25 \\
p &\geqslant 5
\end{aligned}
\qquad
\begin{aligned}
\frac{p(p-3)}{2} &\geqslant 10 \\
p^2 - 3p &\geqslant 20 \\
(p - 1.5)^2 &\geqslant 22.25 \\
p &\geqslant 7
\end{aligned}$$

It is thus clear that an attempt to perform analysis of a cross-classification model assuming interaction will lead to hopelessly aliased quantities. It will be recalled that a usual Latin square, with three factors (rows, columns, letters), is utilized for experiments only where interactions are absent--they are combined into a "residual" sum-of-squares. The present design can be regarded as a Latin square with missing values, so incomplete, however, that application of usual missing-value techniques of analysis would lead to questionable results, for even the estimability conditions are more restricted in the present design than those in the usual Latin square design.

CHAPTER III

ESTIMATION OF PARAMETERS

3.1 Introductory Remarks on Procedures

In this chapter, two methods of estimation of the parameters will be described. We will assume that there is but one observation per cell. Modifications for the case where there are n observations per cell will be discussed in Chapter V. There are several methods which can be employed in order to obtain estimates of the treatment effects under various constraints. It must be noted, however, that these estimates are rather meaningless quantities. It is only when they are combined in estimable functions that unique results are obtained. If only estimation is required, the high-low method should be used. This method is described in complete detail in Section 3.2. If, however, both estimation and tests of hypotheses are required, the modified high-low method should be used. This second method is described in Section 3.3, with the inverse matrix explicitly stated for four and five levels, and a general method for obtaining this inverse presented for the case of more than five levels. Both methods employ four constraints, namely,

$$\sum_i n_{i..} \hat{\alpha}_i + \sum_j n_{.j.} \hat{\beta}_j + \sum_{(k)} n_{..(k)} \hat{\gamma}_{(k)} = 0 \qquad (3.1.1)$$

$$\sum_i \hat{\alpha}_i = \sum_j \hat{\beta}_j = \sum_{(k)} \hat{\gamma}_{(k)} = 0 \ , \qquad (3.1.2)$$

where $n_{i..}$, $n_{.j.}$, and $n_{..(k)}$ refer to the number of observation in the ith row, jth column, and (k)th diagonal, respectively.

The modified high-low method cannot be used with these constraints if the number of levels is three. For, in the special case of three levels, the four constraints can be combined in a linear combination in such a way that an estimable function will result. From the general theory of estimation, see, e.g., Graybill (1961) , it is known that constraints or linear functions of constraints must be non-estimable; otherwise, the resulting constrained normal equations matrix will still be singular. Thus, for the case of three levels we will present, in section 3.3.1, the detailed solution subject to the constraints $\hat{\alpha}_3 = \hat{\beta}_3 = \hat{\gamma}_{(3)} = 0$, in place of the last three constraints in the general case.

Where three, four, and five levels of each factor are involved, the complete inverse and formulas for sums of squares based upon the latter method have been worked out, and are presented in the chapter on "Tests of Hypotheses". Hence, where an experiment is of this size, the detailed formulas in Chapter IV may be used directly, and the results in the present chapter may be disregarded. However, when an experiment has more than five levels, two cases should be distinguished:

(a) only estimation required. In this case, the simpler high-low method should be used.

(b) estimation and tests of hypotheses required. In this case, the modified high-low method should be used throughout, since the formulas producing estimates will be a by-product of the matrix inversion required.

3.2 High-Low Method

The estimation procedure can be reduced to the solution of single equations in one unknown at a time by the following device:

> Let us expand the present design to (p+1) levels by filling in the corner cells. We already know that regardless of what is put into the corner cells, estimable functions involving the non-corner parameters will remain unchanged. The corner parameters, on the other hand, are artifacts, since they are not shared with any other treatment effect combination. Figure 1, with the usual notation and with zeros substituted for convenience in the corner cells, represents the design as considered for the high-low method of estimation.

= Level of B

	1	2	3	...	p-1	p	p+1
1	0	$y_{12(p)}$	$y_{13(p-1)}$	...	$y_{1(p-1)(3)}$	$y_{1p(2)}$	0
2	$y_{21(p)}$	$y_{22(p-1)}$	$y_{23(p-2)}$	...	$y_{2(p-1)(2)}$	$y_{2p(1)}$	
3	$y_{31(p-1)}$	$y_{32(p-2)}$	$y_{33(p-3)}$	...	$y_{3(p-1)(1)}$		
⋮							
p-1	$y_{(p-1)1(3)}$	$y_{(p-1)2(2)}$	$y_{(p-1)3(1)}$				
p	$y_{p1(2)}$	$y_{p2(1)}$					
p+1	0						

Level of A

= Level of C

Figure 1: Symmetrical Complementation Design extended to (p+1) levels.

For illustration purposes, let us consider a p-level design* (p=4) where we will fill in the corner cells with zeros to facilitate the computation. The normal equations, in matrix form, are given by

$$\left[\begin{array}{cccccccccccccccc}
15 & 5 & 4 & 3 & 2 & 1 & 5 & 4 & 3 & 2 & 1 & 5 & 4 & 3 & 2 & 1 \\
5 & 5 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
4 & 0 & 4 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\
3 & 0 & 0 & 3 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\
2 & 0 & 0 & 0 & 2 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
5 & 1 & 1 & 1 & 1 & 1 & 5 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
4 & 1 & 1 & 1 & 1 & 0 & 0 & 4 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\
3 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\
2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
5 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 5 & 0 & 0 & 0 & 0 \\
4 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 4 & 0 & 0 & 0 \\
3 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\
2 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right]
\left[\begin{array}{c}
\hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \hat{\alpha}_4 \\ \hat{\alpha}_5 \\
\hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \\
\hat{\gamma}_{(1)} \\ \hat{\gamma}_{(2)} \\ \hat{\gamma}_{(3)} \\ \hat{\gamma}_{(4)} \\ \hat{\gamma}_{(5)}
\end{array}\right]
=
\left[\begin{array}{c}
Y_{...} \\ Y_{1..} \\ Y_{2..} \\ Y_{3..} \\ Y_{4..} \\ Y_{5..} \\
Y_{.1.} \\ Y_{.2.} \\ Y_{.3.} \\ Y_{.4.} \\ Y_{.5.} \\
Y_{..(1)} \\ Y_{..(2)} \\ Y_{..(3)} \\ Y_{..(4)} \\ Y_{..(5)}
\end{array}\right]
\begin{array}{c}
\\ \\ \\ \\ \\ (=0) \\ \\ \\ . \\ \\ (=0) \\ \\ \\ \\ \\ (=0)
\end{array}
\tag{3.2.1}$$

Let us now solve this system of equations in a systematic manner, by introducing the constraints into some, but not all,

---

* By a p-level design, we mean the Symmetrical Complementation Design with the corner cells empty. This designation will be used throughout the discussion.

of the rows of the normal equations. By introducing the constraint given by equation (3.1.1) into the first row of the normal equations, we immediately find the estimate of the general effect to be

$$\hat{\mu} = Y_{\cdot\cdot\cdot}/15 \quad . \tag{3.2.2}$$

Upon introducing this result into the remaining normal equations, and transposing the terms involving $\hat{\mu}$ to the right-hand side of the equations, we are left with a set of 15 equations. Let us note that these equations can be considered as three subsets of five equations each, to be called the row, column, and diagonal subsets, respectively.

The steps in the computational procedure are as follows:

1. Introduce the constraints into the first equation of each subset, thereby giving the solutions

$$\hat{\alpha}_1 = \frac{Y_{1\cdot\cdot}}{5} - \hat{\mu} \quad ,$$

$$\hat{\beta}_1 = \frac{Y_{\cdot 1\cdot}}{5} - \hat{\mu} \quad , \tag{3.2.3}$$

and

$$\hat{\gamma}_{(1)} = \frac{Y_{\cdot\cdot(1)}}{5} - \hat{\mu} \quad .$$

2. With the results given by equations (3.2.3), solve the last (fifth, in this case) equation in each subset, thereby obtaining

$$\hat{\alpha}_5 = Y_{5..} - \hat{\mu} - \hat{\beta}_1 - \hat{\gamma}_{(1)}$$

$$= Y_{5..} - \frac{Y_{.1.} + Y_{..(1)}}{5} + \hat{\mu}$$

$$= - \frac{Y_{.1.} + Y_{..(1)}}{5} + \hat{\mu} \quad ,$$

$$\hat{\beta}_5 = Y_{.5.} - \hat{\mu} - \hat{\alpha}_1 - \hat{\gamma}_{(1)}$$

$$= Y_{.5.} - \frac{Y_{1..} + Y_{..(1)}}{5} + \hat{\mu}$$

$$= - \frac{Y_{1..} + Y_{..(1)}}{5} + \hat{\mu} \quad , \qquad (3.2.4)$$

and

$$\hat{\gamma}_{(5)} = Y_{..(5)} - \hat{\mu} - \hat{\alpha}_1 - \hat{\beta}_1$$

$$= Y_{..(5)} - \frac{Y_{1..} + Y_{.1.}}{5} + \hat{\mu}$$

$$= - \frac{Y_{1..} + Y_{.1.}}{5} + \hat{\mu} \quad .$$

Note that no constraint has been introduced in these rows.

3. Introduce the constraints given by equations (3.1.2) into the second row of each subset. Here one utilizes the constraints in the form

$$\hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 + \hat{\alpha}_4 = -\hat{\alpha}_5 \quad ,$$

$$\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4 = -\hat{\beta}_5 \quad ,$$

and

$$\hat{\gamma}_{(1)} + \hat{\gamma}_{(2)} + \hat{\gamma}_{(3)} + \hat{\gamma}_{(4)} = -\hat{\gamma}_{(5)} \quad ;$$

since, by equations (3.2.4), the estimates $\hat{\alpha}_5$, $\hat{\beta}_5$, and $\hat{\gamma}_{(5)}$ are known. The solutions are

$$\begin{aligned}
\hat{\alpha}_2 &= \frac{1}{4}\,[Y_{2..} - 4\hat{\mu} + \hat{\beta}_5 + \hat{\gamma}_{(5)}] \\
&= \frac{1}{4}\,[Y_{2..} + Y_{.5.} + Y_{..(5)} - \frac{2Y_{1..} + Y_{.1.} + Y_{..(1)}}{5} - 2\hat{\mu}] \\
&= \frac{1}{4}\,[Y_{2..} - \frac{2Y_{1..} + Y_{.1.} + Y_{..(1)}}{5} - 2\hat{\mu}] \quad ,
\end{aligned}$$

$$\begin{aligned}
\hat{\beta}_2 &= \frac{1}{4}\,[Y_{.2.} - 4\hat{\mu} + \hat{\alpha}_5 + \hat{\gamma}_{(5)}] \\
&= \frac{1}{4}\,[Y_{.2.} + Y_{5..} + Y_{..(5)} - \frac{Y_{1..} + 2Y_{.1.} + Y_{..(1)}}{5} - 2\hat{\mu}] \quad , \\
&= \frac{1}{4}\,[Y_{.2.} - \frac{Y_{1..} + 2Y_{.1.} + Y_{..(1)}}{5} - 2\hat{\mu}] \quad ,
\end{aligned}$$

and

$$\begin{aligned}
\hat{\gamma}_{(2)} &= \frac{1}{4}\,[Y_{..(2)} - 4\hat{\mu} = \hat{\alpha}_5 + \hat{\beta}_5] \\
&= \frac{1}{4}\,[Y_{..(2)} + Y_{5..} + Y_{.5.} - \frac{Y_{1..} + Y_{.1.} + 2Y_{..(1)}}{5} - 2\hat{\mu}] \\
&= \frac{1}{4}\,[Y_{..(2)} - \frac{Y_{1..} + Y_{.1.} + 2Y_{..(1)}}{5} - 2\hat{\mu}] \quad .
\end{aligned} \tag{3.2.5}$$

4. With the results given by equations (3.2.3) and equations (3.2.5), solve the fourth equation in each subset, thereby obtaining

$$\hat{\alpha}_4 = \frac{1}{2} [Y_{4\cdot\cdot} - 2\hat{\mu} - (\hat{\beta}_1 + \hat{\beta}_2) - (\hat{\gamma}_{(1)} + \hat{\gamma}_{(2)})]$$

$$= \frac{1}{2} [Y_{4\cdot\cdot} - \frac{Y_{\cdot 2\cdot} + Y_{\cdot\cdot(2)}}{4} - \frac{2Y_{5\cdot\cdot} + Y_{\cdot 5\cdot} + Y_{\cdot\cdot(5)}}{4}$$

$$+ \frac{2Y_{1\cdot\cdot} - Y_{\cdot 1\cdot} - Y_{\cdot\cdot(1)}}{4.5} + \hat{\mu}]$$

$$= \frac{1}{2} [Y_{4\cdot\cdot} - \frac{Y_{\cdot 2\cdot} + Y_{\cdot\cdot(2)}}{4} + \frac{2Y_{1\cdot\cdot} - Y_{\cdot 1\cdot} - Y_{\cdot\cdot(1)}}{4.5} + \hat{\mu}] ,$$

$$\hat{\beta}_4 = \frac{1}{2} [Y_{\cdot 4\cdot} - 2\hat{\mu} - (\hat{\alpha}_1 + \hat{\alpha}_2) - (\hat{\gamma}_{(1)} + \hat{\gamma}_{(2)}]$$

$$= \frac{1}{2} [Y_{\cdot 4\cdot} - \frac{Y_{2\cdot\cdot} + Y_{\cdot\cdot(2)}}{4} - \frac{Y_{5\cdot\cdot} + 2Y_{\cdot 5\cdot} + Y_{\cdot\cdot(5)}}{4}$$

$$+ \frac{-Y_{1\cdot\cdot} + 2Y_{\cdot 1\cdot} - Y_{\cdot\cdot(1)}}{4.5} - \hat{\mu}]$$

$$= \frac{1}{2} [Y_{\cdot 4\cdot} - \frac{Y_{2\cdot\cdot} + Y_{\cdot\cdot(2)}}{4} + \frac{-Y_{1\cdot\cdot} + 2Y_{\cdot 1\cdot} - Y_{\cdot\cdot(1)}}{4.5} + \hat{\mu}] ,$$

and

$$\hat{\gamma}_{(4)} = \frac{1}{2} [Y_{\cdot\cdot(4)} - 2\hat{\mu} - (\hat{\alpha}_1 + \hat{\alpha}_2) - (\hat{\beta}_1 + \hat{\beta}_2)]$$

$$= \frac{1}{2} [Y_{\cdot\cdot(4)} - \frac{Y_{2\cdot\cdot} + Y_{\cdot 2\cdot}}{4} - \frac{Y_{5\cdot\cdot} + Y_{\cdot 5\cdot} + 2Y_{\cdot\cdot(5)}}{4}$$

$$+ \frac{-Y_{1\cdot\cdot} - Y_{\cdot 1\cdot} + 2Y_{\cdot\cdot(1)}}{4.5} + \hat{\mu}]$$

$$= \frac{1}{2} [Y_{\cdot\cdot(4)} - \frac{Y_{2\cdot\cdot} + Y_{\cdot 2\cdot}}{4} + \frac{-Y_{1\cdot\cdot} - Y_{\cdot 1\cdot} + 2Y_{\cdot\cdot(1)}}{4.5} + \hat{\mu}] .$$

(3.2.6)

5. The results for the remaining estimates are found by again employing the constraints given by equations (3.1.2) in the rows of the third equations in each subset, this time in the form

$$\hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 = -(\hat{\alpha}_4 + \hat{\alpha}_5) \quad ,$$

$$\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 = -(\hat{\beta}_4 + \hat{\beta}_5) \quad ,$$

$$\hat{\gamma}_{(1)} + \hat{\gamma}_{(2)} + \hat{\gamma}_{(3)} = -(\hat{\gamma}_{(4)} + \hat{\gamma}_{(5)}) \quad .$$

The solutions are

$$\hat{\alpha}_3 = \frac{1}{3} [Y_{3..} -3\hat{\mu} + (\hat{\beta}_4 + \hat{\beta}_5) + (\hat{\gamma}_{(4)} + \hat{\gamma}_{(5)})]$$

$$= \frac{1}{3} [Y_{3..} + \frac{Y_{.4.} + Y_{..(4)}}{2} - \frac{2Y_{2..} + Y_{.2.} + Y_{..(2)}}{2.4}$$

$$+ \frac{-2Y_{5..} + 5Y_{.5.} + 5Y_{..(5)}}{2.4} - \frac{18Y_{1..} + 7Y_{.1.} + 7Y_{..(1)}}{2.4.5}]$$

$$= \frac{1}{3} [Y_{3..} + \frac{Y_{.4.} + Y_{..(4)}}{2} - \frac{2Y_{2..} + Y_{.2.} + Y_{..(2)}}{2.4}$$

$$- \frac{18Y_{1..} + 7Y_{.1.} + 7Y_{..(1)}}{2.4.5}] \quad ,$$

$$\hat{\beta}_3 = \frac{1}{3} [Y_{.3.} - 3\hat{\mu} + (\hat{\alpha}_4 + \hat{\alpha}_5) + (\hat{\gamma}_{(4)} + \hat{\gamma}_{(5)})]$$

$$= \frac{1}{3} [Y_{.3.} + \frac{Y_{4..} + Y_{..(4)}}{2} - \frac{Y_{2..} + 2Y_{.2.} + Y_{..(2)}}{2.4}$$

$$+ \frac{5Y_{5..} - 2Y_{.5.} + 5Y_{..(5)}}{2.4} - \frac{7Y_{1..} + 18Y_{.1.} + 7Y_{..(1)}}{2.4.5}]$$

$$= \frac{1}{3} [Y_{.3.} + \frac{Y_{4..} + Y_{..(4)}}{2} - \frac{Y_{2..} + 2Y_{.2.} + Y_{..(2)}}{2.4}$$

$$- \frac{7Y_{1..} + 18Y_{.1.} + 7Y_{..(1)}}{2.4.5}] \quad ,$$

and

$$\hat{\gamma}_{(3)} = \frac{1}{3}\,[Y_{..(3)} - 3\hat{\mu} + (\hat{\alpha}_4 + \hat{\alpha}_5) + (\hat{\beta}_4 + \hat{\beta}_5)]$$

$$= \frac{1}{3}\,[Y_{..(3)} + \frac{Y_{4..} + Y_{.4.}}{2} - \frac{Y_{2..} + Y_{.2.} + 2Y_{..(2)}}{2\cdot 4}$$

$$+ \frac{5Y_{5..} + 5Y_{.5.} - 2Y_{..(5)}}{2\cdot 4} - \frac{7Y_{1..} + 7Y_{.1.} + 18Y_{..(1)}}{2\cdot 4\cdot 5}]$$

$$= \frac{1}{3}\,[Y_{..(3)} + \frac{Y_{4..} + Y_{.4.}}{2} - \frac{Y_{2..} + Y_{.2.} + 2Y_{..(2)}}{2\cdot 4}$$

$$- \frac{7Y_{1..} + 7Y_{.1.} + 18Y_{..(1)}}{2\cdot 4\cdot 5}] \quad .$$

(3.2.7)

As a check, we calculate $\sum_{i} \hat{\alpha}_i$, $\sum_{j} \hat{\beta}_j$, and $\sum_{(k)} \hat{\gamma}_{(k)}$. Each of these quantities should sum to zero.

We have illustrated the high-low procedure for the case where each factor appears at four levels. The procedure can be generalized to the case of p levels by filling in the corner cells and treating the design as a (p+1) level design. The procedure can perhaps best be summarized as follows:

1. Into the first equation of each subset of equations, insert the constraints in the form as given by equations (3.1.2).
2. Into the second equations, insert the constraints in the forms

$$-\hat{\alpha}_{p+1} = \sum_{i=1}^{p} \hat{\alpha}_i \quad ,$$

$$-\hat{\beta}_{p+1} = \sum_{j=1}^{p} \hat{\beta}_j \quad ,$$

and

$$-\hat{\gamma}_{(p+1)} = \sum_{(k)=1}^{p} \hat{\gamma}_{(k)} \quad .$$

3. Into the third equations, insert the constraints in the forms

$$-(\hat{\alpha}_p + \hat{\alpha}_{p+1}) = \sum_{i=1}^{p-1} \hat{\alpha}_i \quad ,$$

$$-(\hat{\beta}_p + \hat{\beta}_{p+1}) = \sum_{j=1}^{p-1} \hat{\beta}_j \quad ,$$

and

$$-(\hat{\gamma}_{(p)} + \hat{\gamma}_{(p+1)} = \sum_{(k)=1}^{p-1} \hat{\gamma}_{(k)} \quad .$$

4. Continue the above procedure until the center row in each subset has been reached, at which time the inserting of the constraints is discontinued. For a subset with p+1 rows, the constraints are omitted from the last p/2 rows if p+1 is odd, and from the last $\frac{p-1}{2}$ rows if p+1 is even.
5. The systematic procedure in solving the system of equations with the constraints inserted is to solve the first equation, (p+1)st equation, second equation, pth equation, etc., in that order, converging toward the center equation.

6. Upon completing the solution, discard the estimates $\hat{\alpha}_{p+1}$, $\hat{\beta}_{p+1}$, and $\hat{\gamma}_{(p+1)}$, since they are just artifacts and were used in the solution as a matter of convenience. Here, as in any other method of solution, quite arbitrary effect estimates may be obtained, but if these estimates are combined in estimable functions, unique expressions result.

As an example, let us consider the following data arranged in a p-level design, where we choose p=4:

	1	2	3	4
1		30	6	23
2	27	8	21	23
3	6	26	21	
4	24	25		

= Level of B

= Level of C

Level of A

Let us extend the design to a (p+1=5) level design by placing a zero in each corner cell, thereby obtaining the following design as shown in the diagram below:

	1	2	3	4	5
1	0	30	6	23	0
2	27	8	21	23	
3	6	26	21		
4	24	25			
5	0				

= Level of B

= Level of C

Level of A

The row, column, and diagonal totals are

$$\begin{array}{lll}
Y_{1..} = 59 & Y_{.1.} = 57 & Y_{..(1)} = 69 \\
Y_{2..} = 79 & Y_{.2.} = 89 & Y_{..(2)} = 94 \\
Y_{3..} = 53 & Y_{.3.} = 48 & Y_{..(3)} = 20 \\
Y_{4..} = 49 & Y_{.4.} = 46 & Y_{..(4)} = 57 \\
Y_{5..} = \ 0 & Y_{.5.} = \ 0 & Y_{..(5)} = \ 0
\end{array}$$

respectively, and the grand total is $Y_{...} = 240$. Utilizing the results derived early in this section, we obtain the following values for the estimates of the population parameters:

$$\begin{array}{lll}
\hat{\mu} = 16 & & \\
\hat{\alpha}_1 = -4.20 & \hat{\beta}_1 = -4.60 & \hat{\gamma}_{(1)} = -\ 2.20 \\
\hat{\alpha}_2 = -0.45 & \hat{\beta}_2 = \ 2.15 & \hat{\gamma}_{(2)} = \ 2.80 \\
\hat{\alpha}_3 = \ 4.425 & \hat{\beta}_3 = \ 3.025 & \hat{\gamma}_{(3)} = -\ 9.45 \\
\hat{\alpha}_4 = \ 9.425 & \hat{\beta}_4 = \ 9.025 & \hat{\gamma}_{(4)} = \ 16.05 \\
\hat{\alpha}_5 = -9.20 & \hat{\beta}_5 = -9.60 & \hat{\gamma}_{(5)} = -\ 7.20
\end{array} \tag{3.2.8}$$

Let us now discard the results for $\hat{\alpha}_5$, $\hat{\beta}_5$, and $\hat{\gamma}_{(5)}$, and compute various contrasts involving the remaining estimates. Let us consider the usual linear, quadratic, and cubic contrasts in one factor only, say α, which we find to be:

Linear Contrast: $(-3\hat{\alpha}_1 - \hat{\alpha}_2 + \hat{\alpha}_3 + 3\hat{\alpha}_4) = 46.75$

Quadratic Contrast: $(\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\alpha}_3 + \hat{\alpha}_4) = 1.25$

Cubic Contrast: $(-\hat{\alpha}_1 + 3\hat{\alpha}_2 - 3\hat{\alpha}_3 + \hat{\alpha}_4) = - 1.00$.

Later we shall solve the system by another method employing a different set of constraints. The estimates will be widely different. However, the quadratic and cubic contrasts will be the same, since, from the theory of linear estimation, these are estimable functions. However, the results for the linear contrast will be widely different, as the linear contrast is a non-estimable function.

3.3 Modified High-Low Method

The previously described method which introduces constraints into the "high" part of the high-low method, but not into the "low" part results in a set of non-symmetric normal equations. For estimation purposes, this will be of no consequence. However, for the testing of hypotheses we need the elements of the inverse matrix in order to obtain "sums-of-squares" (actually quadratic forms) due to the hypothesis. The inverse of a non-symmetric matrix is unnecessarily complicated. We will thus modify the high-low method in such a way as to obtain a resulting symmetric matrix.

As in the previous method, we introduce the constraint given by equation (3.1.1) into the first row of the normal equations, given by euations (2.2.3), so as to obtain the estimate of the general effect as

$$\hat{\mu} = \frac{Y_{\cdots}}{(p-1)(p+4)/2} \quad . \qquad (3.3.1)$$

After introducing the constraints given byequations (3.1.2) into all of the rows of the normal equations, and transposing the terms involving $\hat{\mu}$ to the right-hand sides of the equations, we obtain the following constrained normal equations in matrix form:

$$\begin{bmatrix} D & \bar{B} & \bar{B} \\ \bar{B} & D & \bar{B} \\ \bar{B} & \bar{B} & D \end{bmatrix} \begin{bmatrix} \underline{\hat{\alpha}}_i \\ \underline{\hat{\beta}}_j \\ \underline{\hat{\gamma}}_{(k)} \end{bmatrix} = \begin{bmatrix} Y_{i..} - n_{i..}\,\hat{\mu} \\ Y_{.j.} - n_{.j.}\,\hat{\mu} \\ Y_{..(k)} - n_{..(k)}\,\hat{\mu} \end{bmatrix}, \qquad (3.3.2)$$

where

$$D = \begin{bmatrix} (p-1) & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & p & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & (p-1) & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & (p-2) & \cdots & 0 & 0 \\ & & \cdot & \cdot & \cdot & & \\ 0 & 0 & 0 & 0 & \cdots & 3 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 2 \end{bmatrix},$$

and

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

The $\underline{\hat{\alpha}}_i$, $\underline{\hat{\beta}}_j$, and $\underline{\hat{\gamma}}_{(k)}$ denote (p x 1) vectors of estimates; and the $n_{i..}$, $n_{.j.}$, and $n_{..(k)}$ are the number of observations in the ith row, jth column, and (k)th diagonal, respectively, i.e., $n_{i..} = n_{.j.} = n_{..(k)} = p-1, p, p-1, p-2, \ldots, 3, 2$ for $i = j = (k) = 1, 2, 3, \ldots, p$.

We have immediately, from the second row of each subset, the estimates of α_2, β_2, and $\gamma_{(2)}$ as

$$\hat{\alpha}_2 = Y_{2..}/p - \hat{\mu} ,$$

$$\hat{\beta}_2 = Y_{.2.}/p - \hat{\mu} , \qquad (3.3.3)$$

and

$$\hat{\gamma}_{(2)} = Y_{..(2)}/p - \hat{\mu} .$$

If we now delete the second row and second column in each submatrix in the normal equations matrix, we obtain a modified normal equations matrix which, in partitioned form, can be represented as

$$K = \begin{bmatrix} D & B & B \\ B & D & B \\ B & B & D \end{bmatrix} . \qquad (3.3.4)$$

In the matrix K, $D[(p-1) \times (p-1)]$ is a diagonal matrix of the form

$$D = \begin{bmatrix} (p-1) & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & (p-1) & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & (p-2) & 0 & \cdots & 0 & 0 \\ & \cdot & & \cdot & \cdot & & \\ 0 & 0 & 0 & 0 & \cdots & 3 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 2 \end{bmatrix} ,$$

and B $[(p-1) \times (p-1)]$ is a matrix of the form

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

Assuming the inverse of the matrix K, say L, to be of the general structure

$$L = \begin{bmatrix} V & G & G \\ G & V & G \\ G & G & V \end{bmatrix} , \qquad (3.3.5)$$

we have

$$KL = \begin{bmatrix} D & B & B \\ B & D & B \\ B & B & D \end{bmatrix} \begin{bmatrix} V & G & G \\ G & V & G \\ G & G & V \end{bmatrix}$$

$$= \begin{bmatrix} DV + 2BG & DG + B(V+G) & DG + B(V+G) \\ DG + B(V+G) & DV + 2BG & DG + B(V+G) \\ DG + B(V+G) & DG + B(V+G) & DV + 2BG \end{bmatrix} .$$

Now, equating KL to the identity matrix $I[(3p-3) \times (3p-3)]$, we obtain the following two equations:

$$DV + 2BG = I \qquad (3.3.6)$$

$$DG + B(V+G) = 0 \quad , \qquad (3.3.7)$$

which are to be solved for the matrices V and G, both of size [(p-1) x (p-1)]. We can solve equation (3.3.7) for G, obtaining

$$DG + BV + BG = 0$$

$$(D+B)\ G = -BV$$

$$G = -(D+B)^{-1}\ BV\ . \qquad (3.3.8)$$

Substituting this value for G into equation (3.3.6), we obtain

$$DV - 2B(D+B)^{-1}\ BV = I$$

which, when solved for the matrix V, gives

$$V = [D-2B(D+B)^{-1}\ B]^{-1}\ . \qquad (3.3.9)$$

With the results for the matrices G and V, given by equations (3.3.8) and (3.3.9), respectively, the matrix L is known. Hence, we have as solutions to the modified normal equations:

$$\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_3 \\ \hat{\alpha}_4 \\ \vdots \\ \hat{\alpha}_{p-1} \\ \hat{\alpha}_p \\ \hat{\beta}_1 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \vdots \\ \hat{\beta}_{p-1} \\ \hat{\beta}_p \\ \hat{\gamma}_{(1)} \\ \hat{\gamma}_{(3)} \\ \hat{\gamma}_{(4)} \\ \vdots \\ \hat{\gamma}_{(p-1)} \\ \hat{\gamma}_{(p)} \end{bmatrix} = L \begin{bmatrix} Y_{1..} - (p-1)\hat{\mu} \\ Y_{3..} - (p-1)\hat{\mu} \\ Y_{4..} - (p-2)\hat{\mu} \\ \vdots \\ Y_{(p-1)..} - 3\hat{\mu} \\ Y_{p..} - 2\hat{\mu} \\ Y_{.1.} - (p-1)\hat{\mu} \\ Y_{.3.} - (p-1)\hat{\mu} \\ Y_{.4.} - (p-2)\hat{\mu} \\ \vdots \\ Y_{.(p-1).} - 3\hat{\mu} \\ Y_{.p.} - 2\hat{\mu} \\ Y_{..(1)} - (p-1)\hat{\mu} \\ Y_{..(3)} - (p-1)\hat{\mu} \\ Y_{..(4)} - (p-2)\hat{\mu} \\ \vdots \\ Y_{..(p-1)} - 3\hat{\mu} \\ Y_{..(p)} - 2\hat{\mu} \end{bmatrix} \quad . \qquad (3.3.10)$$

These estimates, together with the estimates for $\hat{\alpha}_2$, $\hat{\beta}_2$, and $\hat{\gamma}_{(2)}$ given by equations (3.3.3), are the solutions to the normal equations for the Symmetrical Complementation design.

To illustrate the modified high-low method, let us consider a Symmetrical Complementation design where each factor appears at p levels, and let us take p = 4. The normal equations, in matrix form, are given by

$$\begin{bmatrix} 12 & 3 & 4 & 3 & 2 & 3 & 4 & 3 & 2 & 3 & 4 & 3 & 2 \\ 3 & 3 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 4 & 0 & 4 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 0 & 0 & 3 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 2 & 0 & 0 & 0 & 2 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 3 & 0 & 1 & 1 & 1 & 3 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 4 & 1 & 1 & 1 & 1 & 0 & 4 & 0 & 0 & 1 & 1 & 1 & 1 \\ 3 & 1 & 1 & 1 & 0 & 0 & 0 & 3 & 0 & 1 & 1 & 1 & 0 \\ 2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 1 & 0 & 0 \\ 3 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 3 & 0 & 0 & 0 \\ 4 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 4 & 0 & 0 \\ 3 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 3 & 0 \\ 2 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \hat{\alpha}_4 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\gamma}_{(1)} \\ \hat{\gamma}_{(2)} \\ \hat{\gamma}_{(3)} \\ \hat{\gamma}_{(4)} \end{bmatrix} = \begin{bmatrix} Y_{\cdots} \\ Y_{1..} \\ Y_{2..} \\ Y_{3..} \\ Y_{4..} \\ Y_{.1.} \\ Y_{.2.} \\ Y_{.3.} \\ Y_{.4.} \\ Y_{..(1)} \\ Y_{..(2)} \\ Y_{..(3)} \\ Y_{..(4)} \end{bmatrix} . \qquad (3.3.11)$$

By introducing the constraint given by equation (3.1.1) into the first row of the normal equations, we obtain as an estimate of the general effect

$$\hat{\mu} = Y_{\cdots}/12 \quad .$$

Further, by introducing the constraints given by equations (3.1.2) into the second row of each of the row, column, and diagonal subsets, we obtain the estimates of α_2, β_2, and γ_2, given by equations (3.3.3), to be

$$\hat{\alpha}_2 = \frac{Y_{2..}}{4} - \hat{\mu} \quad ,$$

$$\hat{\beta}_2 = \frac{Y_{.2.}}{4} - \hat{\mu} \quad ,$$

and

$$\hat{\gamma}_{(2)} = \frac{Y_{..(2)}}{4} - \hat{\mu} \quad .$$

The modified normal equations, with the constraints inserted, and with the terms involving $\hat{\mu}$ transposed to the right-hand sides, then take the form (consider equations (3.3.2) with $p = 4$, and with the second row in each subset deleted):

$$\begin{bmatrix} D & B & B \\ B & D & B \\ B & B & D \end{bmatrix} \begin{bmatrix} \underline{\hat{\alpha}}_i \\ \underline{\hat{\beta}}_j \\ \underline{\hat{\gamma}}_{(k)} \end{bmatrix} = \begin{bmatrix} Y_{i..} - n_{i..}\,\hat{\mu} \\ Y_{.j.} - n_{.j.}\,\hat{\mu} \\ Y_{..(k)} - n_{..(k)}\,\hat{\mu} \end{bmatrix} ,$$

where

$$D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} ,$$

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

$$\hat{\underline{\alpha}}_i' = [\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4] \quad ,$$

$$\hat{\underline{\beta}}_j' = [\hat{\beta}_1, \hat{\beta}_3, \hat{\beta}_4] \quad ,$$

and

$$\hat{\gamma}_{(k)}' = [\hat{\gamma}_1, \hat{\gamma}_3, \hat{\gamma}_4] \quad .$$

To find the inverse matrix L, we must solve equations (3.3.9) and (3.3.8) in that order. Equation (3.3.9) gives

$$\begin{aligned} V &= [D - 2B(D+B)^{-1} B]^{-1} \\ &= \left\{ \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} - 2 \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & -1 \\ 0 & -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & -1 \end{bmatrix} \right\}^{-1} \\ &= \frac{1}{8} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & -2 \\ 0 & -2 & 0 \end{bmatrix} \quad . \end{aligned}$$

Then the matrix G, given by equation (3.3.8), is

$$\begin{aligned} G &= -(D+B)^{-1} BV \\ &= -\frac{1}{8} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & -1 \\ 0 & -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & -2 \\ 0 & -2 & 0 \end{bmatrix} \\ &= \frac{1}{8} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & -1 \\ 0 & -1 & -3 \end{bmatrix} \quad . \end{aligned}$$

With the results for the matrices G and V, we obtain as the matrix L, given by equation (3.3.5),

$$L = \frac{1}{8}\left[\begin{array}{rrrrrrrrr}
4 & 0 & 0 & 2 & 0 & 0 & 2 & 0 & 0 \\
0 & 2 & -2 & 0 & -1 & -1 & 0 & -1 & -1 \\
0 & -2 & 0 & 0 & -1 & -3 & 0 & -1 & -3 \\[6pt]
2 & 0 & 0 & 4 & 0 & 0 & 2 & 0 & 0 \\
0 & -1 & -1 & 0 & 2 & -2 & 0 & -1 & -1 \\
0 & -1 & -3 & 0 & -2 & 0 & 0 & -1 & -3 \\[6pt]
2 & 0 & 0 & 2 & 0 & 0 & 4 & 0 & 0 \\
0 & -1 & -1 & 0 & -1 & -1 & 0 & 2 & -2 \\
0 & -1 & -3 & 0 & -1 & -3 & 0 & -2 & 0
\end{array}\right] .$$

Since, after obtaining the results for $\hat{\alpha}_2$, $\hat{\beta}_2$, and $\hat{\gamma}_{(2)}$, we had removed the second row and second column in each submatrix of the normal equations matrix, we now put these rows and columns back into the matrix L, with each non-zero element in the respective row and column replaced by the reciprocal of that element. Thus, the complete inverse to the normal equations matrix, which we will denote by C, is given by

$$C = \frac{1}{8}\left[\begin{array}{rrrrrrrrrrrr}
4 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & -2 & 0 & 0 & -1 & -1 & 0 & 0 & -1 & -1 \\
0 & 0 & -2 & 0 & 0 & 0 & -1 & -3 & 0 & 0 & -1 & -3 \\[6pt]
2 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & -1 & 0 & 0 & 2 & -2 & 0 & 0 & -1 & -1 \\
0 & 0 & -1 & -3 & 0 & 0 & -2 & 0 & 0 & 0 & -1 & -3 \\[6pt]
2 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\
0 & 0 & -1 & -1 & 0 & 0 & -1 & -1 & 0 & 0 & 2 & -2 \\
0 & 0 & -1 & -3 & 0 & 0 & -1 & -3 & 0 & 0 & -2 & 0
\end{array}\right] . \qquad (3.3.12)$$

Consider, now, the example in Section 3.2, i.e.

