ANALYSIS OF VARIANCE OF A BALANCED INCOMPLETE BLOCK DESIGN WITH MISSING OBSERVATIONS

by

Hugh Robert Baird

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

MASTER OF SCIENCE

in

Statistics

May 1960

Blacksburg, Virginia

TABLE OF CONTENTS

I. INTRODUCTION

The balanced incomplete block design has innumerable applications in all fields of research where each observation may be affected by two criteria. The first of these criteria is what will be called the block effect, arising unavoidably from the nature of the experimental material subjected to various treatments. Each block is a homogeneous milieu for the occurrence of several treatments.

The second criterion is what will be called the treatment effect, deliberately imposed by the experimenter and of especial interest to him.

The distinguishing feature of the balanced incomplete block design is the fact that each block is too small to permit replication of all the treatments under study. However, the designs are so developed that the analyst can obtain orthogonal sums of squares for block, treatment, and error effects, thus permitting a test of treatment effects only.

It may happen that some of the observations are missing. An agricultural plot may be trampled, an animal may die, a test tube may be dropped, or a machine may break down. It is inevitable that situations may arise where some observations are missing in the first place or must be rejected due to some unique effect beyond the control of the experimenter.

When observations are missing, the symmetry of the design is upset, and, consequently, the orthogonality of the sums of squares is destroyed.

The aim of all who have dealt with missing values has been the same: to restore, by estimating the missing values, the properties of symmetry in the design and orthogonality in the sums of squares.

Allen and Wishart (1930) first estimated single missing values in the randomized block and Latin square designs, using a purely statistical approach.

Yates (1933), following the suggestion of R.A. Fisher, developed an iterative procedure for the estimation of several missing values in the randomized block design, the estimates being those which minimized the error sum of squares. He proved that there is a positive bias in the treatment sum of squares obtained from data which has been augmented by estimates of missing values; he also showed how to eliminate this bias.

Cornish (1940) applied the procedures obtained by Yates to the balanced incomplete block design, obtaining a specific formula for a single missing value and an iterative procedure for several missing values. Like Yates, he showed the positive bias in the resulting treatment sum of squares and how to eliminate it.

Wilkinson (1958), following Yates' procedure of minimizing the error sum of squares, obtained a general procedure for estimating missing values in randomized blocks, balanced incomplete blocks, square lattices, simple cubic lattices, Latin squares, Youden squares, and lattice squares. The estimates depend upon the solutions of matrix equations.

In this paper, the established precedent will be followed in obtaining those missing value estimates which minimize the error sum of squares. A general matrix equation for estimating several missing values will be derived which should be easier to apply than the methods now available. Specific formulae for missing values under several particular conditions will also be obtained. In order to eliminate the bias in the treatment sum of squares, a direct method of analysis of the augmented data will be given. It will also be shown how a balanced incomplete block design can be analyzed as a randomized block design with missing values.

It is true that, by estimating missing values, we can restore the symmetry in the design and the orthogonality in the sums of squares. But the estimates of missing values serve only to expedite analysis of the data; they do not, by any means, restore the information lost in the missing observations.

II. THEORETICAL BACKGROUND

2.1 Balanced Incomplete Block Design

An incomplete block design is said to be balanced if it satisfies the following requirements.

1. The experimental material is divided into b blocks of k units each, different treatments being applied to the units in the same block.

2. There are t (> k) treatments, each of which occurs in r blocks.

3. Every treatment appears with every other treatment in the same block an equal number of times, say λ.

4. $\lambda = r(k-1)/(t-1)$

   $rt = kb = N$.

Table 2.1 gives the layout for a balanced incomplete block design where $t = 4$, $b = 4$, $r = 3$, $k = 3$, and $\lambda = 2$.

Table 2.1

Balanced Incomplete Block Design

y_{31}	y_{11}	y_{21}
y_{22}	y_{12}	y_{42}
y_{13}	y_{43}	y_{33}
y_{24}	y_{34}	y_{44}

The experimental model for this design is

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}, \qquad (2.1)$$

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

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

where μ is the overall mean, τ_i is the effect of the $i^{\underline{th}}$ treatment, β_j is the effect of the $j^{\underline{th}}$ block, and ϵ_{ij} is a random effect assumed to be normally distributed with mean zero and variance σ^2; the variates ϵ_{ij} are further taken to be stochastically independent. Restrictions on the parameters in (2.1) are

$$\sum_i \tau_i = \sum_j \beta_j = 0.$$

To test

$$H_o\colon \quad \tau_1 = \tau_2 = \ldots = \tau_t = 0, \qquad (2.2)$$

the following totals are first calculated:

$$\begin{aligned} T_i &= \sum_j y_{ij} \\ B_j &= \sum_i y_{ij} \\ G &= \sum_i T_i = \sum_j B_j \,. \end{aligned} \qquad (2.3)$$

Then

$$Q_i = kT_i - B_{\cdot i} \qquad (2.4)$$

is obtained, where $B_{\cdot i}$ is the sum of the B_j's for those blocks containing treatment i. Using (2.3) and (2.4), the following sums of squares are then calculated:

$$\text{Treatment SS (Adjusted)} = \frac{(t-1)}{rtk(k-1)} \sum_i Q_i^2 \equiv A$$

$$\text{Block SS (Unadjusted)} = \frac{1}{k} \sum_j B_j^2 - \frac{G^2}{N} \equiv B \qquad (2.5)$$

$$\text{Total SS} = \sum_{ij}\sum y_{ij}^2 - \frac{G^2}{N} \equiv T$$

$$\text{Error SS} = \text{Total SS} - \text{Treatment SS} - \text{Block SS} \equiv E.$$

These sums of squares are set up as in Table 2.2.

Table 2.2

Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Blocks (unadjusted)	(b-1)	B		
Treatments (adjusted)	(t-1)	A	s_A^2	s_A^2/s_E^2
Error	(N-t-b+1)	E	s_E^2	
Total	(N-1)	T		

The test of (2.2) is effected by comparing the calculated F with the tabular values of the F distribution with (t-1) and (N-t-b+1) degrees of freedom at a chosen significance level and drawing the appropriate conclusion.

2.2 Missing Observations

The standard least squares procedure used in this paper for obtaining estimates of missing values was originally suggested by R. A. Fisher, was first applied by Yates, and has been favored by most writers.

This procedure obtains values for m, t_i, and b_j (the estimates, respectively, of μ, τ_i, and β_j) such that

$$E' = \sum_{\substack{\text{known}\\ \text{data}}}\sum (y_{ij} - m - t_i - b_j)^2 \tag{2.6}$$

is a minimum. If the missing values are designated by x_{ij}'s, the criterion used to determine the values of the x_{ij}'s is that of minimizing the error sum of squares taken over both the observed and the estimated data. Letting E^* represent the error sum of squares obtained when the data are augmented by the values of the x_{ij}'s, and letting m^*, t_i^*, and b_j^* be the least squares estimates which minimize E^*, one gets

$$E^* = \sum_{\substack{\text{known}\\ \text{data}}}\sum (y_{ij}-m^*-t_i^*-b_j^*)^2 + \sum_{\substack{\text{missing}\\ \text{data}}}\sum (x_{ij}-m^*-t_i^*-b_j^*)^2 \tag{2.7}$$

Partially differentiating (2.7) with respect to x_{ij} and equating the resulting partial derivative with zero yields

$$x_{ij} = m^* + t_i^* + b_j^* \tag{2.8}$$

Substituting (2.8) into (2.7), then comparing the results with (2.6) shows that

$$m^* = m,\ t_i^* = t_i,\ b_j^* = b_j \tag{2.9}$$

and that

$$E^* = E'. \tag{2.10}$$

Since E^* was shown to be equal to the error sum of squares obtained by the correct least squares procedure, the error sum of squares can be obtained from the augmented data, but its degrees of freedom must be reduced by one for each missing value estimated.

