EXPERIMENTAL FIELD TRIALS

Chairman : Prof. R. A. Fisher

8th January, 1938

[Joint Meeting with the Section of Agriculture of the Indian Science Congress.]

ON ESTIMATING INDIVIDUAL YIELDS IN THE CASE OF MIXED-UP YIELDS OF TWO OR MORE PLOTS IN FIELD EXPERIMENT

By S. S. Bose and P. C. Mahalanobis,
Statistical Laboratory, Calcutta.

In conducting modern field experiments, great care is needed for securing accurate data from all the plots included in the design. Owing to accidental causes, or lack of care on the part of the supervising staff, the yields of one or more plots are, however, sometimes found to be missing. Recently another type of difficulty was brought to our notice, namely the mixing up the yields of two or more plots in an important agricultural experiment in India. What happened was this : the yields from individual plots were stored in small labelled bags which were lying side by side, and two of these were accidentally damaged, and the contents got mixed up. The special feature of the contingency in the present case was that although the individual plots yields were not separately available, the total yield of the two plots was known. It would obviously add to the efficiency of the test if this information could be adequately utilised in estimating the errors of the experiment. Further, unless these yields were reconstructed, the experiment could not be analysed by standard methods.

The general solution of the problem of reconstructing the yields of missing plots was given by F. Yates[1] "by minimising the error variance obtained when unknowns are substituted for the missing yields." The validity of this procedure in the scheme of analysis of variance was rigorously proved by Yates in the same paper. This principle has been used in the present paper to obtain the solution of the problem of reconstructing the yields of plots from a knowledge of the total yield of the plots. The relevant formulae for various cases have been given with numerical illustrations based on agricultural data, while the proof of the mathematical expression has been given in the Appendix by S. S. Bose.

Two Mixed-up Plots not in the same Class

Suppose in a field experiment, there is a multiple classification of the f^{th} order, so that every plot is a member of f groups. Let n_a, n_b ...... n_f be the number of classes in each group, the total number of plots being N. Let A'_1 and A'_2 be the totals in A-group involving the two mixed-up plots exclusive, of course, of the mixed-up plots ; similarly B'_1 and B'_2 the totals in the B-group involving the two mixed-up plots and so on for f groups. If u be the total yield of the two mixed-up plots, then the best estimates of the separate yields of the two plots are given by

$$\frac{u}{2} \pm \frac{n_a (A'_1 - A'_2) + n_b (B'_1 - B'_2) + \ldots\ldots + n_f (F'_1 - F'_2)}{2 (N - n_a - n_b - \ldots\ldots - n_f)} \qquad \ldots \quad (1)$$

The positive sign refers to the plot in class $A'_1, B'_1, \ldots\ldots\ldots F'_1$ and the negative sign to the plot in $A'_2, B'_2, \ldots\ldots\ldots F'_2$.

Randomised Block. In a randomised block experiment, we have two groups, namely block and treatment. Let the yields in a particular randomised block experiment with t treatments and b blocks be represented as follows :—

Blocks →	Treatments →					
	1	2	3	...	t	Total
1	$\underline{a_{11}}$ $(=x)$	a_{12}	a_{13}	...	a_{1t}	B'_1+x
2	a_{21}	$\underline{a_{22}}$ $(=u-x)$	a_{23}	...	a_{2t}	B'_2+u-x
3	a_{31}	a_{32}	a_{33}	...	a_{3t}	B_3
...	...	...	...	...	...	...
b	a_{b1}	a_{b2}	a_{b3}	...	a_{bt}	B_b
Total	T'_1+x	T'_2+u-x	T_3	..	T_t	T

If a_{11} and a_{22} are, say, mixed up to form a total $u = (a_{11} + a_{22})$, the estimated yields of a_{11} and a_{22} are easily obtained from (1) by putting $f = 2$, $n_a = b$ (number of blocks) and $n_t = t$ (number of treatments) and $n_c = n_d = \ldots\ldots\ldots = n_f = 0$.

Here

$$A'_1 = a_{21} + a_{31} + \ldots \ldots + a_{b1} = T'_1$$
$$A'_2 = a_{12} + a_{32} + \ldots \ldots + a_{b2} = T'_2$$
$$B'_1 = a_{12} + a_{13} + \ldots \ldots + a_{1t} = B'_1$$
$$B'_2 = a_{21} + a_{23} + \ldots \ldots + a_{2t} = B'_2$$

Hence, we have the formula for the estimate of

$$a_{11} \rightarrow x = \frac{u}{2} + \frac{t\,(T'_1 - T'_2) + b\,(B'_1 - B'_2)}{2\,(bt - b - t)} \qquad \ldots \quad (2\cdot1)$$

$$a_{22} \rightarrow (u-x) = \frac{u}{2} - \frac{t\,(T'_1 - T'_2) + b\,(B'_1 - B'_2)}{2\,(bt - b - t)} \qquad \ldots \quad (2\cdot2)$$

In Table 1 are shown the results of a randomised block experiment with 4 blocks and 5 treatments.

TABLE 1. YIELD DATA OF A RANDOMISED BLOCK EXPERIMENT

Blocks	Treatments 1	2	3	4	5
I	38·2	—	46·5	46·8	49·5
II	37·7	41·0	45·3	47·4	46·6
III	38·9	42·3	45·0	—	48·7
IV	37·9	41·2	45·6	47·1	49·6

$T'_2 = 124\cdot5$ $B'_1 = 181\cdot0$

$T'_4 = 141\cdot3$ $B'_3 = 174\cdot9$

The yield of Treatment 2 in Block I got mixed up with Treatment 4 in Block III and the total yield of these plots was known to be 92·5.