	1	2	3	4	= Level of B
1		30	6	23	
2	27	8	21	23	
3	6	26	21		
4	24	25			
Level of A					

= Level of C

The solutions to the normal equations for this design are given by

$$\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \hat{\alpha}_4 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\gamma}_{(1)} \\ \hat{\gamma}_{(2)} \\ \hat{\gamma}_{(3)} \\ \hat{\gamma}_{(4)} \end{bmatrix} = C \begin{bmatrix} 59-3\hat{\mu} \\ 79-4\hat{\mu} \\ 53-3\hat{\mu} \\ 49-2\hat{\mu} \\ 57-3\hat{\mu} \\ 89-4\hat{\mu} \\ 48-3\hat{\mu} \\ 46-2\hat{\mu} \\ 69-3\hat{\mu} \\ 94-4\hat{\mu} \\ 20-3\hat{\mu} \\ 57-2\hat{\mu} \end{bmatrix} ,$$

where C is given by equation (3.3.12), and $\hat{\mu} = \frac{240}{12} = 20$.

Thus we obtain

$$\begin{array}{lllllllll}
\hat{\mu} & = & 20 & & & & & & \\
\hat{\alpha}_1 & = & 1 & \hat{\beta}_1 & = & .5 & \hat{\gamma}_{(1)} & = & 3.5 \\
\hat{\alpha}_2 & = & -.25 & \hat{\beta}_2 & = & 2.25 & \hat{\gamma}_{(2)} & = & 3.5 \\
\hat{\alpha}_3 & = & -.375 & \hat{\beta}_3 & = & -1.875 & \hat{\gamma}_{(3)} & = & -13.75 \\
\hat{\alpha}_4 & = & -.375 & \hat{\beta}_4 & = & -.875 & \hat{\gamma}_{(4)} & = & 6.75
\end{array} \qquad (3.3.13)$$

Let us consider the linear, quadratic, and cubic contrasts in one factor only, again choosing α, which we find to be

Linear Contrast: $(-3\hat{\alpha}_1 - \hat{\alpha}_2 + \hat{\alpha}_3 + 3\hat{\alpha}_4) = -4.25$

Quadratic Contrast: $(\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\alpha}_3 + \hat{\alpha}_4) = 1.25$

Cubic Contrast: $(-\hat{\alpha}_1 + 3\hat{\alpha}_2 - 3\hat{\alpha}_3 + \hat{\alpha}_4) = -1.00$

Comparing these results with the results given by equations (3.2.8) in section 3.2, we see that the linear contrasts are widely different, but that the higher order contrasts are identical, which serves as an illustration of the already proven fact that the linear contrast is not estimable, while the higher ones are.

3.3.1 Special Case of Three Levels ($p = 3$)

In Section 3.1, we stated that the modified high-low method cannot be used with the constraints (3.1.1) and (3.1.2) if the number of levels is three, since for this case the four constraints can be combined in a linear combination in such a way that an estimable function will result.

In order to find an inverse for this case, we employ the constraint (3.1.1) plus the additional three constraints

$$\hat{\alpha}_3 = \hat{\beta}_3 = \hat{\gamma}_{(3)} = 0. \qquad (3.3.14)$$

Then, as before,

$$\hat{\mu} = Y_{\cdots}/7 \quad .$$

Transposing the terms involving $\hat{\mu}$ to the right-hand side of the equations, and employing the constraints given by equations (3.3.14) we obtain the constrained normal equations, in matrix form, as

$$\begin{vmatrix} 2 & 0 & 0 & 1 & 0 & 1 \\ 0 & 3 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 0 & 0 & 1 \\ 1 & 1 & 0 & 3 & 1 & 1 \\ 0 & 1 & 0 & 1 & 2 & 0 \\ 1 & 1 & 1 & 1 & 0 & 3 \end{vmatrix} \begin{vmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\gamma}_{(1)} \\ \hat{\gamma}_{(2)} \end{vmatrix} = \begin{vmatrix} Y_{1..} - 2\hat{\mu} \\ Y_{2..} - 3\hat{\mu} \\ Y_{.1.} - 2\hat{\mu} \\ Y_{.2.} - 3\hat{\mu} \\ Y_{..(1)} - 2\hat{\mu} \\ Y_{..(2)} - 3\hat{\mu} \end{vmatrix} .$$

Note that the constrained normal equations matrix is symmetric and is of the general structure given by equation (3.3.4); hence we can use equations (3.3.8) and (3.3.9) to find the inverse. Here, we take

$$D = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} ,$$

and

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

Note that the matrices D and B chosen here are not of the same form as those given in equation (3.3.4). However, the results given by equations (3.3.8) and (3.3.9) are valid for finding the inverse to any matrix which can be partitioned and represented as the general structure given by equation (3.3.4).

Solving equations (3.3.8) and (3.3.9), we find

$$V = \frac{1}{18}\begin{bmatrix} 13 & 2 \\ 2 & 10 \end{bmatrix} ,$$

and

$$G = \frac{1}{18}\begin{bmatrix} 1 & -4 \\ -4 & -2 \end{bmatrix} .$$

Thus, we obtain as solutions to the constrained normal equations

$$\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\gamma}_{(1)} \\ \hat{\gamma}_{(2)} \end{bmatrix} = \frac{1}{18}\begin{bmatrix} 13 & 2 & 1 & -4 & 1 & -4 \\ 2 & 10 & -4 & -2 & -4 & -2 \\ 1 & -4 & 13 & 2 & 1 & -4 \\ -4 & -2 & 2 & 10 & -4 & -2 \\ 1 & -4 & 1 & -4 & 13 & 2 \\ -4 & -2 & -4 & -2 & 2 & 10 \end{bmatrix}\begin{bmatrix} Y_{1..} - 2\hat{\mu} \\ Y_{2..} - 3\hat{\mu} \\ Y_{.1.} - 2\hat{\mu} \\ Y_{.2.} - 3\hat{\mu} \\ Y_{..(1)} - 2\hat{\mu} \\ Y_{..(2)} - 3\hat{\mu} \end{bmatrix}$$

with

$$\hat{\mu} = Y_{...}/7 \quad \text{and} \quad \hat{\alpha}_3 = \hat{\beta}_3 = \hat{\gamma}_{(3)} = 0 .$$

3.4 Inverse Matrices for Three, Four, and Five Level Designs (p = 3,4,5)

We now present the inverse matrices for three, four, and five level designs. For the case of three levels, the constraints given by equations (3.1.1) and (3.3.14) have been employed. For the case of four and five levels, the constraints given by equations (3.1.1) and (3.1.2) have been used, and the method of inverting a partitioned matrix, as presented in Section 3.3, has been applied.

3.4.1 Inverse Matrix for Three Level Design (p = 3)

$$\frac{1}{18}\begin{bmatrix} 13 & 2 & 1 & -4 & 1 & -4 \\ 2 & 10 & -4 & -2 & -4 & -2 \\ 1 & -4 & 13 & 2 & 1 & -4 \\ -4 & -2 & 2 & 10 & -4 & -2 \\ 1 & -4 & 1 & -4 & 13 & 2 \\ -4 & -2 & -4 & -2 & 2 & 10 \end{bmatrix} , \qquad (3.4.1)$$

where it is understood that $\hat{\alpha}_3 = \hat{\beta}_3 = \hat{\gamma}_3 = 0$.

3.4.2 Inverse Matrix for Four Level Design (p = 4)

$$\frac{1}{8}\begin{bmatrix} 4 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 & 0 & -1 & -1 & 0 & 0 & -1 & -1 \\ 0 & 0 & -2 & 0 & 0 & 0 & -1 & -3 & 0 & 0 & -1 & -3 \\ 2 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 2 & -2 & 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & -3 & 0 & 0 & -2 & 0 & 0 & 0 & -1 & -3 \\ 2 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & -1 & -1 & 0 & 0 & 2 & -2 \\ 0 & 0 & -1 & -3 & 0 & 0 & -1 & -3 & 0 & 0 & -2 & 0 \end{bmatrix} \qquad (3.4.2)$$

3.4.3 Inverse Matrix for Five Level Design (p = 5)

$$\frac{1}{40}\begin{bmatrix}
12 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\
0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 10 & -2 & -4 & 0 & 0 & -1 & -3 & 0 & 0 & 0 & -1 & -3 & 0 \\
0 & 0 & -2 & 10 & -8 & 0 & 0 & -3 & -1 & -4 & 0 & 0 & -3 & -1 & -4 \\
0 & 0 & -4 & -8 & 8 & 0 & 0 & 0 & -4 & 8 & 0 & 0 & 0 & -4 & 8 \\
4 & 0 & 0 & 0 & 0 & 12 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & -3 & 0 & 0 & 0 & 10 & -2 & -4 & 0 & 0 & -1 & -3 & 0 \\
0 & 0 & -3 & -1 & -4 & 0 & 0 & -2 & 10 & -8 & 0 & 0 & -3 & -1 & -4 \\
0 & 0 & 0 & -4 & 8 & 0 & 0 & -4 & -8 & 8 & 0 & 0 & 0 & -4 & 8 \\
4 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 12 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 \\
0 & 0 & -1 & -3 & 0 & 0 & 0 & -1 & -3 & 0 & 0 & 0 & 10 & -2 & -4 \\
0 & 0 & -3 & -1 & -4 & 0 & 0 & -3 & -1 & -4 & 0 & 0 & -2 & 10 & -8 \\
0 & 0 & 0 & -4 & 8 & 0 & 0 & 0 & -4 & 8 & 0 & 0 & -4 & -8 & 8
\end{bmatrix} \qquad (3.4.3)$$

CHAPTER IV

FORMULATION AND TESTS OF HYPOTHESES

4.1 General Procedures

In this chapter, we shall discuss the testing of estimable hypotheses. A hypothesis H_0 will be called estimable, and hence testable, if there exists a set of linearly independent estimable functions $\underline{\lambda}_1' \underline{\xi}$, $\underline{\lambda}_2' \underline{\xi}$, ..., $\underline{\lambda}_r' \underline{\xi}$ such that H_0 is true if and only if

$$\underline{\lambda}_1' \underline{\xi} = \underline{\lambda}_2' \underline{\xi} = \dots = \underline{\lambda}_r' \underline{\xi} = 0 .$$

It is felt that the more important hypotheses to be investigated are the hypotheses of substitution; which, if accepted, say that applying a factor at one level and another factor at another level does not produce results which are different when the levels of the two factors are exchanged; this exchange of levels holding for a fixed level of the third factor. These hypotheses are described in complete detail in Section 4.2. Thus, it will be possible, by acceptance of the hypothesis of substitution, to test for the equality of the additive effects of any two factors for a fixed level of the third factor. If such hypotheses are accepted for each level of the third factor, i.e., an "overall" hypothesis of substitution being accepted, the design could then be considered as a one-way classification with unequal numbers, in which case the analysis is straight-forward. This seems to be the only way in which it is possible to test in the usual way the effects of all levels of one factor; as the usual linear contrast is a non-estimable function in the Symmetrical Complementation design.

Other hypotheses can be tested, although they may be difficult to interpret. For example, it will be possible to test quadratic contrasts and all contrasts of higher order. These contrasts compare a particular level with several others in the same factor. It was shown in Chapter II that such functions are estimable, and thus testable hypotheses are available for testing purposes.

In summary, the testing of the hypothesis of equality of effects of all levels of one factor is impossible. Thus the user of this design should have a clear concept of this limitation before employing such a design.

4.1.1 Testing Estimable Functions and Sets of Estimable Functions

We will outline in this section the theory necessary for the testing of various hypotheses. Then, in the remainder of this chapter, we will formulate certain hypotheses of interest and show how these hypotheses can be tested.

To test the hypothesis H_o: $\underline{d}'\underline{\xi} = d_o$, where $\underline{d}'\underline{\xi}$ is an estimable function, and d_o is a known constant, we note that the sum of squares due to the hypothesis is given by

$$SS(H_o) = \frac{(\underline{d}'\hat{\underline{\xi}} - d_o)^2}{\underline{d}'(A'A)^{-1}\underline{d}} . \qquad (4.1.1)$$

In this expression, (A'A) is the normal equations matrix made non-singular by constraining.* Therefore, $SS(H_o)/\sigma^2$

---

* It is well known that this expression is unique if $\underline{d}'\underline{\xi}$ is an estimable function, regardless which constraints are used to make (A'A) non-singular (see Graybill (1961), or Kempthorne (1952)).

is distributed as non-central Chi-Square (χ'^2) with p degrees of freedom and non-centrality parameter

$$\lambda = \frac{(\underline{d}'\underline{\xi} - d_o)^2}{2\sigma^2\underline{d}'(A'A)^{-1}\underline{d}} \quad .$$

λ , as used here is the parameter of the Poisson function entering into the expression for the non-central Chi-Square or the non-central $[\frac{m}{n} F]$ distribution. It is the same as that used by Kempthorne (1952).

Thus

$$\frac{SS(Ho)/1}{SS\ E/n_e} = \frac{(\underline{d}'\hat{\underline{\xi}} - d_o)^2}{s^2[\underline{d}'(A'A)^{-1}\ \underline{d}]} \quad , \qquad (4.1.2)$$

where s^2 is the mean-square due to error based on n_e degrees of freedom, is distributed as non-central F(F') with one and n_e degrees of freedom and non-centrality parameter λ , and reduces to central F if H_o: $\underline{d}'\underline{\xi} = d_o$ is true.

In general, suppose we have a set of p vectors, $\underline{d}_i'$ say, which are put into a matrix P so that each element of the vector $P\underline{\xi}$ is estimable. Then, the hypothesis $P\underline{\xi} = \underline{0}$ is testable. Now, a solution to the normal equations is

$$\hat{\underline{\xi}} = (A'A)^{-1}\ A'\underline{y} \quad ,$$

which is non-unique. But

$$P\hat{\underline{\xi}} = P(A'A)^{-1}\ A'\underline{y}$$

is a unique expression. Finally, the sum of squares due to the hypothesis H_o: $P\underline{\xi} = \underline{0}$ is given by

$$SS(H_o) = (P\hat{\underline{\xi}})'[P(A'A)^{-1} P']^{-1} (P\hat{\underline{\xi}}) \quad . \qquad (4.1.3)$$

Also, the sum of squares due to error is given by

$$SSE = \underline{y}'\underline{y} - \hat{\underline{\xi}} \; A'\underline{y} \quad . \qquad (4.1.4)$$

Therefore $SS(H_o)/\sigma^2$ is distributed as non-central Chi-Square (χ'^2) with p degrees of freedom and non-centrality parameter

$$\lambda = \frac{1}{2\sigma^2} (P\underline{\xi})' \; [P(A'A)^{-1}P']^{-1} \; (P\underline{\xi}) \quad .$$

Thus

$$\frac{SS(H_o)/_p}{SS\;E/n_e} \qquad (4.1.5)$$

is distributed as non-central F(F') with p and n_e degrees of freedom and non-centrality parameter λ , and reduces to central F if the hypothesis H_o : $P\underline{\xi} = \underline{0}$ is true.

The remaining sections in this chapter will be arranged in the following standard form:

a) Formulation of hypotheses,

b) Estimation of parameters involved,

c) Sums of squares, and

d) Test statistic.

For up to and including five levels of each factor, the hypotheses and formulas have been completely worked out. All that is required of the user of such a design is to substitute numbers in the expressions developed. If the number of levels exceeds five, numerical inversion will have to be made in each individual case and the theory in this section applied. However, with the method of inversion described in Chapter III, this inversion is reduced to inversion of two [(p-1) x (p-1)] matrices which, incidentally, are of simple form.

4.2 "Hypothesis of Substitution"

In this section, we shall formulate the hypothesis of substitution and develop the sum of squares and the test statistic necessary for testing this hypothesis. If one wishes to test for exchangeability of levels of two factors for a fixed level of the third factor, this is the hypothesis to be used. Due to the inherent symmetry in the design, this hypothesis can be tested for any two of the three factors. We will, however, develop this work for testing for the exchangeability of levels of Factors A and B for a fixed level of Factor C. Recall that Factor C lies along the diagonals of this design. The user of this design can test for the exchangeability of levels of any other two factors by merely substituting into the expressions developed.

For p levels of each factor, we then have p hypotheses of substitution to test. Thus, for example, we may wish to test for exchangeability along the main diagonal of the design. Along this diagonal, Factor C appears at level 2. The hypothesis would be of the form

$$H_o: \alpha_1 + \beta_p = \alpha_2 + \beta_{p-1} = \ldots = \alpha_{p-1} + \beta_2 = \alpha_p + \beta_1 \quad . \tag{4.2.1}$$

This hypothesis, if accepted, says that the composition of Factors A and B, for the second level of Factor C, is immaterial. From the theory on linear estimation (Chapter II), we know that there exists the following set of (p-1) linearly independent estimable functions:

$$\alpha_1 - \alpha_2 - \beta_{p-1} + \beta_p \quad ,$$
$$\alpha_1 - \alpha_3 - \beta_{p-2} + \beta_p \quad ,$$
$$\cdots$$
$$\alpha_1 - \alpha_{p-1} - \beta_2 + \beta_p \quad ,$$

and

$$\alpha_1 - \alpha_p - \beta_1 + \beta_p \quad .$$

Now, the hypothesis H_o is true if and only if these estimable functions are simultaneously equal to zero. Such a hypothesis is conveniently and compactly stated in the form

$$H_o: P\underline{\xi} = \underline{0}$$
$$H_a: P\underline{\xi} \neq \underline{0} \quad ,$$

where P is a predetermined "hypothesis matrix", usually an array of positive and negative integers, and zeros. Actually for the hypothesis of substitution, this array consists only of plus ones, minus ones, and zeros. For a hypothesis to be testable, the elements of the vector $P\underline{\xi}$ must be estimable functions and linearly independent. The test and test statistic for the hypothesis of substitution, which is a testable hypothesis, follows from the theory presented in Section 4.1.1.

The important quantities needed are:

1) Sum of squares due to the hypothesis, which is given by

$$SS(H) = (P\hat{\underline{\xi}})' \; [P(A'A)^{-1} P']^{-1} \; (P\hat{\underline{\xi}}) \quad , \qquad (4.2.2)$$

where $\hat{\underline{\xi}}$ is a solution to the normal equations; and

2) Sum of squares due to error, which is given by

$$SSE = \underline{y}'\underline{y} - \hat{\underline{\xi}} \, A'\underline{y} \quad . \qquad (4.2.3)$$

It should be noted that in the case of one observation per cell, the sum of squares due to error is given by the sum of squares due to lack-of-fit; otherwise, in the case of several observations per cell, the error is the error within cells. This latter case will be discussed in the chapter on extensions.

As an example, let us consider the hypothesis of substitution along the main diagonal of a design where each factor appears at four levels ($p = 4$). In this case, the hypothesis to be tested is

$$H_o: \alpha_1 + \beta_4 = \alpha_2 + \beta_3 = \alpha_3 + \beta_2 = \alpha_4 + \beta_1 \quad .$$

If we consider the functions arising out of this hypothesis, we see that they are given by

$$\alpha_1 - \alpha_2 - \beta_3 + \beta_4 \quad ,$$

$$\alpha_1 - \alpha_3 - \beta_2 + \beta_4 \quad ,$$

and

$$\alpha_1 - \alpha_4 - \beta_1 + \beta_4 \quad .$$

Note that these three functions are estimable; hence H_o is testable. Further note that

$$\alpha_1 - \alpha_2 - \beta_3 + \beta_4 = 0$$

and

$$\alpha_1 - \alpha_3 - \beta_2 + \beta_4 = 0$$

together imply that

$$\alpha_2 - \alpha_3 - \beta_2 + \beta_3 = 0 \quad .$$

Thus, we can conveniently state the hypothesis H_o in the compact form

$$H_o: P\underline{\xi} = \begin{bmatrix} \alpha_1 - \alpha_2 - \beta_3 + \beta_4 \\ \alpha_1 - \alpha_3 - \beta_2 + \beta_4 \\ \alpha_1 - \alpha_4 - \beta_1 + \beta_4 \end{bmatrix} = \underline{0} \quad .$$

The steps and results in the computational procedure for testin the hypothesis are as follows:

1. Calculate $P(A'A)^{-1}P'$. Here P is the (3 x 12) matrix given by

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} ,$$

and $(A'A)^{-1}$ is the inverse matrix stated in Section 3.4.2. Performing the necessary calculations, we obtain

$$P(A'A)^{-1}P' = \frac{1}{4}\begin{bmatrix} 6 & 3 & 3 \\ 3 & 5 & 2 \\ 3 & 2 & 5 \end{bmatrix} \quad .$$

2. Calculate $[P(A'A)^{-1} P']^{-1}$. We thus obtain

$$[P(A'A)^{-1} P']^{-1} = \frac{1}{6} \begin{bmatrix} 7 & -3 & -3 \\ -3 & 7 & -1 \\ -3 & -1 & 7 \end{bmatrix} .$$

3. Calculate the sum of squares due to the hypothesis H_O, which is given by

$$SS(H) = (P\underline{\hat{\xi}})' [P(A'A)^{-1}P']^{-1} (P\underline{\hat{\xi}}) .$$

Here, $$P\underline{\hat{\xi}} = \begin{bmatrix} \hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\beta}_3 + \hat{\beta}_4 \\ \hat{\alpha}_1 - \hat{\alpha}_3 - \hat{\beta}_2 + \hat{\beta}_4 \\ \hat{\alpha}_1 - \hat{\alpha}_4 - \hat{\beta}_1 + \hat{\beta}_4 \end{bmatrix} .$$

The required sum of squares is thus given by

$$\begin{aligned} SS(H) = {} & \frac{7}{6}(\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\beta}_3 + \hat{\beta}_4)^2 + \frac{7}{6}(\hat{\alpha}_1 - \hat{\alpha}_3 - \hat{\beta}_2 + \hat{\beta}_4)^2 \\ & + \frac{7}{6}(\hat{\alpha}_1 - \hat{\alpha}_4 - \hat{\beta}_1 + \hat{\beta}_4)^2 \\ & - (\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\beta}_3 + \hat{\beta}_4)(\hat{\alpha}_1 - \hat{\alpha}_3 - \hat{\beta}_2 + \hat{\beta}_4) \\ & - (\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\beta}_3 + \hat{\beta}_4)(\hat{\alpha}_1 - \hat{\alpha}_4 - \hat{\beta}_1 + \hat{\beta}_4) \\ & - \frac{1}{3}(\hat{\alpha}_1 - \hat{\alpha}_3 - \hat{\beta}_2 + \hat{\beta}_4)(\hat{\alpha}_1 - \hat{\alpha}_4 - \hat{\beta}_1 + \hat{\beta}_4) . \end{aligned}$$

4. Calculate the sum of squares due to error, given by equation (4.2.3). If only one observation is available per cell, this is the contribution due to lack-of-fit.

5. Form the ratio $\frac{SS(H)/3}{SS(E)/2}$, and compare the result with the F-distribution with 3 and 2 degrees of freedom.

In the general case, where there are p levels of each factor, separate hypotheses of substitution can be set up and tested for each level of factor C, giving p hypotheses in all. Further, an "overall" hypothesis of substitution can be tested, to determine whether complete exchangeability of the two factors is possible. Here we must be careful in the formulation of this overall hypothesis, in that there are only 2(p-1) degrees of freedom available for this testing purpose. In general, however, there will be more than 2(p-1) elements in the vector $P\underline{\xi}$ which is needed in the statement of the hypothesis. The elements, however, will not be linearly independent, although they will all be estimable functions. It will thus be possible to find 2(p-1) linearly independent estimable functions, and to proceed with the test of the hypothesis in the usual way.

The sum of squares needed for testing the overall hypothesis of substitution can be found in a more direct way. This hypothesis states, in fact, that the effects due to all levels of factors α and β are simultaneously equal to zero. This being an estimable function, it is possible to set up the full model

$$y_{ij(k)} = \mu + \alpha_i + \beta_j + \gamma_{(k)} + e'_{ij(k)} \quad ,$$

and the reduced model

$$y_{ij(k)} = \tilde{\mu} + \gamma_{(k)} + \tilde{e}'_{ij(k)} \quad .$$

This approach is permissible since the difference between the full model and the reduced model represents a testable hypothesis. The reduced model, in this case, is a one-way classification with unequal cell entries. To obtain the sum of squares due to the overall hypothesis of substitution, we proceed as follows:

1. Obtain the vector of estimates of effects under the original model, and the vector corresponding to the right-hand sides of the normal equations. Call these two vectors $\underline{\hat{\xi}}$ (for estimates) and $\underline{q}$ (for right-hand sides). Then the sum of squares due to treatments is given by

$$SS\ (\text{Treatments}) = \underline{\hat{\xi}}'\,\underline{q}$$

2. Obtain the sum of squares due to treatments under the reduced model. Since the reduced model is a one-way classification with unequal cell entries, the sum of squares due to treatments is the sum of squares between groups, i.e.,

$$SS\ (\text{Between}) = \sum_{(k)} \frac{Y^2_{..(k)}}{n_{..(k)}} - \frac{Y^2_{...}}{n_{...}} ,$$

where $n_{..(k)} = (p-1), p, (p-1), (p-2), \ldots, 3, 2$ for $k = 1, 2, 3, \ldots, p$.

3. Obtain the sum of squares due to the overall hypothesis of substitution as

$$SS\ (\text{Overall Hypothesis}) = SS\ (\text{Treatments}) - SS\ (\text{Between}). \qquad (4.2.4)$$

This sum of squares is the sum of squares due to α and β adjusted for γ. The test statistic is $\frac{SS(H)/_{2(p-1)}}{SS(E)/_{n_e}}$, and this ratio is compared with the F-distribution with $2(p-1)$ and n_e degrees of freedom.

We will now formulate the various hypotheses of substitution, showing the estimable functions that arise out of a particular hypothesis. This will be done for the general case of p levels.

1) For the case where Factor C appears at the lowest level, we have the hypothesis

$$H_o: \alpha_2 + \beta_p = \alpha_3 + \beta_{p-1} = \alpha_4 + \beta_{p-2} = \cdots$$

$$= \alpha_{p-2} + \beta_4 = \alpha_{p-1} + \beta_3 = \alpha_p + \beta_2 \quad .$$

This hypothesis can be written in the compact form

$$H_o: \quad \underline{P\xi} = \begin{bmatrix} \alpha_2 - \alpha_3 - \beta_{p-1} + \beta_p \\ \alpha_2 - \alpha_4 - \beta_{p-2} + \beta_p \\ \cdots \\ \alpha_2 - \alpha_{p-2} - \beta_4 + \beta_p \\ \alpha_2 - \alpha_{p-1} - \beta_3 + \beta_p \\ \alpha_2 - \alpha_p - \beta_2 + \beta_p \end{bmatrix} = \underline{0} \quad ,$$

where each element of the vector $\underline{P\xi}$ is an estimable function. There are $(p-2)$ degrees of freedom associated with this hypothesis.

2) For the case where Factor C appears at the second level, i.e., the main diagonal of the design, we have the hypothesis

$$H_o: \alpha_1 + \beta_p = \alpha_2 + \beta_{p-1} = \alpha_3 + \beta_{p-2} = \cdots$$

$$= \alpha_{p-2} + \beta_3 = \alpha_{p-1} + \beta_2 = \alpha_p + \beta_1 \quad .$$

This hypothesis can be written in the compact form

$$H_o: \quad \underline{P\xi} = \begin{bmatrix} \alpha_1 - \alpha_2 - \beta_{p-1} + \beta_p \\ \alpha_1 - \alpha_3 - \beta_{p-1} + \beta_p \\ \cdots \\ \alpha_1 - \alpha_{p-2} - \beta_3 + \beta_p \\ \alpha_1 - \alpha_{p-1} - \beta_2 + \beta_p \\ \alpha_1 - \alpha_p - \beta_1 + \beta_p \end{bmatrix} = \underline{0} \ ,$$

where each element of the vector $\underline{P\xi}$ is an estimable function. This hypothesis carries with it (p-1) degrees of freedom.

3) For the case where Factor C appears at the third level, we have the hypothesis

$$H_o: \alpha_1 + \beta_{p-1} = \alpha_2 + \beta_{p-2} = \alpha_3 + \beta_{p-3} = \cdots$$
$$= \alpha_{p-3} + \beta_3 = \alpha_{p-2} + \beta_2 = \alpha_{p-1} + \beta_1 \quad .$$

This hypothesis can be written in the compact form

$$H_o: \quad \underline{P\xi} = \begin{bmatrix} \alpha_1 - \alpha_2 - \beta_{p-2} + \beta_{p-1} \\ \alpha_1 - \alpha_3 - \beta_{p-3} + \beta_{p-1} \\ \cdots \\ \alpha_1 - \alpha_{p-3} - \beta_3 + \beta_{p-1} \\ \alpha_1 - \alpha_{p-2} - \beta_2 + \beta_{p-1} \\ \alpha_1 - \alpha_{p-1} - \beta_1 + \beta_{p-1} \end{bmatrix} = \underline{0} \ .$$

Again, each element in the vector $\underline{P\xi}$ is an estimable function. There are (p-2) degrees of freedom associated with this hypothesis.

Continuing these formulations for the other levels of Factor C we finally obtain:

p-1) For the case where Factor C appears at the (p-1)st level, we have the hypothesis

$$H_o: \alpha_1 + \beta_3 = \alpha_2 + \beta_2 = \alpha_3 + \beta_1 \quad .$$

This hypothesis can be written in the compact form

$$H_o: \quad \underline{P\xi} = \begin{bmatrix} \alpha_1 - \alpha_2 - \beta_2 + \beta_3 \\ \alpha_1 - \alpha_3 - \beta_1 + \beta_3 \end{bmatrix} = \underline{0} \quad ,$$

where each element of the vector $\underline{P\xi}$ is an estimable function. There are two degrees of freedom associated with this hypothesis.

p) For the case where Factor C appears at the highest level, we have the hypothesis

$$H_o: \alpha_1 + \beta_2 = \alpha_2 + \beta_1 \quad ,$$

which can be written

$$H_o: \underline{P\xi} = [\alpha_1 - \alpha_2 - \beta_1 + \beta_2] = 0 \quad ,$$

where the element of $\underline{P\xi}$ is an estimable function. There is one degree of freedom associated with this hypothesis.

We have attempted to show the formulation of the various hypotheses of substitution for the general case where each factor appears at p levels. We will now formulate these hypotheses for designs where each factor appears at three, four, and five levels, respectively; and we will show the estimable functions that make such hypotheses testable, and the corresponding sums of squares and test statistics that are necessary for testing purposes.

4.2.1 Three Level Case

For the case where each factor appears at three levels, we first note that in calculating sums of squares, we need the inverse of the constrained normal equations matrix. This has been worked out and presented in Section 3.4.1. In the presentation to follow, we will refer to this matrix as $(A'A)^{-1}$, and we repeat the result here:

$$(A'A)^{-1} = \frac{1}{18}\begin{bmatrix} 13 & 2 & 1 & -4 & 1 & -4 \\ 2 & 10 & -4 & -2 & -4 & -2 \\ 1 & -4 & 13 & 2 & 1 & -4 \\ -4 & -2 & 2 & 10 & -4 & -2 \\ 1 & -4 & 1 & -4 & 13 & 2 \\ -4 & -2 & -4 & -2 & 2 & 10 \end{bmatrix} .$$

1) For the case where Factor C appears at the lowest level, the hypothesis is

$$H_o: \alpha_2 + \beta_3 = \alpha_3 + \beta_2 \quad , \qquad (4.2.5)$$

which, when written in compact form, is

$$H_o: P\underline{\xi} = [\alpha_2 - \alpha_3 - \beta_2 + \beta_3] = 0 ;$$

or, in view of the reparameterization of the model in the three level case, is

$$H_o: P\underline{\xi} = [\alpha_2 - \beta_2] = 0 \quad . \qquad (4.2.6)$$

P is thus the row vector [0 1 0 -1 0 0]. Then $P(A'A)^{-1}P' = \frac{4}{3}$, and $[P(A'A)^{-1}P']^{-1} = \frac{3}{4}$. Thus the sum of squares due to hypothesis is given by

$$SS(H) = (P\hat{\underline{\xi}})'[P(A'A)^{-1} P']^{-1} (P\hat{\underline{\xi}})$$

$$= \frac{(\hat{\alpha}_2 - \hat{\beta}_2)^2}{\frac{4}{3}} \quad . \qquad (4.2.7)$$

The test statistic is $\frac{SS(H)}{SS(E)/n_e}$, and this is compared with the F-distribution with 1 and n_e degrees of freedom. Note that in the three-level case, the error sum of squares cannot be taken as the contribution due to lack-of-fit, as there are no degrees of freedom left for lack-of-fit.