2.3 Bias in the Treatment Sum of Squares

Yates has shown that the treatment sum of squares resulting from the analysis of augmented data has a positive bias. Obviously, this will tend to exaggerate the significance of the F-statistic used to test (2.2).

Yates demonstrates the bias as follows.

In general, it is possible to test a null hypothesis by using the general regression significance test. Find the reduction, say R_o, in the total sum of squares, due to fitting the constants in the original model. Find the reduction, say R_d, in the total sum of squares, due to the model obtained by deleting, from the original model, that parameter which is under study. Get the difference, $R_o - R_d$, and divide this by the number of degrees of freedom, say (t-1), for the effect under study.

Next obtain an error sum of squares, say E, by subtracting R_o from the uncorrected total sum of squares. Then divide E by the degrees of freedom due to error.

The quotient

$$\frac{(R_o-R_d)/(t-1)}{E/(d.f., error)} \tag{2.11}$$

is distributed as F[(t-1), (d.f., error)], and one may test for significance the effect under study.

In the balanced incomplete block, this would mean, generally, that R_o is obtained from the model

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \tag{2.12}$$

and R_d from the model

$$y_{ij} = \mu + \beta_j + \epsilon_{ij} ; \tag{2.13}$$

i.e., the proper sum of squares due to treatment effects is

$$R_o-R_d = \text{regression } (\mu,\tau,\beta) - \text{regression } (\mu,\beta). \tag{2.14}$$

Now suppose that some of the observations are missing.

It has been shown that missing values may be estimated, under the original model, in such a way that

$$\sum_{\substack{\text{known}\\ \text{data}}}\sum (y_{ij} - m - t_i - b_j)^2 \tag{2.15}$$

is minimized.

Similarly, estimates of missing values may be obtained under the diminished model such that the expression

$$\sum_{\substack{\text{known}\\\text{data}}}\sum (y_{ij} - m' - b'_j)^2 \tag{2.16}$$

is minimized, where m' and b'_j are estimates, respectively, of μ and β_j in (2.13).

Including the appropriate missing value estimates and following the analysis of variance procedures for both models, using an o-subscript for values obtained under the original model and a d-subscript for values obtained under the diminished model, yields

Table 2.3

Analyses of Variance

Original Model	Sum of Squares	Diminished Model	Sum of Squares
Block	B_o	Block	B_d
Treatment	A_o	Error	E_d
Error	$E_o \equiv E'$	Total	S_d
Total	S_o		

Now one may obtain the total sum of squares, say S, of the deviations only of the existing values from their mean. The reduction in S due to fitting both block and treatment constants is $S - E_o$; the reduction due to fitting block constants only is $S - E_d$. Then the correct sum of squares for testing treatment effects is

$$(S - E_o) - (S - E_d) = E_d - E_o; \tag{2.17}$$

from Table 2.3,

$$\begin{aligned} E_d - E_o &= (S_d-B_d) - (S_o-B_o-T_o) \\ &= T_o - (S_o-B_o) + (S_d-B_d). \end{aligned} \tag{2.18}$$

Clearly,

$$(S_o-B_o) = \sum_{\substack{\text{known} \\ \text{data}}}\sum (y_{ij}-m-b_j)^2 + \sum_{\substack{\text{missing} \\ \text{data}}}\sum (x_{ij}-m-b_j)^2, \tag{2.19}$$

where m and b_j estimate, respectively, μ and β_j in the original model; also

$$(S_d-B_d) = \sum_{\substack{\text{known} \\ \text{data}}}\sum (y_{ij}-m'-b'_j)^2 + \sum_{\substack{\text{missing} \\ \text{data}}}\sum (x_{ij}-m'-b'_j)^2, \tag{2.20}$$

where m' and b'_j estimate, respectively, μ and β_j in the diminished model.

m' and b'_j were chosen so as to minimize the expression (2.20); i.e., those terms in (2.20) which involve missing values vanish. But m and b were not chosen so as to minimize the expression (2.19).

Hence it must be true that

$$\sum_{\substack{\text{known}\\\text{data}}}\sum (y_{ij}-m-b_j)^2 + \sum_{\substack{\text{missing}\\\text{data}}}\sum (x_{ij}-m-b_j)^2$$

$$> \sum_{\substack{\text{known}\\\text{data}}}\sum (y_{ij}-m'-b_j')^2 + \sum_{\substack{\text{missing}\\\text{data}}}\sum (x_{ij}-m'-b_j')^2, \quad (2.21)$$

so that

$$(S_o - B_o) > (S_d - B_d). \quad (2.22)$$

Consequently, the proper sum of squares for testing the null hypothesis, (2.2), is

$$E_d - E_o = T_o - (S_o - B_o) + (S_d - B_d)$$

where

$$T_o > T_o - (S_o - B_o) + (S_d - B_d). \quad (2.23)$$

2.4 A Direct Method of Analysis

While a specific formula for the bias can be obtained, it is much easier, in the case of the balanced incomplete block, to modify the analysis of variance technique in order to obtain the exact F-test.

After obtaining and inserting estimates of the n missing values, obtain the sums of squares (2.5), modifying the block sum of squares (unadjusted) and total sum of squares as follows:

$$\text{Block S.S. (unadj.)} = \frac{1}{k}\sum_{\substack{\text{known}\\ \text{data}}} B_j^2 + \frac{1}{k}\sum_{\substack{\text{missing}\\ \text{data}}} [B_j' + \Sigma x_{ij}]^2 - \frac{(G' + \sum\sum_{ij} x_{ij})^2}{N}$$

$$\equiv B^* \qquad (2.24)$$

$$\text{Total S.S.} = \sum\sum_{\substack{\text{known}\\ \text{data}}} y_{ij}^2 + \sum\sum_{\substack{\text{missing}\\ \text{data}}} x_{ij}^2 - \frac{(G' + \Sigma x_{ij})^2}{N}$$

$$\equiv T^* \quad . \qquad (2.25)$$

B_j' is the total of existing observations for a block that contains one or more missing values; G' is the grand total of existing values.

Letting the remaining sums of squares calculated from the augmented data be represented by A^* (treatments, adjusted) and E^* (error), the values can be set up as in Table 2.2.

If the F-test is not significant, the analysis is complete. If it is significant, the significance may result from the positive bias in A^*; hence, the bias should be eliminated.

To eliminate the bias, proceed in the following manner:

1. enter the error sum of squares which was obtained from the augmented data (E^*) in a new analysis of variance table;

2. using the first term of (2.25), find a total sum of squares based only on known data,

$$T' = \sum_{\substack{\text{known}\\\text{data}}} \sum y_{ij}^2 - \frac{(G')^2}{N-n} \tag{2.26}$$

3. taking advantage of (2.24), obtain a block sum of squares based only on known data,

$$B' = \frac{1}{k} \sum_{\substack{\text{known}\\\text{data}}} B_j^2 + \sum_{\substack{\text{known}\\\text{data}}} \frac{(B_j')^2}{k_i} - \frac{(G')^2}{N-n} ; \tag{2.27}$$

4. obtain the unbiased treatment sum of squares by subtraction.

Table 2.4 gives the exact analysis of variance.

