THE METHOD OF LEAST SQUARES APPLIED TO THE ANALYSIS OF VARIANCE

Thesis for the Degree of M. A.
MICHIGAN STATE COLLEGE

Miriam M. Geboo

1940

THE METHOD OF LEAST SQUARES
APPLIED TO
THE ANALYSIS OF VARIANCE.

by

Miriam Martha Geboo

A THESIS

Submitted to the Graduate School of Michigan
State College of Agriculture and Applied
Science in partial fulfilment of the
requirements for the degree of

MASTER OF ARTS

Department of Mathematics

1940

I wish to express my appreciation to Dr. William Dowell Baten whose suggestions and guidance made this thesis possible.

Table of Contents

Introduction

One of the newest branches of the science of statistics is that of the analysis of variance. This was first introduced by Dr. R. A. Fisher [5] in 1923, and he and many other statisticians have been working constantly since that time to perfect it.

It is a means for segregating from groups of data being compared the variability arising from known sources, leaving an estimate of the experimental error. It can be utilized in testing significance between means of the groups of data.

The analysis of variance is applicable to large or small samples and to a large number of experimental designs. In an analysis of covariance it is also possible to analyze two or more associated variables.

It is the purpose of this thesis to illustrate with various experimental designs how the experimental error used in the analysis of variance is derived. To do this the method of least squares, as suggested by F. Yates [14] and illustrated by him [13], will be used. A second method, called the method of fundamental identities, will be illustrated in the case of Randomized Blocks.

In conclusion a method for finding values of missing plots occurring in an experiment will be suggested.

Notation

The notation used in this paper is simple; however, a reader who is unaccustomed to it will find it complicated unless this explanation is given.

The letter y will be used consistently through the paper to indicate the yield of a plot in the experiment. It could, however, depending on the experiment, be used as a measurement of height, weight, amount of growth, etc. The letter x will be used in the analysis of covariance as the "stand" while y will be used, as before, as the yield.

The subscripts on the y, such as in $y_{ijk....}$, will be explained in the individual chapters.

The general mean of the y's will be denoted by $\bar{y}$.

Summing will be denoted by dots in the subscripts. For example, let us suppose that the data were arranged in n columns and r rows and that y_{ij} indicates the yield occuring in the i th column and the j th row. The sum of the yields of the first row would be denoted by $Y_{.1}$, which indicated that the yields having a 1 in the second place of the subscript--that is, those in the first row--would be totaled. Similarly, $Y_{2.}$ would be the total of the second column. $Y_{..}$ denotes the grand total.

Throughout the thesis

T_i is used to denote treatment i,

B_i is used to denote block i,

R_i is used to denote replication i,

C_i is used to denote column i,

V_i is used to denote "variety" i.

After the first chapter the sum of squares of the residual errors will be denoted by SS, and the degrees of freedom by D. of F.

Each chapter will have equation indexing independent of any other; that is, the first equation in each chapter will be numbered (1). If it becomes necessary at any time to refer to an equation in a previous chapter, for example, chapter 3, equation (1), this will be done by referring to equation (3.1).

Other notation will be described in the individual chapters.

Chapter 1.

Randomized Blocks

One field arrangement that has been found extremely useful and at the same time specially suited to the application of the analysis of variance, is that of randomized blocks. In this design blocks are set up, each of which contains a complete replication of the treatments (or varieties) arranged at random. The number of blocks used depends upon the desired number of replications of each treatment.

As has been stated previously, there are two ways in which the experimental error used in the analysis of variance can be derived algebraically: 1. the method of fundamental identities, and 2. the method of least squares. The first method may be explained as follows:

Consider the randomized block layout shown in Table 1. This is not the arrangement used in the field layout, but that used for computation purposes. Let n be the number of treatments and r the number of blocks.

Table 1.

Blocks	Treatments T_1	T_2	...	T_n	Totals
B_1	y_{11}	y_{21}	...	y_{n1}	$Y_{.1}$
B_2	y_{12}	y_{22}	...	y_{n2}	$Y_{.2}$
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
B_r	y_{1r}	y_{2r}	...	y_{nr}	$Y_{.r}$
Totals	$Y_{1.}$	$Y_{2.}$	...	$Y_{n.}$	$Y_{..}$

The total sum of squares is equal to the sum of squares of the deviations of the variates from the general mean. In symbols this is:

$$\text{Total sum of squares} = \sum_{\substack{i=1\\j=1}}^{\substack{r\\n}} (y_{ij} - \bar{y})^2 ,$$

where $\bar{y}$ is the general mean and y_{ij} indicates that y appears in the i th treatment and the j th block.

Now consider the identity

$$(y_{ij} - \bar{y}) = (\bar{y}_{b_j} - \bar{y}) + (\bar{y}_{t_i} - \bar{y}) + (d) ,$$

where

(1) $d = y_{ij} - \bar{y}_{b_j} - \bar{y}_{t_i} + \bar{y}$,

and $\bar{y}_{b_j}$ is the mean of block j and $\bar{y}_{t_i}$ is the mean of treatment i.

Squaring and summing,

$$\sum_{ij}(y_{ij} - \bar{y})^2 = n\sum_j(\bar{y}_{b_j} - \bar{y})^2 + r\sum_i(\bar{y}_{t_i} - \bar{y})^2 + \sum_{ij}(d)^2$$
$$+ 2\sum(\bar{y}_{b_j} - \bar{y})(\bar{y}_{t_i} - \bar{y}) + 2\sum(\bar{y}_{b_j} - \bar{y})(d)$$
$$+ 2\sum(\bar{y}_{t_i} - \bar{y})(d) \quad .$$

The fourth term of the right hand side of the equation can be shown to equal zero thus:

$$\sum(\bar{y}_{b_j} - \bar{y})(\bar{y}_{t_i} - \bar{y}) = \sum_j\left[\bar{y}_{b_j}\sum_i(\bar{y}_{t_i} - \bar{y}) - \bar{y}\sum_i(\bar{y}_{t_i} - \bar{y})\right]$$
$$= 0 ,$$

since $\sum_i(\bar{y}_{t_i} - \bar{y}) = 0$, and since the mean of the $\bar{y}_{t_i}$'s is equal to $\bar{y}$.

The fifth and sixth terms equal zero likewise. Therefore we have

$$\sum_{ij}(y_{ij} - \bar{y})^2 = n\sum_j(\bar{y}_{b_j} - \bar{y})^2 + r\sum_i(\bar{y}_{t_i} - \bar{y})^2 + \sum_{ij}(d)^2 .$$

The corresponding equation for the degrees of freedom is:

$$nr - 1 = (r - 1) + (n - 1) + (n - 1)(r - 1) .$$

This has been proven in the literature quoted.

In calculating the sums of squares the following formulae are convenient:

(2) Total $\quad \sum_{ij}(y_{ij} - \bar{y})^2 = \sum_{ij} y_{ij}^2 - \dfrac{Y_{..}^2}{nr}$,

(3) Blocks $\quad n\sum_j(\bar{y}_{b_j} - \bar{y})^2 = \dfrac{\sum_j Y_{.j}^2}{n} - \dfrac{Y_{..}^2}{nr}$,

(4) Treatments $k\sum_{i}(\bar{y}_{t_i} - \bar{y})^2 = \frac{\sum_i Y_{i.}^2}{r} - \frac{Y_{..}^2}{nr}$,

(5) Error $\sum_{ij}(d)^2 = (2) - (3) - (4)$.

The analysis of variance table is shown in Table 2.

Table 2.

Source of Variation	Degrees of Freedom	Sums of Squares	Variance SS/D.of F.
Total	nr - 1	(2)	
Blocks	r - 1	(3)	
Treatments	n - 1	(4)	
Error	(r-1)(n-1)	(5)	

The standard error of the experiment is given by

(6) $$S = \sqrt{\frac{\sum(d)^2}{(r-1)(n-1)}} = \sqrt{\text{error variance}}.$$

It is possible to verify the analyses of variance of other designs in this way; however, it is the purpose of this paper to illustrate the second way, that of the method of least squares.

Let t_i be a factor concerning treatment i which affects yields,

b_j be a factor concerning block j which affects yields,

and m be a constant.

We will assume that $\sum_i t_i = 0$, because, while factors concerning some treatments tend to increase the mean yield of a plot above the general mean for the whole experiment, other factors tend to decrease it. Similarly, $\sum_j b_j = 0$.

We will assume that the t_i's, b_j's, and m are coefficients in the following linear equation:

$$(7) \quad y_{uv} = t_1 G_1 + \ldots + t_n G_n + b_1 H_1 + \ldots + b_r H_r + mK + e_{uv} ,$$

where G_i, H_j, and K are variables which take on the following values for y_{ij} :

$G_u = 1$, when $u = i$; $G_u = 0$, when $u \neq i$;

$H_v = 1$, when $v = j$; $H_v = 0$, when $v \neq j$;

$K = 1$, for all u and v.

In equation (7) the e_{uv} represents the variation due to chance.

After substituting values for G_i, H_j, and K, equation (7) reduces to

$$(8) \quad y_{ij} = t_i + b_j + m + e_{ij} ,$$

or

$$e_{ij} = y_{ij} - t_i - b_j - m \quad .$$

The method of least squares is used to find values of t_i, b_j, and m that make the sum of the squares of the residual errors a minimum. This is done by setting the partial derivatives with respect to t_i, b_j, and m equal to zero and solving the resulting simultaneous equations.

Let

$$F = \sum e^2 = \sum_{ij} (y_{ij} - t_i - b_j - m)^2 .$$

Taking the partial derivatives,gives:

$$\frac{\partial F}{\partial t_i} = \sum^{t_i}(-2)(y_{ij} - t_i - b_j - m) = 0 \ ,$$

or

(9) $\qquad Y_{i\cdot} - r\,t_i - r\,m = 0$;

and

$$\frac{\partial F}{\partial b_j} = \sum^{b_j}(-2)(y_{ij} - t_i - b_j - m) = 0 \ ,$$

or

(10) $\qquad Y_{\cdot j} - n\,b_j - n\,m = 0$;

and

$$\frac{\partial F}{\partial m} = \sum_{ij}(-2)(y_{ij} - t_i - b_j - m) = 0 \ ,$$

or

(11) $\qquad Y_{\cdot\cdot} - N\,m = 0,$ where $N = nr$.

Directly from equation (11), it is seen that

(12) $\qquad m = \dfrac{Y_{\cdot\cdot}}{N} = \bar{y}$.

Substituting this value of m in (9) and solving for t_i, gives

(13) $\qquad t_i = \dfrac{Y_{i\cdot}}{r} - \bar{y}$.

Similarly,

(14) $\qquad b_j = \dfrac{Y_{\cdot j}}{n} - \bar{y}$.

Substituting the values from (12), (13), and (14) in (8), and simplifying,

$$y_{ij} = \frac{Y_{i\cdot}}{r} + \frac{Y_{\cdot j}}{n} - \bar{y} + e_{ij} \ ;$$

this becomes, when predicting the value of y on the average,

$$(15) \qquad \tilde{y} = \frac{Y_{i.}}{r} + \frac{Y_{.j}}{n} - \bar{y} .$$

If an equation is of the form given in (7), the standard error of estimate is given by

$$(16) \qquad s = \sqrt{\frac{w}{\text{degrees of freedom}}} ,$$

where

$$w = \Sigma y^2 - t_1 \Sigma y\, G_1 - t_2 \Sigma y\, G_2 - \dots - t_n \Sigma y\, G_n$$

$$- b_1 \Sigma y\, H_1 - b_2 \Sigma y\, H_2 - \dots - b_r \Sigma y\, H_r - m \Sigma y\, K ,$$

and the degrees of freedom in this case are

$$\text{D. of F.} = N - 1 - (n - 1) - (r - 1) .$$

Substituting the values given in (12), (13), and (14) in (16) and simplifying, gives:

$$w = \Sigma y^2 - \sum_i \left(\frac{Y_{i.}}{r} - \bar{y} \right) Y_{i.} - \sum_j \left(\frac{Y_{.j}}{n} - \bar{y} \right) Y_{.j} - (\bar{y}) Y_{..}$$

$$= \left(\Sigma y^2 - \frac{Y_{..}^2}{N} \right) - \left(\frac{\sum_i Y_{i.}^2}{r} - \frac{Y_{..}^2}{N} \right) - \left(\frac{\sum_j Y_{.j}^2}{n} - \frac{Y_{..}^2}{N} \right) .$$

Comparing this final value for w with that indicated in (5), we see that

$$w = \sum_{ij} (d)^2$$

and (16) reduces to

$$s = \sqrt{\frac{\sum_{ij} (d)^2}{\text{D. of F.}}}$$

and is equivalent to that given in (6), as found by the method of fundamental identities.

The standard error of estimate used in this and the following problems is the experimental error which gives rise to the analysis of variance.

Chapter 2.

Single Criterion of Classification

The simplest set-up to which the analysis of variance can be applied is that of a single criterion of classification with either equal or unequal frequencies in the classes. Such a set-up is illustrated in Table 3.