The total of the known yields in Blocks I and III are respectively 181·0 and 174·9, and the totals of the known yields in Treatments 2 and 4 are respectively 124·5 and 141·3. Hence the yield of Treatment 2 in Block I is given by

$$\frac{92{\cdot}5}{2} + \frac{5\,(124{\cdot}5 - 141{\cdot}3) + 4\,(181{\cdot}0 - 174{\cdot}9)}{2\,(20 - 4 - 5)} = 43{\cdot}54$$

The yield of Treatment 4 in Block III is given by

$$\frac{92{\cdot}5}{2} - \frac{5\,(124{\cdot}5 - 141{\cdot}3) + 4\,(181{\cdot}0 - 174{\cdot}9)}{2\,(20 - 4 - 5)} = 48{\cdot}96$$

This is also easily obtained by subtracting 43·54 from the known sum 92·50 ; but an independent calculation is always desirable in that it gives an arithmetical check. The actual yields in these plots were 43.2 and 49.3 respectively, so that the agreement may be considered fairly satisfactory.

Latin Square. In an $n \times n$ Latin square, there are three groups namely, Rows, Columns, and Treatments. Putting $f=3$, and $n_a = n_b = n_c = n =$ the number of plots in rows, columns and treatments respectively in Equation (1) we have the estimated yields of the mixed-up plots for a Latin Square as follows :

$$\frac{u}{2} \pm \frac{(R'_1 - R'_2) + (C'_1 - C'_2) + (T'_1 - T'_2)}{2\,(n - 3)} \qquad \ldots \quad (3)$$

where R'_1 and R'_2, C'_1 and C'_2 and T'_1 and T'_2 are the totals of the two rows, two columns and two treatments containing the two mixed-up yields.

In Table 2 the results of a 4 × 4 Latin Square experiment are shown with two plot yields mixed up to form a total of 1120.

TABLE 2. YIELD DATA OF A LATIN SQUARE EXPERIMENT
(FOUR VARIATES :—S, C, O and N.)

654 (S)	661 (N)	673 (C)	599 (O)
638 (C)	— (S)	573 (O)	719 (N)
— (O)	581 (C)	639 (N)	752 (S)
557 (N)	479 (O)	591 (S)	669 (C)

$R'_3 = 1972$	$C'_1 = 1849$	$T'_3 = 1651$
$R'_2 = 1930$	$C'_2 = 1721$	$T'_1 = 1997$

By (3), the yield of (O) in column (1) and Row (3)

$$\frac{1120}{2} + \frac{(1972 - 1930) + (1849 - 1721) + (1651 - 1997)}{2 \times 1} = 472$$

Hence the yield of (S) in column (2) and row (2) = 1120 − 472 = 648. The actua yields of these two plots were respectively 499 and 621.

Double Latin square. This is a fairly common design in field experiments. There are two $(n \times n)$ Latin squares in two adjacent blocks with identical treatment sets in each.

Suppose the yield of one plot in one block gets mixed up with the yield of another plot belonging to a different treatment of the second block. Here we have four groups namely, blocks, rows, columns and treatments, so that $f=4$; number of rows = number of columns = $2n$; number of treatments $= n$; number of blocks or Latin squares = 2 . By a slight modification of equation (1) we have

$$\frac{u}{2} \pm \frac{n(R'_1 - R'_2) + n(C'_1 - C'_2) + \tfrac{1}{2} n(T'_1 - T'_2) - (L'_1 - L'_2)}{(2n-1)(n-2)} \qquad \ldots (4)$$

where R'_1 and R'_2 are the two row totals in the two Latin Squares containing the two mixed-up yields, similarly C'_1 and C'_2, T'_1 and T'_2 and L'_1 and L'_2 are the totals of the two columns, two treatments and two blocks containing mixed-up plots.

The results of a double 4×4 Latin square are shown in Table 3. The yields of two corner plots were mixed up to a total of 278.

TABLE 3. YIELD DATA OF A DOUBLE 4×4 LATIN SQUARE EXPERIMEMT. (FOUR VARIATES : — A, B, C AND D.)

I	192 (C)	214 (A)	235 (B)	182 (D)
	228 (A)	182 (C)	173 (D)	232 (B)
	173 (D)	232 (B)	202 (A)	186 (C)
	225 (B)	174 (D)	183 (C)	— (A)
II	137 (B)	74 (C)	112 (A)	— (D)
	123 (A)	75 (D)	79 (C)	128 (B)
	91 (C)	137 (B)	70 (D)	107 (A)
	74 (D)	108 (A)	129 (B)	82 (C)

Here the yield of (A) in column (4) and Row (4) in Block I got mixed up with the yield of D in column (4) and Row (1) of Block II.

We have

	$R'_{4(1)} = 582$	$C'_{4(1)} = 600$	$T'_1 = 1094$	$L'_1 = 3013$
	$R'_{1(2)} = 323$	$C'_{4(2)} = 317$	$T'_4 = 921$	$L'_2 = 1526$
Difference	259	283	173	1487

Hence the yield of (A) in Block I, by (4) is

$$\frac{278}{2} + \frac{4 \times 259 + 4 \times 283 + 2 \times 173 - 1487}{7 \times 2} = 212{\cdot}4$$

The yield of (D) in Block II = 278 − 212·4 = 65·6. The actual yields of these two plots were 215 and 63 respectively.

