ANALYSIS OF NON-ORTHOGONAL EXPERIMENTS

by

JOSEPH ARIMJA ELUKPO

B. Sc. (Hons), University of Ibadan, Nigeria, 1965

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1967

Approved by:

Raja Nassar
Major Professor

TABLE OF CONTENTS

Page

INTRODUCTION 1

THE ANALYSIS OF VARIANCE PROCEDURE AND THE LEAST SQUARE PRINCIPLES . . . 2

1. THE DIRECT SOLUTION TO NORMAL EQUATIONS AND THE ANALYSIS OF VARIANCE 3

Two-way Classifications Without Interaction 3

Analysis of Variance and Sums of Squares 8

Two-way Classification With Interaction 10

Analysis of Variance and Sums of Squares 15

Three-way Classifications 16

Estimation of Variance Components 20

Method I 20

Method III 20

Illustration of Method I 22

Illustration of Method III 31

2. THE GENERAL ANALYTIC METHOD USING VARIANCE-COVARIANCE MATRIX 32

Application to Two-way Non-orthogonal Design 36

Application to Many Non-orthogonal Classifications 41

3. A GEOMETRICAL APPROACH TO THE ANALYSIS OF NON-ORTHOGONAL EXPERIMENTS 43

The Orthogonal Case 43

The Non-orthogonal Case 46

Analysis of Variance Test 49

4. NUMERICAL EXAMPLES 50

(1) Direct Solution to Normal Equations and Analysis of Variance 50

(a) Two-way without interactions 51

(b) Two-way with interactions 57

(2) The General Analytic Method Applied to Two-way Without Interaction 63

(3) The Geometrical Approach (Iterative Procedure) 66

5. DISCUSSION AND CONCLUSION 69

ACKNOWLEDGMENTS 71

REFERENCES 72

INTRODUCTION

Since the introduction of modern experimental design by the late Sir R. A. Fisher in 1925 a great deal of effort and ingenuity went into inventing methods by which heterogeneity could be satisfactorily controlled even with large number of treatments. The non-orthogonality of experimental data has, however, often presented a great deal of difficulty in the analysis of variance. Several non-orthogonal designs have been presented which all possess some degree of symmetry (e.g. partially balanced incomplete blocks), which makes it possible to give a fairly simple formula for statistical treatment of the data. In this presentation such designs will not be discussed.

The type of non-orthogonal experiments which will be the subject of this presentation will be those which have no balance or symmetry whatsoever. The topic of this report may therefore be rightly called "the analysis of messy data." These types of non-orthogonality are often encountered in animal breeding experiments where disproportionate subclass frequencies are often found. They may also be encountered in plant breeding experiments such as experiments with fruit trees where some of the trees may die off. The result of such unequal subclass frequencies is that different classes of effects become non-orthogonal. That is, the different kinds of effects, such as years, sex, sire or dam (in animal breeding), cannot be separated directly without entanglement.

In this presentation an attempt will be made to bring together the current procedures for analyzing these types of messy data.

THE ANALYSIS OF VARIANCE PROCEDURE AND THE LEAST SQUARE PRINCIPLE

The analysis of variance can be considered in various ways, which, although all lead to the same result, place different emphasis on particular points and require different computational techniques.

The analysis of variance could be viewed as a special case of multiple regression (Graybill, 1960). This, although useful as an expository method, has not been favored by those concerned with practical computation of results.

The analysis of variance may be approached by way of the Gauss-Markoff theorem and the method of least squares. This is only multiple regression viewed from a different aspect, but is an aspect from which it is easier to see what the analysis of variance does. The method of least squares leads to a variety of ways by which the analysis of variance can be computed, all of which require, at least implicitly, the solution of a set of normal equations.

All the procedures for the analysis of the non-orthogonal experiments that will be presented here will be based wholly on the least square principle. These procedures will be presented under three headings: (1) the direct solution of normal equations as was given by Harvey (1960), (2) the general analytic method using the variance-covariance matrix in the notation of Tocher (1952) and Plackett (1960), and (3) the iterative procedure using vector spaces language as was given by Kuiper (1952), Justasen and Keuls (1958).

1. THE DIRECT SOLUTION TO NORMAL EQUATIONS AND THE ANALYSIS OF VARIANCE

The normal equations give insight into the way observed values are related to the parameters, and as such, they are of intrinsic interest. For orthogonal designs the equations fall into sets each of which can be solved independent of the others, and in the simpler non-orthogonal (balanced) designs, standard linear operations (which can be found in standard experimental design textbooks) can be performed on the equations to make them soluble. However, in the case of the non-balanced non-orthogonal designs, there is no simple solution to the normal equations particularly when there are several treatments and treatment levels involved. In a simple case with only very few unknown the simultaneous equations can be solved without much difficulty. With several unknown parameters the process becomes quite complicated. A process will be presented which involves using a matrix approach in solving the normal equations after certain restrictions have been imposed on the parameter estimates.

Two-way Classifications Without Interaction

Mathematical model: In the case of a two-way classification with treatments A and B when interaction, AB, is assumed non-existent, the usual model is as follows:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} , \qquad (1)$$

$$i = 1, 2, \ldots p ,$$

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

$$k = 1, 2, \ldots n_{ij} ,$$

y_{ijk} = the k-th observation in the j-th B class and i-th A class,

μ = overall mean when equal number of subclass exists,

α_i = effect of the i-th A class,

β_j = effect of the j-th B class,

e_{ijk} = random error, assumed to be NID(o, σ_e^2).

The α_i and β_j may either be regarded as either fixed random or mixed effects. Whether α_i and β_j are fixed or random would not affect the solution of the normal equations.

<u>Normal equations</u>: By least-square methods it is possible to obtain the following normal equations which will be presented in tabular form (Table 1).

Table 1. Normal Equations

	$\hat{\mu}$		$\hat{\alpha}_i$		$\hat{\beta}_j$		RHM
μ:	$n_{..}$	0	$n_{i.}$	0	$n_{.j}$		$Y_{..}$
α_i:	$n_{i.}$		$n_{i.}$	0	n_{ij}	0	$Y_{i.}$
β_j:	$n_{.j}$		n_{ij}		$n_{.j}$		$Y_{.j}$

The equations may be rewritten as follows:

$$\mu:\ n_{..}\hat{\mu} + \sum_{j=1}^{q}\sum_{i=1}^{p} n_{ij}\hat{\alpha}_i + \sum_{j=1}^{q}\sum_{i=1}^{p} n_{ij}\hat{\beta}_j = \sum_{j=1}^{q}\sum_{i=1}^{p} Y_{ij}\ ,$$

$$\alpha_i:\ n_{i.}\hat{\mu} + \sum_{j=1}^{q} n_{ij}\hat{\alpha}_i + \sum_{j=1}^{q} n_{ij}\hat{\beta}_j = \sum_{j=1}^{q} Y_{ij}\ , \qquad (2)$$

$$\beta_j:\ n_{.j}\hat{\mu} + \sum_{i=1}^{p} n_{ij}\hat{\alpha}_i + \sum_{i=1}^{p} n_{ij}\hat{\beta}_j = \sum_{i=1}^{p} Y_{ij}\ ,$$

where $$n_{..} = \sum_{j=1}^{q} \sum_{i=1}^{p} n_{ij} ,$$

$$n_{i.} = \sum_{j=1}^{q} n_{ij} ,$$

$$n_{.j} = \sum_{i=1}^{p} n_{ij} ,$$

$$Y_{..} = \sum_{j=1}^{q} \sum_{i=1}^{p} Y_{ij} , \; Y_{ij} = \sum_{k=1}^{n_{ij}} Y_{ijk} .$$

Looking at equations (2), it will be noted that the sum of the coefficients for the $\hat{\alpha}_i$ in the μ: equation equals the sum of the coefficients for the $\hat{\beta}_j$ and the coefficients for the $\hat{\mu}$. In addition, the sum of the coefficients for the $\hat{\beta}_j$ in an α_i: equation equals the coefficient for the α_i while the total of the RHM's for the α_i equations and the β_j equations equals the grand total of Y.. . These equalities indicate that there are some linear relations between the rows and columns of the coefficient matrix of the equations. Thus in order to solve these equations or to invert the coefficient matrix it is necessary to impose restrictions on $\hat{\alpha}_i$ and $\hat{\beta}_j$ since the matrix is not of full rank.

Restrictions: A common restriction on these equations is to set $\hat{\alpha}_p = \hat{\beta}_q = 0$ and delete the equations and columns for $\hat{\alpha}_p$ and $\hat{\beta}_q$. The inverse of the resulting reduced coefficient matrix (variance-covariance) must be transformed if the standard errors of the $\hat{\mu} + \hat{\alpha}_i$ or the $\hat{\mu} + \hat{\beta}_j$ are desired, or if the coefficients of the variance components in the expectation of mean-squares are to be obtained by means of a short-cut procedure (Henderson, 1953). Because of these requirements, it is often generally preferred to impose the following restrictions:

$$\sum_{i=1}^{p} \hat{\alpha}_i = \sum_{j=1}^{q} \hat{\beta}_j = 0$$

These restrictions lead to certain subtractions in the variance-covariance matrix before inversion. If say α_p: equation is deleted, then the coefficients of the $\hat{\alpha}_p$ equation must be subtracted from other coefficients by columns and rows within the $\hat{\alpha}_i$ columns. Similarly for the β_q: equation, the coefficient of the $\hat{\beta}_q$ equation is subtracted from within $\hat{\beta}_j$ columns by columns and rows. The same operation is used for the RHM's. The RHM element for the α_p equation is subtracted from the RHM elements of the α_i equation. The same operation is used for RHM's of the β_q and β_j equations. Finally, the resulting reduced variance-covariance matrix is of order (1 + p - 1 + q - 1 =) p + q - 1 and it is symmetric. The number of the remaining equations is also the number of degrees of freedom among the α_i and among the β_j and one additional for μ.

<u>Obtaining parameter estimates</u>: The reduced equations could be rewritten in the following form (Matrix Notation)

$$C\theta^* = P'Y \qquad (3)$$

where C is (p + q - 1) x (p + q - 1) matrix

$$C = \begin{bmatrix} C_{11} & \cdots & C_{1,\, p+q-1} \\ \vdots & & \\ C_{p+q-1,\, 1} & \cdots & C_{p+q-1,\, p+1-1} \end{bmatrix},$$

$$\theta^* = \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\alpha}_{p-1} \\ \hat{\beta}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\beta}_{q-1} \end{bmatrix}$$

P'Y is the $(p + q - 1) \times 1$ matrix, the reduced RHM's.

θ^* is now given by

$$\theta^* = C^{-1} P'Y .$$

The most important part of the solution of these normal equations is the inversion of the C matrix. Several standard methods are available for inverting the matrix C (Fryer, 1965). For the purpose of finding standard errors of the parameter estimates and testing differences among parameter means, it is always useful to obtain the complete inverse elements of the variance-covariance matrix. In order to obtain the inverse elements for $\hat{\alpha}_p$ and $\hat{\beta}_q$, one makes use of the fact that the restrictions, $\sum_{i=1} \hat{\alpha}_i = 0$ and $\sum_j \hat{\beta}_j = 0$ were used in order to obtain the other inverse elements in the reduced matrix.

Thus, for $\hat{\alpha}_p$

$$C^{1,\,p+1} = -[C^{12} + C^{13} + \ldots C^{1p}]$$

$$\vdots$$

$$C^{p+1,\,p+1} = -[C^{1,p+1} + C^{2,p+1} + \ldots C^{p,p+1}]$$

and for $\hat{\beta}_q$

$$C^{1,(p+q+1)} = -[C^{1,p+1} + C^{2,p+1} + \ldots C^{p,p+q}]$$

$$\vdots$$

$$C^{p+q+1,\,p+q+1} = -[C^{i,\,p+q+1} + C^{2,p+q+1} + \ldots C^{p+q,\,p+q+1}]$$

where C^{ij} is i,j-th element of C^{-1}.

Analysis of Variance and Sums of Squares

The total reduction in sum of squares is given by:

$$R(\mu, \alpha_i, \beta_j) = \hat{\mu}\, Y_{..} + \sum_{i=1}^{p-1} \hat{\alpha}_i (Y_{i.} - Y_{p.}) + \sum_{j=1}^{q-1} \hat{\beta}_j (Y_{.j} - Y_{.q}). \tag{4}$$

$$\text{Error sum of squares} = \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{n_{ij}} y_{ijk}^2 - R(\mu, \alpha_i, \beta_j)$$

If these sums of squares are written in matrix notation of equation (3), then it is found that:

$$R(\theta) = \theta^{*\prime} P' Y$$

$$= (C^{-1} P' Y)' Y$$

$$= Y' P C^{-1} P' Y ,$$

and

$$\text{Error S.S.} = Y' Y - Y' P C^{-1} P' Y . \tag{4$'$}$$

In order to obtain the sums of squares for A and B treatments, the following general procedure proposed by Harvey (1960) is used. These are as follows:

$$\begin{aligned} \text{A S.sq.} &= \theta_A^{*'} Z_A^{-1} \theta_A^{*} , \\ \text{B S.sq.} &= \theta_B^{*'} Z_B^{-1} \theta_B^{*} , \end{aligned} \tag{5}$$

where θ_A^* is the row vector of the constant estimates of the α's; that is, $\theta_A^{*'} = (\hat{\alpha}_1, \hat{\alpha}_2, \ldots \hat{\alpha}_{p-1})$, Z_A^{-1} is the inverse of the square symmetrical segment of the reduced inverse variance-covariance matrix corresponding to the $\hat{\alpha}$'s. Similarly, $\theta_B^{*'} = (\hat{\beta}_1, \hat{\beta}_{2-}, \ldots \hat{\beta}_{q-1})$ and Z_B^{-1} is the inverse of the square symmetrical segment of the reduced inverse variance-covariance matrix corresponding to the $\hat{\beta}$'s.

The usual sums of squares for A and B is given by

$$\begin{aligned} \text{A S.sq.} &= R(\mu, \alpha_i, \beta_j) - R(\mu, \beta_j) , \\ \text{B S.sq.} &= R(\mu, \alpha_i, \beta_j) - R(\mu, \alpha_i) . \end{aligned} \tag{5)'}$$

Equations (5) and (5)' will give exactly the same result except for rounding errors.