Table 3.

Treatments			
T_1	T_2	...	T_n
y_{11}	y_{21}	...	y_{n1}
y_{12} . . .	y_{22} . . .	...	y_{n2} . . .
y_{1r_1}	y_{2r_2}	...	y_{nr_n}
$Y_{1.}$	$Y_{2.}$	...	$Y_{n.}$

In the general case we will consider n treatments, and the frequencies in Treatment 1, 2, ... , n will be respectively $r_1, r_2, \ldots, r_n$ where the r's may be equal or unequal.

The equation from the fundamental identity is

$$(1)\quad \Sigma y^2 - \frac{Y_{..}^2}{N} = \left(\Sigma \frac{Y_{i.}^2}{r} - \frac{Y_{..}^2}{N}\right) + \Sigma(d)^2 \quad ,$$

where d is an expression similar to that in (1.1), and $N = \sum_i r_i^2$.

Let t_i be a factor concerning treatment i which affects yields,

and m be a constant.

We will assume that the t_i's and m are coefficients in the following linear equation:

(2) $$y_{uv} = t_1E_1 + t_2E_2 + \dots + t_nE_n + m\theta + e_{uv},$$

where $\sum_i t_i = 0$, and where the variables E and θ take on the following values for y :

$E_u = 1$, when $u = i$; $E_u = 0$, when $u \neq i$;

$\theta = 1$, for all u's and v's.

Then (2) reduces to

(3) $$y_{ij} = t_i + m + e_{ij} ,$$

or

$$e_{ij} = y_{ij} - t_i - m .$$

Then let

$$F = \sum e^2 = \sum_{ij}(y_{ij} - t_i - m)^2 .$$

Taking the partial derivatives and setting them equal to zero, gives

$$\frac{\partial F}{\partial t_i} = \sum^{T_i}(-2)(y_{ij} - t_i - m) = 0 ,$$

or

(4) $$Y_{i.} - r_i t_i - r_i m = 0 ;$$

and

$$\frac{\partial F}{\partial m} = \sum_{ij}(-2)(y_{ij} - t_i - m) = 0 ,$$

or

(5) $$Y_{..} - N m = 0 .$$

Directly from (5) we get

(6) $\quad m = \bar{y}$.

Substituting this value of m in (4) and solving for t_i gives

(7) $\quad t_i = \dfrac{Y_{i.}}{r_i} - \bar{y}$.

Substituting the values in (6) and (7) in (3), and simplifying, gives

(8) $\quad y_{ij} = \dfrac{Y_{i.}}{r_i} + e_{ij}$.

If an equation is of the form (2), the standard error of estimate is given by

(9) $\quad s = \sqrt{\dfrac{w}{\text{Degrees of Freedom}}}$

where

$$w = \sum y^2 - t_1 \sum yE_1 - t_2 \sum yE_2 - \dots - t_n \sum yE_n - m \sum yB ,$$

and

Degrees of Freedom $= N - 1 - (n - 1)$

Substituting the values of t_i and m given in (6) and (7) in the equation for w yields

$$w = \sum y^2 - \sum_i \left(\frac{Y_{i.}^{\,2}}{r} - \bar{y} \right) Y_{i.} - \bar{y}\, Y_{..}$$

$$= \left(\sum y^2 - \frac{Y_{..}^{\,2}}{N} \right) - \left(\sum_i \frac{Y_{i.}^{\,2}}{r_i} - \frac{Y_{..}^{\,2}}{N} \right) ,$$

which is essentially the same as that given in (1) if $\sum d^2$ is replaced by w .

The analysis of variance is given in Table 4.

Table 4.

Source of Variation	Degrees of Freedom	Sums of Squares
Total	N - 1	(a) $\sum y^2 - \frac{Y_{..}^2}{N}$
Treatments	n - 1	(b) $\sum_i \frac{Y_{i.}^2}{r_i} - \frac{Y_{..}^2}{N}$
Error	N-1 -(n-1)	(a) - (b)

Chapter 3.

Multiple Criteria of Classification

Some times an experiment is carried out in which there are more than two criteria of classification, say three; in such an experiment, the problem of interactions occurs. For instance, if we have an arrangement of four varieties and three treatments, replicated three times, we have an interaction appearing between varieties and treatments because of the different responses of the different varieties to the same treatment; similarly, we have interactions between varieties and replication, and treatments and replications.

Suppose in the general case we have n treatments, r varieties, and s blocks or replications. The arrangement for computation purposes is illustrated in Table 5.

Table 5.

Varieties	Treatments				Totals
	T_1	T_2	....	T_n	
V_1	B_1 y_{111}	y_{211}	....	y_{n11}	
	⋮ ⋮	⋮		⋮	
	B_s y_{11s}	y_{21s}	....	y_{n1s}	
	$Y_{11.}$	$Y_{21.}$	....	$Y_{n1.}$	$Y_{.1.}$
V_2	B_1 y_{121}	y_{221}	....	$y_{n2,}$	
	⋮ ⋮	⋮		⋮	
	B_s y_{12s}	y_{22s}	....	y_{n2s}	
	$Y_{12.}$	$Y_{22.}$		$Y_{n2.}$	$Y_{.2.}$
⋮	⋮	⋮		⋮	⋮
V_r	B_1 y_{1r1}	y_{2r1}	....	y_{nr1}	
	⋮ ⋮	⋮		⋮	
	B_s y_{1rs}	y_{2rs}	....	y_{nrs}	
	$Y_{1r.}$	$Y_{2r.}$	....	$Y_{nr.}$	$Y_{.r.}$
Totals	$Y_{1..}$	$Y_{2..}$	....	$Y_{n..}$	$Y_{...}$

By y_{ijk} it is indicated that the yield is in the i th treatment, j th variety, and the k th block. The total for Block k $= Y_{..k}$.

Let t_i be a factor concerning treatment i which affects yields,

Let v_j be a factor concerning variety j which affects yields,

b_k be a factor concerning block k which affects yields,

m be a constant,

$(tv)_{ij}$ be a factor concerning the interaction of treatment i and variety j which affects yields,

$(tb)_{ik}$ be a factor concerning the interaction of treatment i and block k which affects yields,

and $(vb)_{jk}$ be a factor concerning the interaction of variety j and block k which affects yields,

where

$$\sum_i t_i = 0, \sum_j v_j = 0, \sum_k b_k = 0, \sum_i (tv)_{ij} = 0,$$

$$\sum_j (tv)_{ij} = 0, \sum_i (tb)_{ik} = 0, \sum_k (tb)_{ik} = 0,$$

$$\sum_j (vb)_{jk} = 0, \sum_k (vb)_{jk} = 0.$$

We will assume that these factors are coefficients in the following linear equation:

$$\begin{aligned}(1) \quad y_{uvw} = {} & t_1 E_1 + \dots + t_n E_n + v_1 F_1 + \dots \quad v_r F_r + b_1 G_1 + \dots \\ & + b_s G_s + m\,H + (tv)_{11} J_{11} + \dots + (tv)_{nr} J_{nr} \\ & + (tb)_{11} K_{11} + \dots + (tb)_{ns} K_{ns} + (vb)_{11} L_{11} + \dots \\ & + (vb)_{rs} L_{rs} + e_{uvw},\end{aligned}$$

where the variables take on the following values for y_{ijk}:

$E_u = 1$, when $u = i$; $E_u = 0$, when $u \neq i$;

$F_v = 1$, when $v = j$; $F_v = 0$, when $v \neq i$;

$G_w = 1$, when $w = k$; $G_w = 0$, when $w \neq k$;

$H = 1$, for all u, v, and w ;

$J_{uv} = 1$, when $u = i$ and $v = j$;

$J_{uv} = 0$, in all other cases ;

$K_{uw} = 1$, when $u = i$ and $w = k$;

$K_{uw} = 0$, in all other cases ;

$L_{vw} = 1$, when $v = j$ and $w = k$;

$L_{vw} = 0$, in all other cases .

Then (1) reduces to

(2) $y_{ijk} = t_i + y_j + b_k + (tv)_{ij} + (tb)_{ik} + (vb)_{jk} + e_{ijk} + m.$

Then

$$F = \sum e^2 = \sum \left[y_{ijk} - t_i - v_j - b_k - m - (tv)_{ij} - (tb)_{ik} - (vb)_{jk}\right]^2.$$

The partial derivatives result in the following equations:

(3) $\frac{\partial F}{\partial t_i}$; $Y_{i..} - rs\, t_i - rs\, m = 0$,

(4) $\frac{\partial F}{\partial v_j}$; $Y_{.j.} - ns\, v_j - ns\, m = 0$,

(5) $\frac{\partial F}{\partial b_k}$; $Y_{..k} - nr\, b_k - nr\, m = 0$,

(6) $\frac{\partial F}{\partial m}$; $Y_{...} - nrs\, m = 0$,

(7) $\frac{\partial F}{\partial (tv)_{ij}}$; $Y_{ij.} - s\, t_i - s\, v_j - s\, m - s\, (tv)_{ij} = 0$,

(8) $\frac{\partial F}{\partial (tb)_{ik}}$; $Y_{i.k} - r\, t_i - r\, b_k - r\, m - r\, (tb)_{ik} = 0$,

(9) $\frac{\partial F}{\partial (vb)}$; $Y_{.jk} - n\, v_j - n\, b_k - n\, m - n\, (vb)_{jk} = 0$.

From (6) we see immediately that

(10) $m = \bar{y}$.

Substituting this value of m in equations (3), (4), and (5) and solving for t_i, v_j, and b_k respectively, we get

(11) $$t_i = \frac{Y_{i..}}{rs} - \bar{y} ,$$

(12) $$v_j = \frac{Y_{.j.}}{ns} - \bar{y} ,$$

(13) $$b_k = \frac{Y_{..k}}{nr} - \bar{y} .$$

Utilizing the values of (10), (11), (12), and (13) in equations (7), (8), and (9), we get

(14) $$(tv)_{ij} = \frac{Y_{ij.}}{s} - \frac{Y_{i..}}{sr} - \frac{Y_{.j.}}{ns} + \bar{y} ,$$

(15) $$(tb)_{ik} = \frac{Y_{i.k}}{r} - \frac{Y_{i..}}{sr} - \frac{Y_{..k}}{nr} + \bar{y} ,$$

(16) $$(vb)_{jk} = \frac{Y_{.jk}}{n} - \frac{Y_{.j.}}{ns} - \frac{Y_{..k}}{nr} + \bar{y} .$$

After substituting values in (10) through (16) in equation (2), we get the following equation for predicting y_{ijk} on the average:

(17) $$\tilde{y}_{ijk} = \frac{Y_{ij.}}{s} + \frac{Y_{i.k}}{n} + \frac{Y_{.jk}}{r} - \frac{Y_{i..}}{sr} - \frac{Y_{.j.}}{ns} - \frac{Y_{..k}}{nr} .$$

The standard error of estimate is given by

(18) $$S = \sqrt{\frac{W}{D.\ of\ F.}} ,$$

where

$$
\begin{aligned}
W = & \left(\sum y_{ijk}^2 - \frac{Y_{...}^2}{nrs}\right) - \left(\frac{\sum Y_{i..}^2}{rs} - \frac{Y_{...}^2}{nrs}\right) \\
& - \left(\frac{\sum Y_{.j.}^2}{ns} - \frac{Y_{...}^2}{nrs}\right) - \left(\frac{\sum Y_{..k}^2}{nr} - \frac{Y_{...}^2}{nrs}\right) \\
& - \left[\left(\frac{\sum Y_{ij.}^2}{s} - \frac{Y_{...}^2}{nrs}\right) - \left(\frac{\sum Y_{i..}^2}{rs} - \frac{Y_{...}^2}{nrs}\right) - \left(\frac{\sum Y_{.j.}^2}{ns} - \frac{Y_{...}^2}{nrs}\right)\right] \\
& - \left[\left(\frac{\sum Y_{i.k}^2}{r} - \frac{Y_{...}^2}{nrs}\right) - \left(\frac{\sum Y_{i..}^2}{rs} - \frac{Y_{...}^2}{nrs}\right) - \left(\frac{\sum Y_{..k}^2}{nr} - \frac{Y_{...}^2}{nrs}\right)\right] \\
& - \left[\left(\frac{\sum Y_{.jk}^2}{n} - \frac{Y_{...}^2}{nrs}\right) - \left(\frac{\sum Y_{.j.}^2}{ns} - \frac{Y_{...}^2}{nrs}\right) - \left(\frac{\sum Y_{..k}^2}{nr} - \frac{Y_{...}^2}{nrs}\right)\right]
\end{aligned}
$$

= (Total Sum of Squares) - (S.S. for Treatments)

- (S.S. for Varieties) - (S.S. for Blocks)

- [S.S. for Interaction TxV]

- [S.S. for Interaction TxB]

- [S.S. for Interaction VxB]

= (i) - (ii) - (iii) - (iv) - (v) - (vi) - (vii) ,

and

$$
\begin{aligned}
\text{D. of F.} = & (nrs - 1) - (n - 1) - (r - 1) - (s - 1) \\
& - (n - 1)(r - 1) - (n - 1)(s - 1) - (r - 1)(s - 1) .
\end{aligned}
$$

Thus we have derived the experimental error used in the analysis of variance of an experiment with three classifications.