Two plots in the same Class

Formulæ given in (1) – (4) refer to plots not in the same class, *i.e*, to plots belonging to different blocks, treatments, rows etc. When, however, the mixed up plots belong to the same block or same treatment in a randomised block experiment or the same row or column or treatment in a Latin square scheme, the formulæ for estimating the individual yields take a slightly different form.

Randomised Block. If two plots belonging to different treatments but the same block get mixed up, we have the following formula for the individual yields :

$$\frac{u}{2} \pm \frac{T'_1 - T'_2}{2(b-1)} \qquad \dots \quad (5{\cdot}1)$$

where T'_1 and T'_2 are totals of treatments 1 and 2 containing the mixed up plots and b is the number of blocks.

Similarly, if the plots belong to the same treatment but different blocks, the formula for the individual yields are

$$\frac{u}{2} \pm \frac{(B'_1 - B'_2)}{2(t-1)} \qquad \dots \quad (5{\cdot}2)$$

where B'_1 and B'_2 are the totals of the blocks containing the mixed-up plots and t is the number of treatments. In Table 4 are shown the results of a randomised block experiment in which two plots of Treatment 4 in Blocks I and V got mixed up to form a total of 96.3.

Table 4. Yield Data of a Randomised Block Experiment.

	1	2	3	4
I	42·9	42·5	48·8	—
II	42·8	44·1	46·3	47·3
III	37·7	38·4	45·1	47·0
IV	38·8	39·8	44·5	45·5
V	40·1	40·7	45·8	—
	B'_1 = 134·2		B'_5 = 126·6	

By formula (5.1), the yield of Treatment 4 in Block I is

$$\frac{96{\cdot}3}{2} + \frac{134{\cdot}2 - 126{\cdot}6}{2(4-1)} = 49{\cdot}4$$

and that of Block V $= 96{\cdot}3 - 49{\cdot}4 = 46{\cdot}9$. The actual yields were 50.4 and 45.9 respectively.

Latin Square. If the two mixed-up plots belong to the same column (or row) and, necessarily, different treatments and rows (or columns) the estimated yields are given by

$$\frac{u}{2} \pm \frac{(T'_1 - T'_2) + (R'_1 - R'_2)}{2(n-2)} \qquad \dots \quad (6{\cdot}1)$$

Similarly if they belong to the same treatment and, necessarily, different rows and columns, these estimates are

$$\frac{u}{2} \pm \frac{(R'_1 - R'_2) + (C'_1 - C'_2)}{2(n-2)} \qquad \ldots \quad (6{\cdot}2)$$

where C'_1, C'_2, R'_1, R'_2, T'_1, T'_2 are the totals of the two rows, columns or treatments containing the mixed-up plots one in each.

In Table 5 are shown the yields of a 5×5 Latin Square in which the two plot yields in row 1 got mixed up to form a total 547. The figures are shown minus 275, so that the mixed-up yield = 547 − 2 × 275 = − 3.

TABLE 5. YIELD DATA OF A LATIN SQUARE EXPERIMENT.
FIVE VARIATES :—N, U, M, S AND C.

— (N)	— (U)	−24 (M)	−49 (S)	−37 (F)
19 (S)	−2 (F)	19 (N)	−11 (M)	−10 (U)
41 (M)	−4 (S)	13 (U)	11 (F)	1 (N)
28 (F)	31 (N)	20 (S)	−23 (U)	−11 (M)
31 (U)	25 (M)	23 (F)	31 (N)	−19 (S)

$C'_1 = 119$ $\qquad T'_1 = 82$

$C'_2 = 50$ $\qquad T'_2 = 11$

By formula (6.1) the yield of N in row 1 and column (1)

$$= -\frac{3}{2} + \frac{(82 - 11) + (119 - 50)}{2 \times 3} = 21{\cdot}83$$

The yield of U in column (2) row (1) = − 3 − 21·83 = − 24·83. Adding 275, the estimated yields of the plots are 296·83 and 250·17. The actual yields were 302 and 245 respectively.

Double Latin square. With double Latin squares, it is some times found that the yields of two plots in two blocks but belonging to the same treatment get mixed up, owing to carelessness during harvest. As a matter of fact, two actual cases of mixed up plots that were brought to our notice were of this type.

(i) If two plots are under the same treatment but are located in two blocks, we have the formula

$$\frac{u}{2} \pm \frac{n(R'_1 - R'_2) + n(C'_1 - C'_2) - (L'_1 - L'_2)}{2(n-1)^2} \qquad \ldots \quad (7{\cdot}1)$$

(ii) If the two plots belong to different treatments but are in the same block, the formula is of the form

$$\frac{u}{2} \pm \frac{n\,(R'_1 - R'_2) + n\,(C'_1 - C'_2) + \frac{1}{2}\,n(T'_1 - T'_2)}{2\,(2n - 5)} \quad \ldots \quad (7 \cdot 2)$$

(iii) A third contingency may be that plots belong to the same block and same treatment and hence in different rows and columns. Here the formula reduces to the form

$$\frac{u}{2} \pm \frac{(R'_1 - R'_2) + (C'_1 - C'_2)}{2\,(n - 2)} \quad \ldots \quad (7 \cdot 3)$$

More than Two Mixed-up Plots

The method used in the case of two mixed-up plots may be easily extended to estimate the individual yields when three or more plot yields get mixed up; but in this case, explicit formulæ are not available. It is necessary in this case to construct a set of simultaneous linear equations equal in number to the number of unknown values to be estimated by minimising the error variance expressed in terms of the unknowns, and then to solve the equations. The simultaneous linear equations can be solved either by the method of determinants or by the method of iteration.