Table 2.4

Exact Analysis of Variance

Source	d.f.	S.S.	M.S.	F
Blocks (unadjusted)	(b-1)	B'		
Treatments (adjusted)	(t-1)	A' = T'-B'-E*	s_A^2	s_A^2/s_E^2
Error	(N-b-t+1-n)	E*	s_E^2	
Total	(N-1-n)	T'		

III. ESTIMATION OF SEVERAL MISSING VALUES

3.1 General Case

3.1.a Introduction

To obtain estimates of the missing values, the procedure proposed in section 2.2 will be followed; i.e., partially differentiate the error sum of squares with respect to each of the missing values, and equate these partial derivatives with zero, thus obtaining a set of n equations in n unknowns.

The error sum of squares may be written as

$$E^* = \left\{ \Sigma\Sigma y_{ij}^2 - \frac{1}{k}\,\Sigma B_j^2 - \frac{(t-1)}{rtk(k-1)}\,\Sigma Q_i^2 \right\}$$

$$+ \left\{ \Sigma\Sigma x_{ij}^2 - \frac{1}{k}\,\Sigma B_j^2 - \frac{(t-1)}{rtk(k-1)}\,\Sigma Q_i^2 \right\}, \qquad (3.1)$$

where the first quantity is obtained only from existing data and the second quantity is obtained from the data involving missing values.

For purposes of expediency, the following definitions are made:

$$E_1 \equiv \Sigma\Sigma x_{ij}^2 ,$$

$$E_2 \equiv -\frac{1}{k}\,\Sigma B_j^2 ,$$

$$E_3 \equiv -\frac{(t-1)}{rtk(k-1)}\,\Sigma Q_i^2 , \qquad (3.2)$$

where ΣB_j^2 and ΣQ_i^2 are taken only from the second quantity in (3.1). With these definitions, $E_1 + E_2 + E_3$ constitutes that portion of the error sum of squares which involves the missing observations.

Obviously, the first quantity in E* will vanish under differentiation, so that one need consider only the expression $E_1 + E_2 + E_3$ in obtaining estimates of the missing values.

When a particular missing value is being considered, this missing value will be referred to as x_{cd}. Two particular missing values will be referred to as x_{cd} and x_{ef}. n missing values will be denoted by $x_{cd}, x_{ef}, \ldots, x_{uv}$.

In order to deal with the block and treatment totals which involve missing values, define

$$\begin{aligned} T_i &\equiv T_i' + (\text{missing values in treatment } i) \\ B_j &\equiv B_j' + (\text{missing values in block } j). \end{aligned} \tag{3.3}$$

x_{cd} will be involved in two types of Q-values: the Q-value for treatment c and Q-values for the other treatments in the same block as x_{cd}. The former will be denoted by

$$Q_c \equiv Q_c' + (\text{terms involving missing values}), \tag{3.4}$$

the latter by

$$Q_i \equiv Q_{i(c)}' + (\text{terms involving missing values}), \tag{3.5}$$

where Q_c' and $Q_{i(c)}'$ are the numerical values obtained from existing data.

It will also be useful to define M as that collection of terms which involve missing values other than those under consideration.

In general, when differentiating the error sum of squares with respect to the missing value x_{cd}, one obtains an equation of the form

$$\frac{\partial E_1}{\partial x_{cd}} + \frac{\partial E_2}{\partial x_{cd}} + \frac{\partial E_3}{\partial x_{cd}} = (\text{terms involving } x_{cd}) + (\text{constant terms involving } B'_d,\ Q'_c, \text{ and } Q'_{i(c)}) + M. \quad (3.6)$$

Equating this partial derivative with zero yields an equation of the form

$$(\text{terms involving } x_{cd}) + M = (\text{constant terms involving } B'_d,\ Q'_c, \text{ and } Q'_{i(c)}) \quad (3.7)$$

3.1.b <u>Terms Unaffected by Configuration</u>

It is to be expected that the configuration of the missing values within the design will have some bearing on the form of the equation obtained by differentiating with respect to a particular missing value, x_{cd}. However, this configuration affects neither the coefficient of x_{cd} nor the essential form of the constant terms. It is the coefficient of x_{cd} and the form of the constant terms which will be obtained in this section.

If one differentiates E_1 with respect to x_{cd}, the result is

$$\frac{\partial E_1}{\partial x_{cd}} = \frac{\partial(x_{cd}^2 + x_{ef}^2 + \dots + x_{uv}^2)}{\partial x_{cd}}$$

$$= 2x_{cd} \ . \tag{3.8}$$

Obviously, this result is the same under any configuration.

The block total for block d, containing x_{cd}, is of the form

$$B_d = B_d' + x_{cd} + M. \tag{3.9}$$

Differentiating E_2 with respect to x_{cd} one gets

$$\frac{\partial E_2}{\partial x_{cd}} = \frac{\partial \left\{- \frac{1}{k} \left[(B_d' + x_{cd} + M)^2 + \sum_{j \neq d} (B_j' + M)^2 \right] \right\}}{\partial x_{cd}}$$

$$= - \frac{2}{k} B_d' - \frac{2}{k} x_{cd} + M; \tag{3.10}$$

hence, the only expression affected by configuration is M.

There are k distinct treatments replicated in block d, which contains x_{cd}. Since there is a distinct Q-value corresponding to each treatment, the block total B_d, and therefore x_{cd}, must appear in k Q-values - one for each of the k treatments in the block. This follows from (2.4).

By definition,

$$Q_c \equiv kT_c - B_{\cdot c} \ ;$$

but treatment c contains x_{cd}, so that

$$Q_c = k(T_c'+x_{cd}+M) - (B_d'+x_{cd}+M)$$
$$- \sum_{j\neq d} (\text{other involved } B_j)$$
$$\equiv Q_c' + (k-1)x_{cd} + M. \qquad (3.11)$$

The other (k-1) Q-values involving x_{cd} are of the form

$$Q_i = kT_i - (B_d'+x_{cd}+M) - \sum_{j\neq d} (\text{other involved } B_j)$$
$$\equiv Q_{i(c)}' - x_{cd} + M. \qquad (3.12)$$

Differentiating E_3 with respect to x_{cd}, one gets

$$\frac{\partial E_3}{x_{cd}} = \frac{\partial\left\{- \frac{(t-1)}{rtk(k-1)}[(Q_c'+(k-1)x_{cd}+M)^2 + \sum_{i\neq c}(Q_{i(c)}'-x_{cd}+M)^2]\right\}}{x_{cd}}$$
$$+ \frac{\partial\left\{\frac{-(t-1)}{rtk(k-1)} \sum (Q_i' + M)^2\right\}}{\partial x_{cd}}$$

$$= \frac{-2(t-1)}{rtk(k-1)} \left\{[(k-1)^2+(k-1)]x_{cd}+(k-1)Q_c'- \sum_{i\neq c} Q_{i(c)}'\right\} + M;$$
$$(3.13)$$

again, the only expression whose form might be affected by configuration is M.

Consequently, by combining results (3.8), (3.10), and (3.13), one gets

$$\frac{\partial E}{\partial x_{cd}} = \frac{\partial(E_1 + E_2 + E_3)}{\partial x_{cd}}$$

$$= 2x_{cd} - \frac{2}{k} B'_d - \frac{2}{k} x_{cd}$$

$$\frac{-2(t-1)}{rtk(k-1)} \left\{[(k-1)^2+(k-1)]x_{cd}+(k-1)Q'_c - \sum_{i\neq c} Q'_{i(c)}\right\}$$

$$+ M. \qquad (3.14)$$

After equating this partial derivative with zero and simplifying, one obtains the general result

$$mx_{cd}+M = rt(k-1)B'_d+(t-1)(k-1)Q'_c-(t-1) \sum_{i\neq c} Q'_{i(c)}, \qquad (3.15)$$

where

$$m = (k-1)(rtk-rt-tk+k).$$

As a consequence of this derivation, another expedient notation will be introduced:

$$C_{cd} \equiv rt(k-1)B'_d+(t-1)(k-1)Q'_c-(t-1) \sum_{i\neq c} Q'_{i(c)}; \qquad (3.16)$$

i.e., C_{cd} denotes the right-hand-side, or constant term, obtained by differentiating the error sum of squares with respect to the missing observation, x_{cd}.