Finally the analysis of variance table is as follows:

Table 2. Analysis of Variance for Model in (1)

Sources	D.F.	S.S.	M.S.
A	p-1	$\theta_A^{*'} Z_A^{-1} \theta_A^{*}$	$\theta_A^{*'} Z_A^{-1} \theta_{A/p-1}^{*}$
B	q-1	$\theta_B^{*'} Z_B^{-1} \theta_B^{*}$	$\theta_B^{*'} Z_B^{-1} \theta_{B/q-1}^{*}$
Error	n.. - p-q+1	$Y'Y-Y'PC^{-1}P'Y$	$\dfrac{Y'Y-Y'PC^{-1}P'Y}{n.. - p-q+1}$

The standard errors of the least squares means are obtained as follows:

$$S_{\hat{\mu}+\hat{\alpha}_i} = \sqrt{(C^{11} + C^{ii} - 2C^{1i})\ \hat{\sigma}_e^2}\ ,$$

where $\hat{\sigma}_e^2 = (Y'Y - Y'PC^{-1}P'Y)/_{n..\ -\ p-q+1}$.

The standard error between two constant estimates is given by

$$S_{\hat{\alpha}_i - \hat{\alpha}_j} = \sqrt{(C^{ii} + C^{jj} - 2C^{ij})\ \hat{\sigma}_e^2}\ .$$

Two-way Classification With Interaction

Whenever a two-way classification is considered it is always necessary to think of a possibility of interaction. This is of course only possible when there is more than one observation per cell. For two-way classification, when all cells are filled (even though not equally), there are some simplified methods like the weighted method (Snedecor, 1958) which have been described in standard statistical methods books. When all the cells or subclasses are not filled then these standard methods are inapplicable. The least squares principle fits most appropriately.

<u>The mathematical model</u>: The model for this is just like that given in (1) except for the extra interaction $(\alpha\beta)_{ij}$ which is added.

That is,

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}\ .\ .\ .\ , \qquad (6)$$

where

μ, α_i, β_j and e_{ijk} are as defined in (1) ,

$(\alpha\beta)_{ij}$ = effect of the ij-th AB subclass after the averages of A and B have been removed. These are the individual interaction effects expressed as deviations from the general mean μ.

The α's and β's may be regarded as fixed effects (Model I) and so only an analysis of variance for the test of significance will be considered in this section.

<u>Normal equations</u>: Using the least squares principle the following normal equations, presented in tabular form, are obtained.

Table 3. Normal equations for the model in (6).

	$\hat{\mu}$		$\hat{\alpha}_i$		$\hat{\beta}_j$		$\widehat{(\alpha\beta)}_{ij}$		RHM
μ:	$n_{..}$		$n_{i.}$		$n_{.j}$		n_{ij}		$Y_{..}$
				0					
α_i:	$n_{i.}$		$n_{i.}$		n_{ij}		n_{ij}		$Y_{i.}$
		0				0			
β_j:	$n_{.j}$		n_{ij}		$n_{.j}$		n_{ij}		$Y_{.j}$
				0				0	
$(\alpha\beta)_{ij}$	n_{ij}		n_{ij}		n_{ij}		n_{ij}		Y_{ij}
						0			

where

$$n_{..} = \sum_{i=1}^{p} \sum_{j=1}^{q} n_{ij}\ , \qquad Y_{..} = \sum_{i} \sum_{j} \sum_{k} y_{ijk}\ ,$$

$$n_{i.} = \sum_{j=1}^{q} n_{ij}\ , \qquad Y_{i.} = \sum_{j} \sum_{k} y_{ijk}\ ,$$

$$n_{.j} = \sum_{i=1}^{p} n_{ij}\ , \qquad Y_{.j} = \sum_{i} \sum_{k} y_{ijk}\ .$$

The zero beside an element indicates that off-diagonals of that element are zeroes.

<u>Restrictions</u>: In order to obtain a unique solution to the least squares equations it is necessary to impose some restriction on the constants. Numerous restrictions have been proposed but only the one found most appropriate for the purpose of complete analysis of the data will be mentioned here. This restriction is just an extension of that proposed for the "two-way" without interaction. The restrictions are as follows:

$$\sum_{i=1}^{p} \hat{\alpha}_i = \sum_{j}^{q} \hat{\beta}_j = \sum_{i}^{p} (\widehat{\alpha\beta})_{ij} = \sum_{j=1}^{q} (\widehat{\alpha\beta})_{ij} = 0 .$$

These restrictions are equivalent to certain subtractions and additions within the coefficient matrix and the right hand members (RHM's). The subtraction required within the α_i and β_j equations are the same by rows and columns as previously explained. For $(\widehat{\alpha\beta})_{ij}$ coefficients, the subtractions and additions which may be conveniently chosen for the rows are as follows:

$$n_{ij} - n_{iq} - n_{pj} + n_{pq} .$$

Similar subtraction and addition is made on the RHM's. That is,

$$Y_{ij} - Y_{iq} - Y_{pj} + Y_{pq} .$$

The final reduced equations gives a reduced variance-covariance matrix of order $p + q - 1 + (p - 1)(q - 1)$.

As before the reduced equations may be rewritten using matrix notation as

$$C\theta^{*} = P'Y \ldots , \tag{7}$$

where

C is a $[p + q - 1 + (p - 1)(q - 1)] \times [p + q - 1 (p - 1)(q - 1)]$ matrix, and

$$\theta^* = \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_{p-1} \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_{q-1} \\ (\hat{\alpha\beta})_{11} \\ (\hat{\alpha\beta})_{12} \\ \vdots \\ (\hat{\alpha\beta})_{p-1,\, q-1} \end{bmatrix}, \qquad P'Y = \begin{bmatrix} Y_{..} \\ Y_{1.} - Y_{p.} \\ \vdots \\ Y_{p-1.} - Y_{p.} \\ Y_{.1} - Y_{.q} \\ \vdots \\ Y_{.p-1} - Y_{.q} \\ Y_{11} - Y_{1q} - Y_{p1} + Y_{pq} \\ \vdots \\ Y_{p-1,q-1} - Y_{p-1,q} - Y_{p,q-1} + Y_{pq} \end{bmatrix}.$$

The most difficult part of solving these normal equations is finding the inverse C^{-1}. Once the inverse C^{-1} is obtained, then

$$\theta^* = C^{-1} P'Y ,$$

$$\hat{\alpha}_p = - \sum_{i=1}^{p-1} \hat{\alpha}_i ,$$

$$\hat{\beta}_q = - \sum_{q=1}^{q-1} \hat{\beta}_j ,$$

$$(\hat{\alpha\beta})_{p1} = -\sum_{i=1}^{p-1} (\hat{\alpha\beta})_{i1} ,$$

$$\vdots$$

$$(\hat{\alpha\beta})_{pq} = -\sum_{i=1}^{p-1} (\hat{\alpha\beta})_{iq} ,$$

$$(\hat{\alpha\beta})_{1q} = -\sum_{j=1}^{q-1} (\hat{\alpha\beta})_{1j} ,$$

$$(\hat{\alpha\beta})_{p-1,q} = -\sum_{j=1}^{q-1} (\hat{\alpha\beta})_{p-1,j} .$$

The inverse elements for the rows of α_p (or β_q) may be obtained by adding the inverse elements of the $\hat{\alpha}_i$ columns (or rows) and then reversing the signs as was previously explained. The same type of procedure is used for the inverse of the interaction columns and rows which were deleted.

<u>For example</u>:

$$c^{\mu\,\alpha\beta_{iq}} = c^{\alpha\beta_{iq}\mu} = -\sum_{j=1}^{q-1} c^{\mu\alpha\beta_{ij}} ,$$

$$c^{\alpha_i\,\alpha\beta_{iq}} = c^{\alpha\beta_{iq}\,\alpha_i} = -\sum_{j=1}^{q-1} c^{\alpha_i\,\alpha\beta_{ij}} ,$$

$$c^{\beta_j\,\alpha\beta_{iq}} = c^{\alpha\beta_{iq}\,\beta_j} = -\sum_{j=1}^{q-1} c^{\beta_j\,\alpha\beta_{ij}} ,$$

$$c^{\mu\,\alpha\beta_{pj}} = c^{\alpha\beta_{pj}\mu} = -\sum_{i=1}^{p-1} c^{\mu\alpha\beta_{ij}} ,$$

$$c^{\mu\,\alpha\beta_{pq}} = c^{\alpha\beta_{pq}\mu} = -\sum_{i=1}^{p-1} c^{\mu\,\alpha\beta_{iq}} = -\sum_{j=1}^{q-1} c^{\mu\alpha\beta_{ij}} .$$

Analysis of Variance and Sums of Squares

Using the notation of (5), the following sums of squares are obtained:

$$\text{Error S. sq.} = \sum_i \sum_j \sum_k y_{ijk}^2 - R(\theta) ,$$

$$R(\theta) = \theta^{*'} P' Y ,$$

$$\text{A S. sq.} = \theta_A^{*'} Z_A^{-1} \theta_A^* ,$$

$$\text{B S. sq.} = \theta_B^{*'} Z_B^{-1} \theta_B^* ,$$

$$\text{AB S. sq.} = \theta_{AB}^{*'} Z_{AB}^{-1} \theta_{AB}^* .$$

This gives the following analysis of variance table:

Table 4. Analysis of variance.

Sources	D.F.	S.S.	M.S.
A	(p-1)	$\theta_A^{*'} Z_A^{-1} \theta_A^*$	$\theta_A^{*'} Z_A^{-1} \theta_A^*/(p-1)$
B	k-1	$\theta_B^{*'} Z_B^{-1} \theta_B^*$	$\theta_B^{*'} Z_B^{-1} \theta_B^*/(q-1)$
AB	(p-1)(k-1)	$\theta_{AB}^{*'} Z_{AB}^{-1} \theta_{AB}^*$	$\theta_{AB}^{*'} Z_{AB}^{-1} \theta_{AB}^*/(p-1)(q-1)$
Error	n.. - p-q+1 - (p-1)(k-1)	$Y'Y - R(\mu, \alpha_i, \beta_j, (\alpha\beta)_{ij})$	$\dfrac{Y'Y - \theta^{*'}P'Y}{n..-p-q+1-(p-1)(q-1)}$

$$\theta^{*'}P'Y = R(\mu, \alpha_i, \beta_j, (\alpha\beta)_{ij}) = \hat{\mu}\, Y_{..}$$

$$+ \sum_{i=1}^{p-1} \hat{\alpha}_i (Y_{i.} - Y_{p.})$$

$$+ \sum_{j=1}^{q-1} \hat{\beta}_j (Y_{.j} - Y_{.q})$$

$$+ \sum_{i=1}^{p-1} \sum_{j=1}^{q-1} \widehat{(\alpha\beta)}_{ij} (Y_{ij} - Y_{iq} - Y_{pj} + Y_{pq}) .$$

The standard errors for $\hat{\mu}$, $\hat{\mu} + \hat{\alpha}_i$, and $\hat{\mu} + \hat{\beta}_j$ are computed in a similar way as previously shown for the "two-way" classification without interaction. In order to find the interaction effects it would be useful to examine more carefully the least square subclass mean S_{ij}:

$$\hat{S}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \widehat{(\alpha\beta)}_{ij} .$$

$$\therefore \hat{\sigma}_{\hat{S}_{ij}} = \sqrt{\left\{ C^{\mu\mu} + C^{\alpha_i\alpha_i} + C^{\beta_j\beta_j} + C^{(\alpha\beta)_{ij}} + 2C^{\mu\beta_j} + 2C^{\alpha_i\beta_j} + 2C^{\alpha_i(\alpha\beta)_{ij}} + 2C^{\beta_j(\alpha\beta)_{ij}} \right\} \hat{\sigma}_e^2} , \quad (8)$$

where $$\hat{\sigma}_e^2 = \frac{Y'Y - \theta^{*'}P'Y}{n_{..} - p - q + 1 - (p-1)(q-1)} .$$

When all the subclasses are filled up this long formula for the subclass standard error reduces to $\hat{\sigma}^2_{\hat{S}_{ij}} = \frac{1}{n_{ij}} \hat{\sigma}_e^2$. Formula (8) therefore gives the general standard error for the subclass mean $\hat{S}_{ij}$.

Three-way Classifications

Suppose in an experiment with three treatments A, B and C, there exists an unequal number of observations per cell, then the least square procedure for the "two-way" is naturally extended to the "three-way." If two order and three order interactions are considered then the mathematical model is as follows:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + e_{ijkl} \quad (9)$$

$i = 1, 2, \ldots p,$

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

$k = 1, 2, \ldots r,$

$l = 1, 2, \ldots n_{ijk},$

where y_{ijkl} = the l-th observation in the k-th C class, j-th B class and i-th A class,

μ = overall mean with equal subclass number,

α_i = effect of the i-th A class,

β_j = effect of the j-th B class,

γ_k = effect of the k-th C class,

$(\alpha\beta)_{ij}$ = effect of the ij-th (AB) subclass after the effects of A and B have been removed,

$(\alpha\gamma)_{ik}$ = effect of ik-(AC) subclass after the effects of A and C have been removed,

$(\beta\gamma)_{jk}$ = effect of the jk-(BC) subclass after the effects of B and C have been removed,

$(\alpha\beta\gamma)_{ijk}$ = effect of the ijk-th observation within subclass (cell) after the average of A, B and C have been removed. These are individual interaction effects expressed as deviation from the general mean μ,

e_{ijkl} = random error. Assumed NID(o, σ_e^2).

The normal equations for this becomes more involved. It is summarized in the table below.

Table 5. Normal equations for model in (9).