In a rice experiment with 5 varieties of rice replicated in 10 randomised blocks, four plot yields having a total of 1379 were supposed to have been mixed up. The results are shown in Table 6.

The four plots that were mixed up were (*i*) Variety 1 in Block I, (*ii*) Variety 1 in Block II, (*iii*) Variety 3 in Block VII and (*iv*) Variety 4 in Block X. Let the yields be denoted by *a*, *b*, *c*, and *d*.

Table 6. Yield Data of a Randomised Block Experiment.

		Varieties				
		1	2	3	4	5
Blocks	I	*a*	343	324	206	275
	II	*b*	352	348	279	365
	III	390	354	324	272	317
	IV	318	320	285	231	315
	V	348	293	316	201	242
	VI	384	301	254	242	243
	VII	452	357	*c*	257	368
	VIII	272	345	320	223	305
	IX	386	324	309	174	305
	X	340	307	309	*d*	303

The linear equations for the solution of a, b, c for this case are as follows:

$$14a + 6b + 7c = 2(B'_1 - B'_{10}) + (T'_1 - T'_4) + 7u = 10{,}236$$
$$6a + 14b + 7c = 2(B'_2 - B'_{10}) + (T'_1 - T'_4) + 7u = 10{,}628$$
$$7a + 7b + 14c = 2(B'_7 - B'_{10}) + (T'_8 - T'_4) + 7u = 10{,}707$$

On solving these equations, we get the reconstructed values shown in Table 7.

TABLE 7, ACTUAL AND CALCULATED VALUES

Varieties	Calculated	Actual
a	366·2	356
b	415·2	388
c	374·0	372
d	223·6	263
	1379·0	1379

AGREEMENT BETWEEN ACTUAL AND ESTIMATED YIELDS.

The agreement between actual and estimated values is shown in Table 8. In each example the residual standard deviation obtained from the complete set of yields has been given for comparison.

TABLE 8. COMPARISON BETWEEN ACTUAL AND ESTIMATED YIELDS.

	Actual	Calculated	Deviation	Residual S. D.	Ratio	5 p.c. level
Table 1	43·20	43·52	0·32	0·90	0·36	2·20
,, 2	499·00	472·00	27·00	19·10	1·41	2·45
,, 3	215·00	212·40	2·60	5·40	0·48	2·04
,, 4	50·40	49·40	1·00	0·94	1·06	2·18
,, 5	302·00	296·83	5·17	10·44	0·50	2·18
,, 6	356·00	366·20	10·20	28·69	0·36	2·03
,, 6	388·00	415·20	27·20		0·95	2·03
,, 6	372·00	374·00	2·00		0·07	2·03
,, 6	263·00	223·60	39·40		1·38	2·03

It will be noticed that no discrepancy exceeded the 5 p. c. level. In fact in 6 out of 9 cases, the ratio of observed deviation to standard deviation was less than 1. The method given in this paper may therefore be expected to give satisfactory estimates within the limits of experimental errors.

Estimate of Residual Variance

One of the main purposes of estimating yields of missing plots or mixed-up plots is to enable the experimenter to calculate the residual variance; in fact the method of minimising residual variance in making estimates of the missing or mixed-up yields ensures an unbiased estimation of this residual variance. Hence, in order to find out the residual variance of an experiment with missing or mixed-up plots, the residual sum of squares is to be divided by its corresponding degrees of freedom, determined by the number of constants utilised in making the estimates. Thus if the number of degrees of freedom of the residual sum of squares when there is no mixing up is n, and m plots have been mixed up with a known total, then the number of degrees of freedom of the sum of squares of the estimated residual variance will be $(n - m + 1)$.

The residual variances calculated from the actual complete sets of yields as well as from the reconstructed sets are shown in Table 9. The figures in brackets give the degrees of freedom.

Table 9. Comparison Between Residual Variance, Calculated from Actual and Estimated Yields.

	Residual Variance				Ratio	5 p.c. Level
	Actual		Estimated			
Table 1	0·8042	(12)	0·8728	(11)	1·08	2·72
,, 2	361·0000	(6)	365·0000	(5)	1·01	4·39
,, 3	29·2000	(15)	30·8600	(14)	1·06	2·42
,, 4	0·8940	(12)	0·8780	(11)	0·98	2·79
,, 5	109·0000	(12)	116·0000	(11)	1·06	2·72
,, 6	823·0000	(36)	904·0000	(32)	1·10	1·80

The z-test shows that the two variances are in all cases statistically indistinguishable. All tests of significance can thus be conducted with confidence with the residual variance calculated from the estimated values.

One word of caution may be added here. The treatment variance obtained from the reconstructed data will be greater than the appropriate value which should be used in applying tests of significance, (and which can be obtained by a method identical to that used by Yates in connexion with missing plot data). If the approximate and larger value is used and the result turns out to be non-significant, it is clear that no further analysis is needed. If, however, the result is on the verge of significance, it will be necessary to calculate the more accurate and smaller value of the treatment variance by appropriate methods.