Now the equation system involving the n missing values, x_{cd}, x_{ef}, ..., x_{uv}, may be written

$$\begin{aligned} mx_{cd} + M &= C_{cd} \\ mx_{ef} + M &= C_{ef} \\ &\cdots \\ mx_{uv} + M &= C_{uv} ; \end{aligned} \qquad (3.17)$$

i.e., the coefficient of the missing value with respect to which the differentiation is performed is m in every case, and the right-hand-side, or constant term, is of the same form in every case.

3.1.c <u>Terms Affected by Configuration</u>

If one differentiates the error sum of squares with respect to the missing value, x_{cd}, the configuration of the missing values within the design affects neither the term involving x_{cd} nor the essential form of the constant terms, as has been shown in the preceding section. However, the terms involving the other missing values will depend upon their positions, in relation to x_{cd}, in the design. In this section, the coefficients of these other missing values will be obtained under the various possible configurations.

In obtaining these coefficients, it will be most convenient to consider all possible configurations of the two missing values, x_{cd} and x_{ef}, differentiating always with respect to x_{cd}. Unless otherwise indicated, it is assumed that x_{cd} and x_{ef} occur in distinct blocks and distinct treatments; i.e., $c \neq e$, $d \neq f$.

The procedure is as follows. First, the configuration will be identified. Next, those terms in the error sum of squares which are uniquely affected by the configuration will be pointed out. Finally, the resulting equation will be written, showing the term involving x_{ef}.

Configuration 1. No treatments are replicated both in block d and in block f. In this event, x_{cd} and x_{ef} do not occur together in the same block or in the same Q-value. In all events, they form distinct terms in E_1, as was established by (3.8). Hence, the two unknowns will appear in distinct terms in the error sum of squares, and terms in x_{ef} will vanish under differentiation. The resulting equation is

$$mx_{cd} + 0 \cdot x_{ef} + M = C_{cd} \;. \tag{3.18}$$

Configuration 2. x_{cd} and x_{ef} occur in the same block; i.e., $d \equiv f$. The affected terms are

$$\begin{aligned} B_d &= B'_d + x_{cd} + x_{ed} + M \\ Q_c &= Q'_c + (k-1)x_{cd} - x_{ed} + M \\ Q_e &= Q'_e + (k-1)x_{ed} - x_{cd} + M \\ \sum_{i \neq c,e} Q_i &= \sum_{i \neq c,e} (Q'_{i(c,e)} - x_{cd} - x_{ed} + M) \;. \end{aligned} \tag{3.19}$$

The equation resulting from differentiation of the error sum of squares with respect to x_{cd} is

$$mx_{cd} - \frac{m}{(k-1)} x_{ed} + M = C_{cd} \;. \tag{3.20}$$

Configuration 3. x_{cd} and x_{ef} are in the same treatment; i.e., $c \equiv e$. There are no other treatments replicated both in block d and in block f. The affected term is

$$Q_c = Q_c' + (k-1)(x_{cd}+x_{cf}) + M, \tag{3.21}$$

and the final equation is

$$mx_{cd}-(t-1)(k-1)^2x_{cf} + M = C_{cd} \; . \tag{3.22}$$

Configuration 4. There are p treatments (not c or e) common to blocks d and f. The affected terms are

$$\overset{p}{\Sigma}Q_i = \overset{p}{\Sigma}[Q_{i(c,e)}'-x_{cd}-x_{ef} + M]; \tag{3.23}$$

the final equation is

$$mx_{cd} - p(t-1)x_{ef} + M = C_{cd} \; . \tag{3.24}$$

Configuration 5. Treatment c is replicated in block f; no other treatment is replicated both in block d and in block f. (It can be shown that this is identical to the case where treatment e is replicated in block d.) The affected term is

$$Q_c = Q_c' + (k-1)x_{cd} - x_{ef} + M, \tag{3.25}$$

and the final equation is

$$mx_{cd} + (t-1)(k-1)x_{ef} + M = C_{cd}. \tag{3.26}$$

Configuration 6. Treatment c is replicated in block f and treatment e is replicated in block d; no other treatments are replicated in both blocks. The affected terms are

$$Q_c = Q_c' + (k-1)x_{cd} - x_{ef} ,$$

$$Q_e = Q_e' + (k-1)x_{ef} - x_{cd} . \tag{3.27}$$

The final equation is

$$mx_{cd} + 2(t-1)(k-1)x_{ef} + M = C_{cd}. \tag{3.28}$$

These results may be summarized in tabular form, where the relative positions in the design of the two missing values, x_{cd} and x_{ef}, are considered, and where differentiation is performed with respect to x_{cd}.

Before leaving the topic of configurations, it is extremely important to observe that these configurations are not all mutually exclusive.

For instance, x_{cd} and x_{ef} may occur in the same treatment; in addition, there may be three other treatments replicated both in block d and in block f. Thus x_{cd} and x_{ef} would fit both configuration 3 and configuration 4. The coefficients of x_{cf} would be added, and the equation obtained by differentiating the error sum of squares with respect to x_{cd} would be

$$mx_{cd} - (t-1)[p+(k-1)^2]x_{cf} + M = C_{cd}. \tag{3.29}$$

Table 3.1

Off-Diagonal Coefficients

Configuration of x_{cd} and x_{ef}	Coefficient of x_{ef}
1. No treatments common to blocks d and f.	0
2. x_{cd} and x_{ef} occur in the same block.	$-\frac{m}{k-1}$
3. x_{cd} and x_{ef} occur in the same treatment; no other treatments common to both blocks.	$-(t-1)(k-1)^2$
4. There are p treatments (not c or e) common to blocks d and f.	$-p(t-1)$
5. Treatment c is replicated in block f <u>OR</u> treatment e is replicated in block d; no other treatments common to both blocks.	$(t-1)(k-1)$
6. Treatment c is replicated in block f <u>AND</u> treatment e is replicated in block d; no other treatments common to both blocks.	$2(t-1)(k-1)$

3.1.d Matrix Representation and Solution

If there are n missing values, x_{cd}, x_{ef}, ..., x_{uv}, and if the error sum of squares is differentiated with respect to each of the missing values in the order specified, the result, upon equating these partial derivatives with zero, is the equation system

$$\begin{aligned} m x_{cd} + q_{12} x_{ef} + \cdots + q_{1n} x_{uv} &= C_{cd} \\ q_{21} x_{cd} + m x_{ef} + \cdots + q_{2n} x_{uv} &= C_{ef} \\ q_{n1} x_{cd} + q_{n2} x_{ef} + \cdots + m x_{uv} &= C_{uv} \quad , \end{aligned}$$

where $q_{ij} \equiv q_{ji}$, obtained by the appropriate combination of coefficients from Table 3.1.