The resulting analysis of variance is given in Table 6.

Table 6.

Source of Variation	Degrees of Freedom	Sums of Squares
Total	N - 1*	(i)
Treatments	n - 1	(ii)
Varieties	r - 1	(iii)
Blocks	s - 1	(iv)
Interactions		
TxV	(n - 1)(r - 1)	(v)
TxB	(n - 1)(s - 1)	(vi)
VxB	(r - 1)(s - 1)	(vii)
Error (TxVxB)	(N-1)-(n-1)-(r-1) -(s-1)-(n-1)(r-1) -(n-1)(s-1) -(r-1)(s-1)	(i)-(ii)-(iii) -(iv)-(v)-(vi) -(vii)

* N = nrs .

Chapter 4.

Latin Square

An experimental design which is frequently used is the Latin Square. If n varieties (or treatments) are to be tested, a plot of land is divided into a checkerboard arrangement of n rows and n columns, and the n varieties are distributed at random in the plots, but with the double restriction that each variety appear once and only once in each column and each row.* If V_1, V_2, V_3, V_4, V_5 are five varieties, we can form a Latin Square as shown in Table 7.

Table 7.

Rows	Columns				
	C_1	C_2	C_3	C_4	C_5
R_1	V_3	V_1	V_5	V_4	V_2
R_2	V_4	V_2	V_1	V_5	V_3
R_3	V_1	V_4	V_3	V_2	V_5
R_4	V_2	V_5	V_4	V_3	V_1
R_5	V_5	V_3	V_2	V_1	V_4

If n is the number of rows, columns, and varieties, the fundamental identity for the Latin Square, written in terms of summations used in calculations, is

---

* Rider, P. R., An Introduction to Modern Statistical Methods, pp. 166-9.

$$(1)\quad \sum_{ijk} y_{ijk}^2 - \frac{Y_{...}^2}{n^2} = \left(\frac{\sum_i Y_{i..}^2}{n} - \frac{Y_{...}^2}{n^2} \right) + \left(\frac{\sum_j Y_{.j.}^2}{n} - \frac{Y_{...}^2}{n^2} \right)$$

$$+ \left(\frac{\sum_k Y_{..k}^2}{n} - \frac{Y_{...}^2}{n^2} \right) + \sum d^2 \quad ,$$

where d is an expression similar to that for randomized blocks and y_{ijk} is the yield of the variety k which appears in the i th row and the j th column.

Let v_k be a factor concerning variety k which affects yields ,

r_i be a factor concerning row i which affects yields ,

c_j be a factor concerning column j which affects yields ,

and m be a constant,

where

$$\sum_k v_k = 0 \; , \sum_i r_i = 0 \; , \sum_j c_j = 0 \; .$$

We will assume that the v's, r's, and c's and m are coefficients in the following linear equation:

$$(2)\quad y_{uvw} = v_1 G_1 + \dots + v_n G_n + r_1 H_1 + \dots \; r_n H_n + c_1 J_1$$
$$+ \dots + c_n J_n + m\,K + e_{uvw} \; ,$$

where G_w, H_u, J_v, and K are variables which take on the following values for y_{ijk} :

$G_w = 1$, when $w = k$; $G_w = 0$, when $w \neq k$;

$H_u = 1$, when $u = i$; $H_u = 0$, when $u \neq i$;

$J_v = 1$, when $v = j$; $J_v = 0$, when $v \neq j$;

$K = 1$, for all u, v, and w.

Then (2) reduces to

(3) $y_{ijk} = v_k + r_i + c_j + m + e_{ijk}$.

The expression of the sum of squares of the residual errors is

$$F = \sum e^2 = \sum_{ijk} (y_{ijk} - v_k - c_j - r_i - m)^2 .$$

The partial derivatives result in the following equations:

(4) $\frac{\partial F}{\partial v_k}$; $Y_{..k} - n\,v_k - n\,m = 0$;

(5) $\frac{\partial F}{\partial c_j}$; $Y_{.j.} - n\,c_j - n\,m = 0$;

(6) $\frac{\partial F}{\partial r_i}$; $Y_{i..} - n\,r_i - n\,m = 0$;

(7) $\frac{\partial F}{\partial m}$; $Y_{...} - n^2 m = 0$.

Directly from equation (7) it is seen that

(8) $$m = \frac{Y_{...}}{n^2} = \bar{y} .$$

Substituting this value for m into (4), (5), and (6) and solving for v_k, c_j, and r_i respectively, we get

(9) $$v_k = \frac{Y_{..k}}{n} - \bar{y} ,$$

(10) $$c_j = \frac{Y_{.j.}}{n} - \bar{y} ,$$

and

(11) $$r_i = \frac{Y_{i..}}{n} - \bar{y} ,$$

Now substituting these values into (3) and simplifying we have the following equation for predicting y on the average:

$$(12)\quad y_{ijk} = \frac{Y_{..k}}{n} + \frac{Y_{.j.}}{n} + \frac{Y_{i..}}{n} - 2\,\bar{y} \ .$$

The standard error of estimate is given by

$$(13)\qquad S = \sqrt{\frac{W}{\text{D. of F.}}}\ ,$$

where

$$W = \left(\sum y^2 - \frac{Y_{...}^{\,2}}{n^2}\right) - \left(\sum_k \frac{Y_{..k}^{\,2}}{n} - \frac{Y_{...}^{\,2}}{n^2}\right)$$

$$- \left(\sum_i \frac{Y_{i..}^{\,2}}{n} - \frac{Y_{...}^{\,2}}{n^2}\right) - \left(\sum_j \frac{Y_{.j.}^{\,2}}{n} - \frac{Y_{...}^{\,2}}{n^2}\right)\ ,$$

which is essentially the same as (1) is we let

$$W = \sum d^2 \ .$$

Thus we have derived the experimental error used in the analysis of variance for a Latin Square design.

Chapter 5.

Analysis of Covariance

It is the purpose of the analysis of variance to remove from the experimental error all variability except that due to the chance variation within the factor being tested. This is done sometimes through an attempt to hold other factors constant, or through replications. Sometimes, however, a factor enters into the experiment which can not be held constant. For instance, in an experiment which deals with the yields of a certain variety under various treatments, there may be a variation in the yields caused by unequal stands in addition to the variation caused by the treatments. Also, in an experiment regarding the weights of animals fed different diets, original weight may cause variation which should be considered.

Fisher* and Goulden** have discussed problems of this nature and analyzed them by the analysis of covariance.

When an experiment is analyzed by the analysis of covariance, two separate measurements x and y are made on the items. Now, in the first case above, x would refer to stand and y would refer to the yields, while in the second case x would refer to initial weight

---

* Fisher, R. A., Statistical Methods for Research Workers, Sixth Edition, pp. 275-90.

** Goulden, C. H., Methods of Statistical Analysis, pp. 247-60.

and y the final weight.

Randomized Blocks

Let us consider a randomized block experiment in which measurements were made both on stand and yield. Let n be the number of treatments applied and s be the number of replications. Let x = stand and y = yield. Table 8 illustrates such an experiment.

Table 8.

Repli-cation	Treatments				Totals
	T_1	T_2	. . .	T_n	
R_1	x_{11} y_{11}	x_{21} y_{21}	. . . . . .	x_{n1} y_{n1}	$X_{.1}$ $Y_{.1}$
R_2	x_{12} y_{12}	x_{22} y_{22}	. . . . . .	x_{n2} y_{n2}	$X_{.2}$ $Y_{.2}$
. .	. . . .	. . . .		. . . .	. . . .
R_r	x_{1r} y_{1r}	x_{2r} y_{2r}	. . . . . .	x_{nr} y_{nr}	$X_{.r}$ $Y_{.r}$
Totals	$X_{1.}$ $Y_{1.}$	$X_{2.}$ $Y_{2.}$	. . . . . .	$X_{n.}$ $Y_{n.}$	$X_{..}$ $Y_{..}$

Let t_i be a factor concerning treatment i which affects yields;

r_j be a factor concerning replication j which affects yields;

m be a constant;

and b be the regression coefficient between stand and yield;

where

$$\sum_i t_i = 0 \ , \ \sum_j r_j = 0 \ .$$

Then the equation similar to (1.8) is

(1) $y_{ij} = t_i + r_j + m + b\, x_{ij} + e_{ij}$.

The expression for the sum of squares of the residual errors is

$$F = \sum e^2 = \sum_{ij} (y_{ij} - t_i - r_j - m - b\, x_{ij})^2 \ .$$

Setting the partial derivatives equal to zero and simplifying, we have,

(2) $\frac{\partial F}{\partial t_i}$; $Y_{i.} - r\, t_i - r\, m - b\, X_{i.} = 0$;

(3) $\frac{\partial F}{\partial r_j}$; $Y_{.j} - n\, r_j - n\, m - b\, X_{.j} = 0$;

(4) $\frac{\partial F}{\partial m}$; $Y_{..} - N\, m - b\, X_{..} = 0$, where $N = nr$;

(5) $\frac{\partial F}{\partial b}$; $\sum_{ij} x_{ij} y_{ij} - \sum_i t_i X_{i.} - \sum_j r_j X_{.j} - mX_{..} - b\sum_{ij} x_{ij}^2 = 0.$

By straight forward algebraic manipulation on equations (2), (3), (4), and (5), one can arrive at a solution for b ; it is

(6) $b = \dfrac{\text{Sum of products due to error}}{\text{Sum of x squares due to error}}$,

where

(7) Sum of products due to error =

$$\left(\sum xy - \frac{X_{..}Y_{..}}{N}\right) - \left(\frac{\sum_i X_{i.}Y_{i.}}{r} - \frac{X_{..}Y_{..}}{N}\right) - \left(\frac{\sum_j X_{.j}Y_{.j}}{n} - \frac{X_{..}Y_{..}}{N}\right)$$

and

(8) Sum of x squares due to error =

$$\left(\sum x^2 - \frac{X_{..}^2}{N}\right) - \left(\frac{\sum_i X_{i.}^2}{r} - \frac{X_{..}^2}{N}\right) - \left(\frac{\sum_j X_{.j}^2}{n} - \frac{X_{..}^2}{N}\right).$$

This b is used for adjusting y for stand and has the covariance due to treatment and replication removed from it.

It is possible to solve for t, r, and m in terms of b; by doing so we get

(9) $$t_i = \left(\frac{Y_{i.}}{r} - \bar{Y}\right) - b\left(\frac{X_{i.}}{r} - \bar{X}\right);$$

and

(10) $$r_j = \left(\frac{Y_{.j}}{n} - \bar{Y}\right) - b\left(\frac{X_{.j}}{n} - \bar{X}\right);$$

and

(11) $$m = \bar{Y} - b\bar{X};$$

where $\bar{Y}$ is the general mean of the y's and $\bar{X}$ is the general mean of the x's.

Substituting (9), (10), and (11) in (1), we get

(12) $$y = \left(\frac{Y_{i.}}{r} + \frac{Y_{.j}}{n} - \bar{Y}\right) - b\left(\frac{X_{i.}}{r} + \frac{X_{.j}}{n} - \bar{X}\right) + b\,x_{ij} + e_{ij}.$$

The standard error of estimate is given by

(13) $$S = \sqrt{\frac{W}{\text{D. of F.}}},$$

w = Sum of y squares due to error -

b (sum of products due to error) ,

where

(14) Sum of y squares due to error =

$$\left(\sum y^2 - \frac{Y_{..}^2}{N}\right) - \left(\frac{\sum_i Y_{i.}^2}{r} - \frac{Y_{..}^2}{N}\right) - \left(\frac{\sum Y_{.j}^2}{n} - \frac{Y_{..}^2}{N}\right);$$

and

D. of F. = N - 1 - (n - 1) - (r - 1) - 1 .

Hence we have derived the experimental error which gives rise to the analysis of covariance suggested by Fisher.*

* Fisher, R. A., Statistical Methods for Research Workers, Sixth Edition, pp. 275-90.

Chapter 6.

Analysis of Covariance

Latin Square

Let us now consider an experiment to be analyzed by the analysis of covariance which is set up in an n by n Latin Square. Let x = stand and y = yield. Table 9 illustrates such an experiment.

Table 9.