	$\hat{\mu}$	$\hat{\alpha}_i$	$\hat{\beta}_j$	$\hat{\gamma}_k$	$\widehat{(\alpha\beta)}_{ij}$	$\widehat{(\alpha\gamma)}_{ik}$	$\widehat{(\beta\gamma)}_{jk}$	$\widehat{(\alpha\beta\gamma)}_{ijk}$	RHM's
μ	$n_{\cdot\cdot\cdot}$	$n_{i\cdot\cdot}$	$n_{\cdot j\cdot}$	$n_{\cdot\cdot k}$	$n_{\cdot j\cdot}$	$n_{i\cdot k}$	$n_{\cdot jk}$	n_{ijk}	$Y_{\cdot\cdot\cdot}$
α_i	$n_{i\cdot\cdot}$	$n_{i\cdot\cdot}$	$n_{ij\cdot}$	$n_{i\cdot k}$	$n_{ij\cdot}$	$n_{i\cdot k}$	n_{ijk}	n_{ijk}	$Y_{i\cdot\cdot}$
β_j	$n_{\cdot j\cdot}$	$n_{ij\cdot}$	$n_{\cdot j\cdot}$	$n_{\cdot jk}$	$n_{ij\cdot}$	n_{ijk}	$n_{\cdot jk}$	n_{ijk}	$Y_{\cdot j\cdot}$
γ_k	$n_{\cdot\cdot k}$	$n_{i\cdot k}$	$n_{\cdot jk}$	$n_{\cdot\cdot k}$	n_{ijk}	$n_{i\cdot k}$	$n_{\cdot jk}$	n_{ijk}	$Y_{\cdot\cdot k}$
$(\alpha\beta)_{ij}$:	$n_{ij\cdot}$	$n_{ij\cdot}$	$n_{ij\cdot}$	n_{ijk}	$n_{ij\cdot}$	n_{ijk}	n_{ijk}	n_{ijk}	$Y_{ij\cdot}$
$(\alpha\gamma)_{ik}$:	$n_{i\cdot k}$	$n_{i\cdot k}$	n_{ijk}	$n_{i\cdot k}$	n_{ijk}	$n_{i\cdot k}$	n_{ijk}	n_{ij}	$Y_{i\cdot k}$
$(\beta\gamma)_{jk}$:	$n_{\cdot jk}$	n_{ijk}	$n_{\cdot jk}$	$n_{\cdot jk}$	n_{ijk}	n_{ijk}	$n_{\cdot jk}$	n_{ij}	$Y_{\cdot jk}$
$(\alpha\beta\gamma)_{ijk}$:	n_{ijk}	n_{ijk}	n_{ijk}	n_{ijk}	n_{ijk}	n_{ijk}	n_{ijk}	n_{ijk}	Y_{ijk}

$$n_{\cdot\cdot\cdot} = \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{r} n_{ijk} ,$$

$$n_{i\cdot\cdot} = \sum_{j} \sum_{k} n_{ijk} ,$$

$$n_{\cdot jk} = \sum_{i} n_{ijk} ,$$

$$n_{i\cdot k} = \sum_{j} n_{ijk} ,$$

$$Y_{\cdot\cdot\cdot} = \sum_{i} \sum_{j} \sum_{k} \sum_{l} y_{ijkl} ,$$

$$Y_{ijk} = \sum_{l=1}^{n_{ijk}} y_{ijkl} \ ,$$

n_{ijk} is the number of observations in the ijk-th cell.

In order to obtain a unique solution the following restrictions are imposed on the parameter estimates:

$$\sum_i \hat{\alpha}_i = \sum_j \hat{\beta}_j = \sum_k \hat{\gamma}_k = \sum_i (\widehat{\alpha\beta})_{ij} = \sum_j (\widehat{\alpha\beta})_{ij}$$

$$= \sum_i (\widehat{\alpha\gamma})_{ik} = \sum_k (\widehat{\alpha\gamma})_{ik} = \sum_j (\widehat{\beta\gamma})_{jk} = \sum_k (\widehat{\beta\gamma})_{jk}$$

$$= \sum_i (\widehat{\alpha\beta\gamma})_{ijk} = \sum_j (\widehat{\alpha\beta\gamma})_{ijk} = \sum_k (\widehat{\alpha\beta\gamma})_{ijk} = 0 \ .$$

A similar operation of subtraction and addition as before is carried on the coefficient (variance-covariance) matrix and the RHM's before the reduced variance-covariance matrix is inverted to obtain the parameter estimates.

The analysis of variance procedure is quite similar to that discussed for the two-way with interaction. This least squares procedure can be extended to more than three-ways but interactions of order greater than three may not be of much meaning. As was seen from the normal equations in Table 5, the matrix becomes quite large. For example when p = 5, q = 3, n = 2, the coefficient matrix of the normal equations would be of order 72 x 72. After the restrictions are imposed the reduced matrix would be of order 30 x 30 (non-singular) which is still quite large. Most computers may not be able to handle more than 40 x 40 matrix. With the availability of large storage computers, a much larger size of experiment with unbalanced data could be handled.

Estimation of Variance Components

All the presentations given so far have assumed that the treatment effects were fixed. When there are random or mixed treatment effects in the type of non-orthogonal experiments under consideration the estimation of variance components is not quite simple. Henderson (1953) has proposed three techniques, two of which will be presented here. Both of these two methods are based on least-squares principles.

Method I (Henderson, 1953). This involves the computation of sums of squares as in a standard analysis of variance of corresponding orthogonal data. Then the sums of squares are equated to their expectations under the assumption of Model II (random effects) to solve for the unknown variance components.

Method III (Henderson, 1953). This method which Henderson calls Method III involves the computation of mean squares analysis of non-orthogonal data as already discussed. These mean squares are then equated to their expectations and solve for the unknown variance components. Method III was much more laborious than method II but now with the availability of digital computers the method can be used with less difficulty.

Illustration of methods using Henderson's example.

Table 6. Butter Fat Records.

Herd	Sire	Year 1	2	3	4	Total
1	1	3-1414	2-981			5-2395
1	2		4-1766	2-862		2-2628
1	3				5-1609	5-1609
2	1	1-404	3-1270			4-1674
2	2			5-2109		5-2109
2	3			4-1563	2-740	6-2303
3	1		3-1705			3-1705
3	2		4-2310	2-1134		6-3444
4	1	3-1113	5-1951			8-3064
4	3			3-1291	6-2457	9-3748
Total		7-2931	21-9983	16-6959	13-4806	57-24679

The data in Table 6 shows the number of butterfat records in each of the year x herd x sire subclasses and also the sum of the records for each of these subclasses.

It is noted that two major classifications of the data are sire and herd. The number of observations per herd-sire subclass varies; the majority being 0. This condition of the data is bound to present difficulties when an attempt is made to estimate the pertinent variance components.

The linear model proposed for this data is as follows:

$$y_{hijk} = \mu + a_h + h_i + s_j + (hs)_{ij} + e_{hijk} \qquad (10)$$

$h = 1, \ldots p ,$

$$i = 1, \ldots q ,$$

$$k = 1, \ldots n_{hij} , \qquad j = 1, \ldots r ,$$

$$N = \sum_h \sum_i \sum_j n_{hij} .$$

Total number of filled subclasses = s.

y_{hijk} = the record made in the h-th year by k-th daughter of the j-th sire in the i-th herd,

μ = overall mean common to all observations,

a_h = effect common to all observations in the h-year,

h_i = effect common to all observations in the i-th herd,

s_j = effect common to all records made by daughters of the j-th sire,

$(hs)_{ij}$ = effects peculiar to all records made by daughters of the j-th sire in the i-th herd,

e_{hijk} = random error element assumed to have mean zero and variance σ_e^2.

Illustration of Method I. This method is proposed for the case when it is assumed that except for the μ, all elements of the model are uncorrelated variables with mean zero and variances σ_a^2, σ_h^2, σ_s^2, σ_{hs}^2 and σ_e^2. This is the usual Eisenhart Model II (Random).

The following quantities are computed

$$T = \sum_h \sum_i \sum_j \sum_k y_{hijk}^2 ,$$

$$H = \sum_i \left(\frac{y_{\cdot i \cdot \cdot}^2}{n_{\cdot i \cdot}}\right) , \qquad (11)$$

$$A = \sum_h \left(\frac{y_{h\cdots}^2}{n_{h\cdot\cdot}}\right),$$

$$S = \sum_j \left(\frac{y_{\cdot\cdot j\cdot}^2}{n_{\cdot j\cdot}}\right),$$

$$HS = \sum_i \sum_j \left(\frac{y_{\cdot ij\cdot}^2}{n_{\cdot ij}}\right), \qquad (11)$$

$$CF = \frac{y_{\cdots\cdot}^2}{N}.$$

Using the assumptions of model II the expectation of the above quantities are computed. For example,

$$E\,(HS) = \left[E \sum_i \sum_j \left(\frac{y_{\cdot ij\cdot}^2}{n_{\cdot ij}}\right)\right],$$

$$= \sum_i \sum_j E\,[n_{\cdot ij}\,\mu + n_{1ij}a_1 + \ldots + n_{pij}a_p +$$

$$n_{\cdot ij}h_i + n_{\cdot ij}\,S_j + \ldots + n_{\cdot ij}\,(hs)_{ij}$$

$$+ \sum_h \sum_k e_{hijk}]^2/n_{\cdot ij}\,,$$

$$= \sum_i \sum_j E\,[n_{\cdot ij}^2\,\mu^2 + n_{1\,ij}^2\,a_1^2 + \ldots$$

$$+ n_{pij}^2\,a_p^2 + n_{\cdot ij}^2\,h_i^2 + n_{\cdot ij}^2\,S_j^2 + n_{\cdot ij}^2\,(hs)_{ij}^2$$

$$+ \sum_h \sum_k e_{hijk}^2 + \text{Cross products all of which have}$$

$$\text{zero expectation}]/n_{\cdot ij}\,,$$

$$= \sum_i \sum_j [n_{\cdot ij}^2 \mu^2 + \sum_{h=1}^{p} n_{hij} \sigma_a^2 + n_{\cdot ij}^2 \sigma_h^2 + n_{\cdot ij}^2 \sigma_s^2$$

$$+ n_{\cdot ij}^2 \sigma_{hs}^2 + \sum_h \sum_k \sigma_e^2]/n_{\cdot ij} ,$$

$$\therefore E[HS] = N\mu^2 + \sum_i \sum_j \left(\frac{\sum_h n_{hij}}{n_{\cdot ij}}\right) \sigma_a^2$$

$$+ N (\sigma_h^2 + \sigma_{hs}^2 + \sigma_s^2) + S \sigma_e^2 . \qquad (12)$$

(where S is the number of subclasses filled).

Similarly, the others could be found. The coefficient of μ^2 and the variance components in the expectations are as shown in the table below.

Table 7. Variance-component Coefficients.

	μ^2	σ_a^2	σ_h^2	σ_s^2	σ_{hs}^2	σ_e^2
T	N	N	N	N	N	N
A	N	N	K_1	K_2	K_3	P
HS	N	K_4	N	N	N	S
H	N	K_5	N	K_6	K_6	q
S	N	K_7	K_8	N	K_8	r
C	N	K_9	K_{10}	K_{11}	K_{12}	1

From the expectations as shown by relation (12) values of K_1, K_2, . . ., K_{12} are as follows:

$$
\begin{aligned}
K_1 &= \sum_h \left(\frac{\sum_i n_{hi\cdot}^2}{n_{h\cdot\cdot}}\right), \\
K_2 &= \sum_h \left(\frac{\sum_j n_{h\cdot j}^2}{n_{h\cdot\cdot}}\right), \\
K_3 &= \sum_h \left(\frac{\sum_i \sum_j n_{hij}^2}{n_{h\cdot\cdot}}\right), \\
K_4 &= \sum_i \sum_j \left(\frac{\sum_h n_{hij}^2}{n_{\cdot ij}}\right), \\
K_5 &= \sum_i \left(\frac{\sum_h n_{hi\cdot}^2}{n_{\cdot i\cdot}}\right), \\
K_6 &= \sum_i \left(\frac{\sum_j n_{\cdot ij}^2}{n_{\cdot i\cdot}}\right), \\
K_7 &= \sum_j \left(\frac{\sum_h n_{h\cdot j}^2}{n_{\cdot\cdot j}}\right), \\
K_8 &= \sum_j \left(\frac{\sum_i n_{\cdot ij}^2}{n_{\cdot j\cdot}}\right), \\
K_9 &= \sum_h (n_{h\cdot\cdot}^2)/N, \\
K_{10} &= \sum_i (n_{\cdot i\cdot}^2)/N, \\
K_{11} &= \sum_j (n_{\cdot\cdot j}^2)/N, \\
K_{12} &= \sum_i \sum_j (n_{\cdot ij}^2)/N .
\end{aligned}
\tag{13}
$$

If the data were orthogonal, the sum of squares in the analysis of variance would be

Among Years = A-CF ,

Among Herds = H-CF ,

Among Sires = S-CF ,

Herds x Sires = HS-H-S + CF ,

Error = T-A-HS + CF .

If these same quantities are computed in spite of the non-orthogonality and are equated to their expectations, unbiased estimates of the variances can be obtained by solving the resulting equations. The necessary expectations are derived from Table 7.

In order to compute the K's, other tables have to be reconstructed from Table 6. The following two way tables give the subclass numbers.

Table 8. Herd x Year.

Herd	Year 1	2	3	4	Total
1	3	6	2	5	16
2	1	3	9	2	15
3	0	7	2	0	9
4	3	5	3	6	17
Totals	7	21	16	13	57

Table 9. Sire x Year.

Sire	Year 1	2	3	4	Total
1	7	13	0	0	20
2	0	8	9	0	17
3	0	0	7	13	20
Totals	7	21	16	13	57

Table 10. Herd x Sire.

Herd	Sire 1	2	3	Total
1	5	6	5	16
2	4	5	6	15
3	3	6	0	9
4	8	0	9	17
Totals	20	17	20	57

The following table shows certain totals computed from Table 6.

Table 11. Year x Herd.