This equation system may be written in matrix form as

$$A\underline{x} = \underline{C} \qquad (3.30)$$

where

$$A = \begin{bmatrix} m & q_{12} & \cdots & q_{1n} \\ q_{21} & m & \cdots & q_{2n} \\ \cdot & \cdot & & \cdot \\ q_{n1} & q_{n2} & \cdots & m \end{bmatrix} , \qquad (3.31)$$

$$\underline{x} = \begin{bmatrix} x_{cd} \\ x_{ef} \\ \vdots \\ x_{uv} \end{bmatrix} , \qquad (3.32)$$

and

$$\underline{C} = \begin{bmatrix} C_{cd} \\ C_{ef} \\ \vdots \\ C_{uv} \end{bmatrix} \tag{3.33}$$

The estimates of the missing values are then obtained from

$$\underline{x} = A^{-1}\underline{C} \ . \tag{3.34}$$

3.2 Special Cases

3.2.a n Missing Values in n Distinct Blocks and n Distinct Treatments; No Treatment Replicated in Any Two of the Involved Blocks

Clearly, every possible pair of missing observations satisfies configuration 1, so that the equation system reduces to

$$\begin{array}{cccccc} mx_{cd} & 0 & \cdots & 0 & = & C_{cd} \\ 0 & mx_{ef} & \cdots & 0 & = & C_{ef} \\ \cdot & & \cdot & \cdot & & \\ 0 & 0 & \cdots & mx_{uv} & = & C_{uv} \ . \end{array} \tag{3.35}$$

In this case, the estimate of the missing value is treatment i and block j is simply

$$x_{ij} = \frac{1}{m} C_{ij}$$
$$= \frac{rt(k-1)B_j' + (t-1)(k-1)Q_i' - (t-1)\sum_{i' \neq i} Q_{i'(i)}'}{(k-1)(rtk-rt-tk+k)} , \qquad (3.36)$$

which is identical to the estimate of a single missing value, as first shown by Cornish (1940).

3.2.b <u>n Missing Values in the Same Block</u>

Here, each pair of missing values satisfies configuration 2, leading to the equation system

$$\begin{aligned} mx_{cd} - \frac{m}{k-1} x_{ed} - \cdots - \frac{m}{k-1} x_{ud} &= C_{cd} \\ -\frac{m}{k-1} x_{cd} + mx_{ed} - \cdots - \frac{m}{k-1} x_{ud} &= C_{ed} \qquad (3.37) \\ \cdot \qquad \cdot \qquad \cdot \\ -\frac{m}{k-1} x_{cd} - \frac{m}{k-1} x_{ed} - \cdots + mx_{ud} &= C_{ud} \, . \end{aligned}$$

In matrix form, this becomes

$$A\underline{x} = \underline{C} \qquad (3.38)$$

where

$$A = \begin{bmatrix} m & -\frac{m}{k-1} & \cdots & -\frac{m}{k-1} \\ -\frac{m}{k-1} & m & \cdots & -\frac{m}{k-1} \\ \cdot & \cdot & \cdot & \\ -\frac{m}{k-1} & -\frac{m}{k-1} & \cdots & m \end{bmatrix}$$

$$\underline{x} = \begin{bmatrix} x_{cd} \\ x_{ed} \\ \cdot \\ \cdot \\ \cdot \\ x_{ud} \end{bmatrix}, \tag{3.40}$$

and

$$\underline{C} = \begin{bmatrix} C_{cd} \\ C_{ed} \\ \cdot \\ \cdot \\ \cdot \\ C_{ud} \end{bmatrix}. \tag{3.41}$$

The solution is, of course,

$$\underline{x} = A^{-1}\underline{C} , \tag{3.42}$$

where

$$A^{-1} = \frac{1}{mk(k-n)} \begin{bmatrix} k(k-n)+(n-1) & (k-1) & \ldots & (k-1) \\ (k-1) & k(k-n)+(n-1) & \ldots & (k-1) \\ \cdot & \cdot & \cdot & \\ (k-1) & (k-1) & \ldots & k(k-n)+(n-1) \end{bmatrix} \tag{3.43}$$

Performing the operation, $A^{-1}\underline{C}$, one obtains the missing value estimates

$$x_{cd} = \frac{1}{mk(k-n)} \left\{[k(k-n)+(n-1)]C_{cd}+(k-1) \sum_{i\neq c} C_{id}\right\}$$

$$x_{ed} = \frac{1}{mk(k-n)} \left\{[k(k-n)+(n-1)]C_{ed}+(k-1) \sum_{i\neq e} C_{id}\right\} \qquad (3.44)$$

$$\cdot \quad \cdot \quad \cdot$$

$$x_{ud} = \frac{1}{mk(k-n)} \left\{[k(k-n)+(n-1)]C_{ud}+(k-1) \sum_{i\neq u} C_{id}\right\} .$$

3.2.c <u>n Missing Values in the Same Treatment; No other Treatment Replicated in Any Two of the Involved Blocks</u>

In this case, every pair of missing values satisfies configuration 3, and the equation system is

$$mx_{cd} - (t-1)(k-1)^2x_{cf} \cdots -(t-1)(k-1)^2x_{cv} = C_{cd}$$

$$-(t-1)(k-1)^2x_{cd} + mx_{cf}\cdots-(t-1)(k-1)^2x_{cv} = C_{cf} \qquad (3.45)$$

$$\cdot \quad \cdot \quad \cdot$$

$$-(t-1)(k-1)^2x_{cd} - (t-1)(k-1)^2x_{cf} \cdots +mx_{cv} = C_{cv} .$$

Again using matrix notation, these equations may be written as

$$A\underline{x} = \underline{C}, \qquad (3.46)$$

where

$$A = \begin{bmatrix} m & -(t-1)(k-1)^2 \ldots & -(t-1)(k-1)^2 \\ -(t-1)(k-1)^2 & m & \ldots -(t-1)(k-1)^2 \\ \cdot & \cdot & \cdot \\ -(t-1)(k-1)^2 & -(t-1)(k-1)^2 & m \end{bmatrix}, \tag{3.47}$$

and where $\underline{x}$ and $\underline{C}$ are defined as before.

Then

$$A^{-1} =$$

$$\frac{1}{K}\begin{bmatrix} m-(n-2)(t-1)(k-1)^2 & (t-1)(k-1)^2 & \ldots & (t-1)(k-1)^2 \\ (t-1)(k-1)^2 & m-(n-2)(t-1)(k-1)^2 & \ldots & (t-1)(k-1)^2 \\ \cdot & \cdot & & \cdot \\ (t-1)(k-1)^2 & (t-1)(k-1)^2 & \ldots & m-(n-2)(t-1)(k-1)^2 \end{bmatrix} \tag{3.48}$$

where

$$K = m^2 - m(n-2)(t-1)(k-1)^2 - (n-1)(t-1)^2(k-1)^4; \tag{3.49}$$

the solution,

$$\underline{x} = A^{-1}\underline{C}, \tag{3.50}$$

yields the missing value estimates

$$x_{cd} = \frac{1}{K}\left\{[m-(n-2)(t-1)(k-1)^2]C_{cd} + (t-1)(k-1)^2 \sum_{j\neq d} C_{cj}\right\}$$

$$x_{cf} = \frac{1}{K}\left\{[m-(n-2)(t-1)(k-1)^2]C_{cf} + (t-1)(k-1)^2 \sum_{j\neq f} C_{cj}\right\}$$

$$\cdot \qquad \cdot \qquad \cdot \tag{3.51}$$

$$x_{cv} = \frac{1}{K}\left\{[m-(n-2)(t-1)(k-1)^2]C_{cv} + (t-1)(k-1)^2 \sum_{j\neq v} C_{cj}\right\}$$

3.2.d Two Missing Values

By applying the various results obtained in the sections immediately preceding this, it is quite easy to obtain explicit formulae for every possible arrangement, in the balanced incomplete block design, of two missing values.

The procedure will be as follows: (1) the configuration of the two missing values, x_{cd} and x_{ef}, will be identified, and (2) the estimates of these missing values will be given.