Reference

1. Yates, F.: The Analysis of Replicated Experiments when the Field Results are Incomplete. *Emp. Jour. Expt. Agri.* Vol. 1 (2), 1933, pp. 129-142.

APPENDIX. THE ESTIMATION OF MIXED-UP YIELDS AND THEIR STANDARD ERRORS

By S. S. BOSE

In the analysis of variance divided into r columns and s rows it is assumed that each observation is an additive function of the means of the row and column in which it is situated, together with an error term, so that,

$$x_{pq} = K + f_{p\cdot} + f_{\cdot q} + e_{pq}$$

where $f_{p\cdot}$ and $f_{\cdot q}$ are constants for the p^{th} column and q^{th} row respectively.

Yates[1] has shown "that the analysis of variance may be regarded as the process of finding the most likely values of the constants K, $f_{1\cdot}$, $f_{2\cdot}$,......$f_{r\cdot}$, $f_{\cdot 1}$, $f_{\cdot 2}$,......$f_{\cdot s}$ and the errors associated with them ; that is the values such that $S\,(e^2_{pq})$ is a minimum".

Following Fisher, he has also shown that when a number of observations are missing, these may be estimated by "minimising the error variance obtained when unknowns are substituted for the missing observations".

A new contingency arises when a number of observations are missing but a linear relation connecting these observations is known. Thus let the p values x_{11}, x_{23}, x_{35}.........x_{pq}, be missing but we know the relation :—

$$a_1\, x_{11} + a_2\, x_{23} + a_3\, x_{35} + \ldots\ldots\ldots\ldots = u$$

which reduces to the simple sum when $a_1 = a_2 = \ldots\ldots\ldots = 1$

$$x_{11} + x_{23} + x_{35} + \ldots\ldots\ldots\ldots = u$$

so that, although the observations are not known individually, their sum (or mean) is known.

Method of Estimating Individual Observations.

Suppose m observations are missing but a linear relation of them of the form

$\sum_{1}^{m} a_m\, x_{mq} = u$ is known.

Denote these observations by u_1, u_2,........ , u_m, so that

$$a_1\, u_1 + a_2\, u_2 + \ldots\ldots\ldots\ldots + a_m\, u_m = u$$

In the table of Analysis of Variance, the marginal means can now be expressed in terms of $u_1, u_2, \ldots\ldots\ldots u_m$.

$$\text{Now } S(e^2_{pq}) = S(x_{pq} - \bar{x}_{p\cdot} - \bar{x}_{\cdot q} + \bar{x})^2$$

where $\bar{x}_{p\cdot}$, $\bar{x}_{\cdot q}$ are functions of $u_1, u_2, \ldots\ldots\ldots u_m$ and the remaining $(N - m)$ are known observations.

Since the linear relation between $u_1, u_2, \ldots\ldots u_m$ is known, we require $(m - 1)$ more independent equations for estimating the values of $u_1, u_2, \ldots\ldots\ldots u_m$. Expressing, say,

$$a_m u_m = u - (a_1 u_1 + a_2 u_2 + \ldots\ldots\ldots + a_{m-1} u_{m-1}),$$

$S(e^2_{pq})$ contains $(m - 1)$ unknown quantities $u_1, u_2, \ldots\ldots u_{m-1}$. Now, differentiating $S(e^2_{pq})$ in turn with respect to $u_1, u_2, \ldots\ldots u_{m-1}$ we get a system of $(m - 1)$ linear equations from which values of $u_1, u_2, \ldots\ldots u_{m-1}$ may be solved uniquely. It is evident that with these estimates (1) the linear relation between $u_1, u_2, \ldots\ldots u_m$ is satisfied, and (2) $S(e^2_{pq})$ = minimum and as such the estimates will allow of unbiased estimates of error variance being made by analysis of variance.

The actual formulæ for estimating the individual observations may now be obtained when the total of them is known. In the first section, cases of two observations will be discussed and in the next section, the general case of any number of observations will be taken up.

Case of Two Observations.

Let us consider a two-way table with m columns and n rows. If the observations X_{rp} and X_{sq} are missing, where $X_{rp} + X_{sq} = u$, we have the marginal totals as shown in the following Table. The incomplete totals are dashed.

Column

Row	1	2	...	p	...	q	...	m	Total
1	x_{11}	x_{12}	...	x_{1p}	...	x_{1q}	...	x_{1m}	R_1
2	x_{21}	x_{22}	...	x_{2p}	...	x_{2q}	...	x_{2m}	R_2
⋮	⋮	⋮		⋮		⋮		⋮	⋮
r	x_{r1}	x_{r2}	...	X_{rp}	...	x_{rq}	...	x_{rm}	$R'_r + X_{rp}$
⋮	⋮	⋮		⋮		⋮		⋮	⋮
s	x_{s1}	x_{s2}	...	x_{sp}	...	$u - X_{rp}$	...	x_{sm}	$R'_s + (u - X_{rp})$
⋮	⋮	⋮		⋮		⋮		⋮	⋮
n	x_{n1}	x_{n2}	...	x_{np}	...	x_{nq}	...	x_{nm}	R_n
Total	C_1	C_2	...	$C'_p + X_{rp}$	...	$C'_q + (u - X_{rp})$	...	C_m	T