Rows	Columns C_1	C_2	. .	C_n	Totals
R_1	T_1 x_{111} y_{111}	T_2 x_{212} y_{212}	. . . . . .	T_n x_{n1n} y_{n1n}	$X_{.1.}$ $Y_{.1.}$
R_2	T_n x_{12n} y_{12n}	T_1 x_{221} y_{221}	. . . . . .	T_{n-1} $x_{n2(n-1)}$ $y_{n2(n-1)}$	$X_{.2.}$ $Y_{.2.}$
. . .	. . .	. . .		. . .	. . .
R_n	T_2 x_{1n2} y_{1n2}	T_3 x_{2n3} y_{2n3}	. . . . . .	T_1 x_{nn1} y_{nn1}	$X_{.n.}$ $Y_{.n.}$
Totals	$X_{1..}$ $Y_{1..}$	$X_{2..}$ $Y_{2..}$	. . . .	$X_{n..}$ $Y_{n..}$	$X_{...}$ $Y_{...}$

The sum of the stand for plots with treatment k will be denoted by $\overline{T}_{x_k}$. y_{ijk} denotes the yield of the plot in the i th column and the j th row, where treatment k was applied.

Let c_i be a factor concerning column i which affects yields;

r_j be a factor concerning row j which affects yields;

t_k be a factor concerning treatment k which affects yields;

m be a constant;

and b be the regression coefficient between stand and yield;

where $\sum c_i$, $\sum r_j$, $\sum t_k$ are all equal to zero.

Then we have the following linear equation

(1) $y_{ijk} = c_i + r_j + t_k + m + b\,x_{ijk} + e_{ijk}$.

The expression for the sum of the squares of the residual errors is

$$F = \sum e^2 = \sum_{ijk}(y_{ijk} - c_i - r_j - t_k - m - b\,x_{ijk})^2 .$$

Setting the partial derivatives equal to zero and simplifying, we have

(2) $\frac{\partial F}{\partial c_i}$; $Y_{i..} - n^2 c_i - n\,m - b\,X_{i..} = 0$;

(3) $\frac{\partial F}{\partial r_j}$; $Y_{.j.} - n^2 r_j - n\,m - b\,X_{.j.} = 0$;

(4) $\frac{\partial F}{\partial t_k}$; $\bar{T}_{y_k} - n^2 t_k - n\,m - b\bar{T}_{x_k} = 0$;

(5) $\frac{\partial F}{\partial m}$; $Y_{...} - n^2 m - b\,X_{...} = 0$;

(6) $\frac{\partial F}{\partial b}$; $\sum_{ijk} xy - n\sum_i c_i X_{i..} - n\sum_j r_j X_{.j.} - n\sum_k t_k \bar{T}_{x_k}$

$$- n\,m\,X_{...} - b\sum x^2 = 0 .$$

Again as in the preceding chapter, by a straight forward algebraic manipulation on equations (2), (3), (4), (5), and (6), we can arrive at a solution for b ; it is

(7) $$b = \frac{\text{Sum of products due to error}}{\text{Sum of squares of x's due to error}} ,$$

where

(8) Sum of products due to error =

$$\left(\sum xy - \frac{X_{...}Y_{...}}{N}\right) - \left(\frac{\sum_i X_{i..}Y_{i..}}{n} - \frac{X_{...}Y_{...}}{N}\right)$$

$$-\left(\frac{\sum_j X_{.j.}Y_{.j.}}{n} - \frac{X_{...}Y_{...}}{N}\right) - \left(\frac{\sum_k \overline{T}_{x_k}\overline{T}_{y_k}}{n} - \frac{X_{...}Y_{...}}{N}\right)$$

and

(9) Sum of squares of x's for error =

$$\left(\sum x^2 - \frac{X_{...}^2}{N}\right) - \left(\frac{\sum_i X_{i..}^2}{n} - \frac{X_{...}^2}{N}\right)$$

$$-\left(\frac{\sum_j X_{.j.}^2}{n} - \frac{X_{...}^2}{N}\right) - \left(\frac{\sum_k \overline{T}_{x_k}}{n} - \frac{X_{...}^2}{N}\right) .$$

We can solve equations (2), (3), (4), and (5) for c_i, r_j, t_k, and m respectively in terms of b as follows:

(10) $$c_i = \left(\frac{Y_{i..}}{n} - \overline{Y}\right) - b\left(\frac{X_{i..}}{n} - \overline{X}\right) ;$$

(11) $$r_j = \left(\frac{Y_{.j.}}{n} - \overline{Y}\right) - b\left(\frac{X_{.j.}}{n} - \overline{X}\right) ;$$

(12) $$t_k = \left(\frac{T_{y_k}}{n} - \overline{Y}\right) - b\left(\frac{T_{x_k}}{n} - \overline{X}\right) ;$$

and

(13) $m = \bar{Y} - b\,\bar{X}$;.

Equation (1) becomes, after substitution of values given in (7), (10), (11), (12), and (13) ,

$$(14)\quad y_{ijk} = \frac{Y_{i..}}{n} + \frac{Y_{.j.}}{n} + \frac{\bar{T}_{..k}}{n} - 2\,\bar{Y}$$

$$- b\left(\frac{X_{i..}}{n} + \frac{X_{.j.}}{n} + \frac{\bar{T}_{x_k}}{n} - 2\,\bar{X}\right) + b\,x_{ijk} + e_{ijk}\,.$$

The standard error of estimate is given by

$$(15)\qquad s = \sqrt{\frac{w}{\text{D. of F.}}}$$

where

w = Sum of squares of y's for error

- b (Sum of products for error) ,

where Sum of products for error is given by (8) and the Sum of squares of y's for error is the same as (9) except that x is replaced by y ;

and

D. of F. = N - 1 - 3(n - 1) - 1 .

Hence we have derived the experimental error which gives rise to the analysis of variance for a Latin Square experiment.

Chapter 7.

Incomplete Blocks

If the number of varieties being tested is very large, it often becomes impossible to use the complete randomized block layout because of the amount of field space needed. Yates(14) and Weiss and Cox(10) have developed and extended methods by which a large number of varieties can be tested economically. Two designs which are useful in this respect are: 1. Incomplete blocks, balanced or unbalanced, and 2. Lattice Squares or Quasi-Latin Squares.

The requirement of the balanced incomplete block design is that every variety occur with every other variety in the same number of blocks. Since the number of replications and block size must be kept within practical limits, it is possible to arrange such designs for only specific numbers of varieties. If n is the number of varieties, s is the number of varieties in each block, r is the number of replications of each variety, N or nr or sp is the total number of items, and w is the number of times two varieties appear together, it is seen that certain relationships must exist between these numbers, since they must all be integers. Two of these relationships are

1. $(n - 1)\, w = r\, (s - 1)$,
2. $p = \frac{n\, r}{s}$.

For any value of n and s, there do exist values of w and r which do satisfy the first relation, but it is desirable to keep the number of blocks (p) to a small number or all benefit of such a design is lost.

There are simple devices for deriving the combinations which are possible, yet which retain balance. Yates[14] [redacted]( ) gives some of these arrangements. Structure of balanced incomplete block arrangements is discussed thoroughly by Yates[14], [redacted]( ), and by Goulden*.

Let us consider in connection with the derivation of the experimental error used in the analysis of variance, the special case in which $n = 7$ and $s = 3$. By carrying through this problem, yet keeping in mind the general case, we will arrive at a result which will be applicable to any experiment set up in balanced incomplete blocks.

Let the blocks be set up as indicated in Table 10.

Table 10.

Blocks				Total
B_1	y_{11}	y_{21}	y_{31}	$\bar{B}_1$
B_2	y_{12}	y_{42}	y_{52}	$\bar{B}_2$
B_3	y_{13}	y_{63}	y_{73}	$\bar{B}_3$
B_4	y_{24}	y_{44}	y_{64}	$\bar{B}_4$
B_5	y_{25}	y_{55}	y_{75}	$\bar{B}_5$
B_6	y_{36}	y_{56}	y_{66}	$\bar{B}_6$
B_7	y_{37}	y_{47}	y_{77}	$\bar{B}_7$

In this figure, y_{ij} is the yield of the i th variety in

* Goulden, C. H., <u>Methods of Statistical Analysis</u>, pp. 175-202.

j th block. $\overline{B}_j$ denotes the sum of the yields of the plots in block j . $\overline{V}_i$ denotes the sum of the yields of variety i .

This is a balanced incomplete block design because variety one occurs in a block with each other variety once and only once. Similarly for each other variety.

Let v_i be a factor concerning variety i which affects yields ,

b_j be a factor concerning block j which affects yields ,

and m be a constant ,

where

$$\sum_i v_i = 0 \ ; \quad \sum_j b_j = 0 \ .$$

We will assume that the v's, b's, and m are coefficients in the following linear equation:

$$(1) \quad y_{uv} = v_1 E_1 + \dots + v_n E_n + b_1 F_1 + \dots + b_p F_p + m\,G + e_{uv},$$

where E_u, F_v, and G are variables which take on the following values for y_{ij} :

$E_u = 1$, when $u = i$; $E_u = 0$, when $u \neq i$;

$F_v = 1$, when $v = j$; $F_v = 0$, when $v \neq i$;

$G \ = 1$, for all u and v.

Then (1) reduces to

$$(2) \quad y_{ij} = v_i + b_j + m + e_{ij} \ .$$

Let $F = \sum e^2 = \sum (y_{ij} - v_i - b_j - m)^2$.

Setting the partial derivatives equal to zero and simplifying, we have :

$$(3)\quad \frac{\partial F}{\partial v_i}\ ; \quad \bar{V}_i - r\,v_i - \sum^{V_i} b_j - r\,m = 0\ ;$$

$$(4)\quad \frac{\partial F}{\partial b_j}\ ; \quad \bar{B}_j - \sum^{B_j} v_i - s\,b_j - s\,m = 0\ ;$$

$$(5)\quad \frac{\partial F}{\partial m}\ ; \quad Y_{..} - N\,m = 0\ .$$

Directly from equation (5) we see that

$$(6)\quad m = \bar{y}\ .$$

Now considering the special case indicated above, where $n = 7$ and $s = 3$, and writing out the equations indicated by (3) and (4) , we have

(a) $\bar{V}_1 - r\,v_1 - b_1 - b_2 - b_3 - \bar{y} = 0$,

(b) $\bar{V}_2 - r\,v_2 - b_1 - b_4 - b_5 - \bar{y} = 0$,

(c) $\bar{V}_3 - r\,v_3 - b_1 - b_6 - b_7 - \bar{y} = 0$,

(d) $\bar{V}_4 - r\,v_4 - - b_2 - b_4 - b_7 - \bar{y} = 0$,

(e) $\bar{V}_5 - r\,v_5 - b_2 - b_5 - b_6 - \bar{y} = 0$,

(f) $\bar{V}_6 - r\,v_6 - b_3 - b_4 - b_6 - \bar{y} = 0$,

(g) $\bar{V}_7 - r\,v_7 - b_3 - b_5 - b_7 - \bar{y} = 0$,

(h) $\bar{B}_1 - v_1 - v_2 - v_3 - s\,b_1 - \bar{y} = 0$,

(i) $\bar{B}_2 - v_1 - v_4 - v_5 - s\,b_2 - \bar{y} = 0$,

(j) $\bar{B}_3 - v_1 - v_6 - v_7 - s\,b \ - \bar{y} = 0$,

(k) $\bar{B}_4 - v_2 - v_4 - v_6 - s\,b_4 - \bar{y} = 0$,

(l) $\bar{B}_5 - v_2 - v_5 - v_7 - s\,b_5 - \bar{y} = 0$,

(m) $\bar{B}_6 - v_3 - v_5 - v_6 - s\,b \ - \bar{y} = 0$,

(n) $\bar{B}_7 - v_3 - v_4 - v_7 - s\,b_7 - \bar{y} = 0$.

These, together with (5) are called the normal equations.

Subtracting $[(h) + (i) + (j)] / s$ from (a) and simplifying (making use of the fact that $\sum_i v_i = 0$) we have

$$\overline{V}_1 - r v_1 - \frac{\sum^{V_1} \overline{B}_j}{s} + \frac{(r-1)}{s} v_1 = 0 ,$$

or

$$s\,\overline{V}_1 - \sum^{V_1} \overline{B} - v_1 (rs - r + 1) = 0 .$$

Solving this equation for v_1 we have

$$(7) \qquad v_1 = \frac{s\,\overline{V}_1 - \sum^{V_1} \overline{B}_j}{rs - r + 1} ;$$

in general this becomes

$$(8) \qquad v_i = Q_i \frac{(n-1)}{N(s-1)} ,$$

where

$$(9) \qquad Q_i = s\,\overline{V}_i - \sum^{V_i} \overline{B}_j .$$

Substituting the value for v_i given in (8) into (4) and solving for b_j, we have

$$(10) \qquad b_j = \frac{\overline{B}_j}{s} - \frac{(n-1)}{Ns(s-1)} \sum^{B_j} Q_i - \frac{Y_{\cdot\cdot}}{N} .$$

Equation (2) now becomes

$$(11) \quad y_{ij} = \frac{\overline{B}_j}{s} + Q_i \frac{(n-1)}{N(s-1)} - \frac{(n-1)}{Ns(s-1)} \sum^{B_j} Q_i + e_{ij} .$$

If an equation is of the form (1), the standard error of estimate is given by

(12) $$s = \sqrt{\frac{w}{\text{D. of F.}}}$$

where

$$w = \left(\sum y^2 - \frac{Y_{..}^2}{N}\right) - \left(\frac{\sum_j \bar{B}_j^2}{s} - \frac{Y_{..}^2}{N}\right) - \left(\frac{(n-1)}{Ns(s-1)}\sum Q_i^2\right),$$

and

D. of F. $= N - 1 - (p - 1) - (n - 1)$.

Thus we have derived the experimental error used in the analysis of variance of an incomplete block design. This gives rise to the analysis of variance given by Yates(14). The analysis of variance is illustrated in Table 11.