1. No treatments common to blocks d and f. Then

$$x_{cd} = \frac{1}{m} C_{cd} ,$$
$$x_{ef} = \frac{1}{m} C_{ef} . \tag{3.52}$$

2. There are p treatments (not c or e) common to blocks d and f.

$$x_{cd} = \frac{1}{[m^2-p^2(t-1)^2]} \{mC_{cd} + p(t-1)C_{ef}\}$$
$$x_{ef} = \frac{1}{[m^2-p^2(t-1)^2]} \{mC_{ef} + p(t-1)C_{cd}\} . \tag{3.53}$$

3. There are p treatments (not c or e) common to blocks d and f. It is possible that $p = 0$. In addition, treatment c is replicated in block f OR treatment e is replicated in block d.

In this event,

$$x_{cd} = \frac{1}{[m^2-(t-1)^2(k-1-p)^2]} \{mC_{cd}-(t-1)(k-1-p)C_{ef}\}$$
$$x_{ef} = \frac{1}{[m^2-(t-1)^2(k-1-p)^2]} \{mC_{ef}-(t-1)(k-1-p)C_{cd}\}. \tag{3.54}$$

4. There are p treatments (not c or e) common to blocks d and f. It is possible that $p = 0$. In addition, treatment c is replicated in block f AND treatment e is replicated in block d.

Then

$$x_{cd} = \frac{1}{[m^2-(t-1)^2(2k-2-p)^2]} \{mC_{cd}-(t-1)(2k-2-p)C_{ef}\}$$
$$x_{ef} = \frac{1}{[m^2-(t-1)^2(2k-2-p)^2]} \{mC_{ef}-(t-1)(2k-2-p)C_{cd}\} . \tag{3.55}$$

5. x_{cd} and x_{ef} occur in the same block; i.e., $d \equiv f$.

Here,

$$x_{cd} = \frac{(k-1)}{mk(k-2)}\left\{(k-1)C_{cd} + C_{ed}\right\}$$
$$x_{ed} = \frac{(k-1)}{mk(k-2)}\left\{(k-1)C_{ed} + C_{cd}\right\} . \tag{3.56}$$

6. There are p treatments (not c or e) common to blocks d and f. It is possible that $p = 0$. In addition, x_{cd} and x_{ef} occur in the same treatment; i.e., $c \equiv e$.

Then

$$x_{cd} = \frac{1}{m^2-(t-1)^2[(k-1)^2+p]^2}\left\{mC_{cd}+(t-1)[(k-1)^2+p]C_{cf}\right\}$$
$$x_{cf} = \frac{1}{m^2-(t-1)^2[(k-1)^2+p]^2}\left\{mC_{cf}+(t-1)[(k-1)^2+p]C_{cd}\right\} . \tag{3.57}$$

IV. ANALYSIS OF A BALANCED INCOMPLETE BLOCK AS A RANDOMIZED BLOCK WITH MISSING OBSERVATIONS.

The balanced incomplete block design may be thought of as a randomized block with (t-k) missing observations in each block. If these (t-k) missing observations are estimated and inserted in the original incomplete design, a regular randomized block analysis can be used on the data. Glenn and Kramer (1958) have developed a procedure for estimating missing values in a randomized block and for obtaining an exact analysis of the augmented data.

In this section a balanced incomplete block design will be treated as a randomized block with missing values, the procedure of Glenn and Kramer being used to obtain the missing value estimates and the analysis of the augmented data. The exact balanced incomplete block analysis will also be obtained, in order to demonstrate the precision of the randomized block analysis.

It will suffice, for these purposes, to use an arbitrary set of data for a balanced incomplete block design with $r = 3$, $t = 4$, $k = 3$, $b = 4$, and $\lambda = 2$. This data is given in Table 4.1.

Table 4.1

Balanced Incomplete Block Design Considered as a Randomized Block Design with Missing Observations.

		Blocks 1	2	3	4	$Y'_{i\cdot}$ Original Totals	$Y_{i\cdot}$ Augmented Totals
Treatments	1	15.54	12.80	16.59	20.97	44.93	65.90
	2	15.54	13.91	17.22	20.37	49.82	67.04
	3	18.22	18.30	21.81	27.49	67.52	85.82
	4	21.26	20.11	23.97	26.90	70.98	92.24
$Y'_{\cdot j}$ Original Totals		49.30	46.82	62.37	74.76	233.25	
$Y_{\cdot j}$ Augmented Totals		70.56	65.12	79.59	95.73		311.00

The missing values occur in the circled cells; there are four missing values, no two of which are in the same block or treatment. In this event, according to their equation (3.4), Glenn and Kramer have shown that the missing value in treatment h and block k is estimated by

$$x_{hk} = \frac{[(p-1)(q-1)+(n-2)]Z_{hk} - \underset{\substack{\text{missing plots} \\ i\neq h \quad j\neq k}}{\Sigma \quad \Sigma} Z_{ij}}{(p-1)^2(q-1)^2 + (n-2)(p-1)(q-1) - (n-1)},$$

where Z_{ij} is defined by their equation (3.2) as

$$Z_{ij} = pY'_{i.} + qY'_{.j} - Y'_{..} \quad ,$$

where

p = number of treatments
$= 4,$

where

q = number of blocks
$= 4,$

and where

n = number of missing values
$= 4.$

First obtaining the Z_{ij}'s, one gets

$$Z_{41} = 4(70.98) + 4(49.30) - 233.25$$
$$= 247.87$$

$$Z_{32} = 4(67.52) + 4(46.82) - 233.25$$
$$= 224.11$$

$$Z_{23} = 4(49.82) + 4(62.37) - 233.25$$
$$= 215.51$$

$$Z_{14} = 4(44.93) + 4(74.76) - 233.25$$
$$= 245.51 \quad .$$

Next the missing value estimates are obtained:

$$x_{41} = \frac{[(3)(3)+2](247.87)-(224.11 + 215.51 + 245.51)}{[(3^2)(3^2) + (2)(3)(3) - 3]}$$

$$= 21.26$$

$$x_{32} = 18.30$$

$$x_{23} = 17.22$$

$$x_{14} = 20.97 \quad .$$

These estimates are entered in the table, as in Table 4.1, and the augmented totals are obtained.

Following the procedure outlined in the section entitled <u>Analysis of Variance, A Direct Method of Analysis</u> in the paper by Glenn and Kramer, the analysis of variance is then obtained, using the augmented totals. The total sum of squares is partitioned into two parts: that involving existing observations and that involving estimated observations. Ordinarily, the block sum of squares is similarly partitioned, but in this case, every block total involves a missing observation.

In this case, the following results are obtained.

The correction factor is

$$\frac{Y_{..}^2}{16} = \frac{(311.00)^2}{16}$$

$$= 6045.0625.$$

The total sum of squares is

$$T^* = 4796.4039 + 1523.1469 - 6045.0625$$
$$= 274.4883;$$

the block sum of squares is

$$B^* = \frac{24718.1290}{4} - 6045.0625$$
$$= 134.4697;$$

the treatment sum of squares is

$$A^* = \frac{24710.4616}{4} - 6045.0625$$
$$= 132.5529;$$

and the error sum of squares, obtained by subtraction, is

$$E^* = 7.4657.$$

These results are summarized in Table 4.2.

Table 4.2

Approximate Randomized Block Analysis

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Blocks	3	134.4697	44.8232	
Treatments	3	132.5529	44.1843	29.59**
Error	5	7.4657	1.4931	
Total	11	274.4883		

The F-test shows treatment effects significant at the one per cent level. But since this significance may be the result of a positive bias in the treatment sum of squares, the exact analysis must be obtained, treating the original block totals and observations as a one-way classification with respect to blocks.