Year	Herd	Sire
1. 2931	1. 6632	1. 8838
2. 9983	2. 6086	2. 8181
3. 6959	3. 5149	3. 7660
4. 4806	4. 6812	
Total 24,679	Total 24,679	Total 24,679

Using the totals in Table 11 and Table 6,

$$A = \frac{2931^2}{7} + \ldots + \frac{4806^2}{13} = 10,776,451 ,$$

$$HS = \frac{2395^2}{5} + \ldots + \frac{3748^2}{9} = 10,970,369 ,$$

$$H = \frac{6632^2}{16} + \ldots + \frac{6812^2}{17} = 10,893,666 ,$$

$$S = \frac{8838^2}{20} + \frac{8181^2}{17} + \frac{7660^2}{20} = 10,776,278 ,$$

$$CF = 24,679^2/57 = 10,685,141 .$$

The computations of the K's are as follows:

<u>From Table 8</u>:

$$K_1 = \frac{3^2 + 1^2 + 3^2}{7} + \ldots + \frac{5^2 + 2^2 + 6^2}{13} = 19.51 .$$

<u>From Table 9</u>:

$$K_2 = \frac{7^2}{7} + \frac{13^2 + 8^2}{21} + \frac{9^2 + 7^2}{16} + \frac{13^2}{13} = 39.22 .$$

<u>From Table 6</u>:

$$K_3 = \frac{3^2 + 1^2 + 3^2}{7} + \ldots + \frac{5^2 + 2^2 + 6^2}{13} = 15.10 .$$

<u>From Table 6</u>:

$$K_4 = \frac{3^2 + 2^2}{5} + \ldots + \frac{3^2 + 6^2}{9} = 37.35 .$$

<u>From Table 8</u>:

$$K_5 = \frac{3^2 + 6^2 + 2^2 + 5^2}{16} + \ldots + \frac{3^2 + 5^2 + 3^2 + 6^2}{17} = 21.49 .$$

<u>From Table 10</u>:

$$K_6 = \frac{5^2 + 6^2 + 5^2}{16} + \ldots + \frac{8^2 + 9^2}{17} = 24.04 .$$

<u>From Table 9</u>:

$$K_7 = \frac{7^2 + 13^2}{20} + \frac{8^2 + 9^2}{17} + \frac{7^2 + 13^2}{20} = 30.33 .$$

<u>From Table 10</u>:

$$K_8 = \frac{5^2 + 4^2 + 3^2 + 8^2}{20} + \ldots + \frac{5^2 + 6^2 + 9^2}{20} = 18.51 .$$

<u>From Table 8</u>:

$$K_9 = \frac{7^2 + 21^2 + 16^2 + 13^2}{57} = 16.05 .$$

<u>From Table 8</u>:

$$K_{10} = \frac{16^2 + 15^2 + 9^2 + 17^2}{57} = 14.93 .$$

<u>From Table 9</u>:

$$K_{11} = \frac{20^2 + 17^2 + 20^2}{57} = 19.11 .$$

<u>From Table 6</u>:

$$K_{12} = \frac{5^2 + 6^2 + \ldots + 9^2}{57} = 6.19 .$$

The table below gives the expectation of the quantities T, A, HS, . . ., CF.

Table 12. Variance Component Coefficients.

	μ^2	σ_a^2	σ_h^2	σ_s^2	σ_{hs}^2	σ_e^2	
T	57	57	57	57	57	57	11,124,007
A	57	57	19.51	39.22	15.10	4	10,776,451
HS	57	37.35	57	57	57	10	10,970,369
H	57	21.49	57	24.04	24.04	4	10,893,666
S	57	30.33	18.51	57	18.51	3	10,776,278
CF	57	16.05	14.93	19.11	6.19	1	10,685,141

In order to solve the equations the following table must be obtained:

Table 13. Equations for the Variance Components.

	σ_a^2	σ_h^2	σ_s^2	σ_{hs}^2	σ_e^2	
A–CF	40.95	4.58	20.11	8.91	3	91,310
H–CF	5.44	42.07	4.93	17.85	3	208,525
S–CF	14.28	3.58	37.89	12.32	2	91,137
HS–H–S+CF	1.58	-3.58	-4.93	20.64	4	-14,434
T–A–HS+CF	-21.30	-4.58	-20.11	-8.91	44	62,328

Table 13 now becomes five equations in five unknowns (σ_a^2, σ_h^2, σ_s^2, σ_{hs}^2, σ_e^2).

On solving these equations the following estimates of σ^2's are obtained.

$$\sigma_a^2 = 763 ,$$

$$\sigma_h^2 = 4531 ,$$

$$\sigma_s^2 = 1587 ,$$

$$\sigma_{hs}^2 = -164 ,$$

$$\sigma_e^2 = 2950 .$$

If σ_{hs}^2 is set equal to zero then the solution is:

$$\sigma_a^2 = 756, \quad \sigma_h^2 = 4468, \quad \sigma_s^2 = 1542, \quad \sigma_e^2 = 2952 .$$

These estimates have no practical value for p, q, r and s are much too small for accurate estimation of the corresponding variance component. The method is however very important.

The difficulty with this method (Method I) is that it is inappropriate when one of the effects is fixed. For the example considered above it may be inappropriate to regard the years as random variable and so the estimates of σ_h^2, σ_s^2 and σ_{hs}^2 are bound to be biased.

Illustration of Method III. Using the model of equation (10), by the least squares procedure already given the following sum of squares can be computed:

Total reduction $= R(\mu, a_h, h_i, S_j, (hs)_{ij})$,

Among years $= R(\mu, a_h, h_i, S_j, (hs)_{ij}) - R(\mu, h_i, S_j, (hs)_{ij})$,

Among herds $= R(\mu, a_h, h_i, S_j, (hs)_{ij}) - R(\mu, a_h, S_j, (hs)_{ij})$,

Among sires $= R(\mu, a_h, h_i, S_j, (hs)_{ij}) - R(\mu, a_h, h_i, (hS)_{ij})$,

Herds x Sires $= R(\mu, a_h, h_i, S_j, (hS)_{ij}) - R(\mu, a_h, h_i, S_j)$.

$$\text{Error } SS = \sum_h \sum_i \sum_j \sum_k y^2_{hijk} - R\,(\mu,\ a_h,\ h_i,\ S_j,\ (hS)_{ij})\ .$$

The last four of these quantities can be used to estimate σ^2_h, σ^2_s, σ_{hs}^2 and σ^2_e respectively. The present assumption is that the year effects are fixed. As before, taking the expectations of these sums of squares and equating them to their corresponding calculated sums of squares results in the following table of coefficients for the variance components.

Table 14. Variance Component Coefficients.

S.S.	σ^2_h	σ^2_s	σ^2_{hs}	σ^2_e
Among herds	$N-K_4$	0	K_1-K_5	q-1
Among sires	0	$N-K_2$	K_1-K_3	r-1
Herds x sire	0	0	$N-K_1$	s-q-r+1
Residual	0	0	0	N-q-s+1

The difficulty of this method is in obtaining the coefficients $K_1, \ldots, K_5$. Henderson (1953) has shown how the K's can be obtained theoretically and he also gives the application of the method to the data given above.

2. THE GENERAL ANALYTIC METHOD USING VARIANCE-COVARIANCE MATRIX

In this section a more general approach to the least squares method already obtained will be given.

Several general methods have been proposed since Tocher (1952) presented his matrix approach to the analysis of variance. Most of these methods are based on least squares principles. Freeman and Jeffers (1962) produced a general method for the analysis of variance of non-orthogonal experiments for three-way classification. Clarke (1963) has described the analysis of four-way classification design with two independent non-orthogonalities. Bradau (1965) has given a computational scheme for estimating main effects for any number of treatments and finally Rees (1966) has also given a computational scheme for the analysis of variance of designs with many non-orthogonal classifications. An attempt will be made to present these methods in a general form based on a model by Plackett (1960).

Model: The basic model is essentially that adopted by Plackett (1960). An experiment with N observations forming a vector Y, whose values are linear combinations of p unknown parameters forming a vector Θ. Thus

$$Y = A\Theta + \epsilon , \tag{14}$$

where Y is $N \times 1$ vector ,

ϵ is $N \times 1$ vector ,

$E(\epsilon) = \underline{0}$,

$E(\epsilon\epsilon') = \sigma^2 I$.

A is $N \times p$ matrix and assumed to be of rank $p-m$ ($m > 0$). Since A is of rank $p-m$ ($m > 0$), then there must exist a $p \times (p-m)$ matrix D of rank m such that $AD = \underline{0}$. Therefore, an $m \times p$ matrix B must be defined, such that $|BD| \neq 0$, and the restriction $B\Theta = \underline{0}$ imposed. Then the best estimate of Θ is

$$\Theta^* = (A'A + B'B)^{-1} A'Y \text{ with variance-covariance matrix}$$

$$\sigma^2 \left\{ I - D(BD)^{-1} B \right\} (A'A + B'B)^{-1} .$$

That the above statement is true will be shown in the following proof.

Proof: Using the model given in (14), the normal equation is as follows (Graybill, 1960):

$$A'A\,\theta = A'Y \tag{15}$$

Suppose $\theta^* = LY$ is an unbiased estimation of θ, then

$$E(\theta^*) = LE(Y) = LA\theta = \theta$$

only if $LA = I$.

This implies $D = LAD = \underline{0}$ (since $AD = 0$). But this contradicts the assumption that $E(\theta^*) = \theta$. Hence $LA = I$ cannot hold. We therefore add the constraint

$$B\theta = \underline{0} \quad \text{where } B \text{ is } m \times p \text{ and}$$

$$BD \neq 0, \Rightarrow BD \neq \underline{0}\ .$$

Let us again assume that $E(\theta^*) = \theta$, subject to $B\theta = \underline{0}$, where as before $\theta^* = LY$.

Therefore,

$$E(LY) = LA\theta = \theta \quad \text{(if unbiased)}\ ,$$

$$\theta = LA\theta + MB\theta \quad (\text{since } B\theta = \underline{0})\ ,$$

$$\theta = (LA + MB)\,\theta\ ,$$

It follows,

$$I = LA + MB$$

and,

$$D = LAD + MBD = MBD \quad (\text{since } AD = \underline{0})\ .$$

Therefore,

$$M = D(BD)^{-1}$$

Thus,

$$I = LA + D(BD)^{-1}B . \tag{16}$$

Alternatively:

$$(A'A + B'B)D = B'(BD) \quad (\text{since } AD = \underline{0})$$

and

$$D(BD)^{-1} = (A'A + B'B)^{-1}B' .$$

Therefore, using equation 16,

$$\begin{aligned} LA &= I - (A'A + B'B)^{-1} B'B \\ &= (A'A + B'B)^{-1} (A'A + B'B) - (A'A + B'B)^{-1} B'B \\ &= (A'A + B'B)^{-1} A'A . \end{aligned} \tag{17}$$

Now, using the least squares principle and imposing the restriction $B\Theta = \underline{0}$, one finds the minimum of $(Y-A\Theta)'(Y-A\Theta)$.

Let

$$S = (Y-A\Theta)' (Y-A\Theta) + \lambda' B\Theta$$

where λ is a Lagrange vector multiplier.

Then,

$$\frac{\partial S}{\partial \Theta} = -2(Y-A\Theta)'A + \lambda' B = \underline{0} ,$$

and multiplying by D, it is found that

$$-2Y'AD + 2\Theta'A'AD + \lambda' BD = \underline{0} ,$$

or

$$\lambda' BD = \underline{0}$$

But

$$BD \neq 0$$

Thus

$$\lambda = \underline{0}$$

This implies that the absolute minimum of $(Y-A\theta)'(Y-A\theta)$ is equivalent to conditional minimum subject to $B\theta = \underline{0}$. That is,

$$A'A\theta^* = A'Y , \qquad B\theta^* = \underline{0} .$$

Thus,

$$(A'A + B'B)\,\theta^* = A'Y ,$$

and,

$$\theta^* = (A'A + B'B)^{-1} A'Y .$$

Now

$$E\left[(\theta^* - \theta)(\theta^* - \theta)'\right] =$$

$$= \sigma^2 (A'A + B'B)^{-1} A'A (A'A + B'B)^{-1}$$

$$= \sigma^2 LA (A'A + B'B)^{-1} \quad \text{(using (17))}$$

$$= \sigma^2 \left\{I - D(BD)^{-1} B\right\}(A'A + B'B)^{-1} \quad \text{(using (16))} . \qquad (18)$$

Thus the best estimate is that subject to $B\theta^* = 0$, and $\theta^* = (A'A + B'B)^{-1} A'Y$ while (18) gives the variance-covariance of θ^*.

<u>Note</u>: The restriction $B\theta^* = 0$ justifies the restriction used when we were dealing with the direct solution of the normal equations.

Residual sum of squares is given by $Y'Y - Y'A\theta^*$ and

$$E \frac{[Y'Y - Y'A\theta^*]}{n-p+m} = \sigma^2 .$$

Application to Two-way Non-orthogonal Design

Using the notation of Freeman and Jeffers (1962), we suppose that we have two-way classifications, treatments and blocks and there is no general parameter, the rank of A is p-1, so that one constraint is necessary. Let θ be partitioned into

$$\theta = \begin{bmatrix} \underline{t} \\ \hline \underline{b} \end{bmatrix} \tag{19}$$

Let the experiment have incidence matrice $\underline{n}$ whose column corresponds to blocks and rows to treatments. Then the elements of n represent the number of times a particular treatment occurs in a particular block.

Let $\underline{r}$ represent a vector of replication and $\underline{k}$ represent a vector of block sizes. Also let $\underline{1}$ be a unit vector. Then the following relations hold:

$$\begin{aligned} \underline{n}\ \underline{1} &= \underline{r} \\ \underline{n}'\ \underline{1} &= \underline{k} \\ \underline{r}'\ \underline{1} &= \underline{k}'\ \underline{1} = N \end{aligned} \tag{20}$$

We can choose

$$B = [\underline{0}\ \ \underline{K}']$$

and
$$D = \begin{bmatrix} -\underline{1} \\ \hline -\underline{1} \end{bmatrix} \tag{21}$$

$\therefore$ Using last relation in (20)

$$(BD)^{-1} = 1/N$$

$$\therefore\ I - D(BD)^{-1}B = I - 1/N \begin{bmatrix} 0 & -\underline{1}\ \underline{k}' \\ 0 & \underline{1}\ k' \end{bmatrix}$$

$$= \begin{bmatrix} I & \frac{1}{N}\underline{1}\ \underline{k}' \\ 0 & -\frac{1}{N}\underline{1}\ k' \end{bmatrix} \tag{22}$$

Let $\underline{r}^{\delta}$, $\underline{k}^{\delta}$ be the diagonal matrices of $\underline{r}$ and $\underline{k}$ respectively.