Table 11

Source of Variation	Degrees of Freedom	Sums of Squares
Total	N - 1	$\sum y^2 - \frac{Y_{..}^2}{N}$ (i)
Between Blocks	p - 1	$\frac{\sum \bar{B}_j^2}{s} - \frac{Y_{..}^2}{N}$ (ii)
Between Varieties	n - 1	$\frac{n-1}{Ns(s-1)} \sum Q_i^2$ (iii)
Error	N-p-n+1	(i) - (ii) - (iii)

This method of balanced incomplete blocks works very well in the case where the number of varieties is equal to s^2 or $s^2 - s + 1$ where s is the number of items in a block. Goulden* discusses in great detail these two special cases.

* Goulden, C.H., Method of Statistical Analysis, p. 175-202.

Chapter 8.

Unbalanced Incomplete Blocks

The problem of incomplete blocks becomes much more complicated if the blocks are not balanced; that is, if each variety does not appear with every other one the same number of times. Goulden* discusses a simple case of this kind in which the number of varieties is equal to p^2. If the varieties are arranged in the form of a square, the blocks may be set up by first considering the rows as blocks and then the columns as blocks. For instance, consider the following nine varieties arranged in the form of a square:

11	12	13
21	22	23
31	32	33

The blocks which may be set up as indicated above are:

Group X

11	12	13
21	22	23
31	32	33

Group Y

11	21	31
12	22	32
13	23	33

These blocks are unbalanced because variety 21 appears only with 11 , 31 , 22 , and 23 . It does not appear with the other four varieties. Similarly for each other variety.

The groups may be replicated as many times as needed to make the desired number of replications.

* Goulden, C. H., Methods of Statistical Analysis, pp.179-185.

Let p be the number of items in each block, then p^2 equals the number of varieties, $2\ np$ equals the number of blocks, 2 equals the number of groups, and n is the number of times each group is replicated.

Then we can set up summary tables in which each x_{ij} and y_{ij} is the sum of n yields of variety ij in the n replications of group X and group Y respectively, and $T_{ij} = x_{ij} + y_{ij}$.

Group X

				Total
B_{1x}	x_{11}	x_{12}	x_{13}	$X_{1.}$
B_{2x}	x_{21}	x_{22}	x_{23}	$X_{2.}$
B_{3x}	x_{31}	x_{32}	x_{33}	$X_{3.}$
	Group Total			$X_{..}$

Group Y

				Total
B_{1y}	y_{11}	y_{21}	y_{31}	$Y_{.1}$
B_{2y}	y_{12}	y_{22}	y_{32}	$Y_{.2}$
B_{3y}	y_{13}	y_{23}	y_{32}	$Y_{.3}$
	Group Total			$Y_{..}$

Variety Summary

T_{11}	T_{12}	T_{13}
T_{21}	T_{22}	T_{23}
T_{31}	T_{32}	T_{33}

The grand total $X_{..} + Y_{..} = T_{..}$.

It is to be noted in this case that the first subscript of the variety indicates the block in the X-group to which it belongs, and the second subscript indicates the block in the Y-group.

Let v_{ij} be the factor concerning variety ij which affects yields;

b_{ix} and b_{jy} be factors concerning blocks ix and jy respectively which affect yields;

s_x and s_y be factors concerning group X and group Y respectively which affect yields;

m be a constant ;

where $\sum_{ij} v_{ij} = 0$; $\sum_i b_{ix} + \sum_j b_{jy} = 0$; $s_x + s_y = 0$.

Instead of one equation as in previous chapters, we will use the following two for the yields of the X group and Y group respectively:

$$(1) \quad \begin{aligned} x_{ij} &= v_{ij} + b_{ix} + s_x + m + e_{ijx} , \\ y_{ij} &= v_{ij} + b_{jy} + s_y + m + e_{ijy} . \end{aligned}$$

For this experiment the sum of the squares of the residual errors is:

$$(2) \quad F = \sum e^2 = \sum_{ij}(x_{ij} - v_{ij} - b_{ix} - s_x - m)^2 + \sum_{ij}(y_{ij} - v_{ij} - b_{jy} - s_y - m)^2 .$$

Setting the partial derivatives equal to zero and simplifying, we have:

$$(3) \quad \frac{\partial F}{\partial v_{ij}} ; \quad T_{ij} - 2n\, v_{ij} - n\, b_{ix} - n\, b_{jy} - 2n\, m = 0 ;$$

$$(4) \quad \frac{\partial F}{\partial b_{ix}} ; \quad X_{i.} - n\sum^{B_{ix}} v_{ij} - np\, b_{ix} - np\, s_x - np\, m = 0 ;$$

$$(5) \quad \frac{\partial F}{\partial b_{jy}} ; \quad Y_{.j} - n\sum^{B_{jy}} v_{ij} - np\, b_{jy} - np\, s_y - np\, m = 0 ;$$

$$(6) \quad \frac{\partial F}{\partial s_x} ; \quad X_{..} - n\, p^2\, s_x - n\, p^2 m = 0 ;$$

$$(7) \quad \frac{\partial F}{\partial s_y} ; \quad Y_{..} - n\, p^2\, s_y - n\, p^2 m = 0 ;$$

(8) $\frac{\partial F}{\partial m}$; $T_{..} - 2\,n\,p^2\,m = 0$.

Solving (8) for m , recalling that $N = 2\,n\,p^2$, we get

(9) $m = M$,

where M is the general mean.

Substituting this value for m in (6) and (7) and solving for s_x and s_y respectively, we get

(10) $$s_x = \frac{X_{..}}{n\,p^2} - M \;;$$

and

(11) $$s_y \quad \frac{Y_{..}}{n\,p^2} - M \;.$$

By straight forward algebraic manipulation the values of the other factors are determined to be:

(12) $$b_{ix} = \frac{X_{i.} - Y_{i.} - 2\,n\,p\,s_x}{n\,p} \;;$$

(13) $$b_{jy} = \frac{Y_{.j} - X_{.j} - 2\,n\,p\,s_y}{n\,p} \;;$$

(14) $$v_{ij} = \frac{p\,T_{ij} - (X_{i.} - Y_{i.}) - (Y_{.j} - X_{.j}) - 2\,n\,p\,M}{2\,n\,p} \;.$$

By substituting the values of equations (9), (10), (11), (12), (13), and (14) in equations (1), we get the following equations for predicting the yields x_{ij} and y_{ij} on the average:

(15) $$\tilde{x}_{ij} = \frac{p\,T_{ij} + (X_{i.} - Y_{i.}) - (Y_{.j} - X_{.j})}{2\,n\,p} - \frac{X_{..}}{np^2} + M \,,$$

and

(16) $$\tilde{y}_{ij} = \frac{p\,T_{ij} - (X_{i.} - Y_{i.}) + (Y_{.j} - X_{.j})}{2\,n\,p} - \frac{Y_{..}}{np} + M \,.$$

The standard error of estimate is given by

(17) $$s = \sqrt{\frac{w}{\text{D. of F.}}} \,,$$

where

w = Total Sum of Squares - Sum of Squares due to varieties - Sum of Squares due to blocks ,

where

(18) $$\text{Total Sum of Squares} = \sum x_{ij}^2 + \sum y_{ij}^2 - \frac{(X_{..} + Y_{..})^2}{N};$$

and

(19) Sum of Squares due to Varieties =

$$\frac{\sum T_{ij}^2}{2n} + \frac{\sum (X_{i.} - Y_{i.})^2}{2\,np} + \frac{\sum (X_{.j} - Y_{.j})^2}{2\,np} - \frac{(X_{..} - Y_{..})^2}{2\,n\,p^2} - \frac{(\sum X_{i.}^2 + \sum Y_{.j}^2)}{np} \,;$$

and

(20) Sum of Squares due to Blocks =

$$\frac{\sum X_{i.}^2 + \sum Y_{.j}^2}{np} - \frac{T_{..}^2}{N} \,;$$

and

$$\text{D. of F.} = 2\,n\,p^2 - (2\,np - 1) - (p^2 - 1) - 1 \,.$$

Thus we have derived the experimental error used in the analysis of variance of an unbalanced incomplete block design. Equation (17) gives rise to the analysis of variance which Goulden* discusses.

---

* Goulden, C. H., Methods of Statistical Analysis, p. 180.

Chapter 9.

Youden's Square

W. J. Youden[17] has modified the method of incomplete blocks in order to eliminate variations due to replications from the error term. Consider the arrangement in Table 12 in which there are seven treatments (A through G) and three plots in each block, the blocks being the vertical rows:

Table 12.

A	B	C	D	E	F	G
D	E	F	G	A	B	C
B	C	D	E	F	G	A

One will notice that there are three replications of each treatment. The modification which Youden introduced was that of placing the seven blocks side by side and arranging the treatments within the blocks in such a way that each of the horizontal rows contains a complete replication of the treatments. Table 12 shows this. Youden also suggests that this can be done for various combinations of treatments and blocks, when certain restrictions are placed on the number of treatments and replications. It will be noted that the number of incomplete (vertical) blocks is equal to the number of treatments (n); that the number of replications (s) of each treatment is equal to the number of items in each block; that n and s are connected by the relation

$$n - 1 = s\,(s - 1) \quad ;$$

that $ns = N$ is the total number of items .

Let y_{ijk} be the yield of a certain plot, where the plot appears in row i , block j , and treatment k .

Let r_i be the factor concerning row i which affects yields ,

b_j be the factor concerning block j which affects yields ,

t_k be the factor concerning treatment k which affects yields ,

and m be a constant ,

where $\sum_i r_i = 0$, $\sum_j b_j = 0$, and $\sum_k t_k = 0$.

We will assume that the r's , b's , t's , and m are coefficients in the following linear equation:

(1) $$y_{uvw} = \sum_i D_i r_i + \sum_j E_j b_j + \sum_k F_k t_k + G\, m + e_{uvw} ,$$

where D_u, E_v, F_w and G are variables which take on the following values for y_{ijk}:

$D_u = 1$, when $u = i$; $D_u = 0$, when $u \neq i$;

$E_v = 1$, when $v = j$; $E_v = 0$, when $v \neq j$;

$F_w = 1$, when $w = k$; $F_w = 0$, when $w \neq k$;

$G = 1$, for all u, v, and w .

Then (1) reduces to

(2) $$y_{ijk} = r_i + b_j + t_k + e_{ijk} .$$

The sum of the squares of the residual errors is given by

$$F = \sum e^2 = \sum (y_{ijk} - r_i - b_j - t_k)^2 .$$

Setting the partial derivatives with respect to the various constants equal to zero, we have

(3) $\frac{\partial F}{\partial r_i}$; $Y_{i..} - n\, r_i - n\, m = 0$;

(4) $\frac{\partial F}{\partial b_j}$; $Y_{.j.} - s\, b_j - \sum^{B_j} t_k - s\, m = 0$;

(5) $\frac{\partial F}{\partial t_k}$; $Y_{..k} - s\, t_k - \sum^{T_k} b_j - s\, m = 0$;

(6) $\frac{\partial F}{\partial m}$; $Y_{...} - N\, m = 0$.

In (4) $\sum^{B_j} t$ denotes the sum of the factors concerning the treatments appearing in block j . In (5) $\sum^{T_k} b_j$ denotes the sum of the factors concerning the blocks in which treatment k appears.

Directly from (6) we see that

(7) $m = \bar{y}$.

Substituting this value of m in (3) and solving for r_i , we get

(8) $r_i = \frac{Y_{i..}}{n} - \bar{y}$.