2) For the case where Factor C appears at the second level, i.e., the main diagonal of the design, we have the hypothesis

$$H_o: \alpha_1 + \beta_3 = \alpha_2 + \beta_2 = \alpha_3 + \beta_1 \quad , \qquad (4.2.8)$$

which, when written in compact form, is

$$H_o: P\underline{\xi} = \begin{bmatrix} \alpha_1 - \alpha_2 - \beta_2 + \beta_3 \\ \alpha_1 - \alpha_3 - \beta_1 + \beta_3 \end{bmatrix} = \underline{0}; \qquad (4.2.9)$$

or, in view of the reparameterization, is

$$H_o: P\underline{\xi} = \begin{bmatrix} \alpha_1 - \alpha_2 - \beta_2 \\ \alpha_1 \qquad - \beta_1 \end{bmatrix} = \underline{0}.$$

P, here, is the (2 x 6) matrix

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

Then

$$P(A'A)^{-1} P' = \frac{1}{6} \begin{bmatrix} 11 & 4 \\ 4 & 8 \end{bmatrix} ,$$

and

$$[P(A'A)^{-1} P']^{-1} = \frac{1}{12} \begin{bmatrix} 8 & -4 \\ -4 & 11 \end{bmatrix} .$$

The sum of squares due to hypothesis is given by

$$SS(H) = (P\hat{\underline{\xi}})' \, [P(A'A)^{-1} P']^{-1} \, (P\hat{\underline{\xi}})$$

$$= \frac{1}{12} [8(\hat{\alpha}_1-\hat{\alpha}_2-\hat{\beta}_2)^2 + 11(\hat{\alpha}_1-\hat{\beta}_1)^2 - 8(\hat{\alpha}_1-\hat{\alpha}_2-\hat{\beta}_2)(\hat{\alpha}_1-\hat{\beta}_1)]. \tag{4.2.10}$$

The test statistic is $\frac{SS(H)/2}{SS(E)/n_e}$, and this ratio is compared with the F-distribution with 2 and n_e degrees of freedom.

3) For the case where Factor C appears at the highest level, we have the hypothesis

$$H_o: \alpha_1 + \beta_2 = \alpha_2 + \beta_1 \quad , \tag{4.2.11}$$

which, when written in compact form, is

$$H_o: P\underline{\xi} = [\alpha_1 - \alpha_2 - \beta_1 + \beta_2] = 0 \; . \tag{4.2.12}$$

P is merely the row vector [1 -1 -1 1 0 0] .

Then $P(A'A)^{-1} P' = \frac{4}{3}$, and $[P(A'A)^{-1} P']^{-1} = \frac{3}{4}$. Thus the sum of squares due to hypothesis is given by

$$SS(H) = (P\hat{\xi})' \, [P(A'A)^{-1} P']^{-1} \, (P\hat{\underline{\xi}})$$

$$= \frac{(\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\beta}_1 + \hat{\beta}_2)^2}{\frac{4}{3}} \quad , \tag{4.2.13}$$

and the ratio $\frac{SS(H)}{SS(E)/n_e}$ is compared with the F-distribution with 1 and n_e degrees of freedom.

4.2.2 Four Level Case

For the case where each factor appears at four levels, the inverse of the constrained normal equations matrix is given in Section 3.4.2. We repeat the result below:

$$(A'A)^{-1} = \frac{1}{8}\left[\begin{array}{cccccccccccc}
4 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & -2 & 0 & 0 & -1 & -1 & 0 & 0 & -1 & -1 \\
0 & 0 & -2 & 0 & 0 & 0 & -1 & -3 & 0 & 0 & -1 & -3 \\
2 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & -1 & 0 & 0 & 2 & -2 & 0 & 0 & -1 & -1 \\
0 & 0 & -1 & -3 & 0 & 0 & -2 & 0 & 0 & 0 & -1 & -3 \\
2 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\
0 & 0 & -1 & -1 & 0 & 0 & -1 & -1 & 0 & 0 & 2 & -2 \\
0 & 0 & -1 & -3 & 0 & 0 & -1 & -3 & 0 & 0 & -2 & 0
\end{array}\right] .$$

1) For the case where Factor C appears at the lowest level, the hypothesis is

$$H_0: \alpha_2 + \beta_4 = \alpha_3 + \beta_3 = \alpha_4 + \beta_2 \quad , \qquad (4.2.14)$$

which, when written in compact form, is

$$H_0: \mathbf{P}\underline{\xi} = \begin{bmatrix} \alpha_2 - \alpha_3 - \beta_3 + \beta_4 \\ \alpha_2 - \alpha_4 - \beta_2 + \beta_4 \end{bmatrix} = \underline{0} \quad . \qquad (4.2.15)$$

P is given by the (2 x 12) matrix

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

Then

$$P(A'A)^{-1} P' = \frac{5}{8}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} ,$$

and

$$[P(A'A)^{-1} P']^{-1} = \frac{8}{15}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} .$$

The sum of squares due to hypothesis is given by

$$SS(H) = (P\hat{\underline{\xi}})' \, [P(A'A)^{-1} P']^{-1} \, (P\hat{\underline{\xi}})$$

$$= \frac{16}{15} (\hat{\alpha}_2 - \hat{\alpha}_3 - \hat{\beta}_3 + \hat{\beta}_4)^2 + \frac{16}{15} (\hat{\alpha}_2 - \hat{\alpha}_4 - \hat{\beta}_2 + \hat{\beta}_4)^2$$

$$- \frac{16}{15} (\hat{\alpha}_2 - \hat{\alpha}_3 - \hat{\beta}_3 + \hat{\beta}_4)(\hat{\alpha}_2 - \hat{\alpha}_4 - \hat{\beta}_2 + \hat{\beta}_4) \quad .$$

(4.2.16)

The ratio $\dfrac{SS(H)/2}{SS(E)/n_e}$ is compared with the F-distribution with 2 and n_e degrees of freedom. If the contribution due to lack-of-fit is used as the error term, $n_e = 2$.

2) For the case where Factor C appears at the second level, i.e., the main diagonal of the design, the hypothesis is

$$H_o: \alpha_1 + \beta_4 = \alpha_2 + \beta_3 = \alpha_3 + \beta_2 = \alpha_3 + \beta_1 \ ,$$

(4.2.17)

which can then be written as

$$H_o: P\underline{\xi} = \begin{bmatrix} \alpha_1 - \alpha_2 - \beta_3 + \beta_4 \\ \alpha_1 - \alpha_3 - \beta_2 + \beta_4 \\ \alpha_1 - \alpha_4 - \beta_1 + \beta_4 \end{bmatrix} = \underline{0} \ .$$

(4.2.18)

P is given by the (3 x 12) matrix

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

Then

$$P(A'A)^{-1}\ P' = \frac{1}{4}\begin{bmatrix} 6 & 3 & 3 \\ 3 & 5 & 2 \\ 3 & 2 & 5 \end{bmatrix} ,$$

and

$$[P(A'A)^{-1}\ P']^{-1} = \frac{1}{6}\begin{bmatrix} 7 & -3 & -3 \\ -3 & 7 & -1 \\ -3 & -1 & 7 \end{bmatrix} .$$

The sum of squares due to hypothesis is given by

$$SS(H) = (P\hat{\underline{\xi}})'\ [P(A'A)^{-1}\ P']^{-1}\ (P\hat{\underline{\xi}})$$

$$= \frac{7}{6}(\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\beta}_3 + \hat{\beta}_4)^2 + \frac{7}{6}(\hat{\alpha}_1 - \hat{\alpha}_3 - \hat{\beta}_2 + \hat{\beta}_4)^2$$

$$+ \frac{7}{6}(\hat{\alpha}_1 - \hat{\alpha}_4 - \hat{\beta}_1 + \hat{\beta}_4)^2$$

$$-(\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\beta}_3 + \hat{\beta}_4)(\hat{\alpha}_1 - \hat{\alpha}_3 - \hat{\beta}_2 + \hat{\beta}_4)$$

$$-(\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\beta}_3 + \hat{\beta}_4)(\hat{\alpha}_1 - \hat{\alpha}_4 - \hat{\beta}_1 + \hat{\beta}_4)$$

$$- \frac{1}{3}(\hat{\alpha}_1 - \hat{\alpha}_3 - \hat{\beta}_2 + \hat{\beta}_4)(\hat{\alpha}_1 - \hat{\alpha}_4 - \hat{\beta}_1 + \hat{\beta}_4) , \tag{4.2.19}$$

and the ratio $\frac{SS(H)/3}{SS(E)/n_e}$ is compared with the F-distribution with 3 and n_e degrees of freedom.

3) For the case where Factor C appears at the third level, the hypothesis is

$$H_0:\ \alpha_1 + \beta_3 = \alpha_2 + \beta_2 = \alpha_3 + \beta_1 , \tag{4.2.20}$$

which can then be written as

$$H_o: P\underline{\xi} = \begin{bmatrix} \alpha_1 - \alpha_2 - \beta_2 + \beta_3 \\ \alpha_1 - \alpha_3 - \beta_1 + \beta_3 \end{bmatrix} = \underline{0} \; .$$

(4.2.21)

P is given by the (2 x 12) matrix

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

Then

$$P(A'A)^{-1} P' = \frac{5}{8} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \; ,$$

and

$$[P(A'A)^{-1} P']^{-1} = \frac{8}{15} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \; .$$

The sum of squares due to hypothesis is given by

$$SS(H) = (P\hat{\underline{\xi}})' \; [P(A'A)^{-1} P']^{-1} \; (P\hat{\underline{\xi}})$$

$$= \frac{16}{15} (\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\beta}_2 + \hat{\beta}_3)^2 + \frac{16}{15} (\hat{\alpha}_1 - \hat{\alpha}_3 - \hat{\beta}_1 + \hat{\beta}_3)^2$$

$$-\frac{16}{15} (\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\beta}_2 + \hat{\beta}_3)(\hat{\alpha}_1 - \hat{\alpha}_3 - \hat{\beta}_1 + \hat{\beta}_3) \; ,$$

(4.2.22)

and the ratio $\dfrac{SS(H)/2}{SS(E)/n_e}$ is then compared with the F-distribution with 2 and n_e degrees of freedom.

4) For the case where Factor C appears at the highest level, the hypothesis is

$$H_o: \alpha_1 + \beta_2 = \alpha_2 + \beta_1 \quad , \qquad (4.2.23)$$

which can be written as

$$H_o: P\underline{\xi} = [\alpha_1 - \alpha_2 - \beta_1 + \beta_2] = 0 \quad . \qquad (4.2.24)$$

P is given by the row vector

$$[\ 1\ -1\ 0\ 0\quad -1\quad 1\ 0\ 0\quad 0\ 0\ 0\ 0\ ] \quad .$$

Then

$$P(A'A)^{-1} P' = 1.00 \quad ,$$

and

$$[P(A'A)^{-1} P']^{-1} = 1.00 \quad .$$

The sum of squares due to hypothesis is given by

$$\begin{aligned} SS(H) &= (P\hat{\underline{\xi}})' [P(A'A)^{-1} P']^{-1} (P\hat{\underline{\xi}}) \\ &= (\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\beta}_1 + \hat{\beta}_2)^2 \quad , \qquad (4.2.25) \end{aligned}$$

and the ratio $\frac{SS(H)}{SS(E)/n_e}$ is compared with the F-distribution with 1 and n_e degrees of freedom.

4.2.3 Five Level Case

For the case where each factor appears at five levels, the inverse of the constrained normal equations matrix is given in Section 3.4.3. We repeat the result below:

$$(A'A)^{-1} = \frac{1}{40}\begin{bmatrix}
12 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\
0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 10 & -2 & -4 & 0 & 0 & -1 & -3 & 0 & 0 & 0 & -1 & -3 & 0 \\
0 & 0 & -2 & 10 & -8 & 0 & 0 & -3 & -1 & -4 & 0 & 0 & -3 & -1 & -4 \\
0 & 0 & -4 & -8 & 8 & 0 & 0 & 0 & -4 & 8 & 0 & 0 & 0 & -4 & 8 \\
4 & 0 & 0 & 0 & 0 & 12 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & -3 & 0 & 0 & 0 & 10 & -2 & -4 & 0 & 0 & -1 & -3 & 0 \\
0 & 0 & -3 & -1 & -4 & 0 & 0 & -2 & 10 & -8 & 0 & 0 & -3 & -1 & -4 \\
0 & 0 & 0 & -4 & 8 & 0 & 0 & -4 & -8 & 8 & 0 & 0 & 0 & -4 & 8 \\
4 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 12 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 \\
0 & 0 & -1 & -3 & 0 & 0 & 0 & -1 & -3 & 0 & 0 & 0 & 10 & -2 & -4 \\
0 & 0 & -3 & -1 & -4 & 0 & 0 & -3 & -1 & -4 & 0 & 0 & -2 & 10 & -8 \\
0 & 0 & 0 & -4 & 8 & 0 & 0 & 0 & -4 & 8 & 0 & 0 & -4 & -8 & 8
\end{bmatrix}.$$

1) The for case where Factor C appears at the lowest level, the hypothesis is

$$H_o: \alpha_2 + \beta_5 = \alpha_3 + \beta_4 = \alpha_4 + \beta_3 = \alpha_5 + \beta_2 \quad , \tag{4.2.26}$$

which, when written in compact form, is

$$H_o: \quad P\underline{\xi} = \begin{bmatrix} \alpha_2 - \alpha_3 - \beta_4 + \beta_5 \\ \alpha_2 - \alpha_4 - \beta_3 + \beta_5 \\ \alpha_2 - \alpha_5 - \beta_2 + \beta_5 \end{bmatrix} = \underline{0} \, . \tag{4.2.27}$$

P is given by the (3 x 15) matrix

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

Then

$$P(A'A)^{-1} P' = \begin{bmatrix} 1.15 & 0.65 & 0.60 \\ 0.65 & 1.15 & 0.60 \\ 0.60 & 0.60 & 1.20 \end{bmatrix} ,$$

and

$$[P(A'A)^{-1} P']^{-1} = \frac{1}{0.12} \begin{bmatrix} 0.17 & -0.07 & -0.05 \\ -0.07 & 0.17 & -0.05 \\ -0.05 & -0.05 & 0.15 \end{bmatrix} .$$

The sum of squares due to hypothesis is given by

$$\begin{aligned} SS(H) &= (P\hat{\underline{\xi}})' \, [P(A'A)^{-1} P']^{-1} \, (P\hat{\underline{\xi}}) \\ &= \frac{1}{0.12} [(0.17)(\hat{\alpha}_2 - \hat{\alpha}_3 - \hat{\beta}_4 + \hat{\beta}_5)^2 \\ &\quad +(0.17)(\hat{\alpha}_2 - \hat{\alpha}_4 - \hat{\beta}_3 + \hat{\beta}_5)^2 \\ &\quad +(0.15)(\hat{\alpha}_2 - \hat{\alpha}_5 - \hat{\beta}_2 + \hat{\beta}_5)^2 \\ &\quad -(0.14)(\hat{\alpha}_2 - \hat{\alpha}_3 - \hat{\beta}_4 + \hat{\beta}_5)(\hat{\alpha}_2 - \hat{\alpha}_4 - \hat{\beta}_3 + \hat{\beta}_5) \\ &\quad -(0.10)(\hat{\alpha}_2 - \hat{\alpha}_3 - \hat{\beta}_4 + \hat{\beta}_5)(\hat{\alpha}_2 - \hat{\alpha}_5 - \hat{\beta}_2 + \hat{\beta}_5) \\ &\quad -(0.10)(\hat{\alpha}_2 - \hat{\alpha}_4 - \hat{\beta}_3 + \hat{\beta}_5)(\hat{\alpha}_2 - \hat{\alpha}_5 - \hat{\beta}_2 + \hat{\beta}_5)]. \end{aligned} \qquad (4.2.28)$$

The ratio $\frac{SS(H)/3}{SS(E)/n_e}$ is compared with the F-distribution with 3 and n_e degrees of freedom. If the contribution due to lack-of-fit is used as the error term, $n_e = 5$.

2) For the case where Factor C appears at the second level, i.e., the main diagonal of the design, the hypothesis is

$$H_o: \alpha_1 + \beta_5 = \alpha_2 + \beta_4 = \alpha_3 + \beta_3 = \alpha_4 + \beta_2 = \alpha_5 + \beta_1 , \qquad (4.2.29)$$

which can be written as

$$H_o: \quad P\underline{\xi} = \begin{bmatrix} \alpha_1 - \alpha_2 - \beta_4 + \beta_5 \\ \alpha_1 - \alpha_3 - \beta_3 + \beta_5 \\ \alpha_1 - \alpha_4 - \beta_2 + \beta_5 \\ \alpha_1 - \alpha_5 - \beta_1 + \beta_5 \end{bmatrix} = \underline{0} \quad . \tag{4.2.30}$$

P is given by the (4 x 15) matrix

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

Then

$$P(A'A)^{-1} P' = \begin{bmatrix} 1.350 & 0.675 & 0.775 & 0.700 \\ 0.675 & 1.150 & 0.575 & 0.600 \\ 0.775 & 0.575 & 1.150 & 0.500 \\ 0.700 & 0.600 & 0.500 & 1.200 \end{bmatrix} ,$$

and

$$[P(A'A)^{-1} P']^{-1} = \begin{bmatrix} 1.5159 & -0.3250 & -0.6659 & -0.4443 \\ -0.3250 & 1.4250 & -0.3250 & -0.3875 \\ -0.6659 & -0.3250 & 1.5159 & -0.0807 \\ -0.4443 & -0.3875 & -0.0807 & 1.3199 \end{bmatrix} .$$

The sum of squares due to hypothesis is given by

$$SS(H) = (P\hat{\underline{\xi}})' [P(A'A)^{-1} P']^{-1} (P\hat{\underline{\xi}})$$

$$= [1.5159(\hat{\alpha}_1-\hat{\alpha}_2-\hat{\beta}_4+\hat{\beta}_5)^2 + 1.4250(\hat{\alpha}_1-\hat{\alpha}_3-\hat{\beta}_3+\hat{\beta}_5)^2$$

$$+1.5159(\hat{\alpha}_1-\hat{\alpha}_4-\hat{\beta}_2+\hat{\beta}_5)^2 + 1.3199(\hat{\alpha}_1-\hat{\alpha}_5-\hat{\beta}_1+\hat{\beta}_5)^2$$

$$-0.6500(\hat{\alpha}_1-\hat{\alpha}_2-\hat{\beta}_4+\hat{\beta}_5)(\hat{\alpha}_1-\hat{\alpha}_3-\hat{\beta}_3+\hat{\beta}_5)$$

$$-1.3318(\hat{\alpha}_1-\hat{\alpha}_2-\hat{\beta}_4+\hat{\beta}_5)(\hat{\alpha}_1-\hat{\alpha}_4-\hat{\beta}_2+\hat{\beta}_5)$$

$$-0.8886(\hat{\alpha}_1-\hat{\alpha}_2-\hat{\beta}_4+\hat{\beta}_5)(\hat{\alpha}_1-\hat{\alpha}_5-\hat{\beta}_1+\hat{\beta}_5)$$

$$-0.6500(\hat{\alpha}_1-\hat{\alpha}_3-\hat{\beta}_3+\hat{\beta}_5)(\hat{\alpha}_1-\hat{\alpha}_4-\hat{\beta}_2+\hat{\beta}_5)$$

$$-0.7750(\hat{\alpha}_1-\hat{\alpha}_3-\hat{\beta}_3+\hat{\beta}_5)(\hat{\alpha}_1-\hat{\alpha}_5-\hat{\beta}_1+\hat{\beta}_5)$$

$$-0.1614(\hat{\alpha}_1-\hat{\alpha}_4-\hat{\beta}_2+\hat{\beta}_5)(\hat{\alpha}_1-\hat{\alpha}_5-\hat{\beta}_1+\hat{\beta}_5)] \quad .$$

(4.2.31)

The test statistic is $\dfrac{SS(H)/4}{SS(E)/n_e}$, and this ratio is compared with the F-distribution with 4 and n_e degrees of freedom.

3) For the case where Factor C appears at the third level, the hypothesis is

$$H_o: \alpha_1 + \beta_4 = \alpha_2 + \beta_3 = \alpha_3 + \beta_2 = \alpha_4 + \beta_1 \; , \qquad (4.2.32)$$

which can then be written as

$$H_o: P\underline{\xi} = \begin{bmatrix} \alpha_1 - \alpha_2 - \beta_3 + \beta_4 \\ \alpha_1 - \alpha_3 - \beta_2 + \beta_4 \\ \alpha_1 - \alpha_4 - \beta_1 + \beta_4 \end{bmatrix} = \underline{0} \; . \qquad (4.2.33)$$

P is given by the (3 x 15) matrix

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

Then

$$P(A'A)^{-1} P' = \begin{bmatrix} 1.10 & 0.65 & 0.45 \\ 0.65 & 1.15 & 0.50 \\ 0.45 & 0.50 & 0.95 \end{bmatrix} ,$$

and

$$[P(A'A)^{-1} P]^{-1} = \frac{1}{0.585} \begin{bmatrix} 0.8425 & -0.3925 & -0.1925 \\ -0.3925 & 0.8425 & -0.2575 \\ -0.1925 & -0.2575 & 0.8425 \end{bmatrix} .$$

The sum of squares due to hypothesis is given by

$$\begin{aligned} SS(H) &= (P\underline{\hat{\xi}})' \, [P(A'A)^{-1} P']^{-1} \, (P\underline{\hat{\xi}}) \\ &= \frac{1}{0.585} \, [0.8425(\hat{\alpha}_1-\hat{\alpha}_2-\hat{\beta}_3+\hat{\beta}_4)^2 + 0.8425(\hat{\alpha}_1-\hat{\alpha}_3-\hat{\beta}_2+\hat{\beta}_4)^2 \\ &\quad +0.8425(\hat{\alpha}_1-\hat{\alpha}_4-\hat{\beta}_1+\hat{\beta}_4)^2 \\ &\quad -0.7850(\hat{\alpha}_1-\hat{\alpha}_2-\hat{\beta}_3+\hat{\beta}_4)(\hat{\alpha}_1-\hat{\alpha}_3-\hat{\beta}_2+\hat{\beta}_4) \\ &\quad -0.3850(\hat{\alpha}_1-\hat{\alpha}_2-\hat{\beta}_3+\hat{\beta}_4)(\hat{\alpha}_1-\hat{\alpha}_4-\hat{\beta}_1+\hat{\beta}_4) \\ &\quad -0.5150(\hat{\alpha}_1-\hat{\alpha}_3-\hat{\beta}_2+\hat{\beta}_4)(\hat{\alpha}_1-\hat{\alpha}_4-\hat{\beta}_1+\hat{\beta}_4) \;] \; . \end{aligned} \tag{4.2.34}$$

The ratio $\dfrac{SS(H)/3}{SS(E)/n_e}$ is compared with the F-distribution with 3 and n_e degrees of freedom.

4) For the case where Factor C appears at the third level, the hypothesis is

$$H_o: \; \alpha_1 + \beta_3 = \alpha_2 + \beta_2 = \alpha_3 + \beta_1 \quad , \tag{4.2.35}$$

which can be written as

$$H_o: \quad P\underline{\xi} = \begin{bmatrix} \alpha_1 - \alpha_2 - \beta_2 + \beta_3 \\ \alpha_1 - \alpha_3 - \beta_1 + \beta_3 \end{bmatrix} = \underline{0} \, . \qquad (4.2.36)$$

P is given by the (2 x 15) matrix

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

Then

$$P(A'A)^{-1} P' = 0.475 \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} ,$$

and

$$[\, P(A'A)^{-1} P' \,]^{-1} = \frac{1}{1.425} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} .$$

The sum of squares due to hypothesis is given by

$$\begin{aligned} SS(H) &= (P\hat{\underline{\xi}}) \, [\, P(A'A)^{-1} P']^{-1} \, (P\hat{\underline{\xi}}) \\ &= \frac{1}{1.425} \, [2(\hat{\alpha}_1-\hat{\alpha}_2-\hat{\beta}_2+\hat{\beta}_3)^2 + 2(\hat{\alpha}_1-\hat{\alpha}_3-\hat{\beta}_1+\hat{\beta}_3)^2 \\ &\qquad -2(\hat{\alpha}_1-\hat{\alpha}_2-\hat{\beta}_2+\hat{\beta}_3)(\hat{\alpha}_1-\hat{\alpha}_3-\hat{\beta}_1+\hat{\beta}_3) \,] \quad , \end{aligned}$$

(4.2.37)

and the ratio $\dfrac{SS(H)/_2}{SS(E)/_{n_e}}$ is compared with the F-distribution with 2 and n_e degrees of freedom.

5) For the case where Factor C appears at the highest level, the hypothesis is

$$H_o: \alpha_1 + \beta_2 = \alpha_2 + \beta_1 \quad , \qquad (4.2.38)$$

which can be written as

$$H_o: P\underline{\xi} = [\alpha_1 - \alpha_2 - \beta_1 + \beta_2] = 0. \qquad (4.2.39)$$

P is given by the row vector

$$[1\ -1\ 0\ 0\ 0 \quad -1\ 1\ 0\ 0\ 0 \quad 0\ 0\ 0\ 0\ 0\ ] \quad .$$

Then

$$P(A'A)^{-1}\ P' = \frac{4}{5}\ , \text{ and}$$

$$[P(A'A)^{-1}\ P']^{-1} = \frac{5}{4} \quad .$$

The sum of squares due to hypothesis is

$$SS(H) = \frac{(\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\beta}_1 + \hat{\beta}_2)^2}{\frac{4}{5}}\ , \qquad (4.2.40)$$

and the ratio $\frac{SS(H)}{SS(E)/n_e}$ is compared with the F-distribution with 1 and n_e degrees of freedom.

4.3 Testing Quadratic and Higher Order Contrasts

In this section, we shall discuss contrasts involving the levels of one factor only, and certain contrasts involving the levels of two factors; and we will formulate the hypotheses relating to these contrasts, and develop the sums of squares and test statistics necessary for testing these hypotheses. The user of this design should be aware of the fact that the usual linear contrast, which would lead to the hypothesis of equality of effects of all levels of one factor, is not testable in this design. However, the contrasts which compare one level with the mean or weighted mean of others, exist and are testable for each factor separately. Contrasts of higher order are all estimable functions. Thus, for a design where each factor appears at p levels, we would have (p-2) degrees of freedom to be broken down into quadratic and higher order contrasts for each factor. Note that the sums of squares due to these contrasts will not be additive, as the inverse matrix $(A'A)^{-1}$, of the constrained normal equations, is not a diagonal matrix. We thus have three degrees of freedom unaccounted for, one each from the linear contrast of each factor. These three degrees of freedom can be sub-divided into two bilinear contrasts, α-β and α-γ , say, (the β-γ bi-linear contrast being implied by these two contrasts) and a single degree of freedom due to a combination of the three factors.

It is possible to test the hypothesis, say, that an estimable function such as the quadratic contrast is equal to zero or some fixed value. Furthermore, it is possible to test the hypothesis that a vector of estimable functions, made up of quadratic and higher order contrasts, is equal to the zero vector. The testing of such hypotheses follows from the theory outlined in Section 4.1.1. These hypotheses will be formulated for designs where each factor appears at three, four, and five levels, respectively. The format of the following sections will be the same as that in the sections on the hypothesis of substitution.

4.3.1 <u>Three Level Case</u>

For the case where each factor appears at three levels, we can test the hypothesis that the quadratic contrast for each factor is equal to zero, say; and also the hypothesis that each bi-linear contrast is equal to zero. The sum of squares for these hypotheses is given by equation (4.1.3), i.e.,

$$SS(H) = (P\hat{\underset{\sim}{\xi}})' \, [P(A'A)^{-1} \, P']^{-1}(P\hat{\underset{\sim}{\xi}}) \quad ,$$

where $\hat{\underset{\sim}{\xi}}$ is a solution to the normal equations. The inverse matrix, $(A'A)^{-1}$, is given in Section 3.4.1., or in Section 4.2.1.

1) Quadratic Contrast

The quadratic contrast discussed here is the second degree orthogonal polynomial, given by [1, -2, 1]. The hypotheses for the three factors are

$$H_o: \ \alpha_1 - 2\alpha_2 + \alpha_3 = \alpha_1 - 2\alpha_2 = 0$$

$$H_o: \ \beta_1 - 2\beta_2 + \beta_3 = \beta_1 - 2\beta_2 = 0 \qquad (4.3.1)$$

and

$$H_o: \ \gamma_{(1)} - 2\gamma_{(2)} + \gamma_{(3)} = \gamma_{(1)} - 2\gamma_{(2)} = 0 \quad .$$

Each of these hypotheses can be conveniently and compactly stated in the form

$$H_o: \ P\underset{\sim}{\xi} = 0 \quad ,$$

where P is the row vector [1 -2 0 0 0 0], if we are testing α-quadratic contrast; [0 0 1 -2 0 0], if we are testing the β-quadratic contrast; and [0 0 0 0 1 -2], if we are testing the γ-quadratic contrast.

Then

$$P(A'A)^{-1} P' = 2.50 \quad \text{in each case, and}$$

$$[P(A'A)^{-1} P']^{-1} = \frac{1}{2.50} \quad .$$

The sums of squares due to the hypothesis are then given by

$$SS(\alpha) = \frac{(\hat{\alpha}_1 - 2\hat{\alpha}_2)^2}{2.50} \quad ,$$

$$SS(\beta) = \frac{(\hat{\beta}_1 - 2\hat{\beta}_2)^2}{2.50} \quad , \qquad (4.3.2)$$

and

$$SS(\gamma) = \frac{(\hat{\gamma}_{(1)} - 2\hat{\gamma}_{(2)})^2}{2.50} \quad ,$$

since, in the three level case, $\hat{\alpha}_3 = \hat{\beta}_3 = \hat{\gamma}_{(3)} = 0$.

The test statistic, as before, is $\frac{SS(H)}{SS(E)/n_e}$, and this ratio is compared with the F-distribution with 1 and n_e degrees of freedom.

2) <u>Bi-Linear Contrast</u>

The linear contrast is the first degree orthogonal polynomial, given by [-1, 0, 1] . The hypotheses are then given by

$$H_o: -\alpha_1 + \alpha_3 + \beta_1 - \beta_3 = -\alpha_1 + \beta_1 = 0$$

and (4.3.3)

$$H_o: -\alpha_1 + \alpha_3 + \gamma_{(1)} - \gamma_{(3)} = -\alpha_1 + \gamma_{(1)} = 0 .$$

If these are written in the form $H_o: P\underline{\xi} = 0$, P is the row vector $[-1\ 0\ \ 1\ 0\ \ 0\ 0]$, if we are testing the $\alpha - \beta$ bi-linear contrast; and the row vector $[-1\ 0\ \ 0\ 0\ \ 1\ 0]$, if we are testing the $\alpha - \gamma$ bi-linear contrast.

Then

$$P(A'A)^{-1} P' = \frac{4}{3}$$

and

$$[\ P(A'A)^{-1} P']^{-1} = \frac{3}{4} \quad .$$

The sums of squares due to the hypotheses are then given by

$$SS(\alpha-\beta) = \frac{(-\hat{\alpha}_1 + \hat{\beta}_1)^2}{\frac{4}{3}} ,$$

and (4.3.4)

$$SS(\alpha-\gamma) = \frac{(-\hat{\alpha}_1 + \hat{\gamma}_{(1)})^2}{\frac{4}{3}} ,$$

and the ratio $\dfrac{SS(H)}{SS(E)/n_e}$ is compared with the F-distribution with 1 and n_e degrees of freedom.

4.3.2 Four Level Case

For the case where each factor appears at four levels, we can test the hypotheses that the quadratic contrast is equal to zero, the cubic contrast is equal to zero, and each bi-linear contrast is equal to zero. Furthermore, we can test the hypothesis that the quadratic and cubic contrasts are simultaneously equal to zero. For the case of four levels, the inverse matrix, $(A'A)^{-1}$, is given in Section 3.4.2; or in Section 4.2.2.

1) Quadratic Contrast

The quadratic contrast is the second degree orthogonal polynomial, given by [1, -1, -1, 1]. The hypotheses for the three factors are

$$H_o: \alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 = 0 ,$$

$$H_o: \beta_1 - \beta_2 - \beta_3 + \beta_4 = 0 ,$$

and (4.3.5)

$$H_o: \gamma_{(1)} - \gamma_{(2)} - \gamma_{(3)} + \gamma_{(4)} = 0 .$$

These can be written as

$$H_o: P\underline{\xi} = [\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4] = 0 ,$$

where P is the row vector [1 -1 -1 1 0 0 0 0 0 0 0 0];

$$H_o: P\underline{\xi} = [\beta_1 - \beta_2 - \beta_3 + \beta_4] = 0 ,$$

where P is the row vector [0 0 0 0 1 -1 -1 1 0 0 0 0]; and

$$H_o: P\underline{\xi} = [\gamma_{(1)} - \gamma_{(2)} - \gamma_{(3)} + \gamma_{(4)}] = 0 ,$$

where P is the row vector [0 0 0 0 0 0 0 0 1 -1 -1 1].

Then, for each hypothesis

$$P(A'A)^{-1} P' = 1.50 ,$$

and

$$[P(A'A)^{-1} P']^{-1} = \frac{1}{1.50} .$$

The sums of squares due to the hypotheses are then given by

$$SS(\alpha) = \frac{(\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\alpha}_3 + \hat{\alpha}_4)^2}{1.50} \quad ,$$

$$SS(\beta) = \frac{(\hat{\beta}_1 - \hat{\beta}_2 - \hat{\beta}_3 + \hat{\beta}_4)^2}{1.50} \quad ,$$

and (4.3.6)

$$SS(\gamma) = \frac{(\hat{\gamma}_{(1)} - \hat{\gamma}_{(2)} - \hat{\gamma}_{(3)} + \hat{\gamma}_{(4)})^2}{1.50} \quad .$$

The test statistic is $\frac{SS(H)}{SS(E)/n_e}$, and this ratio is compared with the F-distribution with 1 and n_e degrees of freedom.

2) Cubic Contrast

The cubic contrast is the third degree orthogonal polynomial, given by [-1, 3, -3, 1]. The hypotheses for the three factors are

$$H_o: \; -\alpha_1 + 3\alpha_2 - 3\alpha_3 + \alpha_4 = 0 \quad ,$$

$$H_o: \; -\beta_1 + 3\beta_2 - 3\beta_3 + \beta_4 = 0 \quad ,$$

and (4.3.7)

$$H_o: -\gamma_{(1)} + 3\gamma_{(2)} - 3\gamma_{(3)} + \gamma_{(4)} = 0 \quad .$$

These can be written as

$$H_o: P\underline{\xi} = [-\alpha_1 + 3\alpha_2 - 3\alpha_3 + \alpha_4] = 0 \quad ,$$

where P is the row vector [-1 3 -3 1 0 0 0 0 0 0 0 0];

$$H_o: P\underline{\xi} = [-\beta_1 + 3\beta_2 - 3\beta_3 + \beta_4] = 0 \quad ,$$

where P is the row vector [0 0 0 0 -1 3 -3 1 0 0 0 0]; and

$$H_o: P\underline{\xi} = [-\gamma_{(1)} + 3\gamma_{(2)} - 3\gamma_{(3)} + \gamma_{(4)}] = 0 ,$$

where P is the row vector [0 0 0 0 0 0 0 0 -1 3 -3 1] . Then, for each hypothesis,

$$P(A'A)^{-1} P' = 6.50 \quad ,$$

and

$$[P(A'A)^{-1} P']^{-1} = \frac{1}{6.50} \quad .$$

The sums of squares due to the hypotheses are given by

$$SS(\alpha) = \frac{(-\hat{\alpha}_1 + 3\hat{\alpha}_2 - 3\hat{\alpha}_3 + \hat{\alpha}_4)^2}{6.50} \quad ,$$

$$SS(\beta) = \frac{(-\hat{\beta}_1 + 3\hat{\beta}_2 - 3\hat{\beta}_3 + \hat{\beta}_4)^2}{6.50}$$

and (4.3.8)

$$SS(\gamma) = \frac{(-\hat{\gamma}_{(1)} + 3\hat{\gamma}_{(2)} - 3\hat{\gamma}_{(3)} + \hat{\gamma}_{(4)})^2}{6.50}$$

The ratio $\frac{SS(H)}{SS(E)/n_e}$ is then compared with the F-distribution with 1 and n_e degrees of freedom.

