AGRICULTURE AND DESIGN OF EXPERIMENTS

Chairman: PROFESSOR HAROLD HOTELLING

A NOTE ON THE METHOD OF 'FITTING OF CONSTANTS' FOR ANALYSIS OF NON-ORTHOGONAL DATA ARRANGED IN A DOUBLE CLASSIFICATION

By K. R. NAIR

Statistical Laboratory, Calcutta.

§ 1. INTRODUCTION

The analysis of non-orthogonal data in the general case of a multiple classification has been discussed by Yates, using two important methods, setting out the underlying hypothesis appropriate for adopting either method of analysis[1]. The methods are known as 'fitting of constants' and 'weighted squares of means'. When there is a single (one-fold) classification the two methods give identical results.

In this note we shall confine our attention to the application of the method of 'Fitting of Constants' to the analysis of non-orthogonal data arranged in double (two-fold) classification only. For, in this case, convenient expressions can be obtained for the sum of squares due to either classification, which is necessary for making valid tests of significance.

Data accruing from field experiments are in general arranged in multiple classification. The hypothesis adopted being usually that of additiveness in the effects due to the various factors classified, the analysis follows the method of fitting of constants. This method presents absolutely no difficulty when data are orthogonal with respect to blocks and treatments because in every block each treatment occurs once and only once and vice versa.

But sometimes accidental causes bring in non-orthogonality in the data of such experiments, by having to exclude certain plots of the original design. Thus there may be several missing plots. The analysis of such non-orthogonal data has been specially treated by Yates in a paper[2]. He has also tackled the interesting case of non-orthogonality in a Latin Square experiment when a row, column or treatment is missing[3]. Though the fundamental method is fitting of constants he has introduced a technique in the former problem by which the process of analysis is simplified considerably.

Recent researches in the domain of design of experiments have led to deliberate incorporation of non-orthogonality in the design itself, in order to control experimental error, when a large number of treatments has to be included in the lay-out. Thus besides the confounded designs of factorial experiments, we have the quasi-factorial and balanced incomplete designs [Yates,[4, 5]], the partially balanced incomplete designs

[Bose and Nair[6]], and the balanced incomplete designs with blocks of unequal size* [Kishen[7]]. These are all instances of non-orthogonality, in the original lay-out, of a special and systematic type. If these designs are afflicted by some missing plots, that will introduce further non-orthogonality but of an unsystematic type.

Owing to the increasing presence, in field experiments, of non-orthogonality, either accidental or deliberate or both, it is necessary to help the computer with some simple procedure of analysis of variance which is the major test of significance employed. For this purpose, a simple procedure has been developed in this paper for the case in which the experimental data fall in two classes, say, blocks and treatments. The case of Latin Square designs, for example, is not included.

§2. NON-ORTHOGONAL DATA IN DOUBLE CLASSIFICATION

Let the whole data of $n_{..}$ observations on a character y be capable of possessing two attributes A and B. Let $A_1, A_2, \ldots\ldots\ldots A_p$ be the p classes within the attribute A and $B_1, B_2, \ldots\ldots\ldots B_q$ the q classes within the attribute B. Each of the pq sub-classes is identified by the combination $A_i\, B_j (i=1, 2, \ldots\ldots\ldots p,\ j=1, 2, \ldots\ldots\ldots\ldots\ldots q)$ and will be denoted as sub-class (i, j).

Tables (1), (2) and (3) represent the sub-class numbers, the sub-class totals and the sub-class means respectively.

TABLE (1). SUB-CLASS NUMBERS

	A_1	...	A_i	...	A_p	Total
B_1	n_{11}	...	n_{i1}	...	n_{p1}	$n_{\cdot 1}$
⋮	⋮		⋮		⋮	⋮
B_j	n_{1j}	...	n_{ij}	...	n_{pj}	$n_{\cdot j}$
⋮	⋮		⋮		⋮	⋮
B_q	n_{1q}	...	n_{iq}	...	n_{pq}	$n_{\cdot q}$
Total	$n_{1\cdot}$	...	$n_{i\cdot}$	...	$n_{p\cdot}$	$n_{\cdot\cdot}$

TABLE (2). SUB-CLASS TOTALS

	A_1	...	A_i	...	A_p	Total
B_1	T_{11}	...	T_{i1}	...	T_{p1}	$T_{\cdot 1}$
⋮	⋮		⋮		⋮	⋮
B_j	T_{1j}	...	T_{ij}	...	T_{pj}	$T_{\cdot j}$
⋮	⋮		⋮		⋮	⋮
B_q	T_{1q}	...	T_{iq}	...	T_{pq}	$T_{\cdot q}$
Total	$T_{1\cdot}$	...	$T_{i\cdot}$	...	$T_{p\cdot}$	$T_{\cdot\cdot}$

* It is of considerable interest to note the existence of a striking similarity between the parametric relationships involved in the class of designs developed by Bose and Nair and those in the class of designs later developed by Kishen.