Using a method similar to that used to find v_i and b_j in Chapter 7, we find

(9) $b_j = \frac{Y_{.j.}}{s} - \frac{\sum^{B_j} Q_k}{s(s^2 - s + 1)} - \bar{y}$,

and

(10) $t_k = \frac{Q_k}{s^2 - s + 1}$,

where

$$Q_k = s\,Y_{..k} - \sum^{T_k} Y_{.j.} \quad .$$

If we substitute the values of equations (7) through (10) into equation (2), we get the following equations for predicting y_{ijk} on the average:

$$(11) \quad \tilde{y}_{ijk} = \frac{Y_{i..}}{n} - \bar{y} + \frac{Y_{.j.}}{s} - \frac{\sum^{B_j} Q_k}{s(s^2 - s + 1)} + \frac{Q_k}{s^2 - s + 1} \quad .$$

The standard error of estimate is

$$(12) \qquad S = \sqrt{\frac{w}{\text{D. of F.}}}$$

where

$$w = \sum y_{ijk}^2 - \frac{Y_{...}^2}{N} - \frac{1}{s(s^2 - s + 1)} \sum_k Q_k^2$$

$$- \left(\frac{\sum_j Y_{.j.}^2}{s} - \frac{Y_{...}^2}{N} \right) - \left(\frac{\sum_i Y_{i..}^2}{n} - \frac{Y_{...}^2}{N} \right) ,$$

and

$$\text{D. of F.} = N - 1 - (s - 1) - 2\,(n - 1) \; .$$

Thus we have derived the experimental error used in the analysis of variance of an experiment set up in a Youden's Square. The analysis of variance which results from this value of the experimental error is given in Table 13.

Table 13.

Source of Variation	Degrees of Freedom	Sums of Squares
Total	sn - 1	$\sum y_{ijk}^2 - \frac{Y_{...}^2}{N}$ (i)
Between Rows	s - 1	$\frac{\sum_i Y_{i..}^2}{n} - \frac{Y_{...}^2}{N}$ (ii)
Between Blocks	n - 1	$\frac{\sum_j Y_{.j.}^2}{s} - \frac{Y_{...}^2}{N}$ (iii)
Between Treatments	n - 1	$\frac{1}{s(s - s \quad 1)} \sum_k Q_k^2$ (iv)
Error	sn-1-(s-1)-2(n-1)	(i)-(ii)-(iii)-(iv)

Chapter 10.

Lattice Squares

The second design for a large number of varieties that was suggested previously is that of Lattice Squares. For this design to be used it is necessary that the number of varieties being tested be a perfect square. In Table 14 an example of such a design is given for 25 varieties.

Table 14.

1	2	3	4	5	1	10	14	18	22	1	8	15	17	24
6	7	8	9	10	20	24	3	7	11	19	21	3	10	12
11	12	13	14	15	23	2	6	15	19	25	2	9	11	18
16	17	18	19	20	12	16	25	4	8	13	20	22	4	6
21	22	23	24	25	9	13	17	21	5	7	14	16	23	5

In this lattice square arrangement every pair of varieties occurs together once only in either a row or a column of any one of the squares. Also, every variety occurs with every other variety once in one column and one row from each square. Complete discussion of this example has been presented by Weiss and Cox[10]. Other examples have been discussed by Fisher and Yates[ ], and Yates[12].

Let y_{ijkp} be the yield of variety j in square i, column k and row p. Let n^2 be the number of varieties and t the number of squares.

Let s_i be a factor concerning square i which affects yields,

v_j be a factor concerning variety j which affects yields,

c_k be a factor concerning column k which affects yields;

r_p be a factor concerning row p which affects yields;

m be a constant;

where
$$\sum_i s_i = 0\ ;\ \sum_j v_j = 0\ ;\ \sum_{s_i} c_k = 0\ ;\ \sum_{s_i} r_p = 0\ .$$

The yield y_{ijkp} is given by

(1) $$y_{ijkp} = s_i + v_j + c_k + r_p + m + e_{ijkp}\ ,$$

and the expression for the sum of the squares of the residual errors is

(2) $$F = \sum e^2 = \sum (y_{ijkp} - s_i - v_j - c_k - r_p - m)^2\ .$$

Setting the partial derivatives of F with respect to the various factors equal to zero, and simplifying, gives the following equations:

(3) $$\frac{\partial F}{\partial s_i}\ ;\quad Y_{i\ldots} - n^2 s_i - n^2 m = 0\ ;$$

(4) $$\frac{\partial F}{\partial v_j}\ ;\quad Y_{.j..} - t\, v_j - \sum^{V_j} c_k - \sum^{V_j} r_p - t\, m = 0\ ;$$

(5) $$\frac{\partial F}{\partial c_k}\ ;\quad Y_{..k.} - n s_{i_k} - \sum^{C_k} v_j - n\, c_k - n\, m = 0\ ;$$

(6) $$\frac{\partial F}{\partial r_p}\ ;\quad Y_{...p} - n s_{i_p} - \sum^{R_p} v_j - n\, r_p - n\, m = 0\ ;$$

(7) $$\frac{\partial F}{\partial m}\ ;\quad Y_{\ldots.} - t\, n^2 m = 0\ ;$$

where s_{i_k} is the factor concerning the square in which column k occurs; similarly, s_{i_p} is the factor concerning the square in which row p occurs. In equation (4) $\sum^{V_j} c_k$ denotes the sum of the columns in which variety j occurs, and $\sum^{V_j} r_p$ denotes the sum of the rows in which variety j occurs. In equation (5) $\sum^{C_k} v_j$ denotes the sum of the varieties occurring in column k. In equation (6) $\sum^{R_p} v_j$ denotes the sum of the varieties occurring in row p.

Solving equation (7) for m results in

(8) $$m = \bar{y} \, .$$

Substituting this value in (3) and solving for s_i:

(9) $$s_i = \frac{Y_{i\cdots}}{n^2} - \bar{y} \, .$$

We can now solve for the other factors in a manner very similar to that in Chapter 7; when we do so, we get

(10) $$v_j = \frac{Q_i + n\,t\,\bar{y}}{n\,t - 2\,t + 1} \, ,$$

(11) $$c_k = \frac{1}{n}\left[Y_{\cdot\cdot k\cdot} - n\,s_{i_k} - \sum^{C_k}\left(\frac{Q_j + n\,t\,\bar{y}}{nt-2t+1}\right) - n\,\bar{y}\right],$$

(12) $$r_p = \frac{1}{n}\left[Y_{\cdot\cdot\cdot p} - n\,s_{i_p} - \sum^{R_p}\left(\frac{Q_j + n\,t\,\bar{y}}{nt-2t+1}\right) - n\,\bar{y}\right].$$

where

$$Q_i = n\,Y_{\cdot j\cdot\cdot} - \sum^{V_j} Y_{\cdot\cdot k\cdot} - \sum^{V_j} Y_{\cdot\cdot\cdot p} \, .$$

The standard error of estimate is given by

(13) $$s = \sqrt{\frac{w}{\text{D. of F.}}}$$

where

w = Total Sum of Squares - Sum of Squares due to squares

- SS due to columns in squares

- SS due to rows in squares - SS due to varieties ,

where

(14) $$\text{Total SS} = \sum y^2 - \frac{Y_{....}^2}{N} ,$$

(15) $$\text{SS for Squares} = \frac{\sum_i Y_{i...}^2}{n^2} - \frac{Y_{....}}{N} ,$$

(16) SS for Columns in Squares =

$$\frac{\sum_k Y_{..k.}^2}{n} - \frac{Y_{....}^2}{N} - \left(\frac{\sum_i Y_{i...}^2}{n^2} - \frac{Y_{....}^2}{N}\right),$$

(17) SS for Rows in Squares =

$$\frac{\sum Y_{...p}^2}{n} - \frac{Y_{....}^2}{N} - \left(\frac{\sum_i Y_{i...}^2}{n^2} - \frac{Y_{....}^2}{N}\right),$$

and

(18) $$\text{SS for Varieties} = \frac{1}{n(nt-2t+1)}\left[\sum (nQ_j)^2 - \frac{(\sum nQ_j)^2}{n^2}\right];$$

and

$$\text{D. of F.} = n^2t - (t-1) - 2(n\,t - t) - (n^2 - 1) - 1 .$$

Thus we have derived the experimental error used in the analysis of variance. This gives rise to the analysis of variance for a Lattice Square which has been discussed by Weiss and Cox[10].

Chapter 11.

Factorial Design

Let us consider an experiment in which we have three kinds of fertilizer, a', b', c'. We may apply the fertilizers one at a time, two at a time, or all together, so that instead of three treatments we have eight, which may be designated by

a'b'c', a'b', a'c', b'c', a', b', c', (1)

where (1) denotes the absence of each fertilizer and is used as a control "treatment". In the field set-up we would have blocks of land with eight plots in each, the treatments scattered at random in each block, with the one condition that each block contain all the treatments.

The analysis of variance could be considered as in a randomized block experiment with eight treatments; however, the effect A of treatment a' can be found by comparing the yields of all plots containing a', with or without any other ingredient, with the yields of plots not containing a' at all. A comparison may also be made of the effect of a' in the presence of b' with that of a' in the absence of b'. This effect we call AB, the simple interaction between a' and b'. Thus we can make seven different comparisons:

A, B, C, AC, AB, BC, ABC.

The first three of these we shall call the <u>main effects</u> and the remainder we shall call the <u>interactions</u>.

We might also consider that the three fertilizers are applied in two ways -- some fertilizer and no fertilizer; we will call these a_1', b_1', c_1' and a_0', b_0', c_0' respectively where the subscript 1 indicates the presence of the fertilizer and the subscript 0 indicates its absence.

For simplicity we will let y_{ijk} be the yield of the plot in which a_i, b_j, c_k were applied; the subscripts of the y indicate the levels on which the a, b, c, were applied respectively. We will let the number of replications be four, and the fourth subscript on the y will indicate the number of the replication.

Let r_p be the factor concerning block p which affects yields ;

a_0, a_1 be factors concerning main effect A which affect yields ;

b_0, b_1 be factors concerning main effect B which affect yields ;

c_0, c_1 be factors concerning main effect C which affects yields ;

(ab) be the factor concerning simple interaction AB which affects yields ;

(ac) be the factor concerning simple interaction AC which affects yields ;

(bc) be the factor concerning simple interaction BC which affects yields ;

(abc) be the factor concerning double interaction ABC which affects yields ;

and m be a constant.

We will assume that these factors are coefficients in the following linear equation:

$$(1)\quad y_{uvwz} = a_0 D_0 + a_1 D_1 + b_0 E_0 + b_1 E_1 + c_0 F_0 + c_1 F_1 + (ab)G + (ac)H + (bc)I + (abc)J + \sum_p b_p L_p + m K + e_{uvwz},$$

where D_u, E_v, F_w, G, H, I, J, L, and K are variables which take on the following values for y_{ijkp}:

$D_u = 1$, when $u = i$; $D_u = 0$ elsewhere ;

$E_v = 1$, when $v = j$; $E_v = 0$ elsewhere ;

$F_w = 1$, when $w = k$; $F_w = 0$ elsewhere ;

$G = 1$, when $uv = 11$ or 00 ;

$G = -1$, when $uv = 10$ or 01 ;

$H = 1$, when $uw = 11$ or 00 ;

$H = -1$, when $uw = 10$ or 01 ;

$I = 1$, when $vw = 11$ or 00 ;

$I = -1$, when $vw = 10$ Or 01 ;

$J = 1$, when $uvw = 111, 100, 010, 001$;

$J = -1$, when $uvw = 110, 101, 011, 000$;

$K = 1$, for all values of u , v , and w ;

$L_z = 1$, when $z = p$; $L_z = 0$ elsewhere .

The values of the variables G, H, I, and J were determined as follows:

Goulden* states that "algebraically, all the treatments can be represented as follows:

$$N = (N_1 - N_0)(K_1 + K_0)(P_1 + P_0) ,$$
$$P = (N_1 + N_0)(K_1 + K_0)(P_1 - P_0) ,$$
$$K = (N_1 + N_0)(K_1 - K_0)(P_1 + P_0) ,$$
$$N \times P = (N_1 - N_0)(K_1 + K_0)(P_1 - P_0) ,$$
$$N \times K = (N_1 - N_0)(K_1 - K_0)(P_1 + P_0) ,$$
$$P \times K = (N_1 + N_0)(K_1 - K_0)(P_1 - P_0) ,$$
$$N \times P \times K = (N_1 - N_0)(K_1 - K_0)(P_1 - P_0) \text{"} ,$$

where N, P, and K are the three treatments corresponding to the A, B, and C in our experiment. 1 in the subscript of the N, P, and K denotes presence of the fertilizer, while 0 denotes absence.

We shall make use of N x P to determine the values for G . This is the interaction which corresponds to AB, the factor concerning which is the coefficient in (1) . In the notation of our experiment N x P is written:

$$A\,B = (a_1 - a_0)(b_1 - b_0)(c_1 + c_0) .$$

Expanding this we have :

$$A\,B = a_1b_1c_1 + a_1b_1c_0 - a_1b_0c_1 - a_1b_0c_0 - a_0b_1c_1$$
$$- a_0b_1c_0 + a_0b_0c_1 + a_0b_0c_0 .$$

The coeffieients of the terms containing a_1b_1 and a_0b_0 are +1 while the coefficients of the terms containing a_1b_0 and a_0b_1 are - 1 . These are the two values of G. H and I are determined similarly.