The coefficient matrix of the normal equations is:

$$A'A = \begin{bmatrix} \underline{r}^{\delta} & \underline{n} \\ \underline{n}' & \underline{k}^{\delta} \end{bmatrix} \tag{23}$$

$$\therefore \; (A'A + B'B) = \begin{bmatrix} \underline{r}^{\delta} & \underline{n} \\ \underline{n}' & \underline{k}^{\delta} + \underline{k}\,\underline{k}' \end{bmatrix} \tag{24}$$

The most important part of the computational procedure now is to find the inverse of (A'A + B'B). For the two-way classification this will be easily found using a desk calculator if the number of treatment levels is not too large. Of course this can easily be programmed for the digital computers. In order to find $(A'A + B'B)^{-1}$, we make use of the following relations. Let

$$\underline{\Omega}^{-1} = \underline{r} - \underline{n}\,\underline{k}^{-\delta}\,\underline{n}' + (\tfrac{1}{N})\,\underline{r}\,\underline{r}'$$

and

$$\underline{\Omega}^{-1}\,\underline{1} = \underline{r} - \underline{r} + \underline{r} = \underline{r} \tag{25}$$

or

$$\underline{\Omega}\,\underline{r} = \underline{1}$$

The inverse is now given as:

$$(A'A + B'B)^{-1} = \begin{bmatrix} \underline{\Omega} + (\frac{1}{N^2})\,\underline{1}\,\underline{1}' & -\underline{\Omega}\,\underline{n}\,k^{-\delta} + (\frac{N-1}{N^2})\,\underline{1}\,\underline{1}' \\ -\underline{k}^{-\delta}\,\underline{n}'\,\underline{\Omega} + (\frac{N-1}{N^2})\,\underline{1}\,\underline{1}' & \underline{k}^{-\delta} + \underline{k}^{-\delta}\,\underline{n}'\,\underline{\Omega}\,\underline{n}\,k^{-\delta} \\ & - (\frac{2N-1}{N^2})\,\underline{1}\,\underline{1}' \end{bmatrix} \tag{26}$$

Finally, from (22) and (26), the variance-covariance matrix of θ^{*} is given as:

$$\left\{I - D(BD)^{-1}B\right\}(A'A + B'B)^{-1}$$

$$= \begin{bmatrix} \underline{\Omega} & -\underline{\Omega}\,\underline{n}\,\underline{k}^{-\delta} + \frac{1}{N}\underline{1}\,\underline{1}' \\ -\underline{k}^{-\delta}\underline{n}'\underline{\Omega} + \frac{1}{N}\underline{1}\,\underline{1}' & -\frac{1}{N}\underline{1}\,\underline{k}'\,\underline{n}\,\underline{k}^{-\delta} + \underline{k}^{-\delta}\underline{n}'\underline{\Omega}\,\underline{n}\,k^{-\delta} - \left(\frac{N+1}{N^2}\right)\underline{1}\,\underline{1}' \end{bmatrix} \qquad (27)$$

The above gives us the variance-covariance matrix for testing treatment contrasts. Since

$$\theta^* = (A'A + B'B)^{-1}A'Y \,,$$

and

$$\theta^* = \begin{bmatrix} \underline{t}^* \\ \underline{b}^* \end{bmatrix},$$

then,

$$t^* = \left[\underline{\Omega} + \frac{1}{N^2}\underline{1}\,\underline{1}' \;\vdots\; -\underline{\Omega}\,\underline{n}\,\underline{k}^{-\delta} + \frac{N-1}{N^2}\underline{1}\,\underline{1}'\right] \begin{bmatrix} \underline{T} \\ \underline{R} \end{bmatrix},$$

where T and $\underline{R}$ are vectors of treatment and block totals.

Thus

$$\underline{t}^* = \underline{\Omega}(T - \underline{n}\,k^{-\delta}\,R) + \frac{1}{N^2}\underline{1}\,\underline{1}'\,T$$

$$+ \left(\frac{N-1}{N^2}\right)\underline{1}\,\underline{1}\,R\,,$$

$$= \underline{\Omega}(T - \underline{n}\,\underline{k}^{-\delta}\,R) + \left(\frac{G}{N}\right)\underline{1} \qquad (28)$$

where G is grand total.

Let

$$Q = T - \underline{n}\,\underline{k}^{-\delta}\,R\,.$$

Then

$$\underline{t}^* = \underline{\Omega}\,Q + (G/N)\,\underline{1}\,. \qquad (28)'$$

Similarly,

$$\underline{b}^* = \underline{k}^{-\delta} (R - \underline{n}' \underline{t}^*) . \tag{29}$$

The $\underline{\Omega}$ defined in (22) has been shown by Tocher (1952) to be the variance-covariance matrix for $\underline{t}^*$.

Error sum of squares

$$= Y'Y - Y'\underline{A\theta}^* = Y'Y - (Q't^* + R'k^{-\delta} R) . \tag{30}$$

If the main concern is to test differences between treatment parameter estimates, use the variance-covariance matrix for t^* which is given by

$$V(t^*) = \sigma^2 \underline{\Omega} .$$

If

$$\underline{\Omega} = \begin{bmatrix} w_{11} & \cdots & w_{1t} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ w_{t1} & \cdots & w_{tt} \end{bmatrix},$$

then $V(T_i - t_j)$ is given by

$$\sigma^2 (w_{ii} - 2\, w_{ij} + w_{jj}) .$$

The standard error of estimate is:

$$\hat{\sigma}_{t^*i - t^*j} = \sqrt{(w_{ii} - 2\, w_{ij} + w_{jj})\, \hat{\sigma}_e^2} ,$$

where

$$\sigma_e^2 = \frac{Y'Y - (Q't^* + R'k^{-\delta} R)}{N - p + 1} .$$

The complete analysis of variance is given in Table 15.

Table 15. Analysis of Variance.

Sources	S.S.
Mean	G^2/N
Blocks	$R'k^{-\delta}R - G^2/N$
Treatments	$Q'\underline{t}^*$
Residual	$Y'Y - Q'\underline{t}^* - R'k^{-\delta}R$
Total	$Y'Y$

Application to Many Non-orthogonal Classifications

The two-way classification can be extended to the case of t-classifications. Basically the linear model is as given in (14). The parameter vector θ and the matrix A may now be partitioned to correspond with the t-classifications. Thus the model may now read

$$E(Y) = [\underline{1}\ \Delta_1 \Delta_2\ .\ .\ .\ \Delta_t]\ [u\ \theta_1'\ \theta_2'\ .\ .\ .\ \theta_t'] \tag{31}$$

where Δ_i is an n x l_i design matrix, and θ_i an l_i x 1 vector, and l_i is the number of levels of the i-th treatment (classification). μ is the general parameter which was considered null in the last section for the two-way classifications. $\underline{1}$ is the unit vector. Interaction is not considered. It may be assumed that m = t so that certain confounded factorial designs are excluded.

D' may then be defined by

$$D' = \begin{bmatrix} -1 & \underline{1}' & \underline{0}' & \cdots & \underline{0}' \\ -1 & \underline{0}' & \underline{1}' & \cdots & \underline{0}' \\ \vdots \\ -1 & \underline{0}' & \underline{0}' & \cdots & \underline{1}' \end{bmatrix}, \tag{32}$$

and B, by

$$B = \frac{1}{N} \begin{bmatrix} 0 & \underline{r}_1' & \underline{0}' & \underline{0}' & \cdots & \underline{0}' \\ 0 & \underline{0}' & \underline{r}_2' & \cdot & \cdots & \underline{0}' \\ 0 & \underline{0}' & \underline{0}' & \underline{r}_3' & \cdots & \underline{0}' \\ \vdots \\ 0 & \cdot & \cdot & \cdot & \cdots & \underline{r}_t' \end{bmatrix}, \tag{33}$$

where $\underline{r}_i$ is the vector of replications of the i-th classification.

Define $\underline{n}_{ij}$ $(= \Delta_i' \Delta_j)$ as the incidence matrice of the i-th classification with regard to the j-th. Then, $\underline{n}_{ij}$ is analogous to the $\underline{n}$ defined for the two-way classification.

Similarly $\underline{r}_i^{\delta} = (\Delta_i' \Delta_i)$ is $\underline{r}_i$ expressed as diagonal matrix. Let $\underline{r}_i^{-\delta}$ be its inverse, and N the total number of elements in the entire experiment.

From (32) and (33)

$$A'A + B'B = \begin{bmatrix} N & & \underline{r}_2' & \cdots & \underline{r}_t' \\ \underline{r}_1 & r_1^{\delta} + \underline{r}_1\, \underline{r}_1'/N^2 & \underline{n}_{12} & \cdots & n_{1t} \\ \vdots \\ \underline{r}_t & \underline{n}_{1t}' & \underline{n}_{2t}' & \cdots & \underline{r}_t^{\delta} + \underline{r}_t\, \underline{r}_t'/N^2 \end{bmatrix}, \tag{34}$$

and

$$A'Y = T = (G\ T_1' \ldots T_t'),$$

where T_i is the vector of totals for the i-th classification, and G is overall total. The most difficult aspect of the numerical procedure now is the calculation of $(A'A + B'B)^{-1}$. Two methods have been proposed by Bradau (1965) and Rees (1966) respectively. Both of these methods involves a systematic procedure for the inversion of the matrix $(A'A + B'B)^{-1}$.

With a large computer the inversion of this matrix is done without much difficulty. Once $(A'A + B'B)^{-1}$ is found we can find θ^* and then proceed with the analysis as was done for the two-way classification.

3. A GEOMETRICAL APPROACH TO THE ANALYSIS OF NON-ORTHOGONAL EXPERIMENTS

This method was proposed by Kuiper (1952). Basically it involves the method of vector spaces.

The Orthogonal Case

In order to appreciate the geometrical approach, one should first examine its application to the orthogonal experiment. A two-way table with n rows and m columns, of which the element X_{ij} $(i = 1, \ldots n;$ $j = 1, 2, \ldots m)$ represents the yield of treatment i in block j, may be represented as a vector in space E of nm dimension. The usual model for this is

$$X_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \ldots, \tag{35}$$

where the μ, τ_i, β_j and ϵ_{ij} are defined as in equation (1). The vector X can be resolved in the following way. Let

$$X = \mu + \tau + \beta + \epsilon \tag{35'}$$

where μ, τ and β are vectors in subspaces L, T, and B; respectively, and defined as follows:

(a) L is the 1-dimensional space of yield levels, which is spanned by a vector with number 1 in all cells of the table.

(b) The n-dimensional space T of treatment effects is spanned by a set of n vectors each of which contains 1 in all cells of one row and 0 in all other cells.

(c) The m-dimensional space B of block effects is spanned by a set of m vectors containing 1 in the cells of one column and zero in all other cells.

At the same time ϵ is a stochastic vector in E (nm dimension). It has components in L, T, and B as well as in R, where R is the (n-1) (m-1) dimensional space of random effects which is orthogonal to the spaces L, T, and B.

$$E(X) = \mu + \tau + \beta \quad .$$

When the experiment is completed the vector X is obtained. The maximum likelihood estimate (which is the same as a least squares estimate under normal assumption), of the expectation vector $E(X) = \mu + \tau + \beta$ is obtained by an orthogonal projection of X on the subspaces spanned by L, T, and B. This leads to finding the orthogonal projections of X into each subspace. If $\underline{1} = (1, \ldots 1)$ is chosen as a base in L and the projection of X into L is denoted by X_L, then we have $X_L = \lambda(1, \ldots 1) = \lambda\underline{1}$ (where λ is a scalar). Since $X - \lambda\underline{1}$ is perpendicular to 1, then $(X - \lambda\underline{1}) \cdot (\underline{1}') = X1' - \lambda\underline{1}\ \underline{1}' = 0$ (inner product). This implies that $\lambda = \frac{X1'}{11'}$; so that the component for X in L, $X_L = \frac{X1'}{11'} (1, \ldots 1)$ with the vector of means (general mean $\bar{X}$) common to all cells in the table.

In a similar way the arbitrary vector in the n-dimensional space T can be denoted by $\underline{t}$, where $\underline{t} = (t_1, \ldots t_n)$. Here, t_i is the number that occurs in each cell of the row i. The orthogonal projection of X into $T(X_T)$ is characterized by the numbers $t_i = \bar{X}_i$, the mean of the elements of X in row i.

By definition L is a subspace of B as well as of T. If therefore T^* is defined as a subspace of T orthogonal to L, and B^* analogically, then it is found that $X_T = X_L + X_{T^*}$ and $X_B = X_L + X_{B^*}$. From these two relations it is seen that

$$X_{T^*} = X_T - X_L$$

and (35)

$$X_{B^*} = X_B - X_L .$$

Then T^* and B^* are called pure treatment and pure block effects respectively. Also one can speak of the classification being in rows according to treatments and in columns according to blocks. If each class of T (rows) is proportionally represented in each class of B (columns) which is the case when orthogonality holds for the experiment, then $(X_{T^*}, X_{B^*}) = 0$ (inner product), the elements of rows of X_{B^*} add to zero. It therefore follows that T^* and B^* are perpendicular to each other. On subtracting the components of L, T^* and B^* from X, a vector in the space of random effects R is obtained. The equation

$$X = X_L + X_{T^*} + X_{B^*} + X_R \tag{36}$$

expresses the fact that X may be resolved into mutually orthogonal components. This is analogous to the X_{ij} expressed as

$$X_{ij} = \bar{X} + (\bar{X}_{i}. - \bar{X}) + (\bar{X}._{j} - \bar{X}) + (X_{ij} - \bar{X}_{i}. - \bar{X}._{j} + \bar{X})$$

The following illustrates how the relation in (36) can be expressed.