TABLE (3). SUB-CLASS AND MARGINAL MEANS

	A_1	...	A_i	...	A_p	Mean
B_1	$\bar{y}_{11}$	...	$\bar{y}_{i1}$	...	$\bar{y}_{p1}$	$\bar{y}_{\cdot 1}$
⋮	⋮		⋮		⋮	⋮
B_j	$\bar{y}_{1j}$	...	$\bar{y}_{ij}$	...	$\bar{y}_{pj}$	$\bar{y}_{\cdot j}$
⋮	⋮		⋮		⋮	⋮
B_q	$\bar{y}_{1q}$	...	$\bar{y}_{iq}$	...	$\bar{y}_{pq}$	$\bar{y}_{\cdot q}$
Mean	$\bar{y}_{1\cdot}$	...	$\bar{y}_{i\cdot}$	...	$\bar{y}_{p\cdot}$	$\bar{y}_{\cdot\cdot}$

Let n_{ij} represent the number of observations in the sub-class (i, j), $n_{i\cdot}$ and $n_{\cdot j}$ the number of observations in the marginal classes A_i and B_j respectively and $n_{\cdot\cdot}$ the total number of observations in the whole data. If y_{ijk} is the kth observation in sub-class (i, j), let

$$\left.\begin{aligned} T_{ij} &= \sum_{k=1}^{n_{ij}} y_{ijk}, \quad T_{i\cdot} = \sum_{j=1}^{q} T_{ij} \\ T_{\cdot j} &= \sum_{i=1}^{p} T_{ij}, \quad T_{\cdot\cdot} = \sum_{1}^{p} T_{i\cdot} = \sum_{1}^{q} T_{\cdot j} \end{aligned}\right\} \quad \ldots \quad (2\cdot10)$$

T_{ij}, $T_{i\cdot}$, $T_{\cdot j}$ and $T_{\cdot\cdot}$ are given in Table (2). Table (3) gives mean values of y defined by

$$\left.\begin{aligned} \bar{y}_{ij} &= T_{ij}/n_{ij}, \quad \bar{y}_{i\cdot} = T_{i\cdot}/n_{i\cdot} \\ \bar{y}_{\cdot j} &= T_{\cdot j}/n_{\cdot j}, \quad \bar{y}_{\cdot\cdot} = T_{\cdot\cdot}/n_{\cdot\cdot} \end{aligned}\right\} \quad \ldots \quad (2\cdot11)$$

2. Our first problem is to estimate the over-all effects α_i and β_j of the i^{th} class of attribute A and j^{th} class of attribute B on the assumption that the two attributes exert their influence independently of one another. That is to say, individuals belonging to the i^{th} class of A will receive a constant effect α_i whatever be the sub-class of B to which they belong and individuals belonging to the j^{th} class of B will receive a constant effect β_j whatever be the sub-class of A to which they belong. In other words the joint effect of A_i and B_j on the n_{ij} individuals of sub-class (i, j) is the sum of the two separate effects α_i and β_j.

If μ be the hypothetical mean of all possible observations in the whole domain,

$$y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk}$$

where e_{ijk}, the deviation due to uncontrolled factors, is supposed to be independently distributed about zero according to the Normal Law. The problem of getting efficient estimates of μ, α_i and β_j therefore reduces to minimising

$$\sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{n_{ij}} e^2_{ijk} = \Sigma\Sigma\Sigma(y_{ijk} - \mu - \alpha_i - \beta_j)^2 \quad \ldots \quad (2\cdot21)$$

with respect to the unknown parameters μ, α_i and β_j. Two linear restraints are brought in at this stage, namely,

$$\sum_{i=1}^{p} \alpha_i = 0 \quad \text{and} \quad \sum_{j=1}^{q} \beta_j = 0 \qquad \dots \quad (2{\cdot}22)$$

Using Lagrangian multipliers λ_1 and λ_2 we have to minimise without restriction the expression

$$\Sigma\,\Sigma\,\Sigma(y_{ijk} - \mu - \alpha_i - \beta_j)^2 + \lambda_1 \Sigma\, \alpha_i + \lambda_2 \Sigma\, \beta_j \qquad \dots \quad (2{\cdot}23)$$

with respect to μ, α_i, β_j, λ_1 and λ_2.