We shall make use of N x P x K to determine the values of J . In our experiment this relation is written:

* Goulden, C. H., Methods of Statistical Analysis, p.161.

$$A\ B\ C = (a_1 - a_0)(b_1 - b_0)(c_1 - c_0) .$$

Expanding this we have :

$$A\ B\ C = a_1b_1c_1 - a_1b_1c_0 - a_1b_0c_1 + a_1b_0c_0 - a_0b_1c_1$$
$$+ a_0b_1c_0 + a_0b_0c_1 - a_0b_0c_0 .$$

The coefficients of terms $a_1b_1c_1$, $a_1b_0c_0$, $a_0b_1c_0$, and $a_0b_0c_1$ are $+1$ while the coefficients of terms $a_1b_1c_0$, $a_1b_0c_1$, $a_0b_1c_1$, and $a_0b_0c_0$ are -1 . These are the two values of J .

Using the values of these variables (1) reduces to the following equations, where each equation represents four equations since p takes on values from one to four:

(2) $$y_{000p} = a_0 + b_0 + c_0 + (ab) + (ac) + (bc) - (abc)$$
$$+ m + r_p + e_{000p} ;$$

(3) $$y_{001p} = a_0 + b_0 + c_1 + (ab) - (ac) - (bc) + (abc)$$
$$+ m + r_p \quad e_{001p} ;$$

(4) $$y_{010p} = a_0 + b_1 + c_0 - (ab) + (ac) - (bc) + (abc)$$
$$+ m + r_p + e_{010p} ;$$

(5) $$y_{011p} = a_0 + b_1 + c_1 - (ab) - (ac) + (bc) - (abc)$$
$$+ m + r_p + e_{011p} ;$$

(6) $$y_{100p} = a_1 + b_0 + c_0 - (ab) - (ac) + (bc) + (abc)$$
$$+ m + r_p + e_{100p} ;$$

(7) $$y_{101p} = a_1 + b_0 + c_1 - (ab) + (ac) - (bc) - (abc)$$
$$+ m + r_p + e_{101p} ;$$

(8) $$y_{110p} = a_1 + b_1 + c_0 + (ab) - (ac) - (bc) - (abc)$$
$$+ m + r_p + e_{110p} ;$$

(9) $y_{111p} = a_1 + b_1 + c_1 + (ab) + (ac) + (bc) + (abc)$
$+ m + r_p + e_{111p}$.

Now let $F = \sum e_{ijkp}^2$.

Taking the partial derivatives and simplifying we have

(10) $\frac{\partial F}{\partial a_i}$; $Y_{i...} - 16\ a_i - 16\ m = 0$;

(11) $\frac{\partial F}{\partial b_j}$; $Y_{.j..} - 16\ b_j - 16\ m = 0$;

(12) $\frac{\partial F}{\partial c_k}$; $Y_{..k.} - 16\ c_k - 16\ m = 0$;

(13) $\frac{\partial F}{\partial r_p}$; $Y_{...p} - 8\ r_p - 8\ m = 0$;

(14) $\frac{\partial F}{\partial (ab)}$; $(AB) - 32\ (ab) = 0$,

where $(AB) = Y_{000.} + Y_{001.} + Y_{110.} + Y_{111.} - Y_{010.}$
$- Y_{011.} - Y_{100.} - Y_{101.}$;

(15) $\frac{\partial F}{\partial (ac)}$; $(AC) - 32\ (ac) = 0$,

where $(AC) = Y_{000.} + Y_{010.} + Y_{101.} + Y_{111.} - Y_{001.}$
$- Y_{011.} - Y_{100.} - Y_{110.}$;

(16) $\frac{\partial F}{\partial (bc)}$; $(BC) - 32\ (bc) = 0$,

where $(BC) = Y_{000.} + Y_{011.} + Y_{100.} + Y_{111.} - Y_{001.}$
$- Y_{010.} - Y_{101.} - Y_{110.}$;

(17) $\frac{\partial F}{\partial (abc)}$; $(ABC) - 32\ (abc) = 0$,

where $(ABC) = Y_{001.} + Y_{010.} + Y_{100.} + Y_{111.} - Y_{000.}$
$- Y_{011.} - Y_{101.} - Y_{110.}$;

(18) $\frac{\partial F}{\partial m}$; $Y_{....} - 32\,m = 0$.

From (18) we have immediately

(19) $m = \bar{y}$.

Solving the remaining equations for the factors desired, making use of (19), we have :

(20) $a_i = \frac{Y_{i...}}{16} - \bar{y}$;

and

(21) $b_j = \frac{Y_{.j..}}{16} - \bar{y}$;

and

(22) $c_k = \frac{Y_{..k.}}{16} - \bar{y}$;

and

(23) $r_p = \frac{Y_{...p}}{8} - \bar{y}$;

and

(24) $(ab) = \frac{(AB)}{32}$;

and

(25) $(ac) = \frac{(AC)}{32}$;

and

(26) $(bc) = \frac{(BC)}{32}$;

and

(27) $(abc) = \frac{(ABC)}{32}$.;

The standard error of estimate is given by

(28) $S = \sqrt{\frac{W}{D.\ of\ F}}$,

D. of F. = 31 - (4 - 1) - (8 - 1) ;

and

$$w = \sum y^2 - \left(\frac{Y_{0\cdots}}{16} - \bar{y}\right)Y_{0\cdots} - \left(\frac{Y_{1\cdots}}{16} - \bar{y}\right)Y_{1\cdots}$$

$$- \left(\frac{Y_{\cdot 0\cdot\cdot}}{16} - \bar{y}\right)Y_{\cdot 0\cdot\cdot} - \left(\frac{Y_{\cdot 1\cdot\cdot}}{16} - \bar{y}\right)Y_{\cdot 1\cdot\cdot}$$

$$- \left(\frac{Y_{\cdot\cdot 0\cdot}}{16} - \bar{y}\right)Y_{\cdot\cdot 0\cdot} - \left(\frac{Y_{\cdot\cdot 1\cdot}}{16} - \bar{y}\right)Y_{\cdot\cdot 1\cdot}$$

$$- \sum_{p}\left(\frac{Y_{\cdots p}}{8} - \bar{y}\right)Y_{\cdots p} - \frac{(AB)}{32}(AB) - \frac{(AC)}{32}(AC)$$

$$- \frac{(BC)}{32}(BC) - \frac{(ABC)}{32}(ABC) - \bar{y}\,Y_{\cdots\cdot} \;;$$

or

$$w = \left(\sum y^2 - \frac{Y_{\cdots\cdot}^2}{32}\right) - \left(\sum_{p}\frac{Y_{\cdots p}^2}{8} - \frac{Y_{\cdots\cdot}^2}{32}\right)$$

$$- \left[\left(\frac{\sum_i Y_{i\cdots}^2}{16} - \frac{Y_{\cdots\cdot}^2}{32}\right) + \left(\frac{\sum_j Y_{\cdot j\cdot\cdot}^2}{16} - \frac{Y_{\cdots\cdot}^2}{32}\right)\right.$$

$$+ \left(\frac{\sum_k Y_{\cdot\cdot k\cdot}^2}{16} - \frac{Y_{\cdots\cdot}^2}{32}\right) + \frac{(AB)}{32} + \frac{(AC)}{32} + \frac{(BC)}{32}$$

$$\left. + \frac{(ABC)}{32}\right]$$

Total Sum of Squares - Sum of Squares due to Blocks

- Sum of Squares due to Treatments ,

where

Sum of Squares due to Treatments =

SS due to A + SS due to B + SS due to C

+ SS due to AB + SS due to AC + SS due to BC

+ SS due to ABC

= SS due to main effects

+ SS due to simple interactions

+ SS due to double interactions .

Thus we have derived the experimental error which gives rise to the analysis of variance of a factorial experiment. This analysis of variance has been discussed by Yates(15), Fisher*, and Rider** .

---

* Fisher, R. A., <u>The Design of Experiments</u> .

** Rider, P. R., <u>An Introduction to Modern statistical Methods</u>, pp. 169-174.

Chapter 12.

Confounded (3x2x2)

Confounding is a method whereby the necessity of including every treatment combination of a factorial experiment in each block is avoided. This has been discussed in some detail by Yates[15]. By use of this method block size may be kept small even though the number of treatment combinations is large.

The combinations of treatments are divided into two (or more) groups so that the contrasts between the different groups represent some interaction of higher order which is not important in the experiment. In the analysis of variance, information concerning the interaction which is "confounded" is lost but this loss is compensated for in greater accuracy in other comparisons. When an interaction is confounded in only several of the replications, we say that it is partially confounded.

Let us consider a 3 x 2 x 2 experiment in which treatment a is applied in three amounts: 2, 1, 0; treatments b and c in two amounts: 1, 0 . This experiment is one which has been discussed by Yates*. Interactions BC and ABC are partially confounded with block differences. Table 14 shows the design of this experiment.

---

* Yates, F., The Design and Analysis of Factorial Experiments, pp. 58-61.

Table 14.

Blocks

$\overline{I}_a$	$\overline{I}_b$	$\overline{II}_a$	$\overline{II}_b$	$\overline{III}_a$	$\overline{III}_b$
y_{001}	y_{000}	y_{000}	y_{001}	y_{000}	y_{001}
y_{010}	y_{011}	y_{011}	y_{010}	y_{011}	y_{010}
y_{100}	y_{101}	y_{101}	y_{100}	y_{100}	y_{101}
y_{111}	y_{110}	y_{110}	y_{111}	y_{111}	y_{110}
y_{200}	y_{201}	y_{200}	y_{201}	y_{201}	y_{200}
y_{211}	y_{210}	y_{211}	y_{210}	y_{210}	y_{211}

y_{uvw} denotes the yield of a plot on which a was applied on the u level, b on the v level, c on the w level.

We will let the block totals be $[I_a]$, $[I_b]$, etc., and $\{[I_b] - [I_a]\} = g_1$, $\{[II_b] - [II_a]\} = g_2$, $\{[III_b] - [III_a]\} = g_3$. This is the notation which is used by Yates.

Let the factors which affect yield be

for main effect A: a_0, a_1, a_2 ;

B: b_0, b_1 ;

C: c_0, c_1 ;

simple interaction

AB: $(ab)_0$, $(ab)_1$, $(ab)_2$;

AC: $(ac)_0$, $(ac)_1$, $(ac)_2$;

BC: (bc) ;

double interaction

$$ABC: \quad (abc)_0, \ (abc)_1, \ (abc)_2 \ ;$$

replications: $I_a, \ I_b, \ II_a, \ II_b, \ III_a, \ III_b$;

constant: m ;

where the sum of the factors concerning treatments equals zero and the sum of the factors concerning blocks or replications equals zero.

In the factors concerning the simple interaction AB, $(ab)_i$ denotes a contrast between a_i in the presence of b and a_i in the basence of b. Similarly for the factors concerning AC and ABC.

We have the following linear equation for y :

$$(1) \quad y_{uvw} = \sum_u a_u D_u + \sum_v b_v E_v + \sum_w c_w F_w + \sum_u (ab)_u G_u + \sum_u (ac)_u H_u + (bc) J + \sum_u (abc)_u K_u + I_a L_{1a} + I_b L_{1b} + II_a L_{2a} + II_b L_{2b} + III_a L_{3a} + III_b L_{3b} + m P + e_{uvw} .$$

where the variables take on the following values for y_{ijk} :

$D_u = 1$, when $u = i$; $D_u = 0$, when $u \neq i$;

$E_v = 1$, when $v = j$; $E_v = 0$, when $v \neq j$;

$F_w = 1$, when $w = k$; $F_w = 0$, when $w \neq k$;

$G_u = 1$, when $uv = i1$, -1, when $uv = i0$, 0 elsewhere ;

$H_u = 1$, when $uw = i1$, -1, when $uw = i0$, 0 elsewhere ;

$J = 1$, when $vw = 11$ and 00 , -1 when $vw = 01$ and 10 ;

$K_u = 1$, when uvw $=$ 100 and 111, -1 when uvw $=$ 110 and 101, 0 elsewhere ;

$L_{1a} = 1$, when y appears in $\bar{I}_a$, 0 elsewhere ;

similarly for L_{1b}, L_{2a}, L_{2b}, L_{3a}, L_{3b} ;

$P = 1$ everywhere.

These values for the variables were determined in a manner similar to that in the chapter on factorial design.