Let

$$\underset{X = X_E}{\begin{bmatrix} 9 & 3 \\ 1 & 3 \\ 11 & 3 \end{bmatrix}} = \underset{X_L}{\begin{bmatrix} 5 & 5 \\ 5 & 5 \\ 5 & 5 \end{bmatrix}} + \underset{X_{B^*}}{\begin{bmatrix} 2 & -2 \\ 2 & -2 \\ 2 & -2 \end{bmatrix}} + \underset{X_{T^*}}{\begin{bmatrix} 1 & 1 \\ -3 & -3 \\ 2 & 2 \end{bmatrix}} + \underset{X_R}{\begin{bmatrix} 1 & 1 \\ -3 & +3 \\ 2 & -2 \end{bmatrix}},$$

Dimension (= D.F.)	6	1	1	2	2

where X is the vector of data from 3 x 2 table.

Further, because of orthogonality:

$$X^2 = X_L^2 + X_{B^*}^2 + X_{T^*}^2 + X_R^2 .$$

Each term represents the square of the length of the component and equals the sum of squares in the analysis of variance.

$X_{T^*}^2/d_T$ (where d_T is the dimension of the space concerned) is the treatment mean square. On dividing it by the error mean square X_R^2/d_R the resulting variate, will under the null hypothesis, have the F distribution.

The Non-orthogonal Case

Suppose the data presented in the table below is obtained from a randomized block experiment in which several plots are missing. The dashes indicate the missing plots.

Table 16. 3 x 5 Randomized Block Experiment.

B \ T	t_1	t_2	t_3	t_4	t_5
b_1	X_{11}	X_{21}	X_{31}	---	X_{51}
b_2	---	X_{22}	---	X_{42}	X_{52}
b_3	X_{13}	---	X_{33}	---	X_{53}

Table 16 is an element (point) in 10-dimensional space. The aim is now to express X as a sum of components in spaces L(1), $T^*(4)$, $B^*(2)$ and R(3), where the number in brackets gives the dimension (=DF). R is chosen orthogonal to the joint space generated by L, T^* and B^* to obtain least squares estimates of the component τ and β in the spaces T^* and B^* of pure treatment and pure block effects respectively. The spaces T^* and B^* are, however, not orthogonal because each class of T is not proportionally represented in each class of B. Thus these spaces can no longer be obtained as was done for the orthogonal design.

Kuiper has suggested a procedure by which approximate estimates of the components in these spaces could be obtained. He suggested parallel projections (skew components) into T^* and B^*. His procedure for obtaining the parallel projections is by an iterative process which is outlined below.

(1) Project X orthogonally into T (it is convenient to denote X by r_o as the first one in a series of vectors) giving r_{OT} (= X_T). Obtain $r_1 = r_0 - r_{0T}$ as a vector orthogonal to T.

(2) Project r_1 into B giving r_{1B} and obtain $r_2 = r_1 - r_{1B}$.

(3) Project r_2 into T, giving r_{2T} and obtain $r_3 = r_2 - r_{2T}$.

(4) Project r_3 into B giving r_{3B} and obtain $r_4 = r_3 - r_{3B}$.

This process is continued until r_{1T} or $r_{(i+1)B}$ becomes negligible. In that case r_i or r_{i+1} is practically orthogonal to T and B. It will be seen more clearly from the numerical example that will later be presented that the series of the length r_0, r_1, r_2, . . . is now increasing. L. C. A. Corseten (1958) has shown that r_i converges to the vector X_R. The diagram of Fig. 1 helps to illustrate the procedure.

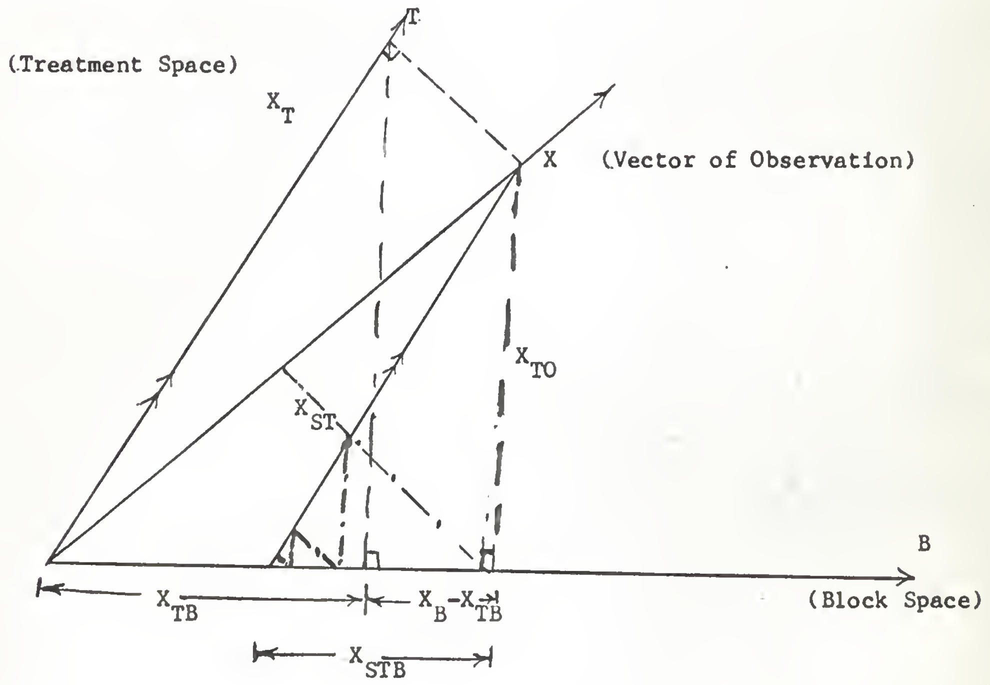

Fig. 1. Geometrical representation of the analysis of variance of two-way classification.

In Fig. 1, X is a vector decomposed into two non-orthogonal spaces T and B, the space T being orthogonal to that of the general mean. There will also be random components in a space orthogonal to B and T combined, but the figure represents the projections in the space of B and T. The vector X is projected on the space T to form X_T and then again onto B to form X_{TB}. The vector X is also projected directly onto B to form X_B. All projections are for the moment assumed orthogonal. The vector $X_B - X_{TB}$ is found by subtraction. By successive projection of $X_B - X_{TB}$ onto the spaces T and B in turn as shown by the dotted broken lines, the vector X_{ST} is eventually found. This represents the skew projection of X on B.

The actual computation is simplified by the following relations:

$$r_{jT} = - r_{j-1 \cdot BT} , \qquad j \geq 2 ,$$

$$r_{jB} = - r_{j-1 \cdot TB} , \qquad j \geq 3 .$$

Writing

$$U_0 = r_{0T} = X_T ,$$

$$V_0 = r_{1B} = X_B - X_{TB} ,$$

then

$$\left.\begin{aligned} X_{ST} &= \sum_0 U_i \\ X_{SB^*} &= \sum_0 V_i \end{aligned}\right\} \quad \text{with} \quad \left\{\begin{aligned} V_i &= - U_{iB} \\ & \qquad i \geq 0 \\ U_i &= - V_{i-1 \cdot T} . \end{aligned}\right.$$

Analysis of Variance Test

One may wish to test the null hypothesis H_0: "difference of treatments does not affect yield."

$$E(X) = \mu + \tau + \beta$$

now reduces to

$$E(X) = \mu + \beta$$

under H_0. If the joint space of L, B^* and T^* is denoted by T_0 the subspace perpendicular to L and B^*, then the orthogonal decomposition of X gives:

$$X = X_L + X_{B*} + X_{To} + X_R$$

where X^2_{To}/d is an unbiased estimate of σ^2 only if the null hypothesis holds. Then

$$F = \frac{X^2_{To}/d_{To}}{X^2_R/dR} \qquad (37)$$

yields a test of treatment effects.

Using Fig. 1, X^2_{To} is computed as follows:

$$X^2_{To} = X^2_{ST} - X^2_{STB} \qquad (38)$$

By subtraction,

$$X^2 - X^2_L - X^2_{B*} - X^2_{To} = X^2_R . \qquad (40)$$

4. NUMERICAL EXAMPLES

(1) Direct Solution to Normal Equation and Analysis of Variance

The table of data below is taken from an example of Harvey (1960).

Table 17. Gains in Weight of Individual Barrows.

Ration No.	Pig No.	Sire No. 1	2	3	
	1	5	2	3	
	2	6	3	-	
1	3	-	5	-	
	4	-	6	-	
	5	-	7	-	
	Subtotals	11(2)	23(5)	3(1)	37(8)
	1	2	8	4	
	2	3	8	4	
2	3	-	9	6	
	4	-	-	6	
	5	-	-	7	
	Subtotals	5(2)	25(3)	27(5)	57(10)
	Totals	16(4)	48(8)	30(6)	94(18)
	Means	4	6	5	

(a) Two-way without interactions

If it is assumed that there was no interaction between the rations and the sires shown in Table 17, then the basic model is of the form

$$y_{ijk} = \mu + S_i + r_j + e_{ijk} ,$$

$$i = 1, 2, 3,$$

$$j = 1, 2,$$

$$k = 1, 2, \ldots n_{ij} ,$$

where

y_{ijk} = the gain of the k-th barrow on the j-th ration by the i-th sire,

μ = the overall mean with equal subclass frequencies,

S_i = the effect of the i-th sire,

r_j = the effect of the j-th ration,

e_{ijk} = random errors which may be assumed NID(0, σ_e^2).

Least squares equations

	$\hat{\mu}$	$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{r}_1$	$\hat{r}_2$	RHM
μ	18	4	8	6	8	10	94
S_1	4	4	0	0	2	2	16
S_2	8	0	8	0	5	3	48
S_3	6	0	0	6	1	5	30
r_1	8	2	5	1	8	0	37
r_2	10	2	3	5	0	10	57

In matrix form:

$$\begin{bmatrix} 18 & 4 & 8 & 6 & 8 & 10 \\ 4 & 4 & 0 & 0 & 2 & 2 \\ 8 & 0 & 8 & 0 & 5 & 3 \\ 6 & 0 & 0 & 6 & 1 & 5 \\ 8 & 2 & 5 & 1 & 8 & 0 \\ 10 & 2 & 3 & 5 & 0 & 10 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{S}_1 \\ \hat{S}_2 \\ \hat{S}_3 \\ \hat{r}_1 \\ \hat{r}_2 \end{bmatrix} = \begin{bmatrix} 94 \\ 16 \\ 48 \\ 30 \\ 37 \\ 57 \end{bmatrix} \qquad (40)$$

$$A'A \qquad \hat{\theta} = A'Y$$

It is now seen that the matrix A (rank A = rank A'A) is not of full rank, since the first column of A'A is a linear combination of the 2nd to 4th columns. It is thus necessary to impose the restrictions

$$\sum_{i=1}^{3} \hat{S}_i = \sum_{j=1}^{2} \hat{r}_j = 0 .$$

The reduced form of equation (4) is of the form

$$\begin{bmatrix} 18 & -2 & 2 & -2 \\ -2 & 10 & 6 & 4 \\ 2 & 6 & 14 & 6 \\ -2 & 4 & 6 & 18 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{S}_1 \\ \hat{S}_2 \\ \hat{r}_1 \end{bmatrix} = \begin{bmatrix} 94 \\ -14 \\ 18 \\ 20 \end{bmatrix}, \qquad (40)'$$

or $$C\,\theta^* = P'Y .$$

On inverting the matrix C, one obtains

$$C^{-1} = \begin{bmatrix} 0.061486 & .021848 & -.022160 & -.009363 \\ .021848 & .144819 & -.061174 & -.009363 \\ -.022160 & -.061174 & .112047 & -.026217 \\ .009363 & -.009363 & -.026217 & .067416 \end{bmatrix}. \qquad (41)$$

$$\therefore \quad \theta^* = C^{-1} P'Y ,$$

and

$$\begin{bmatrix} \hat{\mu} \\ \hat{S}_1 \\ \hat{S}_2 \\ \hat{r}_1 \end{bmatrix} = \begin{bmatrix} 4.8876 \\ -.8876 \\ 1.3146 \\ -.8090 \end{bmatrix}.$$

Then

$$\hat{S}_3 = -\sum_{i=1}^{2} \hat{S}_i = -(-.8876 + 1.3146)$$
$$= -.4270 ,$$

$$\hat{r}_2 = -(-.8090)$$
$$= .8090 .$$

The inverse elements for $\hat{S}_3$ and $\hat{r}_2$ are as follows:

$$C^{uS_3} = -[C^{uS_1} + C^{uS_2}] = -[.021848 - .022160] = .000312 ,$$

$$C^{S_1S_3} = -[C^{S_1S_1} + C^{S_1S_2}] = -[.144819 - .061174] = -.083645 ,$$

$$C^{S_2S_3} = -[C^{S_2S_2}] = -.112047 .$$

- - - - - - - - - - - - - - -

$$C^{r_2r_2} = -[C^{r_2r_1}] = -[-.067416] = .067416$$

The complete inverse is as shown:

$$\begin{array}{c} \\ \mu \\ S_1 \\ S_2 \\ S_3 \\ r_1 \\ r_2 \end{array}
\begin{array}{c}
\begin{array}{cccccc} \hat{\mu} & \hat{S}_1 & \hat{S}_2 & \hat{S}_3 & \hat{r}_1 & \hat{r}_2 \end{array} \\
\begin{bmatrix}
.061486 & .02188 & -.022160 & .000312 & .009363 & -.009363 \\
.021848 & .144819 & -.061174 & -.083645 & -.009363 & +.009363 \\
-.022160 & -.061174 & .112047 & -.050873 & -.026217 & +.026217 \\
.000312 & -.083645 & -.050873 & .134518 & .026217 & -.026217 \\
.009363 & -.009363 & -.026217 & .026217 & .067416 & -.067416 \\
-.009363 & .009363 & .026217 & -.02617 & -.067416 & .067416
\end{bmatrix}
\end{array}$$

The inverse elements for the column and row for S_3 are obtained by adding the inverse elements for the S_i $(i \neq 3)$ column and reversing the sign. Similarly the r_2 columns and rows are found.

Sum of squares for the analysis of variance.