The normal equations involving the estimates m, a_i and b_j of μ, α_i and β_j respectively and λ_1 and λ_2 are easily obtained and are given below:

Leading term	Equations
m	$m\, n_{..} + \Sigma\, a_i\, n_{i.} + \Sigma\, b_j\, n_{.j} = T_{..}$
a_i	$m\, n_{i.} + a_i\, n_{i.} + \Sigma\, b_j\, n_{ij} + \lambda_1 = T_{i.}$ $(i = 1, 2 \dots\dots p)$
b_j	$m\, n_{.j} + \Sigma\, a_i\, n_{ij} + b_j\, n_{.j} + \lambda_2 = T_{.j}$ $(j = 1, 2 \dots\dots q)$
λ_1	$a_1 + a_2 + \dots\dots\dots\dots\dots\, a_p = 0$
λ_2	$b_1 + b_2 + \dots\dots\dots\dots\dots\, b_q = 0$

... (2·24)

We easily see that $\lambda_1 = \lambda_2 = 0$.

3. By the well known analogy between the method of fitting of constants and the problem of partial linear regression, the sum of squares due to the fitted constants m, a_i and b_j is

$$m\, T_{..} + \Sigma\, a_i\, T_{i.} + \Sigma\, b_j\, T_{.j} \qquad \dots \quad (2{\cdot}30)$$

with $p+q-1$ degrees of freedom.

On substituting for b's in terms of a's, this will reduce to

$$\sum_{i=1}^{p} a_i(T_{i.} - \sum_{j=1}^{q} n_{ij}\ \bar{y}_{.j}) + \sum_{j=1}^{q} T_{.j}\, \bar{y}_{.j} \qquad \dots \quad (2{\cdot}31)$$

On the other hand, on substituting for a's in term of b's, (2·30) will reduce to

$$\sum_{j=1} b_j(T_{.j} - \sum_{i=1}^{p} n_{ij}\, \bar{y}_{i.}) + \sum_{i=1} T_{i.}\, \bar{y}_{i.} \qquad \dots \quad (2{\cdot}32)$$

Owing to non-orthogonality in the data the solutions of a_i or b_j in equations (2·24) cannot be expressed independent of one another. For the same reason the sum of squares due to A effects only, or B effects only, does not separate out in (2·30).

The valid estimate of the sum of squares, due to A effects only, is to be obtained by the reduction in (2·30), by fitting constants for μ and β_j $(j = 1, 2\dots\dots\dots q)$ only, on the assumption that

$$y_{ijk} = \mu + \beta_j + e_{ijk} \ \text{and} \ \sum_{j=1}^{q} \beta_j = 0 \qquad \dots, \quad (2{\cdot}33)$$

The sum of squares due to the new fitted values m' and b'_j of μ and β_j with q degrees of freedom, is

$$\sum_{j=1} T_{\cdot j}\, \bar{y}_{\cdot j} \qquad \ldots \quad (2\cdot34)$$

Subtracting (2·34) from (2·31) the *valid* measure of the sum of squares due to A effects is

$$\sum_{i=1}^{p} a_i(T_{i\cdot} - \sum_{j=1}^{q} n_{ij}\, \bar{y}_{\cdot j}) \qquad \ldots \quad (2\cdot35)$$

with $p-1$ degrees of freedom.

Repeating similarly the fitting of constants for μ and α_i alone, *i.e.*, assuming $\beta_j = 0$ $(j=1, 2\ldots\ldots\ldots q)$ the sum of squares due to them, with p degrees of freedom, is

$$\Sigma\, T_{i\cdot}\, \bar{y}_{i\cdot} \qquad \ldots \quad (2\cdot36)$$

Subtracting (2·36) from (2·32) the *valid* measure of the sum of squares due to B effects is

$$\sum_{j=1}^{q} b_j(T_{\cdot j} - \sum_{i=1}^{p} n_{ij}\, \bar{y}_{i\cdot}) \qquad \ldots \quad (2\cdot37)$$

with $q-1$ degrees of freedom.