3) Quadratic and Cubic Contrasts

The hypotheses that the quadratic and cubic contrasts are simultaneously equal to zero can be written in the compact forms

$$H_o: P\underline{\xi} = \begin{bmatrix} \alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 \\ -\alpha_1 + 3\alpha_2 - 3\alpha_3 + \alpha_4 \end{bmatrix} = \underline{0} ,$$

(4.3.9)

where P is the (2 x 12) matrix

$$\begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 3 & -3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} ;$$

$$H_o: P\underline{\xi} = \begin{bmatrix} \beta_1 - \beta_2 - \beta_3 + \beta_4 \\ -\beta_1 + 3\beta_2 - 3\beta_3 + \beta_4 \end{bmatrix} = \underline{0} , \qquad (4.3.10)$$

where P is the (2 x 12) matrix

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 3 & -3 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} ;$$

and

$$H_o: P\underline{\xi} = \begin{bmatrix} \gamma_{(1)} - \gamma_{(2)} - \gamma_{(3)} + \gamma_{(4)} \\ -\gamma_{(1)} + 3\gamma_{(2)} - 3\gamma_{(3)} + \gamma_{(4)} \end{bmatrix} = \underline{0} , \qquad (4.3.11)$$

where P is the (2 x 12) matrix

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 3 & -3 & 1 \end{bmatrix} .$$

Then, for each hypothesis,

$$P(A'A)^{-1} P' = \begin{bmatrix} 1.50 & 0.50 \\ 0.50 & 6.50 \end{bmatrix} ,$$

and

$$[P(A'A)^{-1} P']^{-1} = \frac{1}{9.50} \begin{bmatrix} 6.50 & -0.50 \\ -0.50 & 1.50 \end{bmatrix} .$$

The sums of squares due to the hypotheses are given by

(4.3.12)

$$ss(\alpha) = \frac{1}{9.50}[6.50(\hat{\alpha}_1-\hat{\alpha}_2-\hat{\alpha}_3+\hat{\alpha}_4)^2$$

$$+1.50(-\hat{\alpha}_1+3\hat{\alpha}_2-3\hat{\alpha}_3+\hat{\alpha}_4)^2$$

$$-(\hat{\alpha}_1-\hat{\alpha}_2-\hat{\alpha}_3+\hat{\alpha}_4)(-\hat{\alpha}_1+3\hat{\alpha}_2-3\hat{\alpha}_3+\hat{\alpha}_4)] \quad ,$$

$$ss(\beta) = \frac{1}{9.50}[6.50(\hat{\beta}_1-\hat{\beta}_2-\hat{\beta}_3+\hat{\beta}_4)^2$$

$$+1.50(-\hat{\beta}_1+3\hat{\beta}_2-3\hat{\beta}_3+\hat{\beta}_4)^2$$

$$-(\hat{\beta}_1-\hat{\beta}_2-\hat{\beta}_3+\hat{\beta}_4)(-\hat{\beta}_1+3\hat{\beta}_2-3\hat{\beta}_3+\hat{\beta}_4)] \quad ,$$

and

$$ss(\gamma) = \frac{1}{9.50}[6.50(\hat{\gamma}_{(1)}-\hat{\gamma}_{(2)}-\hat{\gamma}_{(3)}+\hat{\gamma}_{(4)})$$

$$+1.50(-\hat{\gamma}_{(1)}+3\hat{\gamma}_{(2)}-3\hat{\gamma}_{(3)}+\hat{\gamma}_{(4)})^2$$

$$-(\hat{\gamma}_{(1)}-\hat{\gamma}_{(2)}-\hat{\gamma}_{(3)}+\hat{\gamma}_{(4)})(-\hat{\gamma}_{(1)}+3\hat{\gamma}_{(2)}-3\hat{\gamma}_{(3)}+\hat{\gamma}_{(4)})] .$$

The ratio $\frac{SS(H)/2}{SS(E)/n_e}$ is compared with the F-distribution with 2 and n_e degrees of freedom.

4) Bi-Linear Contrast

The linear contrast is the first degree orthogonal polynomial, given by [-3, -1, 1, 3]. The hypotheses are then given by

$$H_o: -3\alpha_1-\alpha_2+\alpha_3+3\alpha_3+3\beta_1+\beta_2-\beta_3-3\beta_4 = 0 \quad ,$$

and

$$H_0: -3\alpha_1-\alpha_2+\alpha_3+3\alpha_3+3\gamma_{(1)}+\gamma_{(2)}-\gamma_{(3)}-3\gamma_{(4)} = 0 .$$

(4.3.13)

If these are written in the form $H_0: P\underset{\sim}{\xi} = 0$, P is the row vector [-3 -1 1 3 3 1 -1 -3 0 0 0 0] if we are testing the $\alpha - \beta$ bi-linear contrast; and the row vector [-3 -1 1 3 0 0 0 0 3 1 -3 -3] , if we are testing the $\alpha-\gamma$ bi-linear contrast. Then, in both cases,

$$P(A'A)^{-1} P' = 11.00 ,$$

and

$$[P(A'A)^{-1} P']^{-1} = \frac{1}{11.00} .$$

The sums of squares due to the hypotheses are then given by

$$SS(\alpha-\beta) = \frac{(-3\hat{\alpha}_1-\hat{\alpha}_2+\hat{\alpha}_3+3\hat{\alpha}_4+3\hat{\beta}_1+\hat{\beta}_2-\hat{\beta}_3-3\hat{\beta}_4)^2}{11.00}$$

and (4.3.14)

$$SS(\alpha-\gamma) - \frac{(-3\hat{\alpha}_1-\hat{\alpha}_2+\hat{\alpha}_3+3\hat{\alpha}_4+3\hat{\gamma}_{(1)}+\hat{\gamma}_{(2)}-\hat{\gamma}_{(3)}+3\hat{\gamma}_{(4)})^2}{11.00} ,$$

and the ratio $\frac{SS(H)}{SS(E)/_{n_e}}$ is compared with the F-distribution with 1 and n_e degrees of freedom.

4.3.3 Five Level Case

For the case where each factor appears at five levels, we can test the hypotheses that the quadratic, cubic, and quartic contrasts are each equal to zero; and each bilinear contrast is equal to zero. Furthermore, we can test the hypothesis that the quadratic, cubic, and quartic contrasts are simultaneously equal to zero. For the five level case, the inverse matrix, $(A'A)^{-1}$, is given in Section 3.4.3, or in Section 4.2.3.

1) Quadratic Contrast

The quadratic contrast is the second degree orthogonal polynomial, given by [2, -1, -2, -1, 2] . The hypotheses for the three factors are

$$H_o: 2\alpha_1 - \alpha_2 - 2\alpha_3 - \alpha_4 + 2\alpha_5 = 0 ,$$

$$H_o: 2\beta_1 - \beta_2 - 2\beta_3 - \beta_4 + 2\beta_5 = 0 ,$$

and (4.3.15)

$$H_o: 2\gamma_{(1)} - \gamma_{(2)} - 2\gamma_{(3)} - \gamma_{(4)} + 2\gamma_{(5)} = 0 .$$

These can be written as

$$H_o: P\underline{\xi} = [2\alpha_1 - \alpha_2 - 2\alpha_3 - \alpha_4 + 2\alpha_5] = 0 ,$$

where P is the row vector [2 -1 -2 -1 2 00000 00000];

$$H_o: P\underline{\xi} = [2\beta_1 - \beta_2 - 2\beta_3 - \beta_4 + 2\beta_5] = 0 ,$$

where P is the row vector [00000 2 -1 -2 -1 2 00000];

and

$$H_o: P\underline{\xi} = [2\gamma_{(1)} - \gamma_{(2)} - 2\gamma_{(3)} - \gamma_{(4)} + 2\gamma_{(5)}] = 0,$$

where P is the row vector [00000 00000 2 -1 -2 -1 2].

Then for each hypothesis

$$P(A'A)^{-1} P' = 4.85 ,$$

and

$$[P(A'A)^{-1} P']^{-1} = \frac{1}{4.85} .$$

The sums of squares due to the hypotheses are then given by

$$SS(\alpha) = \frac{(2\hat{\alpha}_1 - \hat{\alpha}_2 - 2\hat{\alpha}_3 - \hat{\alpha}_4 + 2\hat{\alpha}_5)^2}{4.85} ,$$

$$SS(\beta) = \frac{(2\hat{\beta}_1 - \hat{\beta}_2 - 2\hat{\beta}_3 - \hat{\beta}_4 + 2\hat{\beta}_5)^2}{4.85} ,$$

and (4.3.16)

$$SS(\gamma) = \frac{2\hat{\gamma}_{(1)} -\hat{\gamma}_{(2)} -2\hat{\gamma}_{(3)} -\hat{\gamma}_{(4)} +2\hat{\gamma}_{(5)})^2}{4.85} .$$

The ratio $\frac{SS(H)}{SS(E)/n_e}$ is compared with the F-distribution with 1 and n_e degrees of freedom.

2) Cubic Contrast

The cubic contrast is the third degree orthogonal polynomial, given by [-1, 2, 0, -2, 1]. The hypotheses for the three factors are

$$H_o: -\alpha_1 + 2\alpha_2 - 2\alpha_4 + \alpha_5 = 0 ,$$

$$H_o: -\beta_1 + 2\beta_2 - 2\beta_4 + \beta_5 = 0 ,$$

and (4.3.17)

$$H_o: -\gamma_{(1)} + 2\gamma_{(2)} - 2\gamma_{(4)} + \gamma_{(5)} = 0 .$$

These can be written as

$$H_o: P \underline{\xi} = [-\alpha_1 + 2\alpha_2 - 2\alpha_4 + \alpha_5] = 0 ,$$

where P is the row vector [-1 2 0 -2 1 00000 00000];

$$H_o: P\underline{\xi} = [-\beta_1 + 2\beta_2 - 2\beta_4 + \beta_5] = 0 ,$$

where P is the row vector [00000 -1 2 0 -2 1 00000]; and

$$H_o: P\underline{\xi} = [-\gamma_{(1)} + 2\gamma_{(2)} - 2\gamma_{(4)} + \gamma_{(5)}] = 0 ,$$

where P is the row vector [00000 00000 -1 2 0 -2 1]. Then for each hypothesis

$$P(A'A)^{-1} P' = 3.10 ,$$

and

$$[P(A'A)^{-1} P']^{-1} = \frac{1}{3.10} .$$

The sums of squares due to the hypotheses are given by

$$SS(\alpha) = \frac{(-\hat{\alpha}_1 + 2\hat{\alpha}_2 - 2\hat{\alpha}_4 + \hat{\alpha}_5)^2}{3.10} ,$$

$$SS(\beta) = \frac{(-\hat{\beta}_1 + 2\hat{\beta}_2 - 2\hat{\beta}_4 + \hat{\beta}_5)^2}{3.10} ,$$

and (4.3.18)

$$SS(\gamma) = \frac{(-\hat{\gamma}_{(1)} + 2\hat{\gamma}_{(2)} - 2\hat{\gamma}_{(4)} + \hat{\gamma}_{(5)})^2}{3.10} .$$

The ratio $\frac{SS(H)}{SS(E)/n_e}$ is compared with the F-distribution with 1 and n_e degrees of freedom.

3) Quartic Contrast

The quartic contrast is the fourth degree orthogonal polynomial, given by [1, -4, 6, -4, 1]. The hypotheses for the three factors are

$$H_o: \alpha_1 - 4\alpha_2 + 6\alpha_3 - 4\alpha_4 + \alpha_5 = 0 ,$$

$$H_o: \beta_1 - 4\beta_2 + 6\beta_3 - 4\beta_4 + \beta_5 = 0 ,$$

and (4.3.19)

$$H_o: \gamma_{(1)} - 4\gamma_{(2)} + 6\gamma_{(3)} - 4\gamma_{(4)} + \gamma_{(5)} = 0 .$$

These can be written as

$$H_o: P\underline{\xi} = [\alpha_1 - 4\alpha_2 + 6\alpha_3 - 4\alpha_4 + \alpha_5] = 0 ,$$

where P is the row vector [1 -4 6 -4 1 00000 00000];

$$H_o: P\underline{\xi} = [\beta_1 - 4\beta_2 + 6\beta_3 - 4\beta_4 + \beta_5] = 0 ,$$

where P is the row vector [00000 1 -4 6 -4 1 00000];
and

$$H_o: P\underline{\xi} = [\gamma_{(1)} - 4\gamma_{(2)} + 6\gamma_{(3)} -4\gamma_{(4)} + \gamma_{(5)}] = 0 ,$$

where P is the row vector [00000 00000 1 -4 6 -4 1].
Then for each hypothesis

$$P(A'A)^{-1} P' = 19.50 ,$$

or

$$[P(A'A)^{-1} P']^{-1} = \frac{1}{19.50} .$$

The sums of squares due to the hypotheses are then given by

$$SS(\alpha) = \frac{(\hat{\alpha}_1 - 4\hat{\alpha}_2 + 6\hat{\alpha}_3 - 4\hat{\alpha}_4 + \hat{\alpha}_5)^2}{19.50} ,$$

$$SS(\beta) = \frac{(\hat{\beta}_1 - 4\hat{\beta}_2 + 6\hat{\beta}_3 - 4\hat{\beta}_4 + \hat{\beta}_5)^2}{19.50}$$

and (4.3.20)

$$SS(\gamma) = \frac{(\hat{\gamma}_{(1)} - 4\hat{\gamma}_{(2)} + 6\hat{\gamma}_{(3)} - 4\hat{\gamma}_{(4)} + \hat{\gamma}_{(5)})^2}{19.50} .$$

4) <u>Quadratic, Cubic, and Quartic Contrasts</u>

The hypotheses that the quadratic, cubic, and quartic contrasts are simultaneously equal to zero can be written in the compact forms

$$H_o: P\underline{\xi} = \begin{bmatrix} 2\alpha_1 - \alpha_2 - 2\alpha_3 - \alpha_4 + 2\alpha_5 \\ -\alpha_1 + 2\alpha_2 \qquad - 2\alpha_4 + \alpha_5 \\ \alpha_1 - 4\alpha_2 + 6\alpha_3 - 4\alpha_4 + \alpha_5 \end{bmatrix} = \underline{0},$$

(4.3.21)

where P is the (3 x 15) matrix

$$\begin{bmatrix} 2 & -1 & -2 & -1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & 0 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -4 & 6 & -4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} ;$$

$$H_o: P\underline{\xi} = \begin{bmatrix} 2\beta_1 - \beta_2 - 2\beta_3 - \beta_4 + 2\beta_5 \\ -\beta_1 + 2\beta_2 \qquad - 2\beta_4 + \beta_5 \\ \beta_1 - 4\beta_2 + 6\beta_3 - 4\beta_4 + \beta_5 \end{bmatrix} = \underline{0} ,$$

(4.3.22)

Where P is the (3 x 15) matrix

$$\begin{bmatrix} 0&0&0&0&0 & 2&-1&-2&-1&2 & 0&0&0&0&0 \\ 0&0&0&0&0 & -1&2&0&-2&1 & 0&0&0&0&0 \\ 0&0&0&0&0 & 1&-4&6&-4&1 & 0&0&0&0&0 \end{bmatrix} ;$$

and

$$H_0: P\underline{\xi} = \begin{bmatrix} 2\gamma_{(1)} - \gamma_{(2)} - 2\gamma_{(3)} - \gamma_{(4)} + 2\gamma_{(5)} \\ -\gamma_{(1)} + 2\gamma_{(2)} \qquad - 2\gamma_{(4)} + \gamma_{(5)} \\ \gamma_{(1)} - 4\gamma_{(2)} + 6\gamma_{(3)} - 4\gamma_{(4)} + \gamma_{(5)} \end{bmatrix} = \underline{0} , \tag{4.2.23}$$

Where P is the (3 x 15) matrix

$$\begin{bmatrix} 0&0&0&0&0 & 0&0&0&0&0 & 2&-1&-2&-1&2 \\ 0&0&0&0&0 & 0&0&0&0&0 & -1&2&0&-2&1 \\ 0&0&0&0&0 & 0&0&0&0&0 & 1&-4&6&-4&1 \end{bmatrix} .$$

Then, for each hypothesis,

$$P(A'A)^{-1} P' = \begin{bmatrix} 4.85 & 0.90 & 0.50 \\ 0.90 & 3.10 & 1.50 \\ 0.50 & 1.50 & 19.50 \end{bmatrix} ,$$

and

$$[P(A'A)^{-1}P']^{-1} = \frac{1}{267.05} \begin{bmatrix} 58.200 & -16.800 & -0.200 \\ -16.800 & 94.325 & -6.825 \\ -0.200 & -6.825 & 14.225 \end{bmatrix} .$$

The sum of squares due to the hypotheses are given by

$$\begin{aligned}
SS(\alpha) = \frac{1}{267.05} [& 58.200(2\hat{\alpha}_1-\hat{\alpha}_2-2\hat{\alpha}_3-\hat{\alpha}_4+2\hat{\alpha}_5)^2 \\
& +94.325(\hat{\alpha}_1+2\hat{\alpha}_2-2\hat{\alpha}_4+\hat{\alpha}_5)^2 \\
& +14.225(\hat{\alpha}_1-4\hat{\alpha}_2+6\hat{\alpha}_3-4\hat{\alpha}_4+\hat{\alpha}_5)^2 \\
& -33.600(2\hat{\alpha}_1-\hat{\alpha}_2-2\hat{\alpha}_3-\hat{\alpha}_4+2\hat{\alpha}_5) \\
& x(-\hat{\alpha}_1+2\hat{\alpha}_2-2\hat{\alpha}_4+\hat{\alpha}_5) \\
& -0.400(2\hat{\alpha}_1-\hat{\alpha}_2-2\hat{\alpha}_3-\hat{\alpha}_4+2\hat{\alpha}_5) \\
& x(\hat{\alpha}_1-4\hat{\alpha}_2+6\hat{\alpha}_3-4\hat{\alpha}_4+\hat{\alpha}_5) \\
& -13.650(-\hat{\alpha}_1+2\hat{\alpha}_2-2\hat{\alpha}_4+\hat{\alpha}_5) \\
& x(\hat{\alpha}_1-4\hat{\alpha}_2+6\hat{\alpha}_3-4\hat{\alpha}_4+\hat{\alpha}_5) \quad]
\end{aligned}$$

(4.3.24)

$$\begin{aligned}
SS(\beta) = \frac{1}{267.05} [& 58.200(2\hat{\beta}_1-\hat{\beta}_2-2\hat{\beta}_3-\hat{\beta}_4+2\hat{\beta}_5)^2 \\
& +94.325(\hat{\beta}_1+2\hat{\beta}_2-2\hat{\beta}_4+\hat{\beta}_5)^2 \\
& +14.225(\hat{\beta}_1-4\hat{\beta}_2+6\hat{\beta}_3-4\hat{\beta}_4+\hat{\beta}_5)^2 \\
& -33.600(2\hat{\beta}_1-\hat{\beta}_2-2\hat{\beta}_3-\hat{\beta}_4+2\hat{\beta}_5 \\
& x(-\hat{\beta}_1+2\hat{\beta}_2-2\hat{\beta}_4+\hat{\beta}_5) \\
& -0.400(2\hat{\beta}_1-\hat{\beta}_2-2\hat{\beta}_3-\hat{\beta}_4+2\hat{\beta}_5) \\
& x(\hat{\beta}_1-4\hat{\beta}_2+6\hat{\beta}_3-4\hat{\beta}_4+\hat{\beta}_5) \\
& -13.650(-\hat{\beta}_1+2\hat{\beta}_2-2\hat{\beta}_4+\hat{\beta}_5) \\
& x(\hat{\beta}_1-4\hat{\beta}_2+6\hat{\beta}_3-4\hat{\beta}_4+\hat{\beta}_5) \quad]
\end{aligned}$$

$$SS(\gamma) = \frac{1}{267.05}[58.200(2\hat{\gamma}_1-\hat{\gamma}_2-2\hat{\gamma}_3-\hat{\gamma}_4+2\hat{\gamma}_5)^2$$

$$+94.325(-\hat{\gamma}_1+2\hat{\gamma}_2-2\hat{\gamma}_4+\hat{\gamma}_5)^2$$

$$+14.225(\hat{\gamma}_1-4\hat{\gamma}_2+6\hat{\gamma}_3-4\hat{\gamma}_4+\hat{\gamma}_5)^2$$

$$-33.600(2\hat{\gamma}_1-\hat{\gamma}_2-2\hat{\gamma}_3-\hat{\gamma}_4+2\hat{\gamma}_5)$$

$$x(-\hat{\gamma}_1+2\hat{\gamma}_2-2\hat{\gamma}_4+\hat{\gamma}_5)$$

$$-0.400(2\hat{\gamma}_1-\hat{\gamma}_2-2\hat{\gamma}_3-\hat{\gamma}_4+2\hat{\gamma}_5)$$

$$x(\hat{\gamma}_1-4\hat{\gamma}_2+6\hat{\gamma}_3-4\hat{\gamma}_4+\hat{\gamma}_5)$$

$$-13.650(-\hat{\gamma}_1+2\hat{\gamma}_2-2\hat{\gamma}_4+\hat{\gamma}_5)$$

$$x(\hat{\gamma}_1-4\hat{\gamma}_2+6\hat{\gamma}_3-4\hat{\gamma}_4+\hat{\gamma}_5)\ ]$$

The ratio $\frac{SS(H)/3}{SS(E)/n_e}$ is compared with the F-distribution with 3 and n_e degrees of freedom.

5) Bi-Linear Contrast

The linear contrast is the first degree orthogonal polynomial, given by [-2, -1, 0, 1, 2]. The hypotheses are then given by

$$H_o: -2\alpha_1-\alpha_2+\alpha_4+2\alpha_5+2\beta_1+\beta_2-\beta_4-2\beta_5 = 0 \ ,$$

and (4.3.25)

$$H_o: -2\alpha_1-\alpha_2+\alpha_4+2\alpha_5+2\gamma_{(1)}+\gamma_{(2)}-\gamma_{(4)}-2\gamma_{(5)} = 0 \ .$$

If these are written in the form $H_o: P\underline{\xi} = 0$, P is the row vector [-2 -1 0 1 2 2 1 0 -1-2 0 0 0 0 0] if we are testing the $\alpha - \beta$ bi-linear contrast; and the row vector [-2 -1 0 1 2 00000 2 1 0 -1 -2] if we are testing the α-γ bi-linear contrast. Then, in both cases,

$$P(A'A)^{-1} P' = 10.05 \quad ,$$

and

$$P(A'A)^{-1} P' = \frac{1}{10.05} \quad .$$

The sums of squares due to the hypotheses are then given by

$$SS(\alpha-\beta) = \frac{(-2\hat{\alpha}_1-\hat{\alpha}_2+\hat{\alpha}_4+2\hat{\alpha}_5+2\hat{\beta}_1+\hat{\beta}_2-\hat{\beta}_4-2\hat{\beta}_5)^2}{10.05}$$

and (4.3.26)

$$SS(\alpha-\gamma) = \frac{(-2\hat{\alpha}_1-\hat{\alpha}_2+\hat{\alpha}_4+2\hat{\alpha}_5+2\hat{\gamma}_{(1)}+\hat{\gamma}_{(2)}-\hat{\gamma}_{(4)}-2\hat{\gamma}_{(5)})^2}{10.05} \quad .$$

and the ratio $\frac{SS(H)}{SS(E)/n_e}$ is compared with the F-distribution with 1 and n_e degrees of freedom.

4.4 Intermediate Values

In this section, we will show that the process of linear interpolation within an interval between the levels for which responses are available results in an estimable function. From a knowledge of the responses due to level i of α, j of β, and (k) of γ, where i, j, (k) are integers we can, by linear interpolation, approximate the response due to any combination $[i', j', (k)']$ where $i', j', (k)'$ may be fractional. This linear combination will be an estimable function provided that $i' + j' + (k)' = p + 3$.

We now note that the ith position in the first partition of the Q matrix, given by equation (2.2.6) is

$$[-(p-3+i),\ p-i,\ p-i,\ p-i\ ] \ .$$

The jth element in the second partition of Q is

$$[p-j,\ -(p-3+j),\ p-j,\ p-j\ ] \ ,$$

and the (k)th element in the third partition is

$$[p-(k),\ p-(k),\ -(p-3+(k)),\ p-(k)\ ] \ .$$

Now, the expectation of the response due to any combination of the levels is given by

$$\begin{aligned} E(y_{i'j'(k)'}) = \mu &+ [(1-h_1)\ \alpha_i + h_1\ \alpha_{i+1}] \\ &+ [(1-h_2)\ \beta_j + h_2\ \beta_{j+1}] \\ &+ [(1-h_3)\ \gamma_{(k)} + h_3\ \gamma_{(k+1)}] \ , \end{aligned}$$

where $0 < h_1, h_2, h_3 < 1$,

$$i' = i + h_1, \quad j' = j + h_2, \quad (k)' = (k) + h_3,$$

and $[i, j, (k)] < p - 1$ (since we are dealing with interpolation, and not extrapolation).

Also

$$i' + j' + (k)' = p + 3, \text{ which determines } h_3.$$

Then, in the same notation as in Section 2.2.3.,

$$\underset{-}{\lambda}_1' = [0, \ldots, 0, \underset{(i)}{1-h_1}, \underset{(i+1)}{h_1}, 0, \ldots, \underset{(p-1)}{0}$$

$$0, 0, \ldots, 0, \underset{(j)}{1-h_2}, \underset{(j+1)}{h_2}, 0, \ldots, \underset{(p-1)}{0}$$

$$0, \ldots, 0, \underset{(k)}{1-h_3}, \underset{(k+1)}{h_3}, 0, \ldots, \underset{(p-1)}{0}] ,$$

the numbers in parentheses indicating the position of the non-zero values. We must show that

$$\underset{-}{\lambda}_2' = \underset{-}{\lambda}_1' Q = [0, 0, 0, 1] ,$$

subject to the above conditions.
Hence the first element of $\underset{-}{\lambda}_1'Q$ is

$$-(1-h_1)(p-3+i) - h_1 (p-2+i) + (1-h_2)(p-j) + h_2(p-j-1)$$
$$+ (1-h_3)(p-(k)) + h_3(p-(k)-1)$$
$$= (p+3-i-j-(k)) - h_1 - h_2 - h_3$$
$$= 0 .$$

The second and third elements follow the same pattern. The fourth element is

$$(1-h_1)(p-i) + h_1 (p-i-1) + (1-h_2)(p-j) + h_2 (p-j-1)$$
$$+ (1-h_3)(p-(k)) + h_3 (p-(k)-1)$$
$$= 3p - i - j - (k) - h_1 - h_2 - h_3$$
$$= 3p - (p+3)$$
$$= 2p-3 ,$$

which is the reciprocal of the constant factor in front of the Q-matrix. Hence the fourth element equals 1, which proves estimability of the linear interpolation formula. In actual practice we may thus start with any set of estimates, pick i, j, (k) as the integers immediately below the given (i', j', (k)') and interpolate between α_i, α_{i+1}; β_j, β_{j+1}; and $\gamma_{(k)}$, $\gamma_{(k+1)}$; checking first that, in fact, i'+j'+(k)' = p + 3. To obtain an interpolation based upon the higher degree polynomials we may find the response curves for each of the factors separately, as described in the next section, and interpolate graphically or by inserting intermediate values into the polynomial expressions stated there.

4.5 Fitting of Response Functions for a Single Factor

In the previous section, we discussed linear interpolation involving all three factors simultaneously. Suppose, however, that we want to find the functional form of the response function involving one factor only. We know that it is impossible to find a linear function, since the linear contrast is a non-estimable function. However, all higher degree contrasts are estimable and we would thus like to be able to determine whether a fit is a quadratic, cubic, or a higher degree polynomial. We will construct polynomials of the second, third, and higher degrees in such a way that they are orthogonal with reference to this design. This choice will make the sums-of-squares additive, and we may thus, in one single stage, decide on whether or not higher order contributions to the response function are significant. The polynomials will be developed for the cases where we have four and five levels of the factors. The sums of squares for each contrast will also be given for these cases, and they will be additive.

4.5.1 Four Level Case

Let $\underline{\hat{\alpha}}$ denote the estimates of α_i obtained by the modified high-low method. We may consider the α_i's as pseudo-observations such that $E(\underline{\ell}'\, \underline{\hat{\alpha}}) = \underline{\ell}'\underline{\alpha}$, where $\underline{\ell}'\underline{\alpha}$ is any estimable function, and

$$\text{var}\ (\underline{\ell}'\underline{\hat{\alpha}}) = \sigma^2 \underline{\ell}'\ V\underline{\ell}\ ,$$

where V is that part of the inverse matrix of A'A which is associated with $\underline{\hat{\alpha}}$. For the case of four levels, V is given by

$$V = \frac{1}{8} \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & -2 & 0 \end{bmatrix} .$$

Assume now that $\alpha(z)$ is a function of z for $z = -3, -1, 1, 3$, i.e.,

$$\alpha_1 = \alpha(-3),\ \alpha_2 = \alpha(-1),\ \alpha_3 = \alpha(1),\ \alpha_4 = \alpha(3) \quad .$$

Estimates of the α_i are available from the modified high-low method, let them be $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$, and $\hat{\alpha}_4$. We wish to find two polynomials in z, $P_1(z)$ and $P_2(z)$, say, preferably orthogonal, so that

$$\alpha(z) = B_1P_1(z) + B_2P_2(z) + e \quad . \qquad (4.5.1)$$

Now, B_1 and B_2 are estimated as linear functions of the $\hat{\alpha}_i$, and are the same regardless of the constraints used in the process of estimating the α_i. In other words,

$\hat{B}_1$ and $\hat{B}_2$ must be estimable functions of the pseudo-observations $\hat{\alpha}_i$. Now, since var $(\underline{\ell}'\hat{\alpha})$ = $\sigma^2\underline{\ell}'V\underline{\ell}$ the estimates must be combinations of $\hat{\underline{\alpha}}$, $\underline{g}'V^{-1}\hat{\underline{\alpha}}$ say, such that $\underline{g}'\ V^{-1}$ is a weight vector for estimable functions. We know that any combination

$$\omega\,[1,\ -1,\ -1,\ 1]\quad +\quad [-1,\ 3,\ -3,\ 1]\quad ,$$

where the bracketed terms are the usual quadratic and cubic contrasts, is estimable. Hence $\underline{g}'V^{-1}$ must be a vector of this form, or

$$\underline{g}'V^{-1}\ =\ [\omega-1,\ -\omega+3,\ -\omega-3,\ \omega+1]\quad .$$

Then

$$\begin{aligned}\underline{g}'\ &=\ [\omega-1,\ -\omega+3,\ -\omega-3,\ \omega+1]\ V\\ &=\ [\ \tfrac{1}{2}\ (\omega-1),\ -\ \tfrac{1}{4}\ (\omega-3),\ -\ \tfrac{1}{2}\ (\omega+2),\ \tfrac{1}{4}(\omega+3)\ ]\quad .\end{aligned}$$

Note that

$$\hat{B}_1 = \frac{\underline{g}_1'\ V^{-1}\ \hat{\underline{\alpha}}}{\underline{g}_1'\ V^{-1}\ \underline{g}_1}\quad ,$$

and

$$\hat{B}_2 = \frac{\underline{g}_2'\ V^{-1}\ \hat{\underline{\alpha}}}{\underline{g}_2'\ V^{-1}\ \underline{g}_2}\quad ,$$

provided that

$\underline{g}_1'\ V^{-1}\ \underline{g}_2 = 0$, and that $\underline{g}_1'\ V^{-1}\ \underline{\alpha}$ and $\underline{g}_2'\ V^{-1}\ \underline{\alpha}$ are estimable, the latter being required since $\underline{g}'\ V^{-1}\ \hat{\underline{\alpha}}$ is to be an unbiased estimate of $\underline{g}'\ V^{-1}\ \underline{\alpha}$.