Let

$$R(\mu, S_i, r_j) = (\hat{\mu}, \hat{S}_1\ \hat{S}_2\ \hat{r}_1)\begin{bmatrix}94\\-14\\18\\20\end{bmatrix} = 511.7036 .$$

The total uncorrected sum of squares is

$$Y'Y = 568.0000 .$$

$$\text{Error S. sqs.} = Y'Y - R(\mu, S_i, r_j)$$

$$= 568 - 511.7036$$

$$= 56.2964 .$$

$$\text{S. sqs. Ration (R)} = R(\mu, S_i, r_j) - \sum_i \frac{y_{i\cdot}^2}{n_{i\cdot}}$$

$$= 511.7036 - 502.00 = 9.7036 .$$

$$\text{S. sqs. Sires (S)} = R(\mu, S_i, r_j) - \sum_j \frac{Y_{\cdot j}^2}{n_{\cdot j}}$$

$$= 15.6786 .$$

Using the general method for finding the different sums of squares, one obtains the following:

For Ration:

$$Z_R^{-1} = [.067416]^{-1} = 14.83274 ,$$

$$\therefore\ \theta_R^{*1} Z_R^{-1} \theta_R^{*} = (-.8090)' (14,83274) (-.8090) = \underline{\underline{9.708}} .$$

For Sires:

$$Z_S^{-1} = \begin{bmatrix} .144819 & -.061174 \\ -.061174 & .112047 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 8.975050 & 4.900084 \\ 4.900084 & 11.600112 \end{bmatrix} ,$$

$$\therefore \theta_S^{*'} Z_S^{-1} \theta_S^{*} = [-.8876 \quad 1.3146] \, Z_S^{-1} \, \theta_S^{*} ,$$

$$= [-1.5246 \quad 10.9002] \times \begin{bmatrix} -.8876 \\ 1.3146 \end{bmatrix} ,$$

$$= \underline{\underline{15.683}} .$$

Except for the rounding errors, the two methods give the same results.

Analysis of Variance Table

Source	DF	SS	MS	F
Sires (S)	2	15.6786	7.8393	1.95 NS
Ration (R)	1	9.7036	9.7036	2.45 NS
Error	14	56.2964	4.0212	

For the purpose of illustration some of the standard errors and individual comparisons of the parameter estimates will be given.

Standard error of least square means

$$S_{\hat{\mu}} = .061486 \times (4.20212) = .50 ,$$

$$S_{\hat{\mu}+\hat{S}_1} = (.061486 + .144819) + 2\,(.021848)\,4.0212 = 1.00 ,$$

- - - - - - - - - - -

$$S_{\hat{\mu}+\hat{S}_3} = .061486 + .134518 + 2\,(.000312)\,(4.0212) = .89 .$$

- - - - - - - - - - -

Mean Separation with Duncan's Multiple Range Test (.05 level)

Comparisons	$\bar{y}_i - \bar{y}_j$	$\frac{2}{(C^{ii} + C^{jj} - 2C^{ij})}$	Product of Difference	$\hat{\sigma}_e\ t_p, n_2$
S_2 vs. S_3	1.7416	2.396	4.17	6.08
S_2 vs. S_1	2.2292	2.297	5.12	6.38
S_3 vs. S_1	.4606	2.116	.97	6.08

(b) Two-way with interactions

Suppose that interaction is considered, then the model is as follows:

$$y_{ijk} = \mu + S_i + r_j + (Sr)_{ij} + e_{ijk} ,$$

$$i = 1, 2, 3,$$

$$j = 1, 2,$$

$$k = 1, 2, \ldots n_{ij} ,$$

where μ, S_i and r_j are as defined before and $(Sr)_{ij}$ = the interaction effects.

The Normal equations:

	$\hat{\mu}$	$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{r}_1$	$\hat{r}_2$	$\widehat{(Sr)}_{11}$	$\widehat{(Sr)}_{12}$	$\widehat{(Sr)}_{21}$	$\widehat{(Sr)}_{22}$	$\widehat{(Sr)}_{31}$	$\widehat{(Sr)}_{32}$	RHM
μ	18	4	8	6	8	10	2	2	5	3	1	5	94
S_1	4	4	0	0	2	2	2	2	0	0	0	0	16
S_2	8	0	8	0	5	3	0	0	5	3	0	0	48
S_3	6	0	0	6	1	5	0	0	0	0	1	5	30
r_1	8	2	5	1	8	0	2	0	5	0	1	0	37
r_2	10	2	3	5	0	10	0	2	0	3	0	5	57
$(Sr)_{11}$	2	2	0	0	2	0	2	0	0	0	0	0	5
$(Sr)_{12}$	2	2	0	0	0	2	0	2	0	0	0	0	11
$(Sr)_{21}$	5	0	5	0	5	0	0	0	5	0	0	0	23
$(Sr)_{22}$	3	0	3	0	0	3	0	0	0	3	0	0	25
$(Sr)_{31}$	1	0	0	1	1	0	0	0	0	0	1	0	3
$(Sr)_{32}$	5	0	0	5	0	5	0	0	0	0	0	5	27

or, in matrix form

$$A'A\ \theta = A'Y\ .$$

Impose the restrictions:

$$\sum_{i=1}^{3} \hat{S}_i = 0 = \sum_{j=1}^{2} \hat{r}_j = \sum_{i=1}^{3} \widehat{(Sr)}_{ij} = \sum_{j=1}^{2} \widehat{(Sr)}_{ij} = 0\ .$$

These lead to the following reduced normal equations:

$$\begin{bmatrix} 18 & -2 & -2 & -2 & 4 & 6 \\ -2 & 10 & 6 & 4 & -4 & -4 \\ 2 & 6 & 14 & 6 & -4 & -2 \\ 4 & -4 & 6 & 18 & -2 & 2 \\ 6 & -4 & -2 & 2 & 6 & 14 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{S}_1 \\ \hat{S}_2 \\ \hat{r}_1 \\ (\hat{Sr})_{11} \\ (\hat{Sr})_{21} \end{bmatrix} = \begin{bmatrix} 94 \\ -14 \\ 18 \\ -20 \\ 30 \\ 22 \end{bmatrix},$$

or, in general,

$$C\theta^* = P'Y .$$

The symmetric inverse C^{-1} (excluding the elements to the left of the diagonal) is given as:

$$\begin{bmatrix} .075926 & .007407 & -.031481 & .018519 & -.018519 & -.029630 \\ & .159259 & -.051852 & -.018519 & +.018519 & .029630 \\ & & .120370 & -.029630 & .029630 & .007407 \\ & & & .075926 & .007407 & -.031481 \\ & & & & .159259 & -.051852 \\ & & & & & .120370 \end{bmatrix} .$$

or, in general,

$$\theta^* = C^{-1} P'Y .$$

Then

$$\begin{bmatrix} \hat{\mu} \\ \hat{S}_1 \\ \hat{S}_2 \\ \hat{r}_1 \\ (\hat{Sr})_{11} \\ (\hat{Sr})_{21} \end{bmatrix} = \begin{bmatrix} 4.8889 \\ -.8889 \\ 1.5778 \\ -.5222 \\ 2.0222 \\ -1.3444 \end{bmatrix},$$

and

$$\hat{S}_3 = - \sum_{i=1}^{2} \hat{S}_i = - (.8889 + 1.5778) = -.6889$$

$$\hat{r}_2 = - (-.5222) = .5222 ,$$

$$(\hat{Sr})_{12} = -2.0222 ,$$

$$(\hat{Sr})_{22} = - (-1.3444) = 1.3444 ,$$

$$(\hat{Sr})_{31} = - (2.0222 - 1.3444) = - .6778 ,$$

$$(\hat{Sr})_{32} = - (-2.0222 + 1.3444) = .6778 .$$

<u>Analysis of variance and sums of squares</u>

Now

$$\text{Error S. sqs.} = \sum_{i=1}^{3} \sum_{j=1}^{2} \sum_{k=1}^{n_{ij}} y_{ijk}^2 - R\,(\mu,\ S_i,\ r_j\ (Sr)_{ij}) = 26.0652 .$$

When constants are fitted for all degrees of freedom among the subclasses the error sum of square may be computed from

$$\sum_i \sum_j \sum_k y_{ijk}^2 - \sum_i \sum_j \frac{(\sum_k y_{ijk})^2}{n_{ij}} = 568 - 541.9333 = 26.0667 ,$$

which checks (within rounding error) with the error sum of squares when the general method is used.

$$\text{S. sqs. for Sires (S)} = \theta_S^{*\prime} \, Z_S^{-1} \, \theta_S^{*} ,$$

$$= \begin{bmatrix} -.8889 & 1.5778 \end{bmatrix} \begin{bmatrix} .159259 & -.051852 \\ -.051852 & .120370 \end{bmatrix}^{-1} \theta_S^{*} ,$$

$$= \underline{\underline{21.0015}} .$$

$$\text{S. sqs. for Rations (R)} = \theta_R^{*1} \, Z_R^{-1} \, \theta_R^{*} ,$$

$$= \frac{(-.5222)^2}{.075926} = \underline{\underline{3.5916}} .$$

S. sqs. for Sires x Ration (SR)

$$= \theta_{SR}^{*1} \, Z_{SR}^{-1} \, \theta_{SR}^{*} ,$$

$$= \begin{bmatrix} 2.0222 & -1.3444 \end{bmatrix} \begin{bmatrix} .159259 & -.051852 \\ -.051852 & .120370 \end{bmatrix}^{-1} \begin{bmatrix} 2.0222 \\ -1.3444 \end{bmatrix} ,$$

$$= 30.2245 .$$

This may be found by using the formula

$$R\,(\mu, S_i, r_j, (Sr)_{ij}) - R\,(\mu, S_i, r_j) = 541.9348 - 511.7036$$

$$= 30.2312 .$$

This checks (except for rounding error) with that obtained using the general procedure.

ANOVA Table

Sources	DF	SS	MS	F
S	2	21.0015	10.5008	4.83*
R	1	3.5916	3.5916	1.65 NS
SR	2	30.2245	15.1122	6.96*
Error	12	26.0652	2.1721	

*Indicates significance at 5% level.

In order to find the standard errors and test differences between sire effects, the inverse elements for the parameters deleted before finding the inverse of the reduced variance-covariance matrix, are obtained according to the procedures already given in the last example.

The least-squares means, which may be of interest in an analysis such as this are given below:

$\hat{\mu} + \hat{s}_1$ = 4.0 Sire no. 1 mean,

$\hat{\mu} + \hat{s}_2$ = 6.5 Sire no. 2 mean,

$\hat{\mu} + \hat{s}_3$ = 4.2 Sire no. 3 mean,

$\hat{\mu} + \hat{r}_1$ = 4.4 Ration no. 1 mean,

$\hat{\mu} + \hat{r}_2$ = 5.4 Ration no. 2 mean,

$\hat{\mu} + \hat{s}_1 + \hat{r}_1 + (\widehat{sr})_{11}$ = 5.5 Sire 1 x Ration 1 subclass mean,

$\hat{\mu} + \hat{s}_1 + \hat{r}_2 + (\widehat{sr})_{12}$ = 2.5 Sire 1 x Ration 2 subclass mean,

$\hat{\mu} + \hat{s}_2 + \hat{r}_1 + (\widehat{sr})_{21}$ = 4.6 Sire 2 x Ration 1 subclass mean,

$\hat{\mu} + \hat{s}_2 + \hat{r}_2 + (\widehat{sr})_{22}$ = 8.3 Sire 2 x Ration 2 subclass mean,

$\hat{\mu} + \hat{s}_3 + \hat{r}_1 + (\widehat{sr})_{31}$ = 3.0 Sire 3 x Ration 1 subclass mean,

$\hat{\mu} + \hat{s}_3 + \hat{r}_2 + (\widehat{sr})_{32}$ = 5.4 Sire 3 x Ration 2 subclass mean.

The standard errors for these means may be computed using the same procedure as the last example. All that is needed is the complete inverse of the variance-covariance matrix and $\hat{\sigma}_e^2$ which is 2.1721.

Example:

$$S_{\hat{\mu}} = (.075926)(2.1721) \quad ,$$

$$= 0.44 \ .$$

$$S_{\hat{\mu} + \hat{s}_2} = [(0.75926) + (.120370) + 2(-.03148)](2.1721) \quad ,$$

$$= 0.53 \ .$$

$$S_{\hat{\mu} + \hat{s}_1 + \hat{r}_1 + (\hat{Sr})_{11}}$$

$$= [.075926 + .159259 + .075926 + .159259 + 2(.007407) + 2(.018519) + 2(-.018519) + 2(-.018519) + 2(.018519)]^{1/2} \times (2.1721)^{1/2} \quad ,$$

$$= 0.69 \ .$$

The procedure for pairwise testing of significance among sire means or among subclass means is quite similar to that already given for the two-ways without interaction.

The examples given above can easily be extended to three-way and with the aid of digital computer the procedures are quite straightforward.