4. The simplification secured through the expressions (2·35) and (2·37) is that after solving equations (2·24) and getting estimates a_i and b_j of α_i and β_j, the appropriate sum of squares (*i*) due to A effects and (*ii*) due to B effects can be easily and directly evaluated.

For convenience let us represent

$$T_{i\cdot} - \sum_{j=1}^{q} n_{ij}\, \bar{y}_{\cdot j} \text{ and } T_{\cdot j} - \sum_{i=1}^{p} n_{ij}\, \bar{y}_{i\cdot}$$

by $Q_{i\cdot}$ and $Q_{\cdot j}$ respectively, so that the sum of squares due to A effects and B effects respectively will be $\Sigma a_i\, Q_{i\cdot}$ and $\Sigma b_j\, Q_{\cdot j}$. Since $\Sigma Q_{i\cdot} = 0$ and $\Sigma Q_{\cdot j} = 0$, a_i and b_j may be increased by the observed general mean without affecting the value of the sum of squares.

It is interesting to see that $a_1, a_2, \ldots\ldots a_p$ can be obtained from $Q_{1\cdot}, Q_{2\cdot}, \ldots\ldots Q_{p\cdot}$ by means of the equations

$$a_i\, n_{i\cdot} - \sum_{j=1}^{q} \frac{n_{ij}}{n_{\cdot j}} \sum_{i=1}^{p} a_i\, n_{ij} = Q_{i\cdot} \qquad (i=1, 2\ldots\ldots\ldots p) \qquad \ldots \quad (2\cdot40)$$

and that $b_1, b_2, \ldots\ldots\ldots b_q$ can be obtained from $Q_{\cdot 1}, Q_{\cdot 2}, \ldots\ldots\ldots\ldots Q_{\cdot q}$ by means of the equations

$$b_j\, n_{\cdot j} - \sum_{i=1}^{p} \frac{n_{ij}}{n_{i\cdot}} \sum_{j=1}^{q} b_j\, n_{ij} = Q_{\cdot j} \qquad (j=1, 2\ldots\ldots\ldots q) \qquad \ldots \quad (2\cdot41)$$

Moreover the set of equations involving $Q_{i\cdot}$ and the set involving $Q_{\cdot j}$ each possesses the properties of the normal equations of partial regression coefficients, namely, that (*i*) the variance of the righthand side of any equation is the coefficient of the leading unknown of that equation and (*ii*) the covariance between the right hand sides of the l^{th}

and m^{th} equations is the coefficient, in the l^{th} equation, of the leading unknown of the m^{th} equation or the coefficient, in the m^{th} equation, of the leading unknown of the l^{th} equation, when the variances and covariances are expressed in terms of the variance of y as unit.

Thus, for example,

$$V(Q_{l\cdot}) = \sum_{j=1}^{q} \left(n_{lj} - \frac{n^2_{lj}}{n_{\cdot j}} \right) = \text{Coefficient of } a_l \qquad \dots \quad (2\cdot42)$$

and

$$\text{Cov. } (Q_{l\cdot}\ Q_{m\cdot}) = -\sum_{j=1}^{q} \frac{n_{lj}\ \ n_{mj}}{n_{\cdot j}} = \text{Coefficient of } a_l \text{ in the } m^{th} \text{ equation}$$

$$= \text{Coefficient of } a_m \text{ in the } l^{th} \text{ equation} \qquad \dots \quad (2\cdot43)$$

These properties give us the explanation for the neat form in which the sum of squares due to the a's and due to the b's were obtained as $\Sigma a_i\, Q_{i\cdot}$ and $\Sigma b_j\ Q_{\cdot j}$ respectively.

It will be seen that

$$Q_{i\cdot} = \sum_{j=1}^{q} n_{ij}\,(\bar{y}_{ij} - \bar{y}_{\cdot j}) \text{ and } Q_{\cdot j} = \sum_{i=1}^{p} n_{ij}\,(\bar{y}_{ij} - \bar{y}_{i\cdot}) \qquad \dots \quad (2\cdot44)$$

They can therefore be easily calculated by means of two subsidiary tables prepared from Tables (1) and (3), one having elements $n_{ij}(\bar{y}_{ij} - \bar{y}_{\cdot j})$and the other having elements $n_{ij}(\bar{y}_{ij} - \bar{y}_{i\cdot})$. The totals of the columns of the first table and of the rows of the second table will give $Q_{i\cdot}$ and $Q_{\cdot j}$ respectively.