Now the expansion for $S(e^2_{pq})$ is given by

$$F = S(e^2_{pq}) = S(x_{pq}-\bar{x})^2 - n\,S(\bar{x}_{.q}-\bar{x})^2 - m\,S(\bar{x}_{p.}-\bar{x})^2$$

$$= S(x^2_{pq}) - \frac{S(C^2_q)}{n} - \frac{S(R^2_p)}{m} + \frac{T^2}{N}$$

$$= [X^2_{rp} + (u-X_{rp})^2] - \frac{1}{n}[(C'_p + X_{rp})^2 + (C'_q + u - X_{rp})^2]$$

$$- \frac{1}{m}[(R'_r + X_{rp})^2 + (R'_s + u - X_{rp})^2]$$

$$+ \text{ terms independent of } X_{rp} \qquad \ldots \quad (1)$$

where $C'_{.q}$ = total of q^{th} column after excluding the yield of the mixed-up plot;

$R'_{p.}$ = total of p^{th} row after excluding the yield of the mixed-up plot,

and T = grand total.

Since $N = m \times n$, we have

$$N\frac{\partial F}{\partial X_{rp}} = 2N[X_{rp}-(u-X_{rp})] - 2m[(C'_p + X_{rp}) - (C'_q + u - X_{rp})]$$

$$- 2n[(R'_r + X_{rp}) - (R'_s + u - X_{rp})] = 0 \qquad \ldots \quad (2)$$

which gives X_{rp} in terms of known observations.

Expression (2) represents the case of two-way classification, namely, rows and columns. We can generalise this into a multiple classification of the f^{th} order. Let a, b, c, ... f be the number of classes in the first, second, ... group the total number of observations being N. Without loss of generality, it may be assumed that the two plots mixed-up belong to class 1 and class 2 respectively of each of the f groups. Take X_{rp} to represent the unknown yield of the first plot and $(u-X_{rp})$ that of the second plot. Let A'_1 and A'_2 be the totals in the A group, B'_1 and B'_2 the totals in the B group and so on for f classes.

The generalised expression for $\frac{\partial F}{\partial X_{rp}}$ will be represented by

$$N\frac{\partial F}{\partial X_{rp}} = 2N[(2X_{rp}-u) - 2a[A'_1 - A'_2) + (2X_{rp}-u)]$$

$$- 2b[(B'_1 - B'_2) + (2X_{rp}-u)] - \ldots\ldots\ldots - 2f[(F'_1 - F'_2) + (2X_{rp}-u)] \qquad \ldots \quad (3)$$

Equating this to 0, we have

$$X_{rp} = \frac{u}{2} + \frac{a(A'_1 - A'_2) + b(B'_1 - B'_2) + \ldots\ldots\ldots + f(F'_1 - F'_2)}{2(N-a-b-c- \ldots\ldots\ldots -f)} \qquad \ldots \quad (4\cdot1)$$

This is the estimate of X_{rp}. The other observation is given by

$$X_{sq} = u - X_{rp} = \frac{u}{2} - \frac{a(A'_1 - A'_2) + b(B'_1 - B'_2) + \ldots\ldots + f(F_1' - F'_2)}{2(N-a-b-c- \ldots\ldots\ldots -f)} \qquad \ldots \quad (4\cdot2)$$

Estimates (4·1) and (4·2) are adequate for the case when the two missing observations are in two different classes of each group, as for example, when they are in different rows and in different columns in a two-way grouping.

If they are in the same row (or column), the row (or column) term will not come in the final equation of estimate.

Thus with two groupings we have our equations for estimating the missing values as follows:

$$x = \frac{u}{2} \pm \frac{a(A_1' - A'_1) + b(B'_1 - B'_2)}{2\,(N - a - b)} \qquad \dots \quad (5{\cdot}1)$$

If the observations are in the same row (*i.e.* B group)

$$x = \frac{u}{2} \pm \frac{a(A'_1 - A'_2)}{2\,(N - a)}$$

$$= \frac{u}{2} \pm \frac{A'_1 - A'_2}{2\,(b - 1)} \text{ since } N = ab \qquad \dots \quad (5{\cdot}2)$$

In certain practical experiments, we have a Latin square arrangement, where n^2 observations are arranged in n rows and n columns. Within each row and column n variates are made to occur each only once, the exact position being selected at random.

This is a case of 3-fold grouping with $a = b = c = n$.
Hence from (4·1) and (4·2)

$$x = \frac{u}{2} \pm \frac{(A'_1 - A'_2) + (B'_1 - B'_2) + (C'_1 - C'_2)}{2\,(n - 3)} \qquad \dots \quad (6{\cdot}1)$$

Again, if the observations are both in the same C group, we have omitting C'_1 and C'_2 in (4·1) and (4·2)

$$x = \frac{u}{2} \pm \frac{(A'_1 - A'_2) + (B'_1 - B'_2)}{2\,(n - 2)} \qquad \dots \quad (6{\cdot}2)$$

More than Two Observations Missing.

With more than two observations missing, the estimates may be obtained by an identical procedure. The expression for $F = S(e^2_{pq})$ now contains $m-1$ unknowns if m observations are missing but are connected by a known linear relation. If these are $X_1, X_2, \dots\dots X_{m-1}$,

$$\frac{\partial F}{\partial X_1}, \frac{\partial F}{\partial X_2}, \dots\dots\dots \frac{\partial F}{\partial X_{m-1}}$$

give $(m-1)$ linear equations from which $X_1, X_2, \dots\dots X_{m-1}$ may be readily solved either by determinants or iteration or by straightforward step by step elimination.