Let

$$P_1(z) = a_0 + a_1\ z + a_2\ z^2 + a_3\ z^3\ , \qquad (4.5.2)$$

and

$$P_2(z) = b_0 + b_1\ z + b_2\ z^2 + b_3\ z^3\ . \qquad (4.5.3)$$

Also, let

$$\underset{\sim}{g}_1'V^{-1} = [\omega-1,\ -\omega+3,\ -\omega-3,\ \omega+1] \quad ,$$

and

$$\underset{\sim}{g}_2'V^{-1} = [\eta-1,\ -\eta+3,\ -\eta-3,\ \eta+1] \quad .$$

Then, because orthogonality is desired, we must have that

$$\underset{\sim}{g}_1'V^{-1}\underset{\sim}{g}_2 = 0 \quad ,$$

or

$$[\omega-1,\ -\omega+3,\ -\omega-3,\ \omega+1]\begin{bmatrix} \frac{1}{2}(\eta-1) \\ -\frac{1}{4}(\eta-3) \\ -\frac{1}{2}(\eta+2) \\ \frac{1}{4}(\eta+3) \end{bmatrix} = 0 \quad .$$

We thus find η in terms of ω as

$$\eta = -\frac{\omega + 13}{3\omega + 1} \quad .$$

The two polynomials given by equations (4.5.2) and (4.5.3) should be quadratic and cubic, since linear contrasts are non-estimable. We shall thus choose ω in such a way that $a_3 = 0$, so that $P_1(z)$ is a quadratic polynomial. Since $\alpha(z)$ is a function of z for $z = -3, -1, 1, 3$, we have the following system of equations in the a_i and ω :

$$a_0 - 3a_1 + 9a_2 = \frac{1}{2}(\omega-1)$$

$$a_0 - a_1 + a_2 = -\frac{1}{4}(\omega-3)$$

$$a_0 + a_1 + a_2 = -\frac{1}{2}(\omega+2)$$

$$a_0 + 3a_1 + 9a_2 = \frac{1}{4}(\omega+3) \quad .$$

The solution to this system is easily found to be

$$a_0 = \frac{95}{16}, \quad a_1 = \frac{3}{4}, \quad a_2 = -\frac{19}{16}, \text{ and } \omega = -13 \quad .$$

Hence,

$$g_1'V^{-1} = [-14, 16, 10, -12] \quad ,$$

and

$$g_1' = [-7, 4, \frac{11}{2}, -\frac{5}{2}] \quad ,$$

Thus,

$$\frac{g_1'V^{-1}\hat{\alpha}}{g_1'V^{-1}g_1} = \frac{1}{247}[-14\hat{\alpha}_1 + 16\hat{\alpha}_2 + 10\hat{\alpha}_3 - 12\hat{\alpha}_4] \quad (4.5.4)$$

is an unbiased, minimum-variance estimate of the coefficient B_1 of the polynomial

$$P_1(z) = \frac{1}{16}[95 + 12z - 19z^2] \quad . \quad (4.5.5)$$

Also, the sum of squares due to B_1 is given by

$$SS(B_1) = \frac{(-14\hat{\alpha}_1 + 16\hat{\alpha}_2 + 10\hat{\alpha}_3 - 12\hat{\alpha}_4)^2}{247} \quad . \quad (4.5.6)$$

The second (cubic) polynomial is found by use of the relation

$$\eta = -\frac{\omega + 13}{3\omega + 1} \quad ,$$

and the relations for $\underline{g}_2{}'V^{-1}$ and $\underline{g}_2{}'$. Since, we know that $\omega = -13$, we immediately have that $\eta = 0$. Thus

$$\underline{g}_2{}'V^{-1} = [-1, 3, -3, 1] \quad ,$$

and

$$\underline{g}_2{}' = [-\frac{1}{2}, \frac{3}{4}, -1, \frac{3}{4}] \quad .$$

Now, the polynomial $P_2(z)$ satisfies the relations:

$$\begin{aligned} b_0 - 3b_1 + 9b_2 - 27b_3 &= -\frac{1}{2} \\ b_0 - b_1 + b_2 - b_3 &= \frac{3}{4} \\ b_0 + b_1 + b_2 + b_3 &= -1 \\ b_0 + 3b_1 + 9b_2 + 27b_3 &= \frac{3}{4} \, . \end{aligned}$$

The solution to this system is easily found to be

$$b_0 = -\frac{5}{32}, \quad b_1 = -\frac{97}{96}, \quad b_2 = \frac{1}{32}, \quad b_3 = \frac{13}{96} \quad .$$

Thus

$$\frac{\underline{g}_2{}'V^{-1}\hat{\underline{\alpha}}}{\underline{g}_2{}'V^{-1}\underline{g}_2} = \frac{2}{13}[-\hat{\alpha}_1 + 3\hat{\alpha}_2 - 3\hat{\alpha}_3 + \hat{\alpha}_4] \qquad (4.5.7)$$

is an unbiased, minimum-variance estimate of the coefficient B_2 of the polynomial

$$P_2(z) = \frac{1}{96}[-15 -97z +3z^2 +13z^3] \quad . \qquad (4.5.8)$$

The sum of squares due to B_2 is given by

$$SS(B_2) = \frac{2}{13}[-\hat{\alpha}_1 + 3\hat{\alpha}_2 - 3\hat{\alpha}_3 + \hat{\alpha}_4]^2 \quad , \qquad (4.5.9)$$

4.5.2 Five Level Case

For the case where each factor appears at five levels, V is given by

$$V = \frac{1}{40}\begin{bmatrix} 12 & 0 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 10 & -2 & -4 \\ 0 & 0 & -2 & 10 & -8 \\ 0 & 0 & -4 & -8 & 8 \end{bmatrix} .$$

Assume that $\alpha(z)$ is a function of z for $z = -2, -1, 0, 1, 2$, i.e.

$$\alpha_1 = \alpha(-2), \quad \alpha_2 = \alpha(-1), \; \alpha_3 = \alpha(0), \; \alpha_4 = \alpha(1), \; \alpha_5 = \alpha(2).$$

Estimates of the α_i are available from the modified high-low method; let them be $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$, $\hat{\alpha}_4$, and $\hat{\alpha}_5$. We wish to find three polynomials in z, $P_1(z)$, $P_2(z)$, and $P_3(z)$, say, preferably orthogonal, so that

$$\alpha(z) = B_1P_1(z) + B_2P_2(z) + B_3P_3(z) + e. \qquad (4.5.10)$$

Now, B_1, B_2, and B_3 are estimated as linear functions of the $\hat{\alpha}_i$ and are the same regardless of the constraints used in the process of estimating the α_i . As before, the estimates must be combinations of $\underline{\hat{\alpha}}$, $\underline{g}'V^{-1}\underline{\hat{\alpha}}$ say, such that $\underline{g}'V^{-1}$ is a weight vector for estimable functions. We know that any combination

$$\omega_1[2, -1, -2, -1, 2] + \omega_2[-1, 2, 0, -2, 1] + [1, -4, 6, -4, 1] \quad ,$$

where the bracketed terms are the usual quadratic, cubic, and quartic contrasts, is estimable. Hence $\underline{g}'V^{-1}$ must be a vector of the form

$$\underline{g}'V^{-1} = [2\omega_1 - \omega_2 + 1, -\omega_1 + 2\omega_2 - 4, -2\omega_1 + 6, -\omega_1 - 2\omega_2 - 4, 2\omega_1 + \omega_2 + 1] \quad .$$

Then

$$\underline{g}' = [\frac{3}{10}(2\omega_1 - \omega_2 + 1), \frac{1}{5}(-\omega_1 + 2\omega_2 - 4), -\frac{13}{20}\omega_1 + \frac{8}{5}, -\frac{11}{20}\omega_1 - \frac{7}{10}\omega_2 - \frac{3}{2}, \frac{4}{5}\omega_1 + \frac{3}{5}\omega_2 + \frac{2}{5}] \quad .$$

Note that

$$\hat{\beta}_1 = \frac{\underline{g}_1'V^{-1}\hat{\underline{\alpha}}}{\underline{g}_1'V^{-1}\underline{g}_1} \quad ,$$

$$\hat{\beta}_2 = \frac{\underline{g}_2'V^{-1}\hat{\underline{\alpha}}}{\underline{g}_2'V^{-1}\underline{g}_2} \quad ,$$

and

$$\hat{\beta}_3 = \frac{\underline{g}_3'V^{-1}\hat{\underline{\alpha}}}{\underline{g}_3'V^{-1}\underline{g}_3} \quad ,$$

provided that

$$\underline{g}_1'V^{-1}\underline{g}_2 = \underline{g}_1'V^{-1}\underline{g}_3 = \underline{g}_2'V^{-1}\underline{g}_3 = 0 \quad ,$$

and that

$$\underline{g}_1'V^{-1}\underline{\alpha}\ ,\ \underline{g}_2'V^{-1}\underline{\alpha}\ ,\ \text{and}\ \ \underline{g}_3'V^{-1}\underline{\alpha}$$

are estimable.

Let

$$P_1(z) = a_0 + a_1z + a_2z^2 + a_3z^3 + a_4z^4 \quad , \qquad (4.5.11)$$

$$P_2(z) = b_0 + b_1z + b_2z^2 + b_3z^3 + b_4z^4 \quad , \qquad (4.5.12)$$

and

$$P_3(z) = c_0 + c_1z + c_2z^2 + c_3z^3 + c_4z^4 \quad . \qquad (4.5.13)$$

Further, let

$$\underline{g}_1'V^{-1} = [2\omega_1 - \omega_2 + 1,\ -\omega_1 + 2\omega_2 - 4,\ -2\omega_1 + 6,\ -\omega_1 - 2\omega_2 - 4,\ 2\omega_1 + \omega_2 + 1] \quad ,$$

$$\underline{g}_2'V^{-1} = [2\eta_1 - \eta_2 + 1,\ -\eta_1 + 2\eta_2 - 4,\ -2\eta_1 + 6,\ -\eta_1 - 2\eta_2 - 4,\ 2\eta_1 + \eta_2 + 1] \quad ,$$

and

$$\underline{g}_3\ V^{-1} = [2\rho_1 - \rho_2 + 1,\ -\rho_1 + 2\rho_2 - 4,\ -2\rho_1 + 6,\ -\rho_1 - 2\rho_2 - 4,\ 2\rho_1 + \rho_2 + 1] \ .$$

The three polynomials, given by equations (4.5.11), (4.5.12), and (4.5.13) should be quadratic, cubic, and quartic, since linear contrasts are non-estimable. We shall thus choose ω_1 and ω_2 in such a way that $a_3 = a_4 = 0$, and that $P_1(z)$ is a quadratic polynomial. Since (z) is a function of z for $z = -2, -1, 0, 1, 2$, we have the following system of equations in the a_i and ω_1 and ω_2:

$$\begin{aligned}
a_0 - 2a_1 + 4a_2 &= \frac{3}{5}\omega_1 - \frac{3}{10}\omega_2 + \frac{3}{10} \\
a_0 - a_1 + a_2 &= -\frac{1}{5}\omega_1 + \frac{2}{5}\omega_2 - \frac{4}{5} \\
a_0 &= -\frac{13}{20}\omega_1 + \frac{8}{5} \\
a_0 + a_1 + a_2 &= -\frac{11}{20}\omega_1 - \frac{7}{10}\omega_2 - \frac{3}{2} \\
a_0 + 2a_1 + 4a_2 &= \frac{4}{5}\omega_1 + \frac{3}{5}\omega_2 + \frac{2}{5} .
\end{aligned}$$

The solution to this system is given by

$$a_0 = \frac{763}{4}, \quad a_1 = \frac{35}{8}, \quad a_2 = -\frac{763}{8}, \quad \omega_1 = -291, \quad \omega_2 = 84.$$

Hence

$$\underline{g}_1{}'V^{-1} = [-665,\ 455,\ 588,\ 119,\ -497] \quad ,$$

and

$$\underline{g}_1{}' = [-\frac{399}{2},\ 91,\ \frac{763}{4},\ \frac{399}{4},\ -182] \quad .$$

Thus

$$\frac{\underline{g}_1{}'V^{-1}\hat{\underline{\alpha}}}{\underline{g}_1{}'V^{-1}\underline{g}_1} = \frac{1}{388,557.75}[-665\hat{\alpha}_1 + 455\hat{\alpha}_2 + 588\hat{\alpha}_3 + 119\hat{\alpha}_4 - 497\hat{\alpha}_5] \tag{4.5.14}$$

is an unbiased, minimum-variance estimate of the coefficient B_1 of the polynomial

$$P_1(z) = \frac{1}{8}\,[1526 + 35z - 763z^2] \quad . \qquad (4.5.15)$$

The sum of squares due to B_1 is given by

$$SS(B_1) = \frac{(-665\hat{\alpha}_1 + 455\hat{\alpha}_2 + 588\hat{\alpha}_3 + 119\hat{\alpha}_4 - 497\hat{\alpha}_5)^2}{388,557.75} \quad .$$

(4.5.16)

Now, because orthogonality is desired, we must have that $\underline{a}_1{}'V^{-1}\underline{a}_2 = 0$ or,

$$[2\omega_1 - \omega_2 + 1,\ -\omega + 2\omega_2 - 4,\ -2\omega_1 + 6,\ -\omega_1 - 2\omega_2 - 4,\ 2\omega_1 + \omega_2 + 1]$$

$$\times \begin{bmatrix} \frac{3}{10}\,(2\eta_1 - \eta_2 + 1) \\ \frac{1}{5}\,(-\eta_1 + 2\eta_2 - 4) \\ -\frac{13}{20}\,\eta_1 + \frac{8}{5} \\ -\frac{11}{20}\,\eta_1 - \frac{7}{10}\,\eta_2 - \frac{3}{2} \\ \frac{4}{5}\,\eta_1 + \frac{3}{5}\,\eta_2 + \frac{2}{5} \end{bmatrix} = 0 \quad .$$

Solving, we find $\eta_1 = 0$. We then choose η_2 in such a way that $b_4 = 0$ and that $P_2(z)$ is a cubic polynomial. The polynomial $P_2(z)$ satisfies the following relations:

$$b_0 - 2b_1 + 4b_2 - 8b_3 = \frac{3}{10}\eta_2 + \frac{3}{10}$$

$$b_0 - b_1 + b_2 - b_3 = \frac{2}{5}\eta_2 - \frac{4}{5}$$

$$b_0 = \frac{8}{5}$$

$$b_0 + b_1 + b_2 + b_3 = -\frac{7}{10}\eta_2 - \frac{3}{2}$$

$$b_0 + 2b_1 + 4b_2 + 8b_3 = \frac{3}{5}\eta_2 + \frac{2}{5} \quad ,$$

where η_1 on the right-hand sides has been set equal to zero.

The solution to this system of equations is

$$b_0 = \frac{8}{5}\ ,\ b_1 = \frac{301}{30}\ ,\ b_2 = -\frac{4}{5}\ ,\ b_3 = -\frac{97}{30}\ ,\ \eta_2 = -13 \quad .$$

Hence

$$\underline{g}_2'V^{-1} = [14,\ -30,\ 6,\ 22,\ -12] \quad ,$$

and

$$\underline{g}_2' = \left[\ \frac{21}{5}\ ,\ -6,\ \frac{8}{5}\ ,\ \frac{38}{5}\ ,\ -\frac{37}{5}\ \right] \quad .$$

Thus

$$\frac{\underline{g}_2'V^{-1}\hat{\underline{\alpha}}}{\underline{g}_2'V^{-1}\underline{g}_2} = \frac{1}{504.4}\ [14\hat{\alpha}_1 - 30\hat{\alpha}_2 + 6\hat{\alpha}_3 + 22\hat{\alpha}_4 - 12\hat{\alpha}_5] \tag{4.5.17}$$

is an unbiased, minimum-variance estimate of the coefficient B_2 of the polynomial

$$P_2(z) = \frac{1}{30}\ [\ 48 + 301z - 24z^2 - 97z^3] \quad . \tag{4.5.18}$$

The sum of squares due to B_2 is given by

$$SS(B_2) = \frac{(14\hat{\alpha}_1 - 30\hat{\alpha}_2 + 6\hat{\alpha}_3 + 22\hat{\alpha}_4 - 12\hat{\alpha}_5)^2}{504.4} . \tag{4.5.19}$$

Again, because of orthogonality, the relations

$$\underline{g}_1{}'V^{-1}\underline{g}_3 = 0$$

and

$$\underline{g}_2{}'V^{-1}\underline{g}_3 = 0$$

must be satisfied. The first relation gives $\rho_1 = 0$.

Then

$$\underline{g}_2{}'V^{-1}\underline{g}_3 = [2\eta_1 - \eta_2 + 1,\ -\eta_1 + 2\eta_2 - 4,\ -2\eta_1 + 6,\ -\eta_1 - 2\eta_2 - 4,\ 2\eta_1 + \eta_2 + 1]$$

$$\times \begin{bmatrix} \frac{3}{10}(2\rho_1 - \rho_2 + 1) \\ \frac{1}{5}(-\rho_1 + 2\rho_2 - 4) \\ -\frac{13}{20}\rho_1 + \frac{8}{5} \\ -\frac{11}{20}\rho_1 - \frac{7}{20}\rho_2 - \frac{3}{2} \\ \frac{4}{5}\rho_1 + \frac{3}{5}\rho_2 + \frac{2}{5} \end{bmatrix} .$$

Since, $\eta_1 = 0$, $\eta_2 = -13$, and $\rho_1 = 0$, we find, as a solution to the above equation, that $\rho_2 = 0$. Hence

$$\underline{g}_3'V^{-1} = [1, -4, 6, -4, 1] \quad ,$$

and

$$\underline{g}_3' = [\frac{3}{10}, -\frac{4}{5}, \frac{8}{5}, -\frac{3}{2}, \frac{2}{5}] \quad .$$

The polynomial $P_3(z)$ satisfies the relations

$$\begin{aligned}
c_0 - 2c_1 + 4c_2 - 8c_3 + 16c_4 &= \frac{3}{10} \\
c_0 - c_1 + c_2 - c_3 + c_4 &= \frac{4}{5} \\
c_0 &= \frac{8}{5} \\
c_0 + c_1 + c_2 + c_3 + c_4 &= -\frac{3}{2} \\
c_0 + 2c_1 + 4c_2 + 8c_3 + 16c_4 &= \frac{2}{5} \quad ,
\end{aligned}$$

for which the solution is

$$c_0 = \frac{8}{5}, \quad c_1 = -\frac{19}{40}, \quad c_2 = -\frac{57}{16}, \quad c_3 = \frac{1}{8}, \quad c_4 = \frac{13}{16} \quad .$$

Thus

$$\frac{\underline{g}_3'V^{-1}\hat{\underline{\alpha}}}{\underline{g}_3'V^{-1}\underline{g}_3} = \frac{2}{39}[\hat{\alpha}_1 - 4\hat{\alpha}_2 + 6\hat{\alpha}_3 - 4\hat{\alpha}_4 + \hat{\alpha}_5] \quad (4.5.20)$$

is an unbiased minimum-variance estimate of the coefficient B_3 of the polynomial

$$P_3(z) = \frac{1}{80}[128 - 38z - 285z^2 + 10z^3 + 65z^4] \quad . \quad (4.5.21)$$

The sum of squares due to B_3 is given by

$$SS(B_3) = \frac{2}{39}[\hat{\alpha}_1 - 4\hat{\alpha}_2 + 6\hat{\alpha}_3 - 4\hat{\alpha}_4 + \hat{\alpha}_5]^2 \quad . \quad (4.5.22)$$

CHAPTER V

EXTENSIONS

5.1 Introductory Remarks

In this chapter, we shall concern ourselves with the case where there is more than one experimental unit per cell. In this connection, we will distinguish between three different situations:

a) the replications are strictly repetitions of the experiment under otherwise identical conditions,

b) the experiments within a cell represent repetitions over a period of time, during which some kind of a trend may be present, and

c) the replications constitute several experiments or responses with the same experimental units. Here we may not have independence among the observations within a cell.

In the following three sections of this chapter, we will give the procedures to be followed in testing hypotheses concerning estimable functions. The format of these sections will be similar to that in previous chapters, in that we will discuss estimable functions, tests of hypotheses, sums of squares, and test statistics, in that order. Then, in Section 5.5, we will discuss the mixed model, where we treat the contribution due to lack-of-fit as a random variable. Demonstration studies will be presented in Chapter VI.

The analysis, in the case of one experimental unit per cell, has been presented in detail in the previous chapters. With regard to the situations where we have n experimental units per cell, we might summarize by saying that in situation a), the analysis proceeds in the customary two-way or three-way analysis with n replicates per treatment combination. The situations b) and c) require more thought, however. In situation b), the analysis can be readily extended into an analysis of covariance with one concomitant variable if only linear trend over time is postulated, and more than one if trend of a higher degree is expected. The formulas and tests of hypothesis will be given for a linear trend and also for a trend of higher order. The analysis in situation c) can be performed by multivariate analysis, if the dependence is the same from one experimental unit to another.

One further situation may arise, this being where the levels of the factors cannot be equally spaced. Although this dissertation does not concern itself directly with this question, we feel that it is important enough to warrant special discussion. This situation will be discussed in Section 5.6 of this chapter.

5.2 Case of n Replications

In the case of n experimental units per cell, the model becomes

$$y_{ij(k)m} = \mu + \alpha_i + \beta_j + \gamma_{(k)} + e'_{ij(k)} + e_{ij(k)m} , \tag{5.2.1}$$

where $i = j = (k) = 1, 2, \ldots, p,$

$m = 1, 2, \ldots, n$.

In the model, $e'_{ij(k)}$ is the contribution due to lack-of-fit (fixed) and $e_{ij(k)m}$ is the contribution due to error (within cells). For a discussion of $e'_{ij(k)}$ as a random variable, see Section 5.5.

Table 1 below gives the basic notation for dealing with an analysis of the Symmetrical Complementation design with n observations per cell. For each cell, the totals and the means of the n observations have been computed and recorded. Sums over all the groups of a classification will be denoted by capital letters with the subscript for the classifications overwhich addition is performed replaced by a dot.

TABLE 1

Basic Notation: n Observations Per Cell

	1	2	3	...	(p-1)	p = Level of B
1		$Y_{12(p).}$ $\bar{Y}_{12(p).}$	$Y_{13(p-1).}$ $\bar{Y}_{13(p-1).}$	... ...	$Y_{1(p-1)(3).}$ $\bar{Y}_{1(p-1)(3).}$	$Y_{1p(2).}$ $\bar{Y}_{1p(2).}$
2	$Y_{21(p).}$ $\bar{Y}_{21(p).}$	$Y_{22(p-1).}$ $\bar{Y}_{22(p-1).}$	$Y_{23(p-2).}$ $\bar{Y}_{23(p-2).}$	... ...	$Y_{2(p-1)(2).}$ $\bar{Y}_{2(p-1)(2).}$	$Y_{2p(1).}$ $\bar{Y}_{2p(1).}$
3	$Y_{31(p-1).}$ $\bar{Y}_{31(p-1).}$	$Y_{32(p-2).}$ $\bar{Y}_{32(p-2).}$	$Y_{33(p-3).}$ $\bar{Y}_{33(p-3).}$	... ...	$Y_{3(p-1)(1).}$ $\bar{Y}_{3(p-1)(1).}$	
⋮	⋮	⋮	⋮			
p-1	$Y_{(p-1)1(3).}$ $\bar{Y}_{(p-1)1(3).}$	$Y_{(p-1)2(2).}$ $\bar{Y}_{(p-1)2(2).}$	$Y_{(p-1)3(1).}$ $\bar{Y}_{(p-1)3(1).}$			
p Level of A	$Y_{p1(2).}$ $\bar{Y}_{p1(2).}$	$Y_{p2(1).}$ $\bar{Y}_{p2(1).}$				

where:

$$Y_{ij(k).} = \sum_{m=1}^{n} Y_{ij(k)m} \quad , \quad Y_{ij(k)m} \text{ is the mth observation in the } [i,j,(k)] \text{ th cell,}$$

$$\bar{y}_{ij(k).} = \frac{Y_{ij(k).}}{n} \quad ,$$

$$Y_{i...} = \sum_{j=1}^{p} Y_{ij(k).} \quad , \quad Y_{.j..} = \sum_{i=1}^{p} Y_{ij(k).} \quad ,$$

$$Y_{..(k).} = \sum_{i+j=3}^{p+2} Y_{ij(k).} \quad ,$$

$$Y_{....} = \sum_{i=1}^{p} Y_{i...} = \sum_{j=1}^{p} Y_{.j..} = \sum_{(k)=1}^{p} Y_{..(k).} \quad ,$$

$n_{i...}$ = total number of experimental units in the ith row,

$n_{.j..}$ = total number of experimental units in the jth column,

$n_{..(k).}$ = total number of experimental units in the (k)th diagonal,

$$n_{....} = \sum_{i=1}^{p} n_{i...} = \sum_{j=1}^{p} n_{.j..} = \sum_{(k)=1}^{p} n_{..(k).} \quad ,$$

$$\bar{y}_{i...} = \frac{Y_{i...}}{n_{i...}} \quad , \quad \bar{y}_{.j..} = \frac{Y_{.j..}}{n_{.j..}} \quad , \quad \bar{y}_{..(k).} = \frac{Y_{..(k).}}{n_{..(k).}} \quad ,$$

$$\bar{y}_{....} = \frac{Y_{....}}{n_{....}} \quad .$$

From Table 1 we now compute the following quantities:

1. Sum of squares for total:

$$SS(\text{Total}) = \sum_{\text{all}} y_{ij(k)m} - \frac{\left[\sum_{\text{all}} y_{ij(k)m}\right]^2}{n_{....}}$$

$$= \sum_{\text{all}} y^2_{ij(k)m} - \frac{Y^2_{....}}{n_{....}} \quad , \qquad (5.2.2)$$

where $\sum_{\text{all}}$ denotes the sum over all the observations, and $n_{....}$ is the total number of observations, i.e., is equal to $n\ [\frac{(p+4)(p-1)}{2}]$.

2. Sum of squares for subtotal as shown in Table 1:

$$SS(\text{Subtotal}) = \sum_{i,j,(k)} \frac{Y^2_{ij(k).}}{n} - \frac{Y^2_{....}}{n_{....}} \quad . \qquad (5.2.3)$$

3. Sum of squares for error:

$$SS(E) = SS(\text{Total}) - SS(\text{Subtotal}) \quad . \qquad (5.2.4)$$

A simple preliminary step common to all methods of analysis is to separate the variance within cells from the variance between cells. Table 2 gives the analysis of variance for this preliminary step.

TABLE 2

Preliminary Analysis of Variance

Source	d.f.	S.S.	M.S.	F
Sub-totals	$\frac{(p+4)(p-1)}{2} - 1$	$\sum_{i,j,(k)} \frac{Y^2_{ij(k).}}{n} - \frac{Y^2_{....}}{n_{....}}$	s^2_S	s^2_S/s^2_e
Error	$n_{....} - \frac{(p+4)(p-1)}{2}$	Subtraction	s^2_e	
Total	$n_{....} - 1$ $= n\frac{(p+4)(p-1)}{2} - 1$	$\sum_{all} Y^2_{ij(k)m} - \frac{Y^2_{....}}{n_{....}}$		

A non-significant value of F here would mean that there is no noticeable difference between the combinations, i.e., there is no specific treatment effect.

We may consider the data in Table 1 to be a standard Symmetrical Complementation design with one experimental unit per cell, the experimental unit being the cell total. We would then proceed with the computations necessary for analysis as described in Chapters III and IV, however with the observations there replaced by the totals. We may apply either the high-low method of the modified high-low method to obtain the vector of estimates, say $\underline{\hat{\hat{\xi}}}$; the only difference being that the expressions obtained by those methods must each be divided by n, the number of replications, in order to produce $\underline{\hat{\xi}}$. If we call the vector of right-hand sides of the normal equations $\underline{g}$ say, we can find the sum of squares due to lack-of-fit. To obtain this, we must compute the following quantities:

1. Sum of squares for sub-total:

$$SS(\text{Sub-totals}) = \sum_{i,j,(k)} \frac{Y^2_{ij(k).}}{n} - \frac{Y^2_{....}}{n_{....}} \quad .$$

2. Sum of squares due to treatments;

$$SS(\text{Treatments}) = \underline{\hat{\xi}}'\,\underline{g} \quad . \qquad (5.2.5)$$

3. Sum of squares due to lack-of-fit:

$$SS(\text{Lack-of-fit}) = SS(\text{Sub-totals}) = SS(\text{Treatments}) . \qquad (5.2.6)$$

Table 3 below gives the analysis.

TABLE 3

Analysis of Variance for Lack-of-Fit

Source	d.f	S.S.	M.S.	F
Treatments	3p-3	$\underline{\hat{\xi}}'\underline{g}$	s^2_t	s^2_t / s^2_e
Lack-of-fit	$\frac{p(p-3)}{2}$	Subtraction	s^2_e	
Sub-total	$\frac{(p+4)(p-1)}{2} -1$	$\sum_{i,j,(k)} \frac{Y^2_{ij(k)\cdot}}{n} - \frac{Y^2_{....}}{n_{....}}$		

The case of n replications allows a test for goodness-of-fit of the model. We can use as a test statistic the ratio MS(lack-of-fit)/MS(error). If this ratio is non-significant compared with the F-distribution with $[\frac{p(p-3)}{2}]$ and $[n_{....} - \frac{(p+4)(p-1)}{2}]$ degrees of freedom, we say that the model fits, and we can then test hypotheses about estimable functions and sets of linearly independent estimable functions (see Chapters III and IV). We propose that MS(Error) be used as the denominator for these tests for, even if the lack-of-fit component is non-significant, throwing it into the error term does not seem wise in that MS(Error) is a clean estimate of experimental error. If this test should reveal a significant lack-of-fit, the orthodox recommendation would be to abandon the additive model, and rather treat each cell effect as a separate unit, i.e., to reduce the analysis to a one-way classification analysis with $\frac{(p+4)(p-1)}{2}$ treatments and n replicates per treatment. Following another custom, one might deal with this situation by performing significance tests on treatment effect combinations with the mean square due to lack-of-fit as the denominator term in the F-ratio. For a discussion concerning the treatment of the lack-of-fit effects and errors as variance components, see Section 5.5.

5.3 Repetitions over a Period of Time Subject to Trend

The Symmetrical Complementation design is particularly useful in those situations where the levels are weight percent of certain substances. The response, which is a function of the reaction of these three substances in the stated quantities might be read just once after reaction has proceeded to completion. In that case, we would have one observation per cell, or perhaps, a replication under otherwise identical conditions.

In most instances, however, in chemical or nutritional experiments, the experimenter has the opportunity to record values of the response at various successive times, i.e., before a final state is reached, which, for that matter, may never be attained. In nutrition experiments, where analysis of the retention of certain nutrients may be made, perhaps once a week, there is definite evidence for such underlying time trends to exist. It can be explained by saying that the body has to become adapted to a special type of diet, and it would be a waste of experimental data if results during the initial periods were discarded.

Organic reactions in chemistry proceed, usually, at such a slow rate that measurements are taken as a matter of routine at regular or irregular time intervals. Again, in most instances, the logarithm of the yield of the product is a rather simple function of the time of reaction. It is in these situations, as well as in many well known economic ones, that time is treated as a concomitant variable. As a matter of fact, time is probably the most ideal example of a concomitant variable in that its values are, as a rule, known without error.

In the present section, we will assume such an underlying function of time (or, if applicable, time may be replaced by another concomitant variable). For simplicity of development we will first assume that the function of time is a linear one; later, we will consider more general trends. The model, as stated in the analysis of covariance convention, assumes an underlying time trend which affects all observations in the same way. This assumption will be made in the present context with full understanding that with chemical reaction data, time trend may very well be different for different compositions. In chemical reaction kinetics, the logarithm of the concentration of the product is written as a function of the concentration and another function of the time with little or no interactions between the two, so that here the assumption of an underlying trend independent of the combination is well taken.

In the case where the experiments within a cell represent repetitions over a period of time, during which a general time trend may be present, the model takes the form

$$y_{ij(k)m} = \mu + \alpha_i + \beta_j + \gamma_{(k)} + \sum_d \delta_d f_d[t_{ij(k)m}] + e'_{ij(k)} + e_{ij(k)m}, \qquad (5.3.1)$$

where

$i = j = (k) = 1,2,3,\ldots,p,$

$m = 1,2,3,\ldots,n,$

and

$d = 1,2,\ldots,r$.

In this model, $f_d[t_{ij(k)m}]$ refers to a polynomial function of the dth degree of the concomitant variable t, and δ_d represents the regression coefficient (unknown) of y on $f_d(t)$.

We shall discuss the above situation first by considering that a linear trend is present. The necessary formulas for estimable functions and tests of hypotheses will be given. Then we will generalize these results to the situation where a trend of higher order is present.

5.3.1 Linear Trend

If a linear trend is assumed, the model is

$$y_{ij(k)m} = \mu + \alpha_i + \beta_j + \gamma_{(k)} + \delta_1 t_{ij(k)m}$$

$$+ e'_{ij(k)} + e_{ij(k)m} \quad . \qquad (5.3.2)$$

As before, the unknowns are estimated by least squares, utilizing the same linear restrictions on the parameters as were used in the case of one experimental unit per cell (see equations (3.1), (3.2)). In practice, however, we start with a joint analysis of the sums of squares and products of y and t.

The estimation of the parameters proceeds in exactly the same manner as described in Section 5.2. This procedure yields a vector of estimates of effects under the old model, and a vector corresponding to the right-hand sides of the normal equations. Let us call these two vectors $\underline{\hat{\xi}}_y$ (for estimates) and $\underline{g}_y$ (for the right-hand sides). Conditions of estimability remain unchanged.

The analysis of the values of the concomitant variable t follows, by the same methods as given previously for the random variable y. Here we perform the same analysis, replacing the cell totals of the variable y by those of the variable t. This yields again two vectors, $\underline{\hat{\ell}}_t$ and $\underline{q}_t$, say. To analyze the sum of products we carry out the same operations as for a sum of squares, except that at every stage a square is replaced by the corresponding product.

Table 4 below gives the basic notation for dealing with an analysis of covariance with one concomitant variable. For each cell the totals and means for the random variable, and the totals for the concomitant variable, are computed and recorded.

From Table 4, we now compute:

1. Sum of squares and products for total:

$$\text{SS Total }(y) = \sum_{\text{all}} y_{ij(k)m}^2 - \frac{Y_{....}^2}{n_{....}} ,$$

$$\text{SS Total }(t) = \sum_{\text{all}} t_{ij(k)m}^2 - \frac{T_{....}^2}{n_{....}} ,$$

$$\text{SP Total }(y,t) = \sum_{\text{all}} y_{ij(k)m}\, t_{ij(k)m} - \frac{Y_{....}\, T_{....}}{n_{....}} .$$

(5.3.3)

TABLE 4

Basic Notation: Linear Trend

	1	2	3	...	p	=Level of B
1		$Y_{12(p).}$ $T_{12(p).}$ $\bar{Y}_{12(p)}$.	$Y_{13(p-1).}$ $T_{13(p-1).}$ $\bar{Y}_{13(p-1).}$	... ...	$Y_{1p(2).}$ $T_{1p(2).}$ $\bar{Y}_{1p(2)}$.	
2	$Y_{21(p).}$ $T_{21(p).}$ $\bar{Y}_{21(p).}$	$Y_{22(p-1).}$ $T_{22(p-1).}$ $\bar{Y}_{22(p-1).}$	$Y_{23(p-2).}$ $T_{23(p-2).}$ $\bar{Y}_{23(p-2).}$	... ...	$Y_{2p(1).}$ $T_{2p(1).}$ $\bar{Y}_{2p(1).}$	
3	$Y_{31(p-1).}$ $T_{31(p-1).}$ $\bar{Y}_{31(p-1).}$	$Y_{32(p-2).}$ $T_{32(p-2).}$ $\bar{Y}_{32(p-2).}$	$Y_{33(p-3).}$ $T_{33(p-3).}$ $\bar{Y}_{33(p-3).}$	... ...		
⋮						
p-1	$Y_{(p-1)1(3).}$ $T_{(p-1)1(3).}$ $\bar{Y}_{(p-1)1(3)}$.	$Y_{(p-1)2(2).}$ $T_{(p-1)2(2).}$ $\bar{Y}_{(p-1)2(2).}$	$Y_{(p-1)3(1).}$ $T_{(p-1)3(1).}$ $\bar{Y}_{(p-1)3(1)}$.			
p	$Y_{p1(2).}$ $T_{p1(2).}$ $\bar{Y}_{p1(2).}$	$Y_{p2(1)}$. $T_{p2(1)}$. $\bar{Y}_{p2(1)}$.				
Level of A						

= Level of C

Where:

$Y_{ij(k).}$ = total of all y's in cell [i,j,(k)] = $\sum_{m=1}^{n} y_{ij(k)m}$,

$T_{ij(k).}$ = total of all t's in cell [i,j,(k)] = $\sum_{m=1}^{n} t_{ij(k)m}$,

$Y_{....}$ = grand total of all y's; $T_{....}$ = grand total of all t's,

$n_{....}$ = total number of experimental units = $n\,\frac{(p+4)(p-1)}{2}$.

2. Sum of squares and products for sub-total:

$$\text{SS Sub-total}(y) = \sum_{i,j,(k)} \frac{Y^2_{ij(k).}}{n} - \frac{Y^2_{....}}{n_{....}} ,$$

$$\text{SS Sub-total}(t) = \sum_{i,j,(k)} \frac{T^2_{ij(k).}}{n} - \frac{T^2_{....}}{n_{....}} ,$$

$$\text{SP Sub-total}(y,t) = \sum_{i,j,(k)} \frac{Y_{ij(k).}T_{ij(k).}}{n} - \frac{Y_{....}T_{....}}{n_{....}} . \qquad (5.3.4)$$

3. Sum of squares and products for error:

SS Error (y) = SS Total (y) - SS Sub-total (y) ,

SS Error (t) = SS Total (t) - SS Sub-total (t) ,

SP Error (y,t) = SP Total (y,t) - SP Sub-total (y,t).

(5.3.5)

These computations can be tabulated in the form shown in Table 5 below.

TABLE 5

Computations for Covariance Model

Source	d.f.	y^2	yt	t^2
Sub-Total	$\frac{(p+4)(p-1)}{2} -1$	$\sum_{i,j,(k)} \frac{Y^2_{ij(k).}}{n} - \frac{Y^2_{....}}{n_{....}}$	$\sum_{i,j(k)} \frac{Y_{ij(k).}T_{ij(k).}}{n} - \frac{Y_{....}T_{....}}{n_{....}}$	$\sum_{i,j(k)} \frac{T^2_{ij(k).}}{n} - \frac{T^2_{....}}{n_{....}}$
Error	$n_{....} - \frac{(p+4)(p-1)}{2}$	Subtraction	Subtraction	Subtraction
Total	$n_{....} -1$	$\sum_{all} y^2_{ij(k)m} - \frac{y^2_{....}}{n_{....}}$	$\sum_{all} y_{ij(k)m}t_{ij(k)m} - \frac{Y_{....}T_{....}}{n_{....}}$	$\sum_{all} t^2_{ij(k)m} - \frac{T^2_{....}}{n_{....}}$

If we denote the sums of squares and products for error by E_{yy}, E_{tt}, and E_{yt}, respectively, the regression coefficient of y on t is given by

$$\hat{\delta}_1 = \frac{E_{yt}}{E_{tt}} . \tag{5.3.6}$$

To remove the effect of the regression on t, we first obtain the sum of squares due to regression, which is

$$\text{SS Regression} = \frac{[\text{SP Error}(y,t)]^2}{\text{SS Error }(t)} = \frac{E^2_{yt}}{E_{tt}} . \tag{5.3.7}$$

The error sum of squares is found by subtracting the sum of squares due to regression from the sum of squares for error of the random variable y, i.e.,

$$\text{SS (Error)} = \text{SS (Error) (y)} - \text{SS (Regression)}$$

$$= E_{yy} - \frac{E_{yt}^2}{E_{tt}} \quad . \tag{5.3.8}$$

Then

$$\text{MS (Error)} = \text{SS(Error)} \Big/ \left[n_{....} - \frac{(p+4)(p-1)}{2} - 1\right] \quad .$$

To test the hypothesis H_o: $\delta_1 = 0$, a test criterion is

$$F = \frac{\text{SS (Regression)}}{\text{MS(Error)}} \quad .$$

This is compared with the F-distribution with 1 and $[n_{....} - \frac{(p+4)(p-1)}{2} - 1]$ degrees of freedom. If this is non-significant, we may disregard t and revert to the simple analysis described in Section 5.2.