The meaning of $Q_{i\cdot}$ and $Q_{\cdot j}$ are also easy to understand. Thus $Q_{i\cdot}$ gives the total y belonging to class A_i when each y is corrected by the mean of all y's in the B-class to which it belongs. Similarly $Q_{\cdot j}$ gives the total y belonging to class B_j when each y is corrected by the mean of all y's in the A-class to which it belongs.

5. If more than one character is studied, let y and y' be any two of these characters. Let a'_i, b'_j, $Q'_{i\cdot}$, $Q'_{\cdot j}$ be the values for y' corresponding to a_i, b_j, $Q_{i\cdot}$, $Q_{\cdot j}$, for y. For analysis of covariance of y and y', the valid estimate of the sum of products due to A effects can be shown to be $\Sigma a_i\, Q'_{i\cdot} = \Sigma a'_i\, Q_{i\cdot}$ ... (2·50)

and that for B-effects as $\Sigma b_j\, Q'_{\cdot j} = \Sigma b'_j\, Q_{\cdot j}$ ... (2·51)

affording an independent check for each sum of products.

§3. APPLICATION TO FIELD EXPERIMENTS

1. The data from field experiments in randomized blocks are only special cases of the above two-way classification. When the experiment is orthogonal, n_{ij} remains constant for all values of i and j. When only a single factor is tested, as is the case in varietal trials, the constant value of n_{ij} is 1. But when such single factor experiments become non-orthogonal, n_{ij} no more remains constant but will have the value 1 or 0. There will thus be a number of empty cells.

Let us replace the A-factor classification by the treatments, say, varieties 1, 2, ...v. Let the B-factor be the blocks 1, 2,.......b. As we are in general interested mostly in the test of significance of varieties, the *valid* estimate of the sum of squares due to that alone need be given. But as is evident from the foregoing section, the sum of squares for blocks (eliminating varieties) can also directly be obtained by a process similar to that used below for finding the sum of squares for varieties (eliminating blocks).

Prepare a two-way table with blocks and varieties. Tables (1), (2) and (3) of §2 merge into this Table. There will be some unoccupied cells corresponding to $n_{ij}=0$ and some occupied cells with a single observation, corresponding to $n_{ij}=1$. If B_1, B_2, ... B_b are the block totals, the block means are $B_1/n_{.1}$, $B_2/n_{.2}$, ... $B_j/n_{.j}$, ... $B_b/n_{.b}$ Subtracting from each occupied cell belonging to block j the value $B_j/n_{.j}$ and adding up the corrected values, variety by variety, we get values of Q_i, corresponding to the i^{th} variety. If v_i be the effect of the i^{th} variety estimated by fitting constants for blocks and varieties, the appropriate estimate of the sum of squares due to varieties is

$$\Sigma v_i \, Q_i \qquad \ldots \quad (3{\cdot}01)$$

The expression for the sum of squares 'due to blocks', in order to obtain valid estimate of sum of squares due to error, is

$$\sum_{1}^{b} (B^2_j/n_{.j}) - G^2/n_{..} \qquad \ldots \quad (3{\cdot}11)$$

where G is the total of all plot values of the experiment and $n_{..}$ the total number of plots.

$$\text{The sum of squares due to error is} \quad \Sigma\,\Sigma y^2_{ij} - \sum_{1}^{v} v_i \, Q_i - \sum_{1}^{b} (B^2_j/n_{.j}) \qquad \ldots \quad (3{\cdot}12)$$

The calculation of v_i is not always easy. In every case of non-orthogonality v_i can be obtained as a linear expression of the Q's, as was shown in §(2·4).

2. In symmetrical types of non-orthogonality as in Yates' design, value of v_i is easily obtained as

$$v_i = \frac{k \, Q_i}{\lambda \, v} \qquad \ldots \quad (3{\cdot}20)$$

where k is the number of plots in each block and λ is the number of blocks in which every pair of varieties occurs together.

$$\text{The sum of squares due to varieties is therefore } \Sigma v_i \, Q_i = \frac{k}{\lambda v} \, \Sigma Q^2_i \qquad \ldots \quad (3{\cdot}21)$$

as has been given by Yates.