Using these values for the variables we can get an expression for y_{ijk} in general; this is quite complicated to write down in the general case. However, in a special case, say for y_{111} appearing in block II_a, we would have

$$(2) \quad y_{111} = a_1 + b_1 + c_1 + (ab)_1 + (ac)_1 + (bc) + (abc)_1$$
$$+ II_a + m + e_{111(II_a)} \ .$$

Again we will let F equal the sum of squares of the residual errors. Setting the partial derivatives of F with respect to the various factors equal to zero and simplifying, we have the following equations:

(3) a_i : $\quad Y_{i..} - 12\, a_i - 12\, m = 0$;

(4) b_j : $\quad Y_{.j.} - 18\, b_j - 18\, m = 0$;

(5) c_k : $\quad Y_{..k} - 18\, c_k - 18\, m = 0$;

(6) $(ab)_i$: $\quad Y_{i1.} - Y_{i0.} - 12\,(ab) = 0$;

(7) $(ac)_i$: $\quad Y_{i.1} - Y_{i.0} - 12\,(ac) = 0$;

(8) (bc) : $\quad (BC) - 36\,(bc) + 2\,g_1 + 2\,g_2 + 2\,g_3 = 0$,

where $(BC) = y_{000} + y_{011} + y_{100} + y_{111} + y_{200} + y_{211}$
$- y_{001} - y_{010} - y_{101} - y_{110} - y_{201} - y_{210}$;

(9) $(abc)_0$: $(ABC)_0 - 12\ (bc) - 12\ (abc)_0 - 2\ g_1 + 2\ g_2 + 2\ g_3 = 0$,

where $(ABC)_0 = y_{000} + y_{011} - y_{010} - y_{001}$;

(10) $(abc)_1$: $(ABC)_1 - 12\ (bc) - 12\ (abc)_1 + 2\ g_1 - 2\ g_2 + 2\ g_3 = 0$,

where $(ABC)_1 = y_{100} + y_{111} - y_{110} - y_{101}$;

(11) $(abc)_2$: $(ABC)_2 - 12\ (bc) - 12\ (abc)_2 + 2\ g_1 + 2\ g_2 - 2\ g_3 = 0$,

where $(ABC)_2 = y_{200} + y_{211} - y_{210} - y_{201}$;

(12) I_a : $2\ (bc) - 2\ (abc)_0 + 2\ (abc)_1 + 2\ (abc)_2 + 6\ I_a - [I_a] = 0$;

(13) I_b : $-2\ (bc) + 2\ (abc)_0 - 2\ (abc)_1 - 2\ (abc)_2 + 6\ I_b - [I_b] = 0$;

(14) II_a : $2\ (bc) + 2\ (abc)_0 - 2\ (abc)_1 + 2\ (abc)_2 + 6\ II_a - [II_a] = 0$;

(15) II_b : $-2\ (bc) - 2\ (abc)_0 + 2\ (abc)_1 - 2\ (abc)_2 + 6\ II_b - [II_b] = 0$;

(16) III_a : $2\ (bc) + 2\ (abc)_0 + 2\ (abc)_1 - 2\ (abc)_2 + 6\ III_a - [III_a] = 0$;

(17) III_b : $-2\ (bc) - 2\ (abc)_0 - 2\ (abc)_1 + 2\ (abc)_2 + 6\ III_b - [III_b] = 0$;

(18) m : $Y_{...} - 36\ m = 0$,

from which we see directly that

(19) $m = \bar{y}$.

Knowing this value of m we can solve for the unknown factors in equations (3) to (8); the results are

$$(20) \qquad a_i = \frac{Y_{i..}}{12} - \bar{y} ,$$

$$b_j = \frac{Y_{.j.}}{18} - \bar{y} ,$$

$$c_k = \frac{Y_{..k}}{18} - \bar{y} ,$$

$$(ab)_i = \frac{Y_{i1.} - Y_{i0.}}{12} ,$$

$$(ac)_i = \frac{Y_{i.1} - Y_{i.0}}{12} ,$$

We may solve for (bc) by making the following combination of equations:

(12) - (13) + (14) - (15) + (16) - (17) + 3 (8) ,

making use of the values of g_1, g_2, g_3. Hence we get:

$$(21) \qquad (bc) = \frac{(BC)}{32} - \frac{1}{96}(g_1 + g_2 + g_3) .$$

By a similar process we get:

$$(22) \qquad (abc)_0 = \frac{3}{20}(ABC)_0 + \frac{1}{20}(-g_1 + g_2 + g_3) - \frac{8}{5}(bc) ,$$

$$(23) \qquad (abc)_1 = \frac{3}{20}(ABC)_1 + \frac{1}{20}(g_1 - g_2 + g_3) - \frac{8}{5}(bc) ,$$

$$(24) \qquad (abc)_2 = \frac{3}{20}(ABC)_2 + \frac{1}{20}(g_1 + g_2 - g_3) - \frac{8}{5}(bc) .$$

We will solve for the factors concerning replications in terms of (bc), $(abc)_0$, $(abc)_1$, and $(abc)_2$:

$$(23) \qquad I_a = \frac{[I_a]}{6} - \frac{1}{3}\left[(bc) - (abc)_0 + (abc)_1 + (abc)_2\right] ,$$

$$I_b = \frac{[I_b]}{6} - \frac{1}{3}\Big[(bc) + (abc)_0 - (abc)_1 - (abc)_2\Big],$$

$$II_a = \frac{[II_a]}{6} - \frac{1}{3}\Big[(bc) + (abc)_0 - (abc)_1 + (abc)_2\Big],$$

$$II_b = \frac{[II_b]}{6} - \frac{1}{3}\Big[(bc) - (abc)_0 + (abc)_1 - (abc)_2\Big],$$

$$III_a = \frac{[III_a]}{6} - \frac{1}{3}\Big[(bc) + (abc)_0 + (abc)_1 - (abc)_2\Big],$$

$$III_b = \frac{[III_b]}{6} - \frac{1}{3}\Big[-(bc) - (abc)_0 - (abc)_1 + (abc)_2\Big].$$

The standard error of estimate is given by

$$(24) \qquad S = \sqrt{\frac{w}{\text{D. of F.}}},$$

where

$$w = \text{Total SS} - \text{Block SS} - \left(\frac{\sum_i Y_{i..}^2}{12} - \frac{Y_{...}^2}{36}\right)$$

$$- \left(\frac{\sum_j Y_{.j.}^2}{18} - \frac{Y_{...}^2}{36}\right) - \left(\frac{\sum_k Y_{..k}^2}{18} - \frac{Y_{...}^2}{36}\right)$$

$$- \left(\frac{\sum_i (Y_{i1.} - Y_{i0.})^2}{12}\right) - \left(\frac{\sum_i (Y_{i.1} - Y_{i.0})^2}{12}\right)$$

$$- \frac{(bc)^2}{288} - \frac{\sum_i (abc)_i^2}{60}.$$

and

D. of F. $= 36 - 5 - 11 - 1$.

This is essentially the analysis of variance described by Yates,* and thus we have derived the experimental error for a confounded experiment.

---

* Yates, F., The Design and Analysis of Factorial Experiments, pp. 58-60.

Chapter 13.

Estimation of Missing Plot Values

Sometimes when an experiment is completed, it is seen that a value of the yield in one of the plots is missing. As it would be quite wasteful to discard the entire experiment for this reason, it is better to make an estimation of the value and carry out the analysis of variance using this new value. However, when such an estimation is made, the degrees of freedom for the total must be decreased by the number of estimated values.

Allen and Wishart (1) and Baten (2) have derived formulas for estimating plot values. I am going to show a way, different from theirs, to arrive at the same formulas, and to suggest how it can be extended to any reasonable number of missing plots.

Let us considera randomized block experiment (see Table 1) in which the yield in plot uv is missing. We wish to make an estimation of this value.

From equation (1.15) we know

$$(1) \quad \tilde{y}_{ij} = \frac{Y_{i.}}{r} + \frac{Y_{.j}}{n} - \bar{y} \, ,$$

which is on the average the "best" value of the yield y_{ij} .

In the case when a missing plot occurs in treatment u and block v , $Y_{u.}$ is the sum of the value of the y yield of the missing plot and the sum of the known yields in that treatment; we will denote this sum of known yields by T_{u} . Similarly, $Y_{.v}$ is the sum of the value of the yield of the missing plot and the sum of the known yields in block v which we shall denote by B_v. $Y_{..}$ is the sum of the value of the yield of the missing plot and the sum of the known yields, which we shall denote by T (no subscript).

Substituting these sums in (1) we have

$$(2)\quad y_{uv} = \frac{y_{uv} + T_u}{r} + \frac{y_{uv} + B_v}{n} - \frac{y_{uv} + T}{N} .$$

Solving this equation for y_{uv}, we have

$$(3)\quad y_{uv} = \frac{n\,T_u + r\,B_v - T}{(n-1)(r-1)} .$$

This is essentially the same as that given in the literature cited.

If there are two missing values occuring, we shall assume in two different treatments and two different blocks, the problem becomes one of solving the two following simultaneous equations for y_{uv} and y_{st} :

$$(4)\quad y_{uv} = \frac{y_{uv} + T_u}{r} + \frac{y_{uv} + B_v}{n} - \frac{y_{uv} + y_{st} + T}{N} ,$$

$$(5)\quad y_{st} = \frac{y_{st} + T_s}{r} + \frac{y_{st} + B_t}{n} - \frac{y_{uv} + y_{st} + T}{N} .$$

This gives:

$$(6)\quad y_{uv} = \frac{1}{[(n-1)(r-1)]^2 - 1}\Big[(r-1)(n-1)(n\,T_u + r\,B_v - T)$$

$$- (n\,T_s + r\,B_t - T)\Big] ,$$

and y_{st} is a similar expression.

The reader will readily see that it is possible to solve for any reasonable number of missing plots, no matter where they appear in the layout.

In the case of a Latin Square, equation (4.12) is used. To illustrate this problem let us assume that the values of y_{uvw} and y_{stw} are missing. The values are in different rows and columns, but they are in the same treatment. The resulting simultaneous equations are:

$$(7)\quad y_{uvw} = \frac{y_{uvw} + y_{stw} + T_w}{n} + \frac{y_{uvw} + R_u}{n} + \frac{y_{uvw} + C_v}{n} - 2\left(\frac{y_{uvw} + y_{stw} + T}{N}\right),$$

and

$$(8)\quad y_{stw} = \frac{y_{uvw} + y_{stw} + T_w}{n} + \frac{y_{stw} + R_s}{n} + \frac{y_{stw} + C_t}{n} - 2\left(\frac{y_{uvw} + y_{stw} + T}{N}\right),$$

where T_w is the sum of the known yields in treatment w ; R_u is the sum of the known yields in row u ; C_v is the sum of the known yields in column v ; and T is the grand total of the known yields.

The solution for y_{uvw} is :

$$(9) \quad y_{uvw} = \frac{1}{(n-2)\left[(n-1)^2-1\right]} \Big[(n-1)(n\ T_w + n\ C_v + n\ R_u - 2\ T)$$

$$+ \ (n\ T_w + n\ C_t + n\ R_s - 2\ T) \Big] \ ,$$

and y_{stw} is a similar expression.

Equation (5.12) would be used to estimate the value of missing plots in the analysis of covariance for randomized blocks. Similarly, equations (3.17), (6.14), (7.11), (8.15), and (9.11) may be used in estimating missing plot values for the various experimental designs.

Bibliography

1. Allen, F. E. and Wishart, J. *A method of estimating the yield of a missing plot in field experimental work*. Jour. Agr. Sci., 20:399-406. 1930.
2. Baten, W. D. *Formulas for finding estimates for two and three missing plots in randomized block layouts*. Mich. State College, Tech. Bul. 165. Apr. 1939.
3. Fisher, R. A. *Statistical methods for research workers*. Oliver and Boyd, Londen. 6th Ed. 1936.
4. Fisher, R. A. *The design of experiments*. Oliver and Boyd, Edinburgh. 2nd Ed. 1937.
5. Fisher, R. A. and Mackenzie, W. A. *Studies in crop variation, II. The manurial response of different potato varieties*. Jour. Agr. Sci., 13:311. 1923.
6. Goulden, C. H. *Methods of Statistical analysis*. John Wiley and Sons, New York. 1939.
7. Goulden, C. H. *Modern methods for testing a large number of varieties*. Dominion of Canada, Dept. of Agr., Tech. Bul. 9. 1937.
8. Rider, R. R. *An introduction to modern statistical methods*. John Wiley and Sons, New York. 1939.
9. Snedecor, G. W. *Statistical methods*. Collegiate Press, Ames, Iowa. 1938.
10. Weiss, M. G. and Cox, G. M. *Balanced incomplete*

block and lattice square designs for testing yield differences among large numbers of Soybean varieties. Iowa State College, Research Bul. 257. Apr. 1939.

11. Yates, F. A further note on the arrangement of variety trials: Quasi-Latin squares. Ann. of Eug., 7:319-332. 1936-7.

12. Yates, F. A new method of arranging variety trials involving a large number of varieties. Jour. Agr. Sci., 26:424-55. 1936.

13. Yates, F. Incomplete Latin squares. Jour. Agr. Sci., 26:301-15. 1936.

14. Yates, F. Incomplete randomized blocks. Ann. of Eug., 7:121-140. 1936-7.

15. Yates, F. The design and analysis of factorial experiments.

16. Yates, F. The recovery of interblock information in incomplete blocks and quasi-factorial designs. Ann. of Eug. 1939.

17. Youden, W. J. Use of incomplete block replications in estimating Tobacco-Mosaic virus. Contributions, Boyce-Thompson Institute for Plant Research, 9:41-48. 1937-8.