The error sum of squares from the approximate analysis is retained; i.e.,

$$\begin{aligned} E' &= E^* \\ &= 7.4657. \end{aligned}$$

The correction factor is

$$\frac{(Y'_{..})^2}{12} = 4533.7969.$$

The total sum of squares is

$$\begin{aligned} T' &= 4796.4039 - 4533.7969 \\ &= 262.6070; \end{aligned}$$

the block sum of squares is

$$\begin{aligned} B' &= \frac{14101.6769}{3} - 4533.7969 \\ &= 166.7621; \end{aligned}$$

the treatment sum of squares, obtained by subtraction, is

$A' = 88.3792.$

The exact analysis is given in Table 4.3.

Table 4.3

Exact Randomized Block Analysis

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Blocks	3	166.7621		
Treatments	3	88.3792	29.4597	19.73**
Error	5	7.4657	1.4931	
Total	11	262.6070		

Although the exact analysis considerably reduced the magnitude of the treatment sum of squares, the F-test still shows treatment effects significant at the one per cent level.

The exact balanced incomplete block analysis is obtained as follows, the data being given in Table 4.4.

Table 4.4

Balanced Incomplete Block

1 15.54	2 15.54	3 18.22	B_1 = 49.30
1 12.80	2 13.91	4 20.11	B_2 = 46.82
1 16.59	3 21.81	4 23.97	B_3 = 62.37
2 20.37	3 27.49	4 26.90	B_4 = 74.76
			G = 233.25

One first obtains the block totals, B_1, B_2, B_3, and B_4, and the grand total, G.

Next, the quantities T_i and $B_{\cdot i}$ can be obtained simultaneously by a split dial procedure. Each observation in treatment i is entered on the left side of the dial, while the block total for the block in which that observation occurred is entered on the right side of the dial.

Thirdly, the Q_i's are obtained according to (2.4):

$$Q_i = kT_i - B_{\cdot i} \quad .$$

Treatment	T_i	$B_{\cdot i}$		Q_i
1	44.93	158.49		-23.70
2	49.82	170.88		-21.42
3	67.52	186.43		16.13
4	70.98	183.95		28.99
		699.75	= kG	0
			= 3(233.25)	

The sums of squares are obtained by application of the equations (2.5). The correction factor is

$$\frac{G^2}{12} = \frac{(233.25)^2}{12}$$
$$= 4533.7969.$$

The sum of squares for treatments, adjusted, is

$$\frac{(t-1)}{rtk(k-1)} \sum_i Q_i^2 = \frac{1}{24}(2121.1034)$$
$$= 88.3793;$$

the sum of squares for blocks, unadjusted, is

$$\frac{1}{K} \sum_j B_j^2 - \frac{G^2}{12} = \frac{14101.6769}{3} - 4533.7969$$
$$= 166.7621;$$

the total sum of squares is

$$\sum_i \sum_j y_{ij}^2 - \frac{G^2}{12} = 4796.4039 - 4533.7969$$
$$= 262.6070;$$

the error sum of squares, obtained by subtraction, is

$$E = 7.4656.$$

The exact balanced incomplete block analysis is given in Table 4.5.

Table 4.5

Exact Balanced Incomplete Block Analysis

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Blocks (unadjusted)	3	166.7621	55.5874	
Treatments (adjusted)	3	88.3793	29.4598	19.73**
Error	5	7.4656	1.4931	
Total	11	262.6070		

A comparison of this analysis with the exact randomized block analysis in Table 4.3 shows that the results are in agreement except in the last decimal place in the treatment and error sums of squares and in the treatment mean square.

In a case such as this, the randomized block analysis is very simple to apply. However, if t is considerably greater than k, the estimation of the missing entries for a randomized block analysis may require the inversion of a fairly large matrix. This obviously might take more time than a straightforward balanced incomplete block analysis.

Also, there may be, say, n missing values in the balanced incomplete block design itself. Then, if the randomized block procedure is used, there are $[b(t-k)+n]$ values to estimate, quite possibly involving the inversion of a $[b(t-k)+n]$-square matrix, whereas the procedure developed in this paper for a balanced incomplete block analysis of the same data would involve only estimation of the n values.

V. EXAMPLE

The example presented in this section utilizes data from an experiment of Gish (Virginia Polytechnic Institute, 1949, unpublished results) in which alfalfa was subjected to nine top dressings of fertilizer. The observations represent yields of cured hay in pounds per fifty square feet.

The design was a balanced incomplete block with $r = 4$, $t = 9$, $k = 3$, $b = 12$, and $\lambda = 1$.

In order to demonstrate the method of analysis presented in this paper, two observations were deleted from the data: x_{34}, in treatment 3 and block 4, and x_{37}, in treatment 3 and block 7. These two cells are circled in the data, Table 5.1.

The two missing values both occur in the same treatment, but there are no other treatments replicated in the two involved blocks. Hence, one would apply the formulae in (3.57), with $p = 0$:

$$x_{34} = \frac{mC_{34} + (t-1)(k-1)^2C_{37}}{m^2 - (t-1)^2(k-1)^4} ,$$

$$x_{37} = \frac{mC_{37} + (t-1)(k-1)^2C_{34}}{m^2 - (t-1)^2(k-1)^4} ;$$

where

Table 5.1

Alfalfa Yield in Pounds of Cured Hay

			Block Totals	
			Original Totals	Augmented Totals
4 8.21	5 6.78	6 6.73	21.72	21.72
7 6.01	8 9.13	9 6.89	22.03	22.03
1 7.39	2 6.04	3 3.87	17.30	17.30
3 4.71	6 6.40	9 6.78	13.18	17.89
2 7.75	5 7.23	8 7.57	22.55	22.55
1 7.66	4 7.94	7 6.11	21.71	21.71
3 4.89	4 8.36	8 8.34	16.70	21.59
1 7.67	5 6.07	9 6.51	20.25	20.25
2 7.57	6 6.38	7 5.96	19.91	19.91
1 7.81	6 7.35	8 8.56	23.72	23.72
3 4.82	5 6.82	7 6.11	17.75	17.75
2 7.03	4 8.31	9 6.11	21.45	21.45
			238.27	247.87

$$m = (k - 1)(rtk - rt - tk + k) = 96 ;$$

and where C_{34} and C_{37} are defined, according to (3.16), as

$$C_{34} \equiv rt(k-1)B_4' + (t-1)(k-1)Q_3' - (t-1)[Q_{6(3)}' + Q_{9(3)}'],$$

$$C_{37} \equiv rt(k-1)B_7' + (t-1)(k-1)Q_3' - (t-1)[Q_{4(3)}' + Q_{8(3)}'],$$

the B_j', Q_c', and $Q_{i(c)}'$ being defined, in turn, according to equations (3.3), (3.4), and (3.5), respectively.

In estimating the missing values, one first obtains the original block totals and the original grand total, shown in Table 5.1.

Next, using a split-dial technique, the original quantities T_i and $B_{.i}$ are obtained.

From the latter are obtained the original Q-values.