3. In the less symmetrical, but nevertheless systematic type of non-orthogonality of the partially balanced incomplete block designs developed by Bose and Nair (which includes Yates' two-dimensional Quasi-factorial design in square lattice as a special case) and of the balanced incomplete designs with blocks of unequal size, developed by Kishen, the expressions for v_i take more complicated forms.

4. When there are missing plots in an orthogonal randomized block experiment the values of v_i have to be estimated according to the peculiarities of each case, and will need the direct solving of normal equations. Great simplification has been effected by Yates in this problem, by a process in which values for missing plots are directly estimated. The data are thus apparently reconstructed back to 'orthogonality' and the usual varietal means will supply the values of v_i. By this technique therefore the method of fitting constants is short-circuited.

Yates' technique for dealing with missing plots can be applied to all types of designs including manifold classifications; and to get the appropriate (valid) estimate of the sum of squares for treatment effects, there was no simplification effected. He was however satisfied that the need for such precise estimate would seldom arise and that the (over-estimated) sum of squares due to the treatments obtained from the reconstructed data would be sufficient in most cases. Bartlett has suggested the method of covariance to get the precise estimate of sum of squares, which, though elegant, is tedious when there are a number of missing plots. The method of covariance has been extended by Nair[*] to the case of mixed-up plots, another interesting type of non-orthogonality.

The method given in this paper simplifies the problem of getting precise estimates of sum of squares due to varieties or blocks so long as the data are in double classification. Thus single factor experiments in randomized blocks can be tackled easily. But Latin Square lay-outs cannot be tackled with the formulae given here.

COMPARISON OF PAIRS OF TREATMENT EFFECTS

5. While analysis of variance furnishes the major test of significance of treatment effects, it is usual to carry out, when the major test gives significant results, detailed (minor) tests of significance by comparing any pair of treatment effects, say, v_i and v_j. This will require the knowledge of variance of each pair of treatment effects which is not same for all pairs, in general, in the case of non-orthogonal data. Though the method of finding these separate variances is well known, precise expressions can be easily obtained only when v_i itself can be easily estimated in terms of the Q's. Just to illustrate this point we shall consider the hitherto unsolved problem of missing plots in Yates' balanced incomplete block designs where the non-orthogonality of the original design is symmetrical, ensuring equal variance for comparisons of any two varietal effects.

A SINGLE MISSING PLOT IN BALANCED INCOMPLETE BLOCK DESIGN

6. Let us first consider the simplest case, namely, of a single missing plot. Let block 1 get varieties 1, 2, ... k. Let the plot given to variety 1, in this block be missing.

Let B'_1 be the sum of the $k-1$ plots of block 1 and B_2, B_3 ... . B_b the totals of blocks 2, 3, ... b. Let V'_1 be the sum of the $r-1$ plots of variety 1, and V_2, V_3, ... V_v sum for the r plots of varieties 2, 3, ... v. Denoting by Q''_1, Q'_2, ... Q'_k the Q's corresponding to variety $i(i \leq k)$, we have

$$\left.\begin{aligned}
k\,Q''_1 &= k\,V'_1 - \sum_{(1)}^{(r-1)} B \\
k\,Q'_2 &= k\,V_2 - \frac{k\,B'_1}{k-1} - \sum_{(2)}^{(r-1)} B \\
&\vdots \\
k\,Q'_k &= k\,V_k - \frac{k\,B'_1}{k-1} - \sum_{(k)}^{(r-1)} B \\
\text{and} \quad k\,Q_i &= k\,V_i - \sum_{(i)}^{(r)} B \quad (i \geq k+1)
\end{aligned}\right\} \quad \dots \quad (3{\cdot}60)$$

where $\sum_{(m)}^{(l)} B$ denotes the sum of the totals of the l blocks, each with k plots, in which the m^{th} variety has occurred.