The high-low method or modified high-low method must now be applied to the y-observations and to the set of observations on the concomitant variable t, so that we will have vectors $\hat{\underline{\xi}}_y$ and $\hat{\underline{\xi}}_t$ and right-hand sides $\underline{q}_y$ and $\underline{q}_t$. The sum of squares due to treatments is obtained in the following manner:

1. Obtain the matrix due to treatments, given by

$$T = \begin{bmatrix} \hat{\underline{\xi}}'_y \underline{g}_y & \hat{\underline{\xi}}'_t \underline{g}_y \\ & \hat{\underline{\xi}}'_t \underline{g}_t \end{bmatrix} .$$

2. Add, to the elements of the matrix T, the corresponding elements of the matrix due to error, given by

$$E = \begin{bmatrix} E_{yy} & E_{yt} \\ & E_{tt} \end{bmatrix} .$$

3. Obtain the sum of squares due to treatments plus error as

$$SS(T+E) = [E_{yy} + \hat{\underline{\xi}}'_y \underline{g}_y] - \frac{[E_{yt} + \hat{\underline{\xi}}'_t \underline{g}_y]^2}{[E_{tt} + \hat{\underline{\xi}}'_t \underline{g}_t]} . \qquad (5.3.9)$$

4. Subtract, from SS(T+E), the sum of squares due to error, as given by equation (5.3.8). Thus, the sum of squares due to treatments is given by

$$SS(T) = SS(T+E) - SS(E) . \qquad (5.3.10)$$

This difference, divided by (3p-3), yields the mean square for treatments.

To test for lack-of-fit, the required sum of squares can be obtained in an analogous way with the elements of the matrix (T+E) replaced by the elements of the matrix (L+E), where the elements of the matrix L are obtained by subtracting from the elements of the sums of squares and product matrix for sub-totals, the corresponding elements of the T matrix just described. There are, as before, $p(p-3)/_2$ degrees of freedom due to lack-of-fit. The test statistic would be

$$F = \frac{\text{MS(lack-of-fit)}}{\text{MS(Error)}} ,$$

with $[\frac{p(p-3)}{2}]$ and $[n_{....} - \frac{(p+4)(p-1)}{2} -1]$ degrees of freedom. Interpretation of significance is the same as stated in Section 5.2.

Th estimates of the parameters, corrected for regression, are most easily obtained from the expressions:

$$\begin{aligned} \hat{\alpha}_i &= \hat{\alpha}_{iy} - \hat{\delta}_1 \hat{\alpha}_{it} , \\ \hat{\beta}_j &= \hat{\beta}_{jy} - \hat{\delta}_1 \hat{\beta}_{jt} , \qquad (5.3.11) \\ \hat{\gamma}_{(k)} &= \hat{\gamma}_{(k)y} - \hat{\delta}_1 \hat{\gamma}_{(k)t} , \end{aligned}$$

and

$$\hat{\mu} = \hat{\mu}_y - \hat{\delta}_1 \hat{\mu}_t ,$$

where the subscripts y and t denote estimates of the y- and t- variables, respectively, as described on pages 151 and 152.

To test hypotheses on estimable functions, we proceed as follows:

In each of the sums of squares due to single or multiple contrasts, there are expressions involving $\hat{\alpha}_i$, $\hat{\beta}_j$ and $\hat{\gamma}_{(k)}$. (See formulas in Sections 4.2 and 4.3, Chapter IV.)

As in the usual analysis of covariance, three expressions can be found from these formulas:

1. Sum of squares (y) due to hypothesis:

   Replace each $\hat{\alpha}_i$, $\hat{\beta}_j$, and $\hat{\gamma}_{(k)}$ in the formulas for testable hypotheses by $\hat{\alpha}_{iy}$, $\hat{\beta}_{jy}$, and $\hat{\gamma}_{(k)y}$.

2. Sum of squares (t) due to hypothesis:

   Replace each $\hat{\alpha}_i$, $\hat{\beta}_j$, and $\hat{\gamma}_{(k)}$ in the formulas for testable hypotheses by $\hat{\alpha}_{it}$, $\hat{\beta}_{jt}$, and $\hat{\gamma}_{(k)t}$.

3. Sum of products due to hypothesis:

   Replace one of each of $\hat{\alpha}_i$, $\hat{\beta}_j$, and $\hat{\gamma}_{(k)}$ occurring in the formulas (they always occur in pairs) by $\hat{\alpha}_{iy}$, $\hat{\beta}_{jy}$, and $\hat{\gamma}_{(k)y}$; the other by $\hat{\alpha}_{it}$, $\hat{\beta}_{jt}$, and $\hat{\gamma}_{(k)t}$.

The procedure is shown in Table 6 below:

TABLE 6

Computations for Test of Hypothesis: Linear Trend

Source	(y^2)	(t,y)	(t^2)	$(y/t)_{SS}$	d.f
Hypothesis (H)	H_{yy}	H_{yt}	H_{tt}	SSH	n_h
Error (E)	E_{yy}	E_{yt}	E_{tt}	SSE	$n_{....} - \frac{(p+4)(p-1)}{2} - 1$
H + E	$(H+E)_{yy}$	$(H+E)_{yt}$	$(H+E)_{tt}$	SS(H+E)	

Theexpressions to the left of the vertical line are obtained as already outlined. A new line is formed in the analysis by adding the items for hypothesis and error. From this value for (y^2), we now subtract the contribution due to a regression on t, i.e.

$$\frac{[(H+E)_{yt}]^2}{(H+E)_{tt}} ,$$

and obtain the residual sum of squares for H+E, given by

$$(H+E)_{yy} - \frac{[(H+E)_{yt}]^2}{(H+E)_{tt}} . \qquad (5.3.12)$$

We now subtract from this expression the SS(Error) given by equation (5.3.8); and thus obtain the sum of squares due to the hypothesis with n_h degrees of freedom. This follows the customary analysis of covariance technique.

5.3.2 Trend of Higher Order

If a higher order trend is assumed, the model is

$$y_{ij(k)m} = \mu + \alpha_i + \beta_j + \gamma_{(k)} + \sum_{d=1}^{r} \delta_d f_d (t_{ij(k)m})$$

$$+ e'_{ij(k)} + e_{ij(k)m} \quad , \qquad (5.3.13)$$

where $f_d(t)$ denotes a dth degree polynomial in t. For simplicity, let it be denoted by t_d.

Table 7 below gives the basic notation for dealing with an analysis of this model, i.e., an analysis of covariance with more than one concomitant variable. For each cell, the totals and means for the random variable y and the totals for a typical concomitant variable t_d are computed and recorded.

From Table 7, we now compute the sums of squares and products for total and subtotal, as given by

1. Sums of squares and products for total:

$$\text{SS Total } (y) = \sum_{all} y^2_{ij(k)m} - \frac{Y^2_{....}}{n_{....}}$$

$$\text{SS Total } (t_d) = \sum_{all} t^2_d - \frac{T^2_{....d}}{n_{....}}$$

$$\text{SP Total } (y, t_d) = \sum_{all} yt_d - \frac{Y_{....} T_{....d}}{n_{....}}$$

$$\text{SP Total } (t_c, t_d) = \sum_{all} t_c t_d - \frac{T_{....c} T_{....d}}{n_{....}} \quad (c \neq d),$$

$$(5.3.14)$$

TABLE 7

Basic Notation: Trend of Higher Order

	1	2	3	...	p =Level of B
1		$Y_{12(p).}$ $T_{12(p).d}$ $\bar{y}_{12(p).}$	$Y_{13(p-1).}$ $T_{13(p-1).d}$ $\bar{y}_{13(p-1).}$	... ...	$Y_{1p(2).}$ $T_{1p(2).d}$ $\bar{y}_{1p(2).}$
2	$Y_{21(p).}$ $T_{21(p).d}$ $\bar{y}_{21(p).}$	$Y_{22(p-1).}$ $T_{22(p-1).d}$ $\bar{y}_{22(p-1).}$	$Y_{23(p-2).}$ $T_{23(p-2).d}$ $\bar{y}_{23(p-2).}$	... ...	$Y_{2p(1).}$ $T_{2p(1).d}$ $\bar{y}_{2p(1).}$
3	$Y_{31(p-1).}$ $T_{31(p-1).d}$ $\bar{y}_{31(p-1).}$	$Y_{32(p-2).}$ $T_{32(p-2).d}$ $\bar{y}_{32(p-2).}$	$Y_{33(p-3).}$ $T_{33(p-3).d}$ $\bar{y}_{33(p-3).}$	... ...	
⋮					
p-1	$Y_{(p-1)1(3).}$ $T_{(p-1)1(3).d}$ $\bar{y}_{(p-1)1(3).}$	$Y_{(p-1)2(2).}$ $T_{(p-1)2(2).d}$ $\bar{y}_{(p-1)2(2).}$	$Y_{(p-1)3(1).}$ $T_{(p-1)3(1).d}$ $\bar{y}_{(p-1)3(1).}$		
p Level of A	$Y_{p1(2).}$ $T_{p1(2).d}$ $\bar{y}_{p1(2).}$	$Y_{p2(1).}$ $T_{p2(1).d}$ $\bar{y}_{p2(1).}$			

= Level of C

Where: $Y_{ij(k).}$ = total of all y's in cell i, j, (k) = $\sum_{m=1}^{n} y_{ij(k)m}$,

$T_{ij(k).d}$ = total of all t_d's in cell i, j, (k) = $\sum_{m=1}^{n} t_{ij(k).d}$,

$Y_{....}$ = grand total of all y's; $T_{....d}$ = grand total of all t_d's

$n_{....}$ = total number of experimental units (=total number of y-observations).

where

$$(d = 1, 2, \ldots, r) \quad \text{and}$$

2. Sums of squares and products for sub-total:

$$\text{SS Sub-total } (y) = \sum_{i,j,(k)} \frac{Y^2_{ij(k).}}{n} - \frac{Y^2_{....}}{n_{....}}$$

$$\text{SS Sub-total } (t_d) = \sum_{i,j,(k)} \frac{T^2_{ij(k).d}}{n} - \frac{T^2_{....d}}{n_{....}}$$

$$\text{SP Sub-total } (y, t_d) = \sum_{i,j,(k)} \frac{Y_{ij(k).}T_{ij(k).d}}{n} - \frac{Y_{....}T_{....d}}{n_{....}}$$

$$\text{SP Sub-total } (t_c, t_d) = \sum_{i,j,(k)} \frac{T_{ij(k).c}T_{ij(k).d}}{n} - \frac{T_{....c}T_{....d}}{n_{....}} \quad (c \neq d), \tag{5.3.15}$$

where

$$(d = 1, 2, \ldots, r) \quad .$$

We now compute and record the sums of squares and products due to error in the following matrix form, augmented with a column of sums of products of the random variable with the concomitant variables:

$$E\,\underline{e}_y = \left[\begin{array}{ccccc|c} E_{11} & E_{12} & E_{13} & \cdots & E_{1r} & E_{1y} \\ & E_{22} & E_{23} & \cdots & E_{2r} & E_{2y} \\ & & E_{33} & \cdots & E_{3r} & E_{3y} \\ & & & & \vdots & \vdots \\ & & & & E_{rr} & E_{ry} \end{array}\right] , \tag{5.3.16}$$

where

$$E_{dd} = \text{SS Total } (t_d) - \text{SS Sub-total } (t_d) \quad ,$$

$$E_{cd} = \text{SP Total } (t_c, t_d) - \text{SP Sub-total } (t_c, t_d) \quad ,$$

$$E_{dy} = \text{SP Total } (y, t_d) - \text{SP Sub-total } (y, t_d) \quad .$$

The solutions of this equation system, conveniently obtained by the Doolittle method, are the regression coefficients $\hat{\delta}_1$, $\hat{\delta}_2$, ..., $\hat{\delta}_r$. Then, the sum of squares due to regression is given by

$$\text{SS Regression} = \sum_{d=1}^{r} E_{dy} \hat{\delta}_d \quad , \qquad (5.3.17)$$

with r degrees of freedom. Finally, the sum of squares due to error is given by

$$\text{SS(Error)} = E_{yy} - \text{SS Regression} \quad , \qquad (5.3.18)$$

with $[n_{\cdots\cdot} - \frac{(p+4)(p-1)}{2} - r]$ degrees of freedom.

One may test for significance of regression due to fitting all the concomitant variables, or significance due to any concomitant variable over that obtained for previous ones.

The high-low method or modified high-low method must now be applied to the y-observations and to each set of observations on a concomitant variable, so that we will have vectors $\hat{\xi}_y$, $\hat{\xi}_{t_1}$, $\hat{\xi}_{t_2}$, ... , $\hat{\xi}_{t_r}$, and right-hand sides $\underline{g}_y$, $\underline{g}_t$, $\underline{g}_{t_2}$, ... , $\underline{g}_{t_r}$. The sum of squares due to treatments is obtained in the following manner:

1. Form a matrix T whose elements are $\underline{\hat{\xi}}'_{t_c}\ \underline{g}_{t_d}$. This matrix will be symmetric. Call its elements T_{cd} .

2. Form a right-hand side vector $\underline{\hat{\xi}}'_y\ \underline{g}_{t_d}$, $d = 1,2, \ldots, r$, and call it $\underline{t}_y$.

3. Obtain the matrix (T+E), and the right-hand side $(\underline{t}_y + \underline{e}_y)$, and solve this system of equations in the same way as described for E.

Let the solutions be $\hat{\eta}_d$ $(d = 1,2,\ldots,r)$. Then the sum of squares due to treatments is given by

$$[E_{yy} + \underline{\hat{\xi}}'_y\ \underline{g}_y] - \sum_{d=1}^{r} \hat{\eta}_d\ (\underline{t}_y + \underline{e}_y)_d - SS(Error) \qquad (5.3.19)$$

with (3p-3) degrees of freedom. To test for lack-of-fit the sum of squares can be obtained in an analogous manner with the elements of the matrix (T+E) replaced by the elements of the matrix (L+E), where the elements of L are obtained by subtracting the elements of the T-matrix just described from the elements of the sums of squares and products matrix for sub-totals. There are, as before, $p(p-3)/_2$ degrees of freedom due to lack-of-fit.

The test statistic would be

$$F = \frac{\text{MS(lack-of-fit)}}{\text{MS(error)}} \quad ,$$

with $\frac{p(p-3)}{2}$ and $(n_{\cdots} - \frac{(p+4)(p-1)}{2} - r)$ degrees of freedom.

Interpretation of significance is the same as stated in Section 5.2.

The estimates of the population parameters are most easily obtained from the expressions

$$\begin{aligned} \hat{\alpha}_i &= \hat{\alpha}_{iy} - \sum_{d=1}^{r} \hat{\delta}_d \, \hat{\alpha}_{it_d} \quad , \\ \hat{\beta}_j &= \hat{\beta}_{jy} - \sum_{d=1}^{r} \hat{\delta}_d \, \hat{\beta}_{jt_d} \quad , \\ \hat{\gamma}_{(k)} &= \hat{\gamma}_{(k)y} - \sum_{d=1}^{r} \hat{\delta}_d \, \hat{\gamma}_{(k)t_d} \quad , \end{aligned} \qquad (5.3.20)$$

and

$$\hat{\mu} = \hat{\mu}_y - \sum_{d=1}^{r} \hat{\delta}_d \, \hat{\mu}_{t_d} \quad .$$

To test hypotheses on estimable functions, we proceed as follows:

In each of the sums of squares due to single or multiple contrasts, there are expressions involving $\hat{\alpha}_i$, $\hat{\beta}_j$, and $\hat{\gamma}_{(k)}$ (see formulas in Sections 4.2 and 4.3, Chapter 4). As in the usual analysis of covariance, three types of expressions can be found from these formulas:

1) Sums of squares (y) due to hypothesis: Replace each $\hat{\alpha}_i$, $\hat{\beta}_j$, and $\hat{\gamma}_{(k)}$ in the formulas for testable hypotheses by $\hat{\alpha}_{iy}$, $\hat{\beta}_{jy}$, and $\hat{\gamma}_{(k)y}$.

2) Sums of squares (t_d) due to hypothesis:

Replace each $\hat{\alpha}_i$, $\hat{\beta}_j$, and $\hat{\gamma}_{(k)}$ in the formulas by $\hat{\alpha}_{it_d}$, $\hat{\beta}_{jt_d}$, and $\hat{\gamma}_{(k)t_d}$ $(d = 1,2,\ldots,r)$.

3) Sums of products due to hypothesis: Replace one of each of $\hat{\alpha}_i$, $\hat{\beta}_j$, and $\hat{\gamma}_{(k)}$ occurring in the formulas (they always occur in pairs) by $\hat{\alpha}_{iy}$, $\hat{\beta}_{jy}$, and $\hat{\gamma}_{(k)y}$; the other by $\hat{\alpha}_{it_d}$, $\hat{\beta}_{jt_d}$, and $\hat{\gamma}_{(k)t_d}$ $(d = 1,2,\ldots,r)$. Also replace one of each of $\hat{\alpha}_i$, $\hat{\beta}_j$, and $\hat{\gamma}_{(k)}$ by $\hat{\alpha}_{it_c}$, $\hat{\beta}_{jt_c}$, and $\hat{\gamma}_{(k)t_c}$; the other by $\hat{\alpha}_{it_d}$, $\hat{\beta}_{jt_d}$, and $\hat{\gamma}_{(k)t_d}$ $(c \neq d)$.

The procedure is shown in Table 8 below.

TABLE 8

Computations for Covariance Model: General Trend

Source	y^2	t_1^2	t_2^2	...	t_r^2	t_1t_2	...	$t_{r-1}t_r$	t_{1y}	...	t_{ry}
Hypothesis (H)	H_{yy}	H_{11}	H_{22}	...	H_{rr}	H_{12}	...	H_{r-1r}	H_{1y}	...	H_{ry}
Error (E)	E_{yy}	E_{11}	E_{22}	...	E_{rr}	E_{12}	...	E_{r-1r}	E_{1y}	...	E_{ry}
H+E	$(H+E)_{yy}$	$(H+E)_{11}$	$(H+E)_{22}$	...	$(H+E)_{rr}$	$(H+E)_{12}$	...	$(H+E)_{r-1r}$	$(H+E)_{1y}$	...	$(H+E)_{ry}$

In the table the H_{cd} and H_{dy} represent the sums of squares and products due to the hypothesis H_o, as outlined above; and the E_{cd} and E_{dy} represent the sums of squares and products due to error (actually, these have already been found). The line (H+E) is formed by adding the items for hypothesis and error.

Next, obtain the Doolittle solution for the matrix (H+E) in the same way as described for the matrix E, thereby obtaining the sums of squares for (H+E), corrected. Subtract from this expression, the sum of squares due to error, as given by equation (5.3.18). This difference, divided by n_h degrees of freedom, gives the mean square due to the hypotheses H_o. The test statistic is then

$$F = \frac{MS(Hypothesis)}{MS(Error)} ,$$

with n_h and $[n_{\cdots} - \frac{(p+4)(p-1)}{2} - r]$ degrees of freedom.

This procedure can be conveniently shown in Table 9 below.

TABLE 9

Computations for Test of Hypothesis: General Trend

$^{y}/_{\underline{t}}$	d.f	MS	F
SSH	n_h	$SSH/_{n_h}$	$\frac{MS(H)}{MS(E)}$
SSE	$n_{....} - \frac{(p+4)(p-1)}{2} - r$	$SSE/_{(n_{....} - \frac{(p+4)(p-1)}{2} - r)}$	
SS(H+E)			

5.4 Multivariate Situation

In this section, we will show the analogy between univariate and multivariate analysis of variance (Roy (1957)). In this case, the replications in each cell represent several experiments or responses with the same experimental units. Here we may not have independence among the observations within a cell. On the assumption that the type of dependence is the same from one experimental cell to another, this case is readily extended to a multivariate situation.

The algebra of the multivariate case is essentially the same as that of the univariate case. This leads to estimation theory that is analogous to that of the univariate case and to test criteria that are analogous to Beta tests* in the univariate case. In fact, since we have already presented uni-

---

* Note that $MSH/_{MSE} = F(n_h, n_e)$ (upper tail) implies $\frac{SSE}{SS(H+E)} = B(\frac{n_e}{2}, \frac{n_h}{2})$ (lower tail).

variate test statistics, we will be able to immediately write down a corresponding multivariate test statistic. However, in the multivariate case, we may consider additional tests of significance.

In the traditional methods of multivariate analysis we usually have a matrix of observations Y, of order (n x q), where n is the number of experimental units and q the number of variables. We assume that each row vector of Y represents an observation from a random variable $\underline{y}_i'$ which is normal with some expectation $E(\underline{y}_i')$ and a covariance matrix Σ which is the same for each row vector. Corresponding to the general linear model in the univariate case,

$$E\,(\underline{y})\quad (n \times 1) = A\ (n \times m)\ \ \underline{\xi}\ (m \times 1), \qquad (5.4.1a)$$

(where A is the design matrix, and $\underline{\xi}$ a vector of parameters), and a general linear hypothesis

$$H_o :\ P\underline{\xi} = \underline{k} \quad \text{vs} \quad P\underline{\xi} \neq \underline{k}\quad , \qquad (5.4.2a)$$

[where P is a known matrix of order (s x m) and $\underline{k}$ is a known column vector of order (s x 1)], there exists in multivariate analysis a general linear model

$$E(Y)\quad (n \times q)\ = A\ (\, n \times m)\ \Phi\ (\, m \times q) \qquad (5.4.1b)$$

and a general linear hypothesis

$$H_o :\ P\Phi = K \quad \text{vs} \quad P\Phi \neq K\quad , \qquad (5.4.2b)$$

[where P is a known matrix of order (s x m) and K is a known matrix of order (s x q) usually all zeros].

Notice that the matrix A is the same regardless of whether we deal with a univariate or multivariate situation. It represents a model matrix describing the nature of the experimental units. In univariate analysis, we have one response per unit; in multivariate analysis, q responses per experimental unit. Consequently, all remarks regarding estimability and testability, which are characteristics of the A matrix, apply to the multivariate situations without change. In univariate analysis,

$$\underline{\hat{\xi}} = (A'A)^{-1} A' \underline{y} \quad , \qquad (5.4.3a)$$

where A'A is the (properly constrained) matrix of the normal equations.

The sum of squares due to error was

$$SSE = \underline{y}'\underline{y} - \underline{\hat{\xi}}' A' \underline{y} \quad , \qquad (5.4.4a)$$

while the sum of squares due to hypothesis was

$$SSH = (P\underline{\hat{\xi}})' \, [P(A'A)^{-1}P']^{-1} \, (P\underline{\hat{\xi}}) \; . \qquad (5.4.5a)$$

In the multivariate case, the estimate of Φ is given as the matrix analogue of (5.4.3a),

$$\hat{\Phi} = (A'A)^{-1} A' Y \; . \qquad (5.4.3b)$$

The formulas for the sums of squares are again analogous, but here we must consider sums of squares and products. Consider the two matrices

$$H = (P\hat{\Phi})' \; [\, P(A'A)^{-1}P'] \,^{-1} \; (P\hat{\Phi}) \qquad (5.4.4b)$$

and

$$E = Y'Y - \hat{\Phi}' A' Y . \qquad (5.4.5b)$$

Here the matrix H "due to hypothesis" of order (q x q) has diagonal elements which are the sums of squares due to the hypothesis for each of the variables, and off-diagonal elements which are the sums of products due to the hypotheses between any pair of variables, obtained by the same method. The computation of these is familiar from the analysis of covariance. The matrix E of order (q x q) has diagonal elements which are the sums of squares due to error for each variable, and off-diagonal elements which are the sums of products due to error between any pair of variables. This introduction is merely made in order to state that, whenever we know a method of analyzing a univariate case, we will be able to analyze the corresponding multivariate case by the straightforward extension of numerical techniques used in the former.

We shall discuss two test criteria for testing the general linear hypothesis,

a) the likelihood ratio test, and

b) the union - intersection test.

In general, the correspondence between univariate and multivariate hypothesis testing, no matter what the experimental design, is as follows: the sum of squares due to hypothesis in univariate analysis is replaced by the matrix H consisting of corresponding sums of squares and products of the q variables. The error sum of squares in univariate analysis is replaced by the matrix E whose elements are the error sums of squares and products of the q variables. We may test the general linear hypothesis, by either the likelihood-ratio test, using

$\frac{|E|}{|H+E|}$ as a test statistic, or by the union - intersection test, which employs the largest characteristic root of $E^{-1}H$ as a test statistic. The procedure to be applied in each of these tests will be discussed in detail.

Let us now consider a single replication of a Symmetrical Complementation design experiment, with p levels of each factor. Under each experimental condition (i.e., in each cell) q responses are measured. We may thus express the observations on each experimental unit by a vector with q elements,

$$\underline{y}'_{ij(k)} = [\, y^{(1)}_{ij(k)} , y^{(2)}_{ij(k)}, \ldots, y^{(q)}_{ij(k)}] \quad .$$

The $\underline{y}'_{ij(k)}$ vectors are supposed to be independent and

$N\,[\underline{\mu}' + \underline{\alpha}'_i + \underline{\beta}'_j + \underline{\gamma}'_{(k)} \, ; \Sigma]$. $\underline{\mu}'$ is a (1 x q) vector of unknown parameters due to an overall mean effect, $\underline{\alpha}'_i$ is a (1 x q) vector of unknown row effects, $\underline{\beta}'_j$ is a (1 x q) vector of unknown column effects, and $\underline{\gamma}'_{(k)}$ is a (1 x q) vector of unknown diagonal effects, and Σ is a symmetric (q x q) covariance matrix.

We will denote by $y^{(z)}_{ij(k)}$ the observation on the zth variate ($z = 1,2,\ldots,q$) in the ij(k)th cell. Also, denoting by $\mu^{(z)}$ the overall mean effect associated with the zth variate, by $\alpha^{(z)}_i$ the effect in the ith row on the zth variate, by $\beta^{(z)}_j$ the effect in the jth column on the zth variate, and by $\gamma^{(z)}_{(k)}$ the effect in the kth diagonal on the zth variate, we have the model:

$$E\ (Y) \ = \ A\Phi$$

or

$$\begin{bmatrix} y_{12(p)}^{(1)} & \cdots & y_{12(p)}^{(q)} \\ \vdots & & \vdots \\ y_{1p(2)}^{(1)} & \cdots & y_{1p(2)}^{(q)} \\ y_{21(p)}^{(1)} & \cdots & y_{21(p)}^{(q)} \\ \vdots & & \vdots \\ y_{2p(1)}^{(1)} & \cdots & y_{2p(1)}^{(q)} \\ y_{31(p-1)}^{(1)} & \cdots & y_{31(p-1)}^{(q)} \\ \vdots & & \vdots \\ y_{3(p-1)(1)}^{(1)} & \cdots & y_{3(p-1)(1)}^{(q)} \\ \vdots & & \vdots \\ y_{p1(2)}^{(1)} & \cdots & y_{p1\ (2)}^{(q)} \\ y_{p2\ (1)}^{(1)} & \cdots & y_{p2\ (1)}^{(q)} \end{bmatrix} = A \begin{bmatrix} \mu^{(1)} & \cdots & \mu^{(q)} \\ \alpha_1^{(1)} & \cdots & \alpha_1^{(q)} \\ \vdots & & \\ \alpha_p^{(1)} & \cdots & \alpha_p^{(q)} \\ \beta_1^{(1)} & \cdots & \beta_1^{(q)} \\ \vdots & & \\ \beta_p^{(1)} & \cdots & \beta_p^{(q)} \\ \gamma_{(1)}^{(1)} & \cdots & \gamma_{(1)}^{(q)} \\ \vdots & & \\ \gamma_{(p)}^{(1)} & \cdots & \gamma_{(p)}^{(q)} \end{bmatrix},$$

$$\left(\frac{(p+4)\ (p-1)}{2} \times q \right) \qquad\qquad (3p+1 \times q)$$

(5.4.6)

Where the A matrix of order $[\frac{(p+4)\ (p-1)}{2} \times 3p+1]$ is of the same structure as in the univariate case. Hence all the theory developed there applies in this situation, and all the estimable functions developed are applicable if we replace the scalar estimates by vector estimates.

To make a test of a testable hypothesis, we calculate H and E, where H is the (q x q) matrix whose elements are the sums of squares and products due to the hypothesis for the q variables, and E is the (q x q) matrix whose elements are the sums of squares and products due to error. The elements of the matrix H are found in the same way as those described in Table 8, except that we here disregard the y-observations and replace the t_d's by the q variables $y^{(1)}, y^{(2)}, \ldots, y^{(r)}$. An analogous table can be set up.

The two test criteria for testing a general linear hypothesis are:

1. The Likelihood Ratio Test

To perform this test, we form the ratio of the determinant of E, $|E|$, to that of the determinant of $(H+E)$, i.e., $\frac{|E|}{|H+E|}$. If we denote this ratio by Λ, we see that Λ is an extension of the Beta-statistic in univariate analysis. The test procedure is as follows:

$$\text{Obtain} \quad -m \log \Lambda , \tag{5.4.7}$$

$$\text{where } \Lambda = \frac{|E|}{|H+E|} ,$$

and where

$$m = n_e + \frac{1}{2}(n_h - q - 1) ,$$

$$q = \text{number of variables} \tag{5.4.8}$$

n_h = "degrees of freedom" due to hypothesis,

and

n_e = "degrees of freedom" due to error.*

Then

$$\Pr\{-m \log \Lambda > c/_{H_o \text{ true}}\} = \Pr\{\chi^2_f > c\} + \frac{\gamma}{m^2}[\Pr\{\chi^2_{f+4} > c\} - \Pr\{\chi^2_f > c\}] + o(\frac{1}{m^4}) , \qquad (5.4.9)$$

where

$$f = qn_h \ , \qquad (5.4.10)$$
$$\gamma = qn_h \ (q^2 + n_h^2 - 5)/_{48}$$

2. The Union - Intersection Test

To perform the test, we first form the matrix $E^{-1}H$. Now let the largest characteristic root of $E^{-1}H$ be denoted by c_s, where $s = \min (q, n_h)$. Then

$$\theta = \frac{c_s}{1 + c_s} \qquad (5.4.11)$$

---

* The term "degrees of freedom" may be somewhat ambiguous in this context. In our formulation, n_h and n_e are the same numbers which one would use in a univariate analysis, regardless of the number of responses.

has the standard distribution of the largest characteristic root with parameters

$$s = \min(q, n_h) \quad ,$$

$$m = \frac{n_h - q - 1}{2} \quad , \qquad (5.4.12)$$

and

$$n = \frac{n_e - q - 1}{2} \quad .$$

The test of H_o is to accept H_o at the α-level if $\frac{c_s}{1 + c_s} = \theta \leq x_\alpha$, and reject otherwise, where x_α is determined from the distribution of the largest root such that

$$P\{ \theta \leq x_\alpha \} = 1 - \alpha \quad .$$

If $2 \leq s \leq 5$, x_α , for a given α, may be obtained in entering charts (see Heck (1958)) with the above mentioned parameters. If $s = 1$, then m and n should be increased by unity in order to obtain the parameters for the Incomplete Beta-Function.

In the Likelihood Ratio Test, we need to find the determinants $|H+E|$ and $|E|$. This can most easily be accomplished by obtaining the Doolittle solution for the matrices $H+E$ and E, respectively. (See Section 5.3.2.) The product of the "leading coefficients" in the left-hand side of the Doolittle solution is the value of the determinant. The term "leading coefficient" denotes the first term of the second to last row in each cycle of the Doolittle Solution. However, in the test statistic, we need

$$\log \frac{|E|}{|H+E|} = \log |E| - \log |H+E| \quad .$$

Since $|E|$ is given by the product of the "leading coefficients", $\log |E|$ is given by the sum of the logarithms of the "leading coefficients". The same applies to $(H+E)$.

After rejection of some general linear hypothesis in multivariate analysis it is frequently desirable to ascertain how much each variable contributes to the observed differences. The step-down procedure - i.e., a process analogous to the analysis of covariance in which we study the first variable, the second given the first, etc. - depends on the ordering of the variables. This procedure, (Roy (1958)), employs the ratio of "leading coefficients" in the left-hand side of the Doolittle solution of E to that of (H+E). Under H_o, these ratios have a Beta distribution, and they are all independent of each other. A detailed demonstration study will be presented for illustration.

5.5 Mixed Model

In the following we will assume an experiment with n replications per cell, i.e., the same situation as before. However, it will be assumed that the "lack-of-fit" or "residual" component is a random variable with expectation 0 and variance $\sigma^2_{e'}$. The model will then be

$$y_{ij(k)m} = \mu + \alpha_i + \beta_j + \gamma_{(k)} + e'_{ij(k)} + e_{ij(k)m} \quad . \qquad (5.5.1)$$

In other words, the observations in the [i, j, (k)] cell are assumed to be affected by a common random variable $e'_{ij(k)}$ in addition to the fixed effects. Now, let $\mathrm{var}(e'_{ij(k)}) = \sigma^2_{e'}$; $\mathrm{var}(e_{ij(k)m}) = \sigma^2_e$; and all covariances between the e' 's and the e's be zero. Writing all observations as a column vector, $\underline{y}$, we then have

$$\mathrm{var}(\underline{y}) = \sigma^2_e I + \sigma^2_{e'} MM' , \qquad (5.5.2)$$

which denotes the variance-covariance matrix of all observations. M is a matrix of the form

$$\begin{bmatrix} \underline{j} & \underline{o} & \cdots & \underline{o} \\ \underline{0} & \underline{j} & \cdots & \underline{o} \\ & & \cdots & \\ \underline{o} & \underline{o} & \cdots & \underline{j} \end{bmatrix} ,$$

where $\underline{j}$ denotes a column vector with n elements, all equal to unity. The order of this matrix is

$$\left[\frac{n\ (p+4)(p-1)}{2} \quad x \quad \frac{(p+4)(p-1)}{2} \right] .$$

Since each category (cell or treatment combination) is repeated an equal number of times, and since covariances, all equal to $\sigma^2_{e'}$, occur only within cells, the situation is fully comparable to a randomized block experiment, i.e., estimation procedures will be invariant under the proposed change of model.