	T_i		$B_{\cdot i}$		Q_i	
Treatment	Original	Augmnt.	Original	Augmnt.	Original	Augmnt.
1	30.53	30.53	82.98	82.98	8.61	8.61
2	28.39	28.39	81.21	81.21	3.96	3.96
3	8.69	18.29	64.93	74.53	-38.86	-19.66
4	32.82	32.82	81.58	86.47	16.88	11.99
5	26.90	26.90	82.27	82.27	- 1.57	- 1.57
6	26.86	26.86	78.53	83.24	2.05	- 2.66
7	24.19	24.19	81.40	81.40	- 8.83	- 8.83
8	33.60	33.60	85.00	89.89	15.80	10.91
9	26.29	26.29	76.91	81.62	1.96	- 2.75
	238.27	247.87	714.81	743.61	0	0

From these results, it is determined that

$$C_{34} = (4)(9)(2)(13.18) + (8)(2)(-38.86) - 8(2.05 + 1.96)$$
$$= 295.12$$

and

$$C_{37} = 72(16.70) + 16(-38.86) - 8(16.88 + 15.80)$$
$$= 319.20.$$

Then the missing value estimates are obtained:

$$x_{34} = \frac{(96)(295.12) + (32)(319.20)}{96^2 - (64)(16)}$$

$$= \frac{28545.92}{8192}$$

$$= 4.71 ;$$

$$x_{37} = \frac{(96)(319.20) + (32)(295.12)}{8192}$$

$$= 4.89.$$

These estimates are entered in the data (they are circled in Table 5.1), and the augmented quantities B_j, T_i, $B_{\cdot i}$, and Q_i are obtained.

The approximate analysis of variance and, if necessary, the exact analysis of variance are obtained according to the procedure given in section 2.4.

First the approximate analysis is obtained.

Here, the correction factor is

$$\frac{G^2}{36} = \frac{(247.87)^2}{36}$$

$$= 1706.6538.$$

The sum of squares for blocks, unadjusted, is partitioned into two parts, one part corresponding to block totals not involving missing observations and one part corresponding to block totals involving missing observations:

$$B^* = \frac{4380.4699}{3} + \frac{786.1802}{3} - 1706.6538$$

$$= 15.5629.$$

The total sum of squares is also partitioned into the part involving existing observations and the part involving estimates of missing observations:

$$T^* = 1709.5133 + 46.0962 - 1706.6538$$

$$= 48.9557.$$

The adjusted treatment sum of squares is obtained in the usual manner:

$$A^* = \frac{(t-1)}{rtk(k-1)} \quad \Sigma Q_i^2$$

$$= \frac{1}{27}\,(834.1894)$$

$$= 30.8959\ .$$

The error sum of squares, obtained by subtraction, is

$$E^* = 2.4969.$$

The approximate analysis of variance is given in Table 5.2.

Table 5.2

Approximate Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Blocks (unadjusted)	11	15.5629	1.4148	
Treatments (adjusted)	8	30.8959	3.8620	21.66**
Error	14	2.4969	.1783	
Total	33	48.9557		

The F-test is significant at the one per cent level, but, since this significance may be due to bias, it is necessary to get the exact analysis.

The correction factor for the exact analysis is

$$\frac{(G')^2}{34} = \frac{(238.27)^2}{34}$$
$$= 1669.7821 .$$

For the exact analysis, the error sum of squares from the approximate analysis is retained:

$$E^* = 2.4969.$$

Employing the first term in the partitioned total sum of squares from the augmented data, as indicated by equation (2.26), the total sum of squares is

$$T' = 1709.5133 - 1669.7821$$

$$= 39.7312 \ .$$

In a similar fashion, according to equation (2.27), the sum of squares for blocks, unadjusted, is obtained:

$$B' = \frac{4380.4699}{3} + \frac{452.6024}{2} - 1669.7821$$

$$= 16.6757.$$

Then the unbiased treatment sum of squares, adjusted, is obtained by subtraction:

$$A' = 39.7312 - 2.4969 - 16.6757$$

$$= 20.5586.$$

The exact analysis of variance is shown in Table 5.3.

Table 5.3

Exact Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Blocks (unadjusted)	11	16.6757	1.5160	
Treatments (adjusted)	8	20.5586	2.5698	14.40**
Error	14	2.4969	.1784	
Total	33	39.7312		

VI. SUMMARY

The problem considered in this paper is that of estimating several missing values and analyzing the resulting augmented data in a balanced incomplete block design.

The estimates of missing values are obtained by minimizing the error sum of squares, the procedure first employed by Yates (1933). Explicit formulae for missing value estimates are derived for all cases in which there are not more than two missing values. Formulae for n missing values are given for several particular configurations of the missing observations within the design, and a completely general solution is obtained for n missing values under any conditions. While the latter requires the inversion of a symmetric matrix whose dimension is equal to the number of missing values, it is felt that this procedure involves less effort than the other methods presently available for estimating missing values in a balanced incomplete block design.

It is shown that analysis of data which has been augmented by missing value estimates leads to a positively biased treatment sum of squares. A direct method of analysis is developed which eliminates this bias and precludes the necessity of a tedious bias formula.

It is possible to treat a balanced incomplete block design as a randomized block design with missing values.

Estimates of the missing entries and a randomized block analysis can then be obtained according to the methods of Glenn and Kramer (1958). An example of this procedure is given, and the results are compared with the results obtained by the usual balanced incomplete block analysis.

An example is given illustrating the techniques of missing value estimation and subsequent exact analysis for the balanced incomplete block design.

VII. ACKNOWLEDGEMENTS

The author wishes to express his sincere thanks to the following:

, for his excellent guidance throughout the writing of this paper;

the National Institutes of Health, for making this work financially possible;

, for his administrative aid and encouragement;

, for his encouragement in the field of statistics;

for the excellent final form of this paper.

VIII. BIBLIOGRAPHY

1. Allan, F. E., and Wishart, J. (1930). "A Method of Estimating the Yield of a Missing Plot in Field Experimental Work." Journal of Agricultural Science, Vol. 20, pp. 399 - 406.

2. Cochran, W. G., and Cox, G. M. (1950). Experimental Designs. John Wiley and Sons, Inc., New York.

3. Cornish, E. A. (1940). "The Estimation of Missing Values in Incomplete Randomized Block Experiments." Annals of Eugenics, Vol. 10, pp. 112 - 118.

4. Federer, W. T. (1955). Experimental Design. The Macmillan Company, New York.

5. Glenn, W. A., and Kramer, C. Y. (1958). "Analysis of Variance of a Randomized Block Design with Missing Observations." Applied Statistics, Vol. VII, No. 3, pp. 173 - 185.

6. Wilkinson, G. N. (1958). "Estimation of Missing Values for the Analysis of Incomplete Data." Biometrics, Vol. 14, pp. 257 - 286.

7. Wilkinson, G. N. (1958). "The Analysis of Variance and Derivation of Standard Errors for Incomplete Data." Biometrics, Vol. 14, pp. 360 - 384.

8. Yates, F. (1933). "The Analysis of Replicated Experiments When the Field Results are Incomplete." Empire Journal of Experimental Agriculture, Vol. 1, pp. 129 - 142.

The vita has been removed from the scanned document

ABSTRACT

The problem considered in this paper is that of estimating several missing values and analyzing the resulting augmented data in a balanced incomplete block design.

The estimates are obtained by Yates' procedure of minimizing the error sum of squares.

Explicit formulae are obtained for all cases involving not more than two missing values and for several particular configurations of the missing values within the design. A general solution is obtained which involves the inversion of a symmetric n-square matrix, where n is the number of missing values.

An exact analysis of data augmented by missing value estimates is given which eliminates a positive bias in the treatment sum of squares.

It is possible to treat a balanced incomplete block design as a randomized block design with missing values. Estimates of the missing entries and a randomized block analysis can then be obtained according to the methods of Glenn and Kramer. An example of this procedure is given, and the results are compared with the results obtained by the usual balanced incomplete block analysis.

An example is given illustrating the techniques of missing value estimation and subsequent exact analysis for the balanced incomplete block design.