It may be pointed out that with more and more missing values, the degrees of freedom available for estimating the residual variance are necessarily diminished and the efficiency of the test is thereby reduced.

THE STANDARD ERROR OF TREATMENT MEANS.

If all the data are available, the standard error of treatment means is given by $S_o/\sqrt{b}$ where S_o = residual standard deviation and b = number of replications (or blocks). When, however, some of the data are missing, but can be estimated as a linear function of the other known data, the means of treatments affected by the missing data will be linear expressions involving more plot values than would be the case if the data were complete.

If $y = a_1 x_1 + a_2 x_2 + \dots\dots\dots + a_p x_p$ where $x_1, x_2, \dots\dots\dots x_p$ are uncorrelated, the variance of y is given by

$$V(y) = (a_1^2 + a_2^2 + \dots\dots\dots\dots + a_p^2)\, V(x) = \Sigma\, a^2.\, \sigma^2$$

$$\text{where } V(x_1) = V(x_2) = \dots\dots\dots\dots \dots = V(x_p) = \sigma^2$$

Let us first take the case of a randomised block experiment in which two plot yields say a_{11} and a_{22} have been mixed up.

$$\text{Then } a_{11} \rightarrow \frac{u}{2} + \frac{t(T'_1 - T'_2) + b(B'_1 - B'_2)}{2(bt - b - t)} \quad \dots \quad (4{\cdot}1)$$

$$a_{22} \rightarrow \frac{u}{2} - \frac{t(T'_1 - T'_2) + b(B'_1 - B'_2)}{2(bt - b - t)} \quad \dots \quad (4{\cdot}2)$$

The total of treatment (1)

$$T_1 = T'_1\left(1 + \frac{t}{n}\right) - T'_2\left(\frac{t}{n}\right) + \frac{b}{n}(B'_1 - B'_2) + \frac{u}{2}$$

where $\quad n = 2(bt - b - t)$

If σ^2 = residual variance of a single plot,

$$V(T_1) = \sigma^2\left[\left(1 + \frac{t-b}{n}\right)^2 + \left(\frac{t-b}{n}\right)^2 + (b-2)\left(1 + \frac{t}{n}\right)^2\right.$$

$$\left. + (b-2)\frac{t^2}{n^2} + 2(t-2)\frac{b^2}{n^2} + \tfrac{1}{2}\right]$$

$$= b\,\sigma^2\left[1 + \frac{t}{2(bt - b - t)}\right]$$

The variance of T_2 can be derived in the same way but obviously this is the same as that of T_1.

Latin square. Let the scheme of the arrangement be a Latin square as follows :—

Let $a_{11 \cdot 1}$ and $a_{22 \cdot 2}$ be the two mixed up yields. Here $x_{31 \cdot 2}$ represents yield of the plot in column 3, row 1 which gets treatment 2.

We have $T'_1 = x_{23 \cdot 1} + x_{32 \cdot 1} + \ldots\ldots (n - 1)$ terms.

The total of Treatment (1) is given by

$$T_1 = T'_1 + a_{11}$$

$$= T'_1 \left(1 + \frac{1}{k} \right) - T'_2 . \frac{1}{k} + \frac{R'_1 - R'_2}{k} + \frac{C'_1 - C'_2}{k} + \frac{u}{2}$$

where $k = 2 (n - 3)$

If σ^2 is the residual variance per plot yield,

$$V(T_1) = \sigma^2 \left[2 + (n-3) \{ (1 + \frac{1}{k})^2 + \frac{5}{k^2} + \tfrac{1}{2} \} \right]$$

$$= n \sigma^2 \left[1 + \frac{1}{2(n-3)} \right] \qquad \ldots (6 \cdot 1)$$

The variance of T_2 is also identical.

Sometimes the two plots that are mixed up belong to the same block or treatment ; or if the design is a Latin square, the two plots may belong to the same row, column or treatment. The standard errors of these cases may be calculated as follows :—

(*i*) If the two plots belong to the same block but different treatments, the standard error of the mean of Treatment 1 or 2 is given by

$$\frac{\sigma}{\sqrt{b}} \left[1 + \frac{1}{2(b-1)} \right]^{\frac{1}{2}} \qquad \ldots (7 \cdot 0)$$

(*ii*) If the two plots belong to the same column (or row) but different treatments in a Latin square design, the standard error of the mean of Treatment 1 or 2 is.

$$\frac{\sigma}{\sqrt{n}} \left[1 + \frac{1}{2(n-2)} \right]^{\frac{1}{2}} \qquad \ldots (8 \cdot 0)$$

(*iii*) If the mixed-up plots belong to the same treatment, the treatment yields are known completely from observed values and the standard error $= \sigma / \sqrt{n}$ where the treatments are replicated n times.