(2) The General Analytic Method Applied to Two-way Without Interaction

Using the data in Table 17, we have

$$A'A = \begin{bmatrix} 18 & 4 & 8 & 6 & 8 & 10 \\ 4 & 4 & 0 & 0 & 2 & 2 \\ 8 & 0 & 8 & 0 & 5 & 3 \\ 6 & 0 & 0 & 6 & 1 & 5 \\ 8 & 2 & 5 & 1 & 8 & 0 \\ 10 & 2 & 3 & 5 & 0 & 10 \end{bmatrix} .$$

With

$$N = 18 ,$$

$$\underline{r}_1 = \begin{bmatrix} 4 \\ 8 \\ 6 \end{bmatrix} , \qquad \underline{r}_2 = \begin{bmatrix} 8 \\ 10 \end{bmatrix} ,$$

$$\underline{r}_1^{\delta} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 6 \end{bmatrix} , \qquad \underline{r}_2^{\delta} = \begin{bmatrix} 8 & 0 \\ 0 & 10 \end{bmatrix} ,$$

$$\underline{n}_{12} = \begin{bmatrix} 2 & 2 \\ 5 & 3 \\ 1 & 5 \end{bmatrix} ,$$

$$B = (\frac{1}{18}) \begin{bmatrix} 0 & 4 & 8 & 6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 10 \end{bmatrix} ,$$

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

$$BD = \frac{1}{18} \begin{bmatrix} 18 & 0 \\ 0 & 18 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \neq \underline{0} ,$$

$$B'B = \frac{1}{18^2}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 16 & 32 & 24 & 0 & 0 \\ 0 & 32 & 64 & 48 & 0 & 0 \\ 0 & 24 & 48 & 36 & 0 & 0 \\ 0 & 0 & 0 & 0 & 64 & 80 \\ 0 & 0 & 0 & 0 & 80 & 100 \end{bmatrix},$$

$$A'A + B'B = \begin{bmatrix} 18 & 4 & 8 & 6 & 8 & 10 \\ 4 & \frac{1312}{324} & \frac{32}{324} & \frac{24}{324} & 2 & 2 \\ 8 & \frac{32}{324} & \frac{2688}{324} & \frac{48}{324} & 5 & 3 \\ 6 & \frac{24}{324} & \frac{48}{324} & \frac{1980}{324} & 1 & 5 \\ 8 & 2 & 5 & 1 & \frac{2688}{324} & \frac{80}{324} \\ 10 & 2 & 3 & 5 & \frac{80}{324} & \frac{3340}{324} \end{bmatrix}.$$

We proceed to obtain $(A'A + B'B)^{-1}$. Let

$$A'Y = \begin{bmatrix} 94 \\ 16 \\ 48 \\ 30 \\ 37 \\ 57 \end{bmatrix}.$$

Then

$$\theta^* = (A'A + B'B)^{-1} A'Y .$$

The procedure for the analysis of variance is continued as that in the last section.

(3) The Geometrical Approach (Iterative Procedure)

The data below shows an actual example of what may happen in experiments on tree crops (cocoa in this example) after many years. This experiment was carried on for 16 years to see the effects of certain treatments on yield of cocoa. The experiment was a randomized block experiment but quite a number of the trees died off before the end of the experiment. The data gives the annual mean yield in pounds per tree.

Since the only test required here is on the treatments the geometrical method looks quite reasonable. This example will help to illustrate the usefulness of this approach.

Table 18. Annual Mean Yield of Cocoa per Tree.

	Blocks								
Treatments	1	2	3	4	5	6	7	8	Totals
1	--	--	29	39	31	34	--	27	160 (5)
2	--	--	27	34	30	--	31	29	151 (5)
3	33	30	--	34	30	31	29	28	215 (7)
4	33	31	26	35	26	27	31	--	209 (7)
5	33	--	26	35	27	31	31	27	210 (7)
6	34	32	28	36	27	32	31	30	250 (8)
7	30	--	--	31	26	27	26	--	140 (5)
8	35	33	31	38	29	31	33	29	259 (8)
9	34	33	27	--	--	--	--	--	94 (3)
10	38	34	--	--	--	34	35	30	171 (5)
11	--	34	30	--	--	--	--	--	64 (2)
Totals	270(8)	227(7)	224(8)	282(8)	226(8)	247(8)	247(8)	200(7)	1923(62)

Computational Procedure

	$X_{OT}=U_0$	U_1	U_0-U_1	U_2	$U_0-U_1-U_2$	U_3	X_{ST} $U_0-U_1-U_2-U_3$
1 (5)	32.00	-.73	32.73	-.03	32.76	.02	32.74
2 (5)	30.20	-.68	30.88	-.03	30.91	.02	30.89
3 (7)	30.71	+.40	30.31	.07	30.24	.03	30.21
4 (7)	29.86	+.42	29.44	.05	29.39	.03	29.36
5 (7)	30.00	-.11	30.11	.04	30.07	.03	30.04
6 (8)	31.25	-.01	31.26	.03	31.23	.02	31.21
7 (5)	28.00	1.01	26.99	.13	26.86	.04	26.82
8 (8)	32.38	-.01	32.39	.03	32.36	.02	32.34
9 (3)	31.33	.22	31.11	.01	31.10	.02	31.08
10 (5)	34.20	.09	34.11	.09	34.02	.03	33.99
11 (2)	32.00	-1.06	33.06	-.10	33.16	.01	33.15

	1 (8)	2 (7)	3 (8)	4 (8)	5 (8)	6 (8)	7 (8)	8 (7)
X_B	33.75	32.43	28.25	35.25	28.25	30.88	30.88	28.57
X_{TB}	30.97	31.68	31.13	30.55	30.55	31.05	30.82	31.53
$T=V_0(=X_B-X_{TB})$	2.78	0.75	-2.88	4.70	-2.30	-.17	.06	-2.98
U_{OT}	30.71	31.66	31.37	30.51	30.51	30.92	30.69	31.63
$r_3=V_1(=X_{TB}-U_{OT})$	.26	.02	-.24	.04	.04	.13	.13	-.10
r_{3B}	30.66	31.64	31.37	30.48	30.48	30.87	30.64	31.66
$r_5=V_2$	.05	.02	.00	.03	.03	.05	.05	-.03
r_{5B}	30.63	31.62	31.37	30.45	30.45	30.84	30.66	31.63
$r_7=V_3$	.03	.02	.00	.03	.03	.03	.03	.03
$\sum_{i=0} V_i$	3.12	.81	-3.12	4.80	-2.20	.04	.24	-3.08

Under H_0: $(\tau = 0)$

$$X^2_{ST} = 10662.62 ,$$

$$X^2_{STB} = 7769.94 ,$$

$$X^2_{TO} = X^2_{ST} - X^2_{STB} ,$$

$$= 2892.68 .$$

Error S.S.

$$X^2_R = X^2 - X_L - X^2_{B*} - X^2_{To}$$

$$\equiv X^2 - X^2_{ST} - X^2_{SB*}$$

$$= 60291 - 10662.62 - 57.55$$

$$= 49570.83 .$$

ANOVA (Under H_0)

Sources	DF	S.S.	M.S.	F
Treatments	10	2892.68	289.27	< 1
Error	44	49570.83	1126.60	

In the calculation of U's and V's we note that U_3 and V_3 become negligible. In order to obtain U_1 for example, consider the first treatment and add all the $X_B - X_{TB}$ for which no observation is missing in the first treatment and then divide by 5. Continue in this way until the last treatment is reached. This is what is meant by projecting $X_B - X_{TB}$ onto T. Similar operations go on for the projections onto B.

5. DISCUSSION AND CONCLUSION

Of the three methods presented for handling non-orthogonal data, the first is apparently the most straightforward to use. It can easily be programmed for the computer once the reduced normal equations are found. When the restrictions are imposed the size of the variance-covariance matrix is greatly reduced. The method of finding the sums of squares for the different treatment components is quite easy and interesting in that when a null hypothesis is to be tested it does not require the solution of another set of normal equations under the null hypothesis. All that is needed is the formula $\theta_t^{*1} Z_t^{-1} \theta_t^{*}$ as the sum of squares for the t-th treatment. This greatly reduces the amount of work. Also the method is very valuable in handling any type of non-orthogonal design. It can be used to handle treatments with interactions and it can be extended to handling all types of nested and multiple classifications. The normal equations or the variance-covariance matrices are very easy to write down as long as the appropriate mathematical model is used.

The general analytic method provides a theoretical basis and an understanding for all the least squares procedures we use in handling the non-orthogonal data.

The restriction $B\theta^* = \underline{0}$ is quite essential in order to obtain the best estimate. This is analogous to the restrictions imposed in the first method in order to obtain a unique solution to the normal equations. The general method can be used in handling designs with many non-orthogonal classifications but it is best used for handling balanced non-orthogonal designs. When interaction is considered the first method is definitely preferred. The restriction $B\theta^* = \underline{0}$ does not decrease the size of the variance-covariance matrix. The matrix (A'A + B'B), as shown in the example, is of full rank. For example, when the restrictions were imposed using the first method, the rank of A'A was reduced to 4 but when $B\theta^{*} = 0$ is used in the general method the rank is 6. One great advantage of the general method is when it is applied to randomized block designs and one is only interested in the standard errors of the treatment effects then the most fundamental part of the calculation is that of $\underline{\Omega}$ which is already defined with a simple relationship. Tocher (1952) has shown that $\underline{\Omega}$ is the variance-covariance matrix of t^{*}. The method of computer programming to obtain $\underline{\Omega}$ for two and three way classifications has been given by Freeman and Jeffers (1962).

The iterative (geometric) procedure is very useful for the case of randomized block experiments in which many plots are missing and a test must be made on the treatment effects. The computational procedure is quite easy and fairly fast because no higher terms of U_4 and V_4 are usually needed. This method is quite interesting in that it illustrates how the analysis of variance can be interpreted geometrically. The use of this method is however very limited. The only thing that recommends it is its simplicity. When components of variance are required the first method is definitely the best approach. Except for Henderson's first method, all the other methods of finding components of variance in non-orthogonal design is fairly involved.

ACKNOWLEDGMENTS

I seize this opportunity to express my gratitude to Dr. H. C. Fryer, the Head of Department of Statistics and Director of the Statistical Laboratory, who was my major academic adviser for the most part of my studies at Kansas State University. Also, I am very grateful to Dr. Koh of the Computer Science for his constant advice and guidance on the computer programming of unbalanced data. His guiding influence has enabled the production of this presentation.

Finally, I am grateful to Dr. Nassar who was, for the last semester of my studies at K-State, my major academic adviser.

REFERENCES

Bradau, D. (1966). "Main effect analysis of the general non-orthogonal layout with any number of factors". Ann. Math. Stat., 36:88-97.

Clarke, G. M. (1963). "A second set of treatments in a Yowder square design". Biometrics, 19:98-104.

Corsten, L. C. A. (1958). "Vectors, a tool in statistical regression theory". Meded Landb. Hoogesch, Wageningen, 58:1-92.

Freeman, G. H., and Jeffers, J. N. R. (1962). "Estimation of means and standard errors in the analysis of non-orthogonal experiments by electronic computers". J. R. Stat. Soc. B, 24:435-446.

Fryer, H. C. (1966). "Concepts and Methods of Experimental Statistics". Allyn and Bacon, Inc., Boston.

Graybill, F. A. (1960). "An introduction to Linear Statistical Models". Vol. 1. McGraw-Hill Book Company, Inc., New York.

Harvey, W. R. (1960). "Least squares analysis of data with unequal subclass numbers". U.S.D.A. publication (ARS).

Henderson, C. R. (1953). "Estimation of variance and covariance components". Biometrics, 9:226-252.

Justesen, S. H., and Keuls, M. (1958). "Notes on the use of non-orthogonal designs". Bull. Int. Stat. Inst., 36:269-276.

Kuiper, N. H. (1952). "Variante-Analyse". Statistica, 6:149-194.

Plackett, R. L. (1960). "Principles of Regression Analysis." The Clarendon Press, Oxford, England.

Rees, D. H. (1966). "Analysis of variance of designs with many non-orthogonal classifications". J. R. Stat. Soc. B, 110-117.

Snedecor, G. W. "Statical Methods". Iowa State College Press, Ames, Iowa, 1956.

Tocher, D. K. (1952). "The design and analysis of block experiments". J. R. Stat. Soc. B, 14:45-100.

ANALYSIS OF NON-ORTHOGONAL EXPERIMENTS

by

JOSEPH ARIMJA ELUKPO

B. Sc. (Hons), University of Ibadan, Nigeria, 1965

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1967

Standard methods for analyzing balanced non-orthogonal experiments have been given in several experimental design text books. In non-orthogonal designs where there is no balance whatsoever, there has been a great difficulty in presenting a general method of analysis.

The most appropriate method so far is the least squares procedure. This may be given in three parts. The first method involves writing down a set of normal equations which gives an insight into the way observed values are related to the parameters. In order to obtain unique solutions to the parameter estimates a set of restrictions are imposed on the parameters. This leads to a reduced variance-covariance matrix which can now be inverted in order to obtain the parameter estimates. The various treatment sums of squares for the analysis of variance test are quite easy to obtain by the general formula $\theta_t^{*'} Z_t^{-1} \theta_t^*$. θ_t^* is a column vector of the t-th treatment parameter estimates and Z_t is the square symmetrical segment of the inverse reduced variance-covariance matrix corresponding to the t-th treatment. This method can be applied to all types of unbalanced data and is most useful when interactions are of interest. The inverse of the variance-covariance matrix is used to obtain the standard errors and pairwise comparisons of the treatment effects.

The second method is based on Plackett's (1960) notation. It is given that $Y = A\theta + \epsilon$, where Y and ϵ are $n \times 1$ vectors, $E(\epsilon) = \underline{0}$, and A an $n \times p$ matrix. The matrix A is assumed of rank $p - m$ so that a matrix D exists, of rank m, such that $AD = \underline{0}$. Therefore an $m \times p$ matrix B must be defined, such that $|BD| \neq 0$, and the restriction $B\theta = \underline{0}$ imposed. Then the best estimate of θ is given by

$$\theta^* = (A'A + B'B)^{-1} A'Y ,$$

with variance-covariance matrix

$$\sigma^2 \left\{ I - D(BD)^{-1} B \right\} (A'A + B'B)^{-1} .$$

The greatest computational problem is to obtain $(A'A + B'B)^{-1}$. With a large digital computer this will no longer be a problem.

The third method is a geometrical iterative approach based on the concept of vector spaces which was first proposed by Kuiper (1952). It is very useful in analyzing a randomized block experiment in which several plots are missing and a test must be made on the treatment effects. The model in terms of vectors is of the form:

$$X = \mu + \tau + \beta + \epsilon \quad .$$

X is the vector of elements in the two-way table. Kuiper suggested a method whereby approximate estimates of the components of τ and β in T^* and B^* spaces respectively, could be found. This involves parallel projections of X into T^* and B^* spaces by iterative procedures to give:

$$X = X_L + X_{B^*} + X_{To} + X_R ,$$

where

X_L = the general mean space (1 dimensional) component,

X_{B^*} = the block space component,

X_{To} = the treatment space component (under the null hypothesis),

X_R = random effect space component.

An unbiased estimate of σ^2 is given by X_{To}^2/d_{To} while the variate $(X_{To}^2/d_{To})/(X_R^2/d_R)$ gives the F-test under the null hypothesis "that treatment effects are nill," where d_{To} and d_R are the degrees of freedom for treatments and error respectively.