If x be the estimate of the missing value (which can be obtained by either of the three methods: fitting of constants, Yates' method of minimising error sum of squares; or Bartlett's covariance method), it is given by

$$x = \frac{(v\lambda - k)B'_1 + k\left[(k-1)Q''_1 - Q'_2 - Q'_3 \,\ldots\ldots\ldots\, Q'_k\right]}{(k-1)\;(v\lambda - k)} \quad \dots \quad (3{\cdot}61)$$

After substituting this value of x and reconstructing the data, it can be easily seen, that the varietal effects v''_1, v'_2, ... v'_k; $v_i(i = k+1,\ k+2,\ \ldots\ldots\ v)$ are

$$\left.\begin{aligned}
v''_1 &= \frac{1}{r\lambda}\left[\,k(V'_1 + x) - (B'_1 + x) - \sum_{(1)}^{(r-1)} B\,\right] \\
v'_2 &= \frac{1}{r\lambda}\left[\,k\,V_2 - (B'_1 + x) - \sum_{(2)}^{(r-1)} B\,\right] \\
&\vdots \\
v'_k &= \frac{1}{r\lambda}\left[\,k\,V_k - (B'_1 + x) - \sum_{(k)}^{(r-1)} B\,\right] \\
\text{and} \quad v_i &= \frac{1}{r\lambda}\left[\,k\,V_i - \sum_{(i)}^{(r)} B\,\right] (i \geq k+1)
\end{aligned}\right\} \quad \dots \quad (3{\cdot}62)$$

The analysis of variance will give the following sum of squares for blocks and varieties respectively

$$B'^2_1/(k-1) + (B^2_2 + \,\ldots\ldots\ldots\, B^2_b)/k - G'^2/(N-1)$$

$$Q''_1 + v'_2 . Q'_2 + \,\ldots\ldots\ldots\ldots\, + v'_k\,Q'_k + k\sum_{k+1} Q_i^2/r\lambda$$

7. Coming to the minor tests of significance between pairs of varieties, it will be recognised that the total $v(v-1)/2$ comparisons fall in five groups each presumably with a different error variance. Thus for any two varieties i and j, we have

Group:	No. of Comparisons:	Variance $/\sigma^2$:
(i) $i=1;\ j=2 \ldots\ldots k$	$k-1$	$\dfrac{k[2v\lambda(k-1)-k^2+2k]}{v\lambda(v\lambda-k)(k-1)}$
(ii) $i=1;\ j=k+1, \ldots\ldots v$	$v-k$	$\dfrac{k(2v\lambda-k-1)}{v\lambda(v\lambda-k)}$
(iii) $i=2,\ldots k;\ j=2,\ldots\ldots k$	$(k-1)(k-2)/2$	$2k/v\lambda$
(iv) $i=2,\ldots k;\ j=k+1,\ldots v$	$(v-k)(k-1)$	$\dfrac{k[2(k-1)(v\lambda-k)+1]}{v\lambda(v\lambda-k)(k-1)}$
(v) $i=k+1,\ldots v;\ j=k+1 \ldots v$	$(v-k)(v-k-1)/2$	$2k/v\lambda$

The mean variance for the whole experiment is

$$V_m = \frac{2k\sigma^2}{v\lambda}\left[1+\frac{k}{(v-1)(v\lambda-k)}\right] \qquad \ldots (3\cdot70)$$

If there were no missing plot, every comparison would have had the same variance, namely, $2k\sigma^2/v\lambda$. Therefore loss of efficiency due to a single plot missing, is

$$\frac{1}{1+(v-1)\left(\dfrac{v\lambda}{k}-1\right)} \qquad \ldots (3\cdot71)$$

By putting $k=v$ and $\lambda=b$ in the above results we get the parallel solutions for an ordinary randomized block experiment with a single missing plot.

If more than one plot are missing, the number of groups of comparisons with different variances will increase; and the determination of these variances will be a laborious task. In such cases it is better to leave out the minor tests of significance and to confine attention to the major tests of significance available in the analysis of variance.

A Single Missing Block in Balanced Incomplete Block Design

8. If there are several missing plots in a balanced incomplete design we have of course no convenient means of writing down beforehand the expressions for the error variance of various comparisons. But there are interesting special cases such as what happens when all the plots of one of the blocks are missing.

In ordinary randomized block experiments the number of blocks used bears no mathematical relationship with the number of plots (or treatments) within the block. If one block of such an experiment is entirely lost or damaged the orthogonality of the design is not lost and hence there is no difficulty in the analysis of the data, although the efficiency of treatment comparisons will be lowered from the pre-arranged level through the loss of one replication.