To obtain an estimate of σ_e^2, , we need to know the expected mean squares. To facilitate notation, let

$$D = \begin{bmatrix} (p-1) & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & p & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & (p-1) & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & (p-2) & \cdots & 0 & 0 \\ & & & \cdots & & & \\ 0 & 0 & 0 & 0 & \cdots & 3 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 2 \end{bmatrix} ,$$

i.e., the diagonal part of the normal equations. Also let

$$E = \begin{bmatrix} 0 & 1 & 1 & 1 & \cdots & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & \cdots & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & \cdots & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & \cdots & 1 & 0 & 0 \\ & & & \cdots & & & & \\ 1 & 1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix} ,$$

i.e., the off-diagonal part of the system of normal equations. Also let

$$\underline{z}' = [\ (p-1) \quad p \quad (p-1) \quad (p-2) \quad \cdots \quad 3 \quad 2\ ] \quad .$$

We will now assume the constraint made in the modified high-low method, namely,

$$\underline{z}'\,\hat{\underline{\alpha}}_i + \underline{z}'\,\hat{\underline{\beta}}_j + \underline{z}'\,\hat{\underline{\gamma}}_{(k)} = 0 \quad .$$

Subject to this, the expected grand total is

$$E(Y_{....}) = n_{....} \mu .$$

Hence

$$E\left(\frac{Y^2_{....}}{n_{....}}\right) = \frac{1}{n_{....}} \{ \text{var}(Y_{....}) + [E(Y_{....})]^2\}$$

$$= \sigma_e^2 + n \sigma_{e'}^2 + n_{....} \mu^2 . \qquad (5.5.3)$$

Then

$$E(\text{SS Total}) = E(\underline{y}'\underline{y} - \frac{Y^2_{....}}{n_{....}})$$

$$= (n_{....} -1) \sigma_e^2 + (n_{....} -n) \sigma_{e'}^2$$

$$+ n(\underline{\alpha}' D\underline{\alpha} + \underline{\beta}'D\underline{\beta} + \underline{\gamma}'D\underline{\gamma}$$

$$+ 2\underline{\alpha}'E\underline{\beta} + 2\underline{\alpha}'E\underline{\gamma} + 2\underline{\beta}'E\underline{\gamma}) .$$

Let $\underline{C} = M'\underline{y}$ denote the vector of cell totals. Then

$$\text{SS Subtotal} = \frac{1}{n} \underline{C}' \underline{C} - \frac{Y^2_{....}}{n_{....}} .$$

This yields

$$\text{var}(\underline{C}) = M'[\sigma_e^2 + MM' \sigma_{e'}^2] M$$

$$= n [\sigma_e^2 I + n \sigma_{e'}^2 I] .$$

Thus, recalling that

$$E(\underline{C}'\underline{C}) = \text{tr var } (\underline{C}) + E(\underline{C}')\ E\ (\underline{C}) \quad ,$$

we have

$$E\ [\text{SS Subtotal}]$$

$$= [\frac{(p+4)(p-1)}{2} -1]\ \sigma_e^2 + (n_{....} - n)\ \sigma_{e'}^2$$

$$+ n(\underline{\alpha}'D\underline{\alpha} + \underline{\beta}'D\underline{\beta} + \underline{\gamma}'D\underline{\gamma} + 2\underline{\alpha}'E\underline{\beta}$$

$$+ 2\underline{\alpha}'\ E\underline{\gamma} + 2\underline{\beta}'E\underline{\gamma}\ ) \quad .$$

Thus, as expected,

$$E[\text{SS Error}] = E[\text{SS Total}] - E[\text{SS Subtotal}]$$

$$= [n_{....} - \frac{(p+4)(p-1)}{2}]\ \sigma_e^2$$

$$= (n-1)\ \frac{(p+4)(p-1)}{2}\ \sigma_e^2 \quad . \qquad (5.5.4)$$

and

$$E\ (\text{MS Error}) = \sigma_e^2 \quad . \qquad (5.5.5)$$

Subject to the constraint

$$\underline{z}'\ \hat{\underline{\alpha}}_i + \underline{z}\ \hat{\underline{\beta}}_j + \underline{z}'\ \hat{\underline{\gamma}}_{(k)} = 0 \quad ,$$

$$E(\text{SS Treatments}) = \frac{1}{n}\ E\ [\ \hat{\underline{g}}'\ \underline{g}\ ] \quad ,$$

where $\hat{\underset{\sim}{\xi}}$ denotes the estimates $\begin{bmatrix} \hat{\underset{\sim}{\alpha}}_i \\ \hat{\underset{\sim}{\beta}}_j \\ \hat{\underset{\sim}{\gamma}}_{(k)} \end{bmatrix}$,

and $\underset{\sim}{q}$ is the right-hand side of the system of normal equations. Then

$$\frac{1}{n} E \ [\hat{\underset{\sim}{\xi}}' \ \underset{\sim}{q} \ ] = (3p-3) \ [\ \sigma_e^2 + n\sigma_{e'}^2 \ ] + n \ [\underset{\sim}{\alpha}'D\underset{\sim}{\alpha} + \underset{\sim}{\beta}'D\underset{\sim}{\beta} + \underset{\sim}{\gamma}'D\underset{\sim}{\gamma} + 2\underset{\sim}{\alpha}'E\underset{\sim}{\beta} + 2\underset{\sim}{\alpha}'E\underset{\sim}{\gamma} + 2\underset{\sim}{\beta}'E\underset{\sim}{\gamma} \ ] \ .$$

Hence

$$\begin{aligned} E \ [\text{SS Lack-of-fit}] &= E \ [\text{SS Subtotal}] - E \ [\text{SS Treatments}] \\ &= [\frac{(p+4)(p-1)}{2} -3p+2] \ \sigma_e^2 + [n_{....} - n - n(3p-3)] \ \sigma_{e'}^2 \\ &= \frac{p(p-3)}{2} \sigma_e^2 + n \ [\frac{(p+4)(p-1)}{2} -1 -3p+3]\sigma_{e'}^2 \\ &= \frac{p(p-3)}{2} \sigma_e^2 + \frac{np(p-3)}{2} \sigma_{e'}^2 \quad . \end{aligned} \tag{5.5.6}$$

Hence

$$E \ [\text{MS Lack-of-fit}] = \sigma_e^2 + n \ \sigma_{e'}^2 \quad , \tag{5.5.7}$$

and, as expected, an unbiased estimate of $\sigma^2_{e'}$, $s^2_{e'}$ say, is available as

$$s^2_{e'} = \frac{1}{n} \{[\text{MS Lack-of-fit}] - [\text{MS Error}]\} \quad .$$

5.6 Levels of the Factors Not Equally Spaced

In the case where the levels of the three factors are referred to a common unit of measurement, but are not equally spaced, symmetry in the design will be lost in that the third factor will require more levels. If such a case does arise, we propose that the levels of the third factor be treated as a concomitant variable, and that an analysis of covariance be performed. The design, in this case, would be considered as a two-way classification with one concomitant variable. The analysis can be performed in the usual way, e.g., see Kempthorne (1952) or Graybill (1961). However, since the normal equations, which result in this case, fall into a pattern similar to that in the Symmetrical Complementation design, we will give a brief description of the analysis.

Let us first consider the method of analysis suitable for data which may be arranged in a two-way classification such as that set out diagrammatically below. In the diagram, data are available for the cells marked x. It has been assumed that we have four levels of each of the two factors.

	1	2	3	4	= Level of B
1		x	x	x	
2	x	x	x	x	
3	x	x	x		
4	x	x			
Level of A					

In this situation, we would have the model

$$Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \quad , \qquad (5.6.1)$$

where $i = j = 1, 2, \ldots, p$.

We will assume one observation per cell for simplicity. As before, μ , α_i , and β_j are the general effect, row effects, and column effects, respectively. The normal equations for this model, in matrix notation, are

$$\begin{bmatrix} \frac{(p-1)(p+4)}{2} & \underline{z}' & \underline{z}' \\ \underline{z} & D & E \\ \underline{z} & E & D \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\underline{\alpha}}_i \\ \hat{\underline{\beta}}_j \end{bmatrix} = \begin{bmatrix} Y_{..} \\ Y_{i.} \\ Y_{.j} \end{bmatrix} , \qquad (5.6.2)$$

where

$\underline{z}'$ is a $(1 \times p)$ row vector of the form

$$\underline{z}' = \left[(p-1) \quad p \quad (p-1) \quad (p-2) \quad \ldots \quad 4 \quad 3 \quad 2 \right] ;$$

D is a $(p \times p)$ diagonal matrix of the form

$$D = \begin{bmatrix} p-1 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & p & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & p-1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & p-2 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 2 \end{bmatrix} ;$$

and E is a matrix of the form

$$E = \begin{bmatrix} 0 & 1 & 1 & 1 & \cdots & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & \cdots & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & \cdots & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & \cdots & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix} .$$

$\underline{\hat{\alpha}}_i$ denotes a (p x 1) vector of unknown row effects, and $\underline{\hat{\beta}}_j$ a (p x 1) vector of unknown column effects. Row totals, column totals, and the grand total are denoted by $Y_{i.}$, $Y_{.j}$, and $Y_{..}$, respectively. Note the similarity of the normal equations in this case to those given by equation (2.2.3) in Chapter II.

Several methods can be employed to obtain the estimates of the treatment effects. These methods are modifications of the high-low method and modified high-low method described in Chapter III. Both methods employ two constraints, namely,

$$\sum_i \hat{\alpha}_i = 0 \quad ,$$

and (5.6.3)

$$\sum_j \hat{\beta}_j = 0 \quad .$$

We will, however, discuss only the latter of these two methods, since the formulas producing estimates will be a by-product of the matrix inversion required for testing hypotheses.

5.6.1 Modified High-Low Method

Let us, for the moment, disregard the very first row, i.e., the equation for the mean effect, in the normal equations. Now, let us introduce the constraints given by equations (5.6.3) into all the remaining rows of the normal equations, after transposing the terms involving $\hat{\mu}$ to the right-hand sides of the equations. Thus, we obtain the following constrained normal equations, in matrix form:

$$\begin{bmatrix} D & B \\ B & D \end{bmatrix} \begin{bmatrix} \underline{\hat{\alpha}}_i \\ \hat{\beta}_j \end{bmatrix} = \begin{bmatrix} Y_{i.} - n_{i.} & \hat{\mu} \\ Y_{.j} - n_{.j} & \hat{\mu} \end{bmatrix}, \qquad (5.6.4)$$

where

$$D = \begin{bmatrix} (p-1) & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & p & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & (p-1) & 0 & \ldots & 0 & 0 \\ 0 & & 0 & (p-2) & \ldots & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 3 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 2 \end{bmatrix},$$

and

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

The $\underline{\hat{\alpha}}_i$ and $\underline{\hat{\beta}}_j$ denote (p x 1) vectors of estimates and the $n_{i.}$ and $n_{.j}$ are the number of observations in

the ith row and jth column, respectively, i.e.
$n_{i.} = n_{.j}$ = p-1, p, p-1, p-2, ..., 3, 2 for

$$i = j = 1, 2, 3, \ldots, p \quad .$$

We have immediately, from the second row in each subset of equations, the estimates of α_2 and β_2 as

$$\hat{\alpha}_2 = \frac{Y_{2.}}{p} - \hat{\mu} \quad ,$$

and (5.6.5)

$$\hat{\beta}_2 = \frac{Y_{.2}}{p} - \hat{\mu} \quad .$$

If we now delete the second row and second column in each submatrix of the normal equations matrix, we obtain a modified normal equations matrix, which, in partitioned form, can be represented as

$$K = \begin{bmatrix} D & B \\ B & D \end{bmatrix} \quad . \tag{5.6.6}$$

In the matrix K, [D p-1 x p-1] is a diagonal matrix of the form

$$D = \begin{bmatrix} (p-1) & 0 & 0 & \ldots & 0 & 0 \\ 0 & (p-1) & & & & \\ 0 & 0 & (p-2) & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & 3 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 2 \end{bmatrix} \quad ,$$

and B[p-1 x p-1] is a matrix of the form

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

Note that the matrices D and B given here are identical to those given in Section 3.3. Assuming the inverse of the matrix $\bar{K}$, say $\bar{L}$, to be of the general structure

$$\bar{L} = \begin{bmatrix} \bar{V} & \bar{G} \\ \bar{G} & \bar{V} \end{bmatrix} , \qquad (5.6.7)$$

we have

$$\bar{K}\,\bar{L} = \begin{bmatrix} D & B \\ B & D \end{bmatrix} \begin{bmatrix} \bar{V} & \bar{G} \\ \bar{G} & \bar{V} \end{bmatrix}$$

$$= \begin{bmatrix} D\bar{V} + B\bar{G} & D\bar{G} + B\bar{V} \\ D\bar{G} + B\bar{V} & D\bar{V} + B\bar{G} \end{bmatrix} .$$

Now, equating $\bar{K}\bar{L}$ to the identity matrix I [2p-2 x 2p-2] , we obtain the following two equations:

$$D\bar{V} + B\bar{G} = I$$
$$D\bar{G} + B\bar{V} = \underline{0} ,$$

which can be solved for the matrices $\bar{V}$ and $\bar{G}$, giving

$$\bar{V} = [D - BD^{-1}B]^{-1} \quad , \qquad (5.6.8)$$

and

$$\bar{G} = -D^{-1}B\,\bar{V} \quad . \qquad (5.6.9)$$

With these results for the matrices $\bar{V}$ and $\bar{G}$, the matrix $\bar{L}$ is known, and we have as solutions to the modified normal equations

$$\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_3 \\ \vdots \\ \hat{\alpha}_p \\ \hat{\beta}_1 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = \bar{L} \begin{bmatrix} Y_{1.} - (p-1)\,\hat{\mu} \\ Y_{3.} - (p-1)\,\hat{\mu} \\ \vdots \\ Y_{p.} - 2\,\hat{\mu} \\ Y_{.1} - (p-1)\,\hat{\mu} \\ Y_{.2} - (p-1)\,\hat{\mu} \\ \vdots \\ Y_{.p} - 2\,\hat{\mu} \end{bmatrix} \quad . \qquad (5.6.10)$$

These estimates, together with the estimates for $\hat{\alpha}_2$ and $\hat{\beta}_2$ given by equations (5.6.5), are the solutions for the normal equations for the two-way classification. They are all functions of $\hat{\mu}$. The estimate for the mean effect, $\hat{\mu}$, can be found by substitution back into the very first equation of the normal equations.

5.6.2 Analysis of Covariance

If we treat the levels of third factor as a concomitant variable, and consider the design as a two-way classification with one concomitant variable, we would have the model

$$y_{ij} = \mu + \alpha_i + \beta_j + \delta x_{ij} + e_{ij} , \qquad (5.6.11)$$

where $i = j = 1, 2, \ldots, p$.

In the model, x_{ij} refers to the concomitant variable, and δ represents the regression coefficient (unknown) of y on x. This effect may be of a higher order; however, we will assume the effect to be linear. As before, the unknowns are estimated by least squares, utilizing the same linear restrictions on the parameters as were used in the previous section, i.e,

$$\sum_i \hat{\alpha}_i = \sum_j \hat{\beta}_j = 0 .$$

In practice, however, we start with a joint analysis of the sums of squares and products of y and x .

The estimation of the parameters proceeds in exactly the same manner as described in Section 5.6.1. This procedure yields a vector of estimates of effects under the old model, and a vector corresponding to the right-hand sides of the normal equations. Let us call these two vectors $\hat{\underline{\xi}}_y$ (for estimates) and $\underline{q}_y$ (for the right-hand sides).

The analysis of the values of the concomitant variable x follows by the same methods as given previously for the random variable y. Here we perform the same analysis, replacing the observations on the variable y by those on the variable x. This yields again two vectors, $\underline{\hat{\xi}}_x$ and $\underline{g}_x$, say. To analyze the sum of products, we carry out the same operations as for a sum of squares, except that at every stage a square is replaced by the corresponding product.

We now compute the following quantities:

1. Sum of squares and products for total:

$$\text{SS Total }(y) = \sum_{\text{all}} y_{ij}^2 - \frac{Y_{..}^2}{n_{..}} ,$$

$$\text{SS Total }(x) = \sum_{\text{all}} x_{ij}^2 - \frac{X_{..}^2}{n_{..}} , \qquad (5.6.12)$$

$$\text{SP Total }(y,x) = \sum_{\text{all}} y_{ij}\, x_{ij} - \frac{Y_{..}\, X_{..}}{n_{..}} .$$

2. Sum of squares and products due to treatments:

$$\text{SS Treatments }(y) = \underline{\hat{\xi}}'_y\, \underline{g}_y ,$$

$$\text{SS Treatments }(x) = \underline{\hat{\xi}}'_x\, \underline{g}_x , \qquad (5.6.13)$$

$$\text{SP Treatments }(y,x) = \underline{\hat{\xi}}'_x\, \underline{g}_y .$$

3. Sum of squares and products for error:

SS Error (y) = SS Total (y) - SS Treatments (y)

SS Error (x) = SS Total (y) - SS Treatments (y)

SP Error (y,x) = SP Total (y,x) - SP Treatments (y,x).

The computations can be tabulated in the form shown in Table 10 below.

TABLE 10

Computations for Covariance Model

Source	y^2	yx	x^2
Treatments	$\underline{\hat{\xi}}'_y \underline{q}_y$	$\underline{\hat{\xi}}'_x \underline{q}_y$	$\underline{\hat{\xi}}'_t \underline{q}_t$
Error	Subtraction	Subtraction	Subtraction
Total	$\sum_{all} y_{ij}^2 - \frac{Y_{..}^2}{n_{..}}$	$\sum_{all} y_{ij} x_{ij} - \frac{Y_{..} X_{..}}{n_{..}}$	$\sum_{all} x_{ij}^2 - \frac{X_{..}^2}{n_{..}}$

If we denote the sums of squares and products for error by E_{yy}, E_{xx}, and E_{yx}, respectively, the regression coefficient of y on x is given by

$$\hat{\delta} = \frac{E_{yx}}{E_{xx}} . \tag{5.6.15}$$

To remove the effect of the regression on x, we first obtain the sum of squares due to regression, which is

$$\text{SS Regression} = \frac{[\text{SP Error } (y,x)]^2}{\text{SS Error } (x)} = \frac{E_{yx}^2}{E_{xx}} . \qquad (5.6.16)$$

The error sum of squares is found by subtracting the sum of squares due to regression from the sum of squares for error of the random variable y, i.e.,

$$\begin{aligned}\text{SS (Error)} &= \text{SS Error } (y) - \text{SS Regression} \\ &= E_{yy} - \frac{E_{yx}^2}{E_{xx}} . \qquad (5.6.17)\end{aligned}$$

Then $\text{MS (Error)} = \text{SS (Error)} \Big/ \frac{(p-2)(p+1)}{2} - 1$.

To test the hypothesis $H_o : \delta = 0$, a test criterion is

$$F = \frac{\text{SS Regression}}{\text{MS (Error)}} .$$

This is compared with the F-distribution with 1 and $\left[\frac{(p-2)(p+1)}{2} - 1\right]$ degrees of freedom. If this is non-significant, we may disregard x.

The sum of squares due to treatments can be obtained in the following manner:

1. Obtain the matrix due to treatments, given by

$$T = \begin{bmatrix} \hat{\underline{\xi}}'_y \underline{g}_y & \hat{\underline{\xi}}'_x \underline{g}_y \\ & \hat{\underline{\xi}}'_x \underline{g}_x \end{bmatrix} .$$

2. Add, to the elements of the matrix T, the corresponding elements of the matrix due to error, given by

$$E = \begin{bmatrix} E_{yy} & E_{yx} \\ & E_{xx} \end{bmatrix} .$$

3. Obtain the sum of squares due to treatments plus error as

$$SS(T+E) = [E_{yy} + \hat{\underline{\xi}}'_y \underline{g}_y] - \frac{[E_{yx} + \hat{\underline{\xi}}'_x \underline{g}_y]^2}{[E_{xx} + \hat{\underline{\xi}}'_x \underline{g}_x]} . \qquad (5.6.18)$$

4. Subtract, from SS (T+E), the sum of squares due to error, as given by equation (5.6.17). Thus, the sum of squares due to treatments is given by

$$SS\ (T) = SS\ (T+E) - SS(E) .$$

This difference, divided by (2p-2), yields the mean square for treatments.

The estimates of the parameters, corrected for regression, are most easily obtained from the expressions:

$$\hat{\alpha}_i = \hat{\alpha}_{iy} - \hat{\delta}\,\hat{\alpha}_{ix} ,$$

$$\hat{\beta}_j = \hat{\beta}_{jy} - \hat{\delta}\,\hat{\beta}_{jx} , \qquad (5.6.19)$$

and

$$\hat{\mu} = \hat{\mu}_y - \hat{\delta}\,\hat{\mu}_x .$$

To test any other hypothesis, the computations can be tabulated in the form as shown in Table 11 below:

TABLE 11

Computations for Test of Hypothesis

Source	y^2	x,y	x^2	$(y/x)_{SS}$	d.f
Hypothesis (H)	H_{yy}	H_{xy}	H_{xx}	SSH	n_h
Error (E)	E_{yy}	E_{xy}	E_{xx}	SEE	$\frac{(p-2)(p+1)}{2} -1$
H+E	$(H+E)_{yy}$	$(H+E)_{xy}$	$(H+E)_{xx}$	SS(H+E)	

The residual sum of squares for H+E is then given by

$$SS(H+E) = (H+E)_{yy} - \frac{[(H+E)_{xy}]^2}{(H+E)_{xx}} .$$

If we subtract from this expression the sum of squares of error, as given by equation (5.6.17), we obtain the sum of squares due to the hypothesis with n_h degrees of freedom. This follows the customary analysis of covariance technique.

CHAPTER VI

Demonstration Studies

6.1 Univariate Analysis

The illustration givenbelow, of the application of the techniques described in previous chapters, is based on a Symmetrical Complementation design experiment where each factor appears at four levels (p = 4). The experiment was concerned with the study of the effects on pressure due to changes in the composition of a casting solvent involving: nitroglycerin (NG) plus one percent 2-nitrodiphenylamine (2-NDPA), adiponitrile (ADN), and triacetin (TA). Nitroglycerin is, of course, the base substance while the latter two are stabilizers.

Data for the analysis were supplied by the Hercules Power Company at Salt Lake City, Utah. The design used in this study is set out diagrammatically below:

	0	2	4	6	8	= % ADN
17	83	81	79	77	75	
19	81	79	77	75		
21	79	77	75			
23	77	75				
25	75					
% TA						

+ % NG + 1% 2-NDPA

Fifteen different mixtures were predetermined in such a way that they would supply adequate coverage of the range of composition encountered in the process. Each casting was then subdivided into three parts, so that data were furnished for each of the three parts of the 15 mixtures. Thus, the experiment is considered as a Symmetrical Complementation design with three observations per cell. The results of the physical analyses have been recorded in Table 12.

TABLE 12

Experimental Data on Effects of
Changes in Composition of Mixture
on Pressure.

	0	2	4	6	8	= % ADN
17	443.6	420.8*	377.2	351.7	320.2	
	443.1	420.4	380.9*	350.3	331.7	
	439.7	422.3	373.5	345.1*	308.7	
19	405.9	385.8	351.3	325.9		
	406.9	390.3*	358.8	323.6		
	408.6*	332.2	348.5*	318.9*		
21	375.0	350.2	341.0			
	379.8	354.0*	331.7			
	376.4	357.9	329.6*			
23	352.2	321.6*				
	356.8*	320.4				
	355.2	327.4				
25	325.8					
	327.7					
	329.3					
% TA						

= % NG + 1% 2 NDPA

* The data marked with the asterisk will be used to illustrate the analysis with one observation per cell. It was felt by Hercules Powder Company personnel that these data were the most reliable. For the cell [21, 0, (79)], no particular "most reliable" value was given, so that we used the mean in this instance.

6.1.1 One Observation Per Cell

The data are given in Table 13 below.

TABLE 13

Experimental Data: One Observation Per Cell

	0	2	4	6	= % ADN
17		420.8	380.9	345.1	
19	408.6	390.3	348.5	318.9	
21	377.1	354.0	329.6		
23	356.8	321.6			
% TA					

= %NG + 1% 2-NDPA

The row, column, and diagonal totals, and the grand total are

$Y_{...} = 4352.2$

$Y_{1..} = 1146.8$	$Y_{.1.} = 1142.5$	$Y_{..(1)} = 970.1$
$Y_{2..} = 1466.3$	$Y_{.2.} = 1486.7$	$Y_{..(2)} = 1404.4$
$Y_{3..} = 1060.7$	$Y_{.3.} = 1059.0$	$Y_{..(3)} = 1148.3$
$Y_{4..} = 678.4$	$Y_{.4.} = 664.0$	$Y_{..(4)} = 829.4$

Hence $\hat{\mu} = 362.68$. The right-hand side of the normal equations, written as a row vector, is

$$\underline{g}' = [58.75,\ 15.57,\ -27.35,\ -46.97,\ 54.45,\ 35.97,\ -29.05,\ -61.37,\ -117.95,\ -46.33,\ 60.25,\ 104.03]. \qquad (6.1.1)$$

Hence, by premultiplying the column vector of right-hand sides by the inverse matrix given in Section 3.4.2, we obtain the parameter estimates as

$$\begin{array}{llllll}
\hat{\alpha}_1 & = 13.50 & \hat{\beta}_1 & = 12.42 & \hat{\gamma}_{(1)} & = -30.68 \\
\hat{\alpha}_2 & = 3.89 & \hat{\beta}_2 & = 8.99 & \hat{\gamma}_{(2)} & = -11.58 \\
\hat{\alpha}_3 & = -4.33 & \hat{\beta}_3 & = -3.16 & \hat{\gamma}_{(3)} & = 9.65 \\
\hat{\alpha}_4 & = -13.06 & \hat{\beta}_4 & = -18.25 & \hat{\gamma}_{(4)} & = 32.61
\end{array} \tag{6.1.2}$$

The uncorrected total sum of squares is 1,590,578.34. Next we compute the following corrected sums of squares:

$$\begin{aligned}
SS(\text{Total}) &= \sum_{\text{all}} y^2_{ij(k)} - \frac{y^2_{\cdots}}{n_{\cdots}} = 1{,}590{,}578.34 - 1{,}578{,}470.40 \\
&= 12{,}107.94
\end{aligned}$$

$$SS(\text{Treatments}) = \underline{\hat{\xi}}'\,\underline{q} = 11{,}926.02$$

$$\begin{aligned}
SS(\text{Lack-of-fit}) &= SS(\text{Total}) - SS(\text{Treatments}) \\
&= 12{,}107.94 - 11{,}926.02 \\
&= 181.92
\end{aligned}$$

The preliminary analysis of variance is shown in Table 14 below.

TABLE 14

Preliminary ANOVA: One Observation Per Cell

Source	d.f	Sum of Squares	Mean Square
Treatments	9	11,926.02	1325.11
Lack-of-fit	2	181.92	90.96
Total	11	12,107.94	

Of the nine degrees of freedom due to treatments, two each can be assigned to single factors, and another two can be expressed in terms of two bi-linear comparisons (the third is implied). The ninth degree of freedom would be a composite of effects from all factors, which is hard to interpret. As described in Section 4.3.2, a convenient method for obtaining the two degrees of freedom in single factors is to express them in terms of Tchebycheff polynomials of the second and third degree (actually two arbitrary reference components can be used). We then obtain, from equations (4.3.12),

SS(TA) (quadratic and cubic) = 1.27 ,
SS(ADN) (quadratic and cubic) = 105.09 ,

and

SS(NG) (Quadratic and cubic) = 10.41 ;

each with two degrees of freedom. Then, from equations (4.3.14), we obtain

SS(TA-ADN) = 24.09 ,
SS(TA-NG) = 8119.21 ,

and

SS(ADN-NG) = 9038.21 ,

this last function being linearly dependent upon the first two.

In the absence of replication or an independent error variance, it is necessary to use the lack-of-fit mean square as the error variance. Even compared with this contribution due to lack-of-fit, i.e., $^{8119.21}/_{90.49}$ = 89.72, the bilinear contrast (TA-NG) is highly significant.

Hypotheses of substitution can be tested for each level of Factor NG, or an "overall" hypothesis of substitution can be tested to determine whether levels of Factors TA and ADN are completely exchangeable. The sum of squares for the "overall" hypothesis of substitution is given by equation (4.2.4), and the result is

$$\begin{aligned} \text{SS (Overall Hypothesis)} &= 11{,}926.02 - 11{,}795.58 \\ &= 130.44 \end{aligned}$$

The ratio $\dfrac{\text{SS(Overall Hypothesis)}/6}{\text{SS(Lack-of-fit)}/2} = \dfrac{21.74}{90.96} = 0.24$

is thencompared with the F-distribution with 6 and 2 degrees of freedom, and is not significant. The sums of squares and mean squares for the hypotheses of substitution for the various levels of Factor NG are calculated by using the equations in Section 4.2.2, and the results are:

SS (75% NG) = 87.78 MS (75% NG) = 43.89
SS (77% NG) = 70.99 MS (77% NG) = 23.66
SS (79% NG) = 18.29 MS (79% NG) = 9.15
SS (81% NG) = 38.19 MS (81% NG) = 38.19

These results are non-significant when compared with the contribution due to lack-of-fit.

The foregoing was merely intended as an illustration. With two degrees of freedom due to error, a contrast must indeed be enormously large to show up significantly. If only one observation is available per cell, tests of significance would probably be of limited value unless an independent estimate of the error of measurement is available.

6.1.2 Three Observations Per Cell

For this case, the data have been given in Table 12. The cell totals are first calculated and arranged in Table 15 below.

TABLE 15

Cell Totals, Three Observations Per Cell

	0	2	4	6	= % ADN
17		1263.5	1131.6	1047.1	
19	1221.4	1158.3	1058.6	968.4	
21	1131.2	1072.1	1002.3		
23	1064.2	969.4			
% TA					

We have considered the corner cells as being empty. The row, column, and diagonal totals, and the grand total are:

$Y_{....} = 13{,}088.1$

$Y_{1...} = 3{,}442.2 \quad Y_{.1..} = 3{,}416.8 \quad Y_{..(1).} = 2{,}940.1$

$Y_{2...} = 4{,}406.7 \quad Y_{.2..} = 4{,}463.3 \quad Y_{..(2).} = 4{,}242.0$

$Y_{3...} = 3{,}205.6 \quad Y_{.3..} = 3{,}192.5 \quad Y_{..(3).} = 3{,}421.1$

$Y_{4...} = 2{,}033.6 \quad Y_{.4..} = 2{,}015.5 \quad Y_{..(4).} = 2{,}484.9$

Hence $\hat{\mu} = 363.56$. The right-hand side of the normal equations, written as a row vector, is

$$\underline{q}' = [170.18, 44.00, -66.42, -147.75, 144.78, 100.60 \\ -79.52, -165.85, -331.92, -120.70, \\ 149.08, 303.55] \quad . \qquad (6.1.3)$$

Hence, by premultiplying the column vector of right-hand sides by the inverse matrix given in Section 3.4.2, and dividing each element of the vector of estimates by three, the number of observations per cell, we obtain the parameter estimates as

$$\begin{array}{lllll}
\hat{\alpha}_1 = 12.77 & & \hat{\beta}_1 = 10.65 & & \hat{\gamma}_{(1)} = -29.07 \\
\hat{\alpha}_2 = 3.67 & & \hat{\beta}_2 = 8.38 & & \hat{\gamma}_{(2)} = -10.06 \\
\hat{\alpha}_3 = -1.86 & & \hat{\beta}_3 = -2.74 & & \hat{\gamma}_{(3)} = 6.27 \\
\hat{\alpha}_4 = -14.58 & & \hat{\beta}_4 = -16.29 & & \hat{\gamma}_{(4)} = 32.86
\end{array} \qquad (6.1.4)$$

The uncorrected total sum of squares is 4,790,701.39. Next we compute the following corrected sums of squares:

$$\begin{aligned}
SS(\text{Total}) &= \sum_{\text{all}} Y^2_{ij(k)m} - \frac{Y^2_{....}}{n_{....}} = 4{,}790{,}\ 701.39 \\
&\qquad\qquad\qquad\qquad\qquad -4{,}758{,}\ 287.82 \\
&\qquad\qquad\qquad\qquad = 32{,}413.57
\end{aligned}$$

$$\begin{aligned}
SS(\text{Sub-total}) &= \sum_{i,j,(k)} \frac{Y^2_{ij(k).}}{n} - \frac{Y^2_{....}}{n_{....}} = 4{,}790{,}\ 384.58 \\
&\qquad\qquad\qquad\qquad\qquad -4{,}758{,}\ 287.82 \\
&\qquad\qquad\qquad\qquad = 32{,}096.76
\end{aligned}$$

$$\begin{aligned}
SS(\text{Error}) = SS(\text{Total}) - SS(\text{Sub-total}) &= 32{,}413.57 \\
&\ \ -32{,}096.76 \\
&= 316.81
\end{aligned}$$

The preliminary analysis of variance is shown in Table 16 below.

TABLE 16

Preliminary ANOVA: Three Observations Per Cell

Source	d.f	Sum of Squares	Mean Square	F
Sub-total	11	32,096.76	2917.89	221.05**
Error	24	316.81	13.20	
Total	35	32,413.57		

The sum of squares due to treatments can be found by using equation (5.2.5), and the result is

$$\text{SS (Treatments)} = \underline{\hat{\xi}}' \, \underline{g} = 31,689.47 \quad .$$

Then, the sum of squares due to lack-of-fit is found by subtracting the sum of squares due to treatments from the sum of squares due to sub-total, and the result is:

$$\begin{aligned} \text{SS(Lack-of-fit)} &= \text{SS(Sub-totals)} - \text{SS(Treatments)} \\ &= 32,096.76 \; -31,689.47 \\ &= 407.29 \quad . \end{aligned}$$

Thus, we obtain the following analysis of variance as shown in Table 17 below.

TABLE 17

ANOVA: Three Observations Per Cell

Source	d.f	Sum of Squares	Mean Square
Treatments	9	31,689.47	3521.05
Lack-of-Fit	2	407.29	203.64
Sub-total	11	32,096.76	2917.89
Error	24	316.81	13.20
Total	35	32,413.57	

The contribution due to lack-of-fit is significant, and some question may arise regarding the applicability of the purely additive model. For illustration purposes, however, we will continue the analysis.

Of the nine degrees of freedom due to treatments, two each can be assigned to single factors, and another two can be expressed in terms of two bilinear comparisons. As described in Section 4.5.1, the two degrees of freedom in single factors can be expressed in terms of polynomials of the second and third degree, in such a way that they are orthogonal with reference to this design. This choice will make the sums of squares additive. Thus, from equations (4.5.6) and (4.5.9), we obtain

SS (TA) (quadratic) = 5.33
SS (ADN) (quadratic) = 94.85
SS (NG) (quadratic) = 29.67

SS (TA) (cubic) = 17.81
SS (ADN) (cubic) = 6.34
SS (NG) (cubic) = 25.76 ,

each with one degree of freedom.
By summing the quadratic and cubic sums of squares, we obtain

SS (TA) (quadratic and cubic) = 23.14
SS (ADN) (quadratic and cubic) = 101.19
SS (NG) (quadratic and cubic) = 55.43 ,

each with two degrees of freedom. For the bilinear contrasts, we obtain

SS (TA-ADN) = 1.73
SS (TA-NG) = 7629.64

and

SS (ADN-NG) = 7861.03 ,

the last function being linearly dependent upon the first two.

Since in this case we have three observations per cell, we will use the within cells mean square as the error mean square. Compared with this mean square,

$$\frac{\text{SS(ADN)(quadratic and cubic)}/2}{\text{MS Error}} = \frac{50.60}{13.20} = 3.83,$$

which, when compared with the F-distribution with 2 and 24 degrees of freedom, is significant at the 5% level. The quadratic contrast in Factor ADN is also significant at the 5% level. The bi-linear contrast (TA-NG) is highly significant.