If the experiment is in the form of a Double Latin square, the estimated yield is given by

$$a = \frac{u}{2} \pm \frac{n\,(R'_1 - R'_2) + n\,(C'_1 - C'_2) + \frac{1}{2}n\,(T'_1 - T'_2) - (L'_1 - L'_2)}{(2n - 1)\,(n - 2)}$$

where the two mixed-up yields belong to different treatments and to different blocks. Then

$$T_1 = T'_1 + \frac{u}{2} + \frac{n(R'_1 - R'_2) + n\,(C'_1 - C'_2) + \frac{1}{2}n\,(T'_1 - T'_2) - (L'_1 - L'_2)}{(2n - 1)\,(n - 2)}$$

$$= T'_1 \left(1 + \frac{n}{2k}\right) - T'_2\,\frac{n}{2k} + (R' - R'_2)\,\frac{n}{k} + (C'_1 - C'_2)\,\frac{n}{k} - \frac{L'_1 - L'_2}{k} + \frac{u}{2}$$

where R'_1, C'_1, L'_1 refer to the row, column and block of the first Latin square containing one of the mixed-up plots $a_{11\cdot 1}$, R'_2, C'_2, L'_2 refer to the row, column and block of the second Latin square containing the other plot $a_{22\cdot 2}$ and $k = (2n - 1)\,(n - 2)$

Proceeding as before, we get the variance of the total of Treatment 1 :

$$V\,(T_1) = 2n\,\sigma^2 \left[1 + \frac{n}{(2n - 1)\,(n-2)}\right] \ldots\ldots\ldots\ldots \quad \ldots \quad 9{\cdot}0$$

LOSS OF EFFICIENCY.

When two or more plot yields have been mixed up, there is a loss of efficiency of the experiment owing to the loss of information regarding the yield. The variances of a treatment-mean including one of the mixed-up plots and excluding any such plot are shown in Table 1 for various types of replications.

TABLE 1. VARIANCE OF TREATMENT MEANS.

	Including Mixed Plot	Without Mixed Plot
Randomised Block	$\frac{\sigma^2}{b}\left\{1 + \frac{t}{2(bt - b - t)}\right\}$	$\frac{\sigma^2}{b}$
Latin Square	$\frac{\sigma^2}{n}\left\{1 + \frac{1}{2(n-3)}\right\}$	$\frac{\sigma^2}{n}$
Double Latin Square	$\frac{\sigma^2}{2n}\left\{1 + \frac{n}{(2n-1)(n-2)}\right\}$	$\frac{\sigma^2}{2n}$

It is possible to obtain an exact estimate of the loss of information due to the mixing up of the plots. If a is the loss in replication of the treatments including mixed-up plots, we may equate the variances of mean of the affected treatments in a Latin square experiment as follows:

$$\frac{\sigma^2}{n-a} = \frac{\sigma^2}{n}\left[1 + \frac{1}{2(n-3)}\right]$$

$$\text{or}, \; a = \frac{n}{2n-5}$$

The values of a for different values of n are shown in Table 2.

TABLE 2. VALUE OF 'a' FOR DIFFERENT VALUES OF n

$n =$	4	5	10	20	∞
$a =$	1·333	1·000	0·667	0·571	0·500

Thus, the mixing up of one plot of one treatment with another plot of a second treatment results in a loss of $1\frac{1}{3}$ replication for each of the affected treatments when the experiment is in the form of a 4×4 Latin square. The loss is exactly 1 with 5 × 5 arrangement and as n increases, the loss gets lesser and lesser in magnitude till with infinite replication, the loss of replications is exactly $\frac{1}{2}$ for each treatment.

In the case of randomised blocks,

$$a = \frac{bt}{2\,bt - 2b - t}$$

the values of a for different values of b and t are shown in Table 3.

TABLE 3. VALUES OF 'a' FOR DIFFERENT VALUES OF b and t.

b	t=3	t=4	t=5	t=10	t=∞
4	0·9231	0·8000	0·7408	0·6452	0·5714
5	·8824	·7692	·7143	·6250	·5556
10	·8108	·7143	·6667	·5882	·5263
20	·7792	·6897	·6452	·5714	·5128
∞	·7500	·6667	·6250	·5556	·5000

For a given value of t, a diminishes as b is increased; when $b \rightarrow \infty$, $a = \frac{t}{2(t-1)}$, and when b and t both tend to infinity, $a = \frac{1}{2}$.

In practice it is sometimes very laborious to calculate the appropriate standard errors; but it is possible to form estimates of the upper and lower limits of the error. The upper limit is obtained by omitting the two replications containing the mixed-up plots. Thus for a randomised block, we can show that

$$\frac{\sigma^2}{b-2} > \frac{\sigma^2}{b}\left[1 + \frac{t}{2(bt - b - t)}\right]$$

if $\quad b > \frac{2t}{3t-4} > 2$, where $t = 2$ or more, and for a Latin square

$$\frac{\sigma^2}{n-2} > \frac{\sigma^2}{n}\left[1 + \frac{1}{2(n-3)}\right], \text{ if } n > 4$$

i.e., the experiment is rendered more precise by including the estimated yields than by rejecting the replications containing the mixed-up plots.

The lower limit is obtained by deducting $\frac{1}{2}$ from the number of replications. That this is the lower limit can be shown by the following inequalities.

For randomised blocks

$$\frac{\sigma^2}{b-\frac{1}{2}} < \frac{\sigma^2}{b}\left[1 + \frac{t}{2(bt-b-t)}\right]$$

and for Latin squares

$$\frac{\sigma^2}{n-\frac{1}{2}} < \frac{\sigma^2}{n}\left[1 + \frac{1}{2(n-3)}\right]$$

for all positive values of b, t and n. Thus if in a 6×6 Latin square, we have a pair of mixed-up plots belonging to different rows, columns and treatments, the limits of the the standard errors of mean of affected treatments are $\sigma/2$ and $\sigma\sqrt{2}/\sqrt{11}$.