From equation (4.5.4), the minimum-variance estimate of the coefficient B_1 of the polynomial

$$P_1(z) = \frac{1}{16}[95 + 12z - 19z^2]$$

is

$$\hat{B}_1 = \frac{153.06}{247} = 0.62 \quad .$$

Then, for the values $z = -3, -1, 1, 3$, the values of the quadratic polynomial are

$$\beta(-3) = -4.34 \quad ,$$
$$\beta(-1) = 2.48 \quad ,$$
$$\beta(1) = 3,43 \quad ,$$

and

$$\beta(3) = -1.55 \quad .$$

In predicting intermediate values of the total response function, we used linear interpolation. The above values are corrections which would have to be applied to Factor β to improve the interpolation.

Hypotheses of substitution can be tested for each level of Factor NG, or an "overall" hypothesis of substitution can be tested to determine whether levels of Factors TA and ADN are completely exchangeable. The sum of squares for the "overall" hypothesis of substitution is given by equation (4.2.4), and the result is

$$\begin{aligned} \text{SS(Overall Hypothesis)} &= 31,689.47 - 31,281.98 \\ &= 407.49 \quad . \end{aligned}$$

The ratio $$\frac{\text{SS(Overall Hypothesis)}/6}{\text{MS Error}} = \frac{67.92}{13.20} = 5.14$$
is compared with the F-distribution with 6 and 24 degrees of freedom, and is significant.

The sums of squares and mean squares for the hypotheses of substitution for the various levels of Factor NG are calculated by using the equations in Section 4.2.2, and the results are:

SS (75% NG) = 57.65 MS (75% NG) = 28.82
SS (77% NG) = 99.48 MS (77% NG) = 33.16
SS (79% NG) = 8.67 MS (79% NG) = 4.34
SS (81% NG) = 46.64 MS (81% NG) = 46.64

This breakdown shows that the significance in the departure from the hypothesis of substitution can be traced to the highest level of Factor NG, i.e., Factors ADN and TA are exchangeable for all but the highest level of Factor NG; at the highest level, exchangeability breaks down. None of the F-ratios are significant at 0.05, but the last one is significant at 0.10.

6.2 Multivariate Analysis

To illustrate the multivariate analysis of a Symmetrical Complementation design, we construct an artificial example. We consider a design with five levels of each factor under each experimental condition with three response variables, i.e., in each cell, three responses are given, constructed from the model

$$E(Y) = A\,\phi \,, \qquad (6.2.1)$$

where A is of the same structure as in the univariate case. We may express the observations on each experimental unit by a vector with three elements

$$\underline{y}'_{ij(k)} = [y^{(1)}_{ij(k)}, y^{(2)}_{ij(k)}, y^{(3)}_{ij(k)}] .$$

The $\underline{y}'_{ij(k)}$ vectors are independent and $N[\underline{\mu}' + \underline{\alpha}'_i + \underline{\beta}'_j + \underline{\gamma}'_{(k)} ; \Sigma]$, where $\underline{\mu}'$ is a (1 x 3) vector of unknown parameters due to an overall mean effect; $\underline{\alpha}'_i$, $\underline{\beta}'_j$, and $\underline{\gamma}'_{(k)}$ are (1 x 3) vectors of unknown row, column, and diagonal effects, respectively. Σ is a symmetric (3 x 3) covariance matrix. For the purpose of this illustration, we will assume the following parameters

$$\begin{array}{llllll}
 & \alpha_1^{(1)} = 0 & \alpha_2^{(1)} = 2 & \alpha_3^{(1)} = 5 & \alpha_4^{(1)} = 10 & \alpha_5^{(1)} = 20 \\
\mu^{(1)}=10 & \beta_1^{(1)} = 10 & \beta_2^{(1)} = 3 & \beta_3^{(1)} = 3 & \beta_4^{(1)} = 3 & \beta_5^{(1)} = 0 \\
 & \gamma_{(1)}^{(1)} = 0 & \gamma_{(2)}^{(1)} = 5 & \gamma_{(3)}^{(1)} = 5 & \gamma_{(4)}^{(1)} = 5 & \gamma_{(5)}^{(1)} = 10 \\
 & & & & & \\
 & \alpha_1^{(2)} = 10 & \alpha_2^{(2)} = 5 & \alpha_3^{(2)} = 10 & \alpha_4^{(2)} = 5 & \alpha_5^{(2)} = 10 \\
\mu^{(2)}=20 & \beta_1^{(2)} = 0 & \beta_2^{(2)} = 15 & \beta_3^{(2)} = 20 & \beta_4^{(2)} = 15 & \beta_5^{(2)} = 0 \\
 & \gamma_{(1)}^{(2)} = 5 & \gamma_{(2)}^{(2)} = 20 & \gamma_{(3)}^{(2)} = 25 & \gamma_{(4)}^{(2)} = 20 & \gamma_{(5)}^{(2)} = 5
\end{array}$$

$$\mu^{(3)}=10 \qquad \alpha_i^{(3)} = \beta_j^{(3)} = \gamma_{(k)}^{(3)} = 10 . \qquad (6.2.2)$$

The errors will be simulated in such a way that they would constitute a random sample with covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix} . \qquad (6.2.3)$$

In order to produce such a sample, we take three sets of 18 numbers x_1, x_2, x_3 from a table of random normal numbers. Then we take the following linear combinations

$$\begin{aligned} \ell_1 &= x_1 \\ \ell_2 &= x_1 + x_2 \\ \ell_3 &= x_1 + x_2 + x_3 \end{aligned}$$

in order to produce a sample of errors from a normal distribution which has the variances and covariances given in the theoretical error covariance matrix (6.2.3).

The model (6.2.1) is used to construct the three observations on each experimental unit as given in the design set out diagrammatically below.

Response Variable 1: $y^{(1)}$

	1	2	3	4	5	= Level of B
1		23.464	18.060	19.486	16.022	
2	33.394	20.906	21.179	18.499	11.310	
3	31.372	22.518	21.624	16.990		
4	34.995	29.393	21.213			
5	44.895	31.661				
Level of A						

= Level of C

Response Variable 2: $y^{(2)}$

	1	2	3	4	5	= Level of B
1		30.601	47.534	51.132	30.550	
2	30.839	60.393	70.124	58.011	30.066	
3	51.597	71.196	68.474	49.588		
4	49.096	60.230	47.952			
5	29.538	30.488				
Level of A						

= Level of C

Response Variable 3: $y^{(3)}$

	1	2	3	4	5	= Level of B
1		43.056	37.003	40.498	41.829	
2	40.885	39.868	41.131	37.849	38.448	
3	41.975	41.139	39.830	38.670		
4	39.108	39.319	39.189			
5	38.154	39.529				
Level of A						

= Level of C

The row, column, and diagonal totals, and the grand total for each of the three response variables are:

Response Variable 1

$Y_{...} = 436.981$

$Y_{1..} = 77.032$ $Y_{.1.} = 144.656$ $Y_{..(1)} = 81.174$
$Y_{2..} = 105.288$ $Y_{.2.} = 127.942$ $Y_{..(2)} = 130.433$
$Y_{3..} = 92.504$ $Y_{.3.} = 82.076$ $Y_{..(3)} = 98.178$
$Y_{4..} = 85.601$ $Y_{.4.} = 54.975$ $Y_{..(4)} = 70.338$
$Y_{5..} = 76.556$ $Y_{.5.} = 27.332$ $Y_{..(5)} = 56.858$

Response Variable 2

$Y_{...} = 867.409$

$Y_{1..} = 159.817$ $Y_{.1.} = 161.070$ $Y_{..(1)} = 158.094$
$Y_{2..} = 249.433$ $Y_{.2.} = 252.908$ $Y_{..(2)} = 246.803$
$Y_{3..} = 240.855$ $Y_{.3.} = 234.084$ $Y_{..(3)} = 241.548$
$Y_{4..} = 157.278$ $Y_{.4.} = 158.731$ $Y_{..(4)} = 159.524$
$Y_{5..} = 60.026$ $Y_{.5.} = 60.616$ $Y_{..(5)} = 61.440$

Response Variable 3

$Y_{...} = 717.480$

$Y_{1..} = 162.386$ $Y_{.1.} = 160.122$ $Y_{..(1)} = 155.836$
$Y_{2..} = 198.181$ $Y_{.2.} = 202.911$ $Y_{..(2)} = 196.981$
$Y_{3..} = 161.614$ $Y_{.3.} = 157.153$ $Y_{..(3)} = 161.876$
$Y_{4..} = 117.616$ $Y_{.4.} = 117.017$ $Y_{..(4)} = 118.846$
$Y_{5..} = 77.686$ $Y_{.5.} = 80.277$ $Y_{..(5)} = 83.941$

Hence $\hat{\mu}^{(1)} = 24.277$, $\hat{\mu}^{(2)} = 41.189$, and $\hat{\mu}^{(3)} = 39.860$. The right-hand side of the normal equations, written as a row vector, is given below for each of the three response variables.

$$\underline{g}^{(1)'} = [-20.076, -16.097, -4.604, 12.770, 28.002, 47.548, 6.557, -15.032, -17.856, -21.222, -15.934, 9.048, 1.070, -2.493, 8.304] \quad ,$$

$$\underline{g}^{(2)'} = [-32.939, 8.488, 48.099, 12.711, -36.352, -31.686, 11.963, 41.328, 14.164, -35.762, -34.662, 5.858, 48.792, 14.957, -34.938] \quad ,$$

and

$$\underline{g}^{(3)'} = [2.946, -1.119, 2.174, -1.964, -2.037, 0.682, 3.611, -2.287, -2.563, 0.557, -3.604, -2.319, 2.436, -0.734, 4.221] \quad .$$

(6.2.4)

The inverse normal equations matrix is given in Section 3.4.3. It is the same as that in the univariate case. Hence, by premultiplying each column vector of right-hand sides by this inverse matrix, we obtain the parameter estimates as

$$[\ \hat{\underline{\alpha}}_i^{(1)'} , \hat{\underline{\beta}}_j^{(1)'} , \hat{\underline{\gamma}}_{(k)}^{(1)'}\ ] = [-2.861400, -3.219400, -2.714475, 0.669975, 8.125300, 10.663400, 1.311400, -1.425425, -3.090475, -7.458900, -2.033000, 1.809600, 0.434100, -1.415700, 1.205000\ ] \quad ,$$

$$[\ \underline{\hat{\alpha}}_i^{(2)'}\ ,\ \underline{\hat{\beta}}_j^{(2)'}\ ,\ \underline{\hat{\gamma}}_{(k)}^{(2)'}\ ] = [-16.516500,\ 1.697600,\ 10.587325,\ 7.626175,\ -3.394600,\ -16.265900,\ 2.392600,\ 8.702625,\ 7.797475,\ -2.626800,\ -16.861100,\ 1.171600,\ 10.692650,\ 8.119750,\ -3.122900]\ ,$$

and

$$[\ \underline{\hat{\alpha}}_i^{(3)'}\ ,\ \underline{\hat{\beta}}_j^{(3)'}\ ,\ \underline{\hat{\gamma}}_{(k)}^{(3)'}\ ] = [0.591600,\ -0.223800,\ 1.088950,\ -0.598850,\ -0.857900,\ 0.138800,\ 0.722200,\ -0.412200,\ -1.13450,\ 0.685700,\ -0.718400,\ -0.463800,\ 0.565950,\ -0.879850,\ 1.496100\ ]\ .$$

(6.2.5)

The sums of squares and products for total are given by

$$\text{SS Total }(y^{(1)}) = \sum_{\text{all}} y_{ij(k)}^{(1)\ 2} - \frac{Y_{\cdots}^{(1)\ 2}}{n_{\cdots}}$$

$$= 11{,}788.120763 - \frac{(436.981)^2}{18}$$

$$= 1{,}179.654410$$

$$\text{SS Total }(y^{(2)}) = \sum_{\text{all}} y_{ij(k)}^{(2)\ 2} - \frac{Y_{\cdots}^{(2)\ 2}}{n_{\cdots}}$$

$$= 45{,}546.915757 - \frac{(867.409)^2}{18}$$

$$= 3747.006131$$

$$\text{SS Total } (y^{(3)}) = \sum_{\text{all}} y_{ij(k)}^{(3)\,2} - \frac{Y_{\cdots}^{(3)\,2}}{n_{\cdots}}$$

$$= 28,641.972154 - \frac{(717.480)^2}{18}$$

$$= 43.219354$$

$$\text{SP Total } (y^{(1)}, y^{(2)}) = \sum_{\text{all}} y_{ij(k)}^{(1)} y_{ij(k)}^{(2)} - \frac{Y_{\cdots}^{(1)} \; Y_{\cdots}^{(2)}}{n_{\cdots}}$$

$$= 20,595.202197 - \frac{(436.981)(867.409)}{18}$$

$$= -462.645149$$

$$\text{SP Total } (y^{(1)}, y^{(3)}) = \sum_{\text{all}} y_{ij(k)}^{(1)} y_{ij(k)}^{(3)} - \frac{Y_{\cdots}^{(1)} \; Y_{\cdots}^{(3)}}{n_{\cdots}}$$

$$= 17,424.342013 - \frac{(436.981)(717.480)}{18}$$

$$= 6.279353$$

$$\text{SP Total } (y^{(2)}, y^{(3)}) = \sum_{\text{all}} y_{ij(k)}^{(2)} y_{ij(k)}^{(3)} - \frac{Y_{\cdots}^{(2)} \; Y_{\cdots}^{(3)}}{n_{\cdots}}$$

$$= 34,568.761103 - \frac{(867.409)(717.480)}{18}$$

$$= -7.029046$$

The sums of squares and products due to treatments are given by

$$\text{SS (Treatments}^{(1)}) = \hat{\underline{g}}^{(1)'}\,\underline{g}^{(1)} = 1{,}171.138530$$

$$\text{SS (Treatments}^{(2)}) = \hat{\underline{g}}^{(2)'}\,\underline{g}^{(2)} = 3{,}739.665341$$

$$\text{SS (Treatments}^{(3)}) = \hat{\underline{g}}^{(3)'}\,\underline{g}^{(3)} = 26.223400$$

$$\text{SP (Treatments}^{(1),(2)}) = \hat{\underline{g}}^{(1)'}\,\underline{g}^{(2)} = -\ 469.649307$$

$$\text{SP (Treatments}^{(1),(3)}) = \hat{\underline{g}}^{(1)'}\,\underline{g}^{(3)} = 0.751986$$

$$\text{SP(Treatments }^{(2),(3)}) = \hat{\underline{g}}^{(2)'}\,\underline{g}^{(3)} = -15.319318\ .$$

The sums of squares and products due to lack-of-fit are given by

$$\begin{aligned}\text{SS (Lack-of-fit}^{(1)}) &= \text{SS Total } (y^{(1)}) - \text{SS (Treatments}^{(1)}) \\ &= 1179.654410 - 1171.138530 \\ &= 8.515880\end{aligned}$$

$$\begin{aligned}\text{SS(Lack-of-fit}^{(2)}) &= \text{SS Total } (y^{(2)}) - \text{SS (Treatments}^{(2)}) \\ &= 3747.006131 - 3739.665341 \\ &= 7.340790\end{aligned}$$

$$\begin{aligned}\text{SS (Lack-of-fit}^{(3)}) &= \text{SS Total } (y^{(3)}) - \text{SS (Treatments}^{(3)}) \\ &= 43.219354 - 26.223400 \\ &= 16.995954\end{aligned}$$

$$\text{SP (Lack-of-fit}^{(1),(2)}) = \text{SP Total }(y^{(1)}, y^{(2)}) - \text{SP(Treatments}^{1,2})$$

$$= -462.645149 - (-469.49307)$$

$$= 7.004158$$

$$\text{SP (Lack-of-fit}^{(1),(3)}) = \text{SP Total }(y^{(1)}, y^{(3)}) - \text{SP(Treatments }^{1,3})$$

$$= 6.279353 - 0.751986$$

$$= 5.527367$$

$$\text{SP (Lack-of-fit}^{(2),(3)}) = \text{SP Total }(y^{(2)}, y^{(3)}) - \text{SP(Treatments }^{2,3})$$

$$= -7.029046 - (-15.319318)$$

$$= 8.290272$$

Suppose we wish to estimate and test the significance of the estimable contrast $2\beta_1-\beta_2-2\beta_3-\beta_4+2\beta_5$. This contrast compares the central level with a weighted combination of the mean of the two adjoining ones and the mean of the two extreme levels. The estimate of this contrast, for each of the variables separately, is

$$[2, -1, -2, -1, 2] \; [\hat{\underset{\sim}{\beta}}_j^{(1)}, \hat{\underset{\sim}{\beta}}_j^{(2)}, \hat{\underset{\sim}{\beta}}_j^{(3)}]$$

$$= [11.035925, -65.380725, 2.875700]$$

The same contrast on the original parameters would be [8, -70, 0]. The above illustrates the discrepancy due to sampling. The sums of squares and products due to this hypothesis (see Section 5.3.2) are then given, in matrix form by

$$H = \begin{bmatrix} 121.857865 & -721.732920 & 31.744637 \\ -721.732920 & 4274.639202 & -188.015351 \\ 31.744637 & -188.015351 & 8.269650 \end{bmatrix},$$

where the diagonal elements are sums of squares, and the off-digaonal elements sums of products.

The matrix E, whose diagonal elements are sums of squares due to lack-of-fit, and whose off-diagonal elements are sums of products due to lack-of-fit, is given by

$$E = \begin{bmatrix} 8.515880 & 7.004158 & 5.527367 \\ 7.004158 & 7.340790 & 8.290272 \\ 5.527367 & 8.290272 & 16.995954 \end{bmatrix}.$$

We then form the matrix H+E, by adding corresponding elements in matrix H and matrix E, respectively, and obtain

$$H{+}E = \begin{bmatrix} 130.373745 & -714.728762 & 37.272004 \\ -714.728762 & 4281.979992 & -179.725079 \\ 37.272004 & -179.725079 & 25.265604 \end{bmatrix}.$$

The leading coefficients of the forward Doolittle solution (first term in second to last row - before normalizing to unity - in each cycle) for the lack-of-fit matrix E, and for the matrix H+E are given below.

1) Leading coefficients of E

$\underline{8.515880}$ $\underline{1.580003}$ $\underline{4.535935}$

2) Leading coefficients of H+E: Quadratic Contrast

$\underline{130.373745}$ $\underline{363.728280}$ $\underline{12.945498}$.

To test the hypothesis that the quadratic contrast in β is zero, we use the likelihood ratio test, using $\log |E| - \log |H+E|$ as a test statistic. The determinants $|E|$ and $|H+E|$ are given by the products of the leading coefficients of E and H+E, and thus $\log |E|$ and $\log |H+E|$ are given by the sum of the logarithms of the "leading coefficients". We then obtain

$$- m \log_e \Lambda = - m \log_e \frac{|E|}{|H+E|} = 32.256 ,$$

where $m = n_e + \frac{1}{2} (n_h - q - 1) = 5 + \frac{1}{2} (1 -3 -1) = 3.5$.

The test is performed by using equation (5.4.9) to obtain

$$\Pr \{ - m \log_e \Lambda > 32.256 \} < 0.00001 .$$

The correction in equation (5.4.9) would be of even smaller value, so that we may decide right here that the over-all departure from the null hypothesis is highly significant.

For the step-down analysis, we note that the ratio of any two corresponding leading coefficients for the matrix E, and for the matrix H+E is a Beta variable. The results of this step-down analysis, and the associated probabilities, are:

$$B_{y(1)} = \frac{8.515880}{130.373745} = 0.065318 , \qquad \beta_{.01} (\tfrac{5}{2}, \tfrac{1}{2}) = 0.23520$$

$$B_{y(2)/y(1)} = \frac{1.580003}{363.728280} = 0.004343 , \quad \beta_{.01} (2, \tfrac{1}{2}) = 0.15875$$

$$B_{y^{(3)}/y^{(1)},y^{(2)}} = \frac{4.535935}{12.945498} = 0.350387\ ,\quad \beta_{.01}\left(\frac{3}{2},\frac{1}{2}\right) = 0.080827$$

$$\beta_{.05}\left(\frac{3}{2},\frac{1}{2}\right) = 0.22852$$

The associated probability values are the lower tail percentage points of the Beta-distribution (Table 16, Biometrika Tables (1956)). They clearly show very high significance for variables 1 and 2; whereas variable 3, given variables 1 and 2, does not contribute to the significance, even at the 5% level.

Suppose we wish to test the significance of the equality of the additive effects of Factors β and γ for fixed levels of Factor α , i.e., an overall hypothesis of substitution. The sums of squares and products due to this hypothesis (see Section 4.2) are then given in matrix form, by

$$H = \begin{bmatrix} 604.296964 & -372.225235 & -50.565869 \\ -372.225235 & 1578.901246 & 27.818253 \\ -50.565869 & 27.818253 & 6.962179 \end{bmatrix} ,$$

where the diagonal elements are sums of squares, given by $\Sigma \frac{Y_{i..}^2}{n_{i..}} - \frac{Y_{...}^2}{n_{...}}$ for each variable; and the off-diagonal elements are sums of products, given by $\Sigma \frac{Y_{i..}^{(c)}Y_{i..}^{(d)}}{n_{i..}} - \frac{Y_{...}^{(c)}Y_{...}^{(d)}}{n_{...}}$ $(c \neq d)$.

We now form the matrix H+E, where E is the same as that given before, and obtain

$$
H+E = \begin{bmatrix} 612.812844 & -365.221077 & -45.038502 \\ -365.221077 & 1586.242036 & 36.108525 \\ -45.038502 & 36.108525 & 23.958133 \end{bmatrix}
$$

The leading coefficients of the forward Doolittle solution for the matrix H+E are given below, along with the leading coefficients of the matrix E.

1) Leading coefficients of E

$$\underline{8.515880} \qquad \underline{1.580003} \qquad \underline{4.535935}$$

2) Leading coefficients of H+E: "Overall" Hypothesis

$$\underline{612.812844} \qquad \underline{1368.579405} \qquad \underline{20.585285}$$

To test for significance of the "overall" hypothesis of substitution, we use the likelihood ratio test, using $\log |E| - \log |H+E|$ as a test statistic. We then obtain

$$
- m \log_e \Lambda = - m \log_e \frac{|E|}{|H+E|} = 87.868 ,
$$

where $m = n_e + \frac{1}{2} (n_h - q - 1) = 5 + \frac{1}{2} (8 -3 -1) = 7.$

The test is performed by using equation (5.4.9), to obtain

$$
\Pr \{ - m \log_e \Lambda > 87.868 \} < 0.00001
$$

The overall departure from the null hypothesis is highly significant.

The results and associated probabilities for the step-down analysis are:

$$B_{y^{(1)}} = \frac{8.515880}{612.812844} = 0.013896 \,, \quad \beta_{.01}\left(\frac{5}{2},4\right) = 0.057264$$

$$B_{y^{(2)}/y^{(1)}} = \frac{1.580003}{1368.579405} = 0.001154 \,, \quad \beta_{.01}\,(2,4) = 0.032682$$

$$B_{y^{(3)}/y^{(1)},y^{(2)}} = \frac{4.535935}{20.585285} = 0.220348, \quad \beta_{.01}\left(\frac{3}{2},4\right) = 0.013458$$

$$\beta_{.05}\left(\frac{3}{2},4\right) = 0.040671$$

This clearly shows that substitutability of Factors β and γ is not possible for variables 1 and 2, but is possible for variable 3. However, substitutability is possible in the case of variable 1 if we only consider the hypothesis of substitution for levels two, three, and four of Factor α. In this case, the step-down analysis gives $B_{y^{(1)}} = 0.83498$, which is not significant when compared with $\beta\,(\frac{5}{2},1)$. This clearly shows that Factors β and γ are substitutable in the case of variable 1, when we consider only the intermediate values of Factor α.

Summary

In this investigation methods of estimation and tests of significance have been developed for the Symmetrical Complementation design, which is intended for those experimental situations where the levels of three factors always sum to the same constant. Here the level of the third factor will be strictly determined in terms of the levels of the first two factors. If the levels of the factors are referred to a common unit of measurement, e.g., weight percent, calory equivalents, the levels of the factors must be equally spaced. Certain cell entries are omitted to insure complete interchangeability of the three factors. The usual additive model is assumed, i.e.,

$$y_{ij(k)m} = \mu + \alpha_i + \beta_j + \gamma_{(k)} + e'_{ij(k)} + e_{ij(k)m} \quad ,$$

where $i = j = (k) = 1,2,3,\ldots,p$ = number of levels,

and $m = 1,2,3,\ldots,n$ = number of observations per cell.

The subscript (k) is dependent on the values chosen for the subscripts i and j. In the model, $e'_{ij(k)}$ is referred to as the contribution due to lack-of-fit, and $e_{ij(k)m}$ the error within cells.

Estimability in the Symmetrical Complementation design has been investigated, and this study has been presented in the chapter on linear estimation. The study shows that this design differs from the usual types of designs, in that the number of contrasts that can be estimated is limited. For example, the usual linear contrast, which would lead to the hypothesis of equality of effects of all levels of one factor,

is not testable in this design. On the other hand, contrasts which compare one level with the mean or weighted mean of others, are testable for each factor separately. Quadratic and higher order contrasts are testable. These contrasts are combined into different hypotheses. Estimable functions in one factor only and in two factors have been presented for the general case of p levels. It is this latter case which led to the "hypothesis of substitution", which tests whether exchangeability of the levels of two factors is possible.

Several methods have been presented in order to obtain estimates of the treatment effects under various constraints. It must be noted, however, that these estimates are rather meaningless quantities, as it is only when they are combined in estimable functions that unique results are obtained. If only estimation is required, the simpler "high-low" method should be used. If, however, both estimation and tests of hypotheses are required, the "modified high-low" method should be used throughout, since the formulas producing estimates will be a by-product of the matrix inversion required. These methods have been described in complete detail in the chapter on estimation of parameters, with the complete inverse matrix, or a method of obtaining this patterned inverse, being presented.

A general technique, based upon the inversion of the matrix in the modified high-low method, can be used for the testing of any testable hypothesis. Such a hypothesis is conveniently and compactly stated in the form

$$H_o: \; P\underline{\xi} = \underline{0}$$

$$H_a: \; P\underline{\xi} \neq \underline{0} \;;$$

where P is a predetermined "hypothesis matrix", usually an array of positive and negative integers and zeros. For a hypothesis to be testable, the elements of the vectors $P\underline{\xi}$ must be estimable functions and linearly independent. The sums of squares and test statistics have been presented for the various hypotheses formulated. The important quantity needed here is the sum of squares due to hypothesis, and this is given by

$$SS(H) = (P\hat{\underline{\xi}})' \, [P(A'A)^{-1}P']^{-1}(P\hat{\underline{\xi}}) \quad ,$$

where $(A'A)$ is the normal equations matrix made non-singular by constraining, and $\hat{\underline{\xi}}$ is a solution to the normal equations. This dissertation shows that the hypothesis of substitution is one of the most important ones to consider. Sections have also been included, in the chapter on testing of hypotheses, which indicate how one might obtain the response for intermediate levels of the factors, and how one might obtain response functions for single factors.

After the various degrees of freedom available for each factor and factor combinations have been considered, there will be a certain residual which has been referred to as "lack-of-fit". The degrees of freedom due to lack-of-fit may serve to assess the magnitude of the error, if only one experimental unit is available per cell. A better way to obtain this error of measurement would be by replication. In this connection, three different cases have been presented;

a) the replications are strictly repetitions of the experiment under otherwise identical conditions. In this case, the analysis proceeds in the customary three-way analysis with n replicates per treatment combination;

b) the experiments within a cell represent repetitions over a period of time, during which some kind of a trend may be present. In this case, the analysis has been readily extended into an analysis of covariance with one concomitant variable if only a linear trend is postulated and more than one if trend of a higher degree is expected;

c) a case, similar to case b), where the replications constitute several experiments with the same experimental units, so that the observations within a cell are dependent. On the assumption that the covariance matrix of observations is the same from one cell to another, this method has been extended to a multivariate situation.

The problem of estimation is essentially the same in these three methods. However, different methods are necessary for the testing of hypotheses.

If the levels of the three factors are referred to a common unit of measurement, but are not equally spaced, symmetry will be lost in that the third factor will require more levels. Special discussion for this case has been presented in the chapter on extensions. Also included in this chapter is a discussion of the situation where the contribution due to lack-of-fit is treated as a random variable.

Several demonstrations studies of the design and the extensions have been presented in the chapter on applications.

In conclusion, it has been found that this type of design requires a rather careful consideration of the types of functions that can be estimated and the types of hypotheses that can be tested. For the case of three, four, and five levels, the formulas have been completely worked out. All that is required of the user of such a design is to substitute numbers in the expressions developed.

ACKNOWLEDGEMENTS

The author wishes to express his deep appreciation to Dr. Rolf E. Bargmann for his constant encouragement and helpful guidance which contributed so greatly to the successful completion of this dissertation. Special thanks are offered for the many hours he has spent trying to acquaint the author with the various statistical theory necessary to solve a difficult, and at times, seemingly hopeless, problem. It was indeed a privilege to have worked under his direction.

To Dr. Boyd Harshbarger, special thanks are offered for the initial suggestion out of which this problem developed, and also for his encouragement and careful consideration at all times in the research and writing of this dissertation.

Appreciation is also expressed to Professors H. A. David and C. Y. Kramer who have read the entire manuscript very carefully and have contributed much to the value of this dissertation by their useful suggestions and constructive criticisms; and to members of the graduate committee and Dr. William H. Glenn for their suggestions and criticisms.

Finally, heartfelt appreciation is offered to my wife , who undertook the arduous task of typing the rough draft in its various stages, and the manuscript in its final form. Most assuredly, thanks are due her for her patience and untiring support throughout.

BIBLIOGRAPHY

Allan, F. E. and Wishart, J. (1930). A Method of Estimating the Yield of a Missing Plot in Field Experimental Work. Journal of Agricultural Science, 20, 399.

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

Bargmann, R. E. (1960). On the Problem of Ordering Variables in the Tracing of Significant Contributions. Address presented at the 120th Annual Meeting of the American Statistical Association, Stanford, California.

Bose, R. C. (1944). The Fundamental Theorem of Linear Estimation. Proceedings of the Indian Science Congress.

Cochran, William G. and Cox, Gertrude M. (1957). Experimental Designs. New York: John Wiley and Sons, Inc.

Davies, Owen L. (1956). Design and Analysis of Industrial Experiments. New York: Hafner Publishing Company.

DeLury, D. B. (1946). The Analysis of Latin Squares When Some Observations are Missing. Journal of the American Statistical Association, 41, 370.

Ditchburne, Nell (1955). A Method of Analysis for a Double Classification Arranged in a Triangular Table. Biometrics, 11, 453.

Graybill, F. A. (1961). An Introduction to Linear Statistical Models. New York: McGraw-Hill Book Company, Inc.

Heck, D. L. (1958). Some Uses of the Distribution of the Largest Root in Multivariate Analysis. Institute of Statistics, University of North Carolina, Mimeograph Series No. 194.

Kempthorne, Oscar (1952). The Design and Analysis of Experiments. New York: John Wiley and Sons, Inc.

Kramer, C. Y. and Glass, Suzanne (1961). Analysis of Variance of a Latin Square Design with Missing Observations. Applied Statistics 9, 43.

Pearson, E. S. and Hartley, H. O. (1954). Biometrika Tables for Statisticians, Vol. I. Cambridge University Press.

Perlis, S. (1952). Theory of Matrices. Cambridge, Mass.: Addison-Wesley Press, Inc.

Rao, C. R. (1945a). Generalization of Markoff's Theorem and Tests of Linear Hypotheses. Sankhya, 7, 9.

Rao, C. R. (1945b). Markoff's Theorem with Linear Restrictions on Parameters. Sankhya, 7, 16.

Rao, C. R. (1946). On the Linear Combination of Observations and the General Theory of Least Squares. Sankhya, 7, 237.

Rao, C. R. (1948). Tests of Significance in Multivariate Analysis. Biometrika, 35, 58.

Rao, C. R. (1952). Advanced Statistical Methods in Biometric Research. New York: John Wiley and Sons, Inc.

Roy, J. (1958). Step-Down Procedure in Multivariate Analysis. Annals of Mathematical Statistics, 29, 1177.

Roy, S. N. and Sarhan, A. E. (1955). On Inverting a Class of Partitioned Matrices, Part I. Institute of Statistics, University of North Carolina, Mimeograph Series No. 123.

Roy, S. N. (1957). Some Aspects of Multivariate Analysis. New York: John Wiley and Sons, Inc.

Stevens, W. L. (1948). Statistical Analysis of a Non-Orthogonal Tri-Factorial Experiment. Biometrika, 35, 346.

Wilkinson, G. W. (1958). Estimation of Missing Values for the Analysis of Incomplete Data. Biometrics, 14, 257.

Wilkinson, G. W. (1958). The Analysis of Variance and Derivation of Standard Errors for Incomplete Data. Biometrics, 14, 360.

Yates, F. (1933). The Analysis of Replicated Experiments When the Field Results are Incomplete. Empire Journal of Experimental Agriculture, 1, 129.

Yates, F. (1933). The Principles of Orthogonality and Confounding in Replicated Experiments. Journal of Agricultural Science, 23, 108.

Yates, F. (1936). Incomplete Latin Squares. Journal of Agricultural Science, 26, 301.

Yates, F. and Hale, R. W. (1939). The Analysis of Latin Squares When Two or More Rows, Columns, or Treatments are Missing. Supplement of Journal of the Royal Statistical Society, 6, 67.

Abstract

The "Symmetrical Complementation Design" which is discussed in this dissertation is intended for those experimental situations where the levels of three factors always sum to the same constant. The levels of the factors, if referred to a common unit of measurement, must be equally spaced. Certain cell entries are omitted to insure complete interchangeability of the three factors. The usual additive model is assumed.

A detailed study about the types of functions which are estimable in this design is presented in the chapter on linear estimation. The study shows that the number of contrasts that can be estimated is limited. For example, the usual linear contrast, which would lead to the hypothesis of equality of effects of all levels of one factor, is not testable in this design. On the other hand, quadratic and higher order contrasts are estimable for each factor separately. These contrasts are combined into different hypotheses. Estimable functions in one factor only and in two factors are presented for the general case of p levels.

There are several methods which can be employed in order to obtain estimates of the treatment effects under various constraints. It must be noted, however, that these estimates are rather meaningless quantities. It is only when they are combined in estimable functions that unique results are obtained. Two methods are described in complete detail; the "high-low" method if only estimation is required, and the "modified high-low" method if both estimation and tests of

hypotheses are required. The complete inverse matrix required for this latter method, or a method of obtaining this patterned inverse, is presented.

For testing hypotheses, a general technique, based upon the inversion of the matrix in the modified high-low method, is presented. Sums of squares and test statistics are presented for the various hypotheses formulated. Sections are also included which indicate how one might obtain the response for intermediate levels of the factors, and how one might obtain response functions for single factors.

A chapter on extensions is presented, where n observations are available per treatment combination. In this connection, three different cases are considered;

a) the replications are strictly repetitions of the experiment under otherwise identical conditions. In this case, the analysis proceeds in the customary three-way analysis with n replicates per treatment combination;

b) the experiments within a cell represent repetitions over a period of time, during which some kind of trend may be present. In this case the analysis is readily extended into an analysis of covariance;

c) the experiments within a cell represent several experiments with the same experimental units, so that the observations within a cell are dependent. On the assumption that the covariance matrix of observations in a cell is the same for every cell, a multivariate analysis can be performed.

The problem of estimation is essentially the same in these methods. However different methods are necessary for the testing of hypotheses. Special discussion is also presented for the case where the levels of the factors are not equally spaced; and the case where the model is considered as a mixed model.

In conclusion, it has been found that this type of design requires a rather careful consideration of the types of functions that can be estimated and the types of hypotheses that can be tested. Recommendations for interpretation and statement of limitations are made in detail.