LATIN SQUARES WITH SEVERAL MISSING PLOTS

By M. N. Das

Indian Council of Agricultural Research, New Delhi

1. Introduction

Yates (1933) dealt with the problem of missing plots in Latin Squares. Though his method, which consists in carrying out the analysis by estimating the missing observations, provides a valid error sum of squares, only an approximate value of the treatment sum of squares can be obtained from this method except in the case of one missing plot for which a correction term has been given. Except for the case of one missing plot, the exact standard errors between affected treatments or between affected and non-affected treatments cannot also be obtained. Later (1936) he gave a method of exactly analysing Latin Squares when a row, column or a treatment is completely missing or when both a row and a column or either of them and a treatment, are missing completely. Afterwards, Pearce (1953) deduced expressions for standard errors for all possible differences between treatments when two observations are missing in a Latin Square.

An attempt has been made in this paper to deduce an exact method of analysis of Latin Squares when (1) the number of plots missing is less than the number of treatments and each belongs to a different row, column and treatment, and (2) a row, column or a treatment is partly but not completely missing.

2. General Method of Solution

Let there be x missing plots in an $r \times r$ Latin Square such that n_i^c plots are missing in the ith column, n_j^t replicates are missing from the jth treatment and n_k^r plots are missing in the kth row, so that

$$\Sigma n_i^c = \Sigma n_j^t = \Sigma n_k^r = x.$$

On the usual additive model, the observation belonging to the ith column, the jth treatment and the kth row may be represented as:

$x_{ijk.} = \mu + c_i + t_j + r_k + \epsilon_{ijk}$, where c_i, t_j and r_k are the column, treatment and row effects respectively. It follows from the method of fitting constants by least squares that the best estimates of the effects can be obtained from the following normal equations:

$$C_i = (r - n_i^c)\ (\mu + c_i) + \sum_{j\,(i)} t_j + \sum_{k\,(i)} r_k$$

$$T_j = (r - n_j^t)\ (\mu + t_j) + \sum_{i\,(j)} c_i + \sum_{k\,(j)} r_k$$

$$R_k = (r - n_k^r)\ (\mu + r_k) + \sum_{i\,(k)} c_i + \sum_{j\,(k)} t_j$$

where C_i, T_j and R_k denote respectively the totals of the i-th column, j-th tretament and the kth row, $\sum_{j\,(i)}$ and $\sum_{k\,(i)}$ in the first equation denote respectively the summation over those (1) treatments and (2) rows for which observations are available in the i-th column; $\sum_{i\,(j)}$ and $\sum_{k\,(j)}$ in the second equation denote respectively the summation over those (1) columns and (2) rows in which the j-th treatment is not missing and lastly $\Sigma_{i\,(k)}$ and $\Sigma_{j\,(k)}$ in the third equation denote similarly the summation over those (1) columns and (2) treatments for which observations are available in the kth row.

By eliminating r_k from these equations and any one of each of the unaffected column effects and treatment effects with the help of the additional restrictions $\Sigma c = 0$ and $\Sigma t = 0$, the resultant equations can be divided into two sets. In one set there will be only those equations which correspond to the affected columns and treatments such that the equation corresponding to any such column or treatment will contain only the affected treatment and column effects. The other set will contain equations corresponding to the unaffected treatments and columns such that an equation corresponding to any such column (treatment) will contain only that column (treatment) effect together with all the affected treatment and column effects. Thus once the solution of the first set of equations is available that of the second set can be obtained easily from them.

If the plot (i, j, k) be missing and P_{ijk} stands for $\dfrac{c_i + t_j}{r - n_k^r}$, then the resultant equations can be written as:

$$rc_i - \Sigma'_i\ (r - n_k^r)\ P_{ijk} + \Sigma'_{k\,(i)}\ P_{ijk} = Q_i^{\mathrm{C}} \qquad (1)$$

$$rt_j - \Sigma'_j\ (r - n_k^r)\ P_{ijk} + \Sigma'_{k\,(j)}\ P_{ijk} = Q_j^{\mathrm{T}} \qquad (2)$$

$$rc_a + \Sigma'\ P_{ijk} = Q_a^{\mathrm{C}} \qquad (3)$$

$$rt_b + \Sigma'\ P_{ijk} = Q_b^{\mathrm{T}} \qquad (4)$$

where Σ'_i denotes summation over those missing plots which are in the i-th column and $\Sigma'_{k\,(i)}$ over those missing plots which are in rows other than those in which the ith column is missing. Similarly $\sum_{j}'$ denotes summation over those plots where the jth treatment is missing and $\sum_{k(j)}'$ over those missing plots which are in rows other than those in which the jth treatment is missing, and Σ' denotes summation over all the missing plots.

Here c_i and t_j stand respectively for the affected columns and treatments and c_a and t_b for unaffected columns and treatments.

Q_i^{C} stands for the row-adjusted ith column total and is equal to $C_i - \sum_{k(i)} \bar{x}_{rk}$, where $\bar{x}_{rk}$ is the kth row average.

Similarly

$$Q_j^{\mathrm{T}} = T_j - \sum_{k(j)}' \bar{x}_{rk}$$

$$\left.\begin{aligned} Q_a^{\mathrm{C}} &= C_a - \Sigma\, \bar{x}_{rk} \\ Q_b^{\mathrm{T}} &= T_b - \Sigma\, \bar{x}_{rk} \end{aligned}\right\} \quad \Sigma \text{ indicates summation over all rows}$$

By adding the corresponding equations in (1) and (2), *i.e.*, the equations belonging to the same plot—one for the column effect and the other for the treatment effect—we get

$$r\,(r-n_k^{r})\,P_{ijk} - \sum_{i}'\,(r-n_k^{r})\,P_{ijk} - \sum_{j}'\,(r-n_k^{r})\,P_{ijk} + \sum_{k(i)}' P_{ijk} + \sum_{k(j)}' P_{ijk} = Q_i^{c} + Q_j^{\mathrm{T}} \quad (5)$$

Thus there will be as many equations in the first set as there are missing plots. The solution of these equations and consequently of (1), (2), (3) and (4) is not sufficient for the complete analysis. Another set of equations corresponding to rows and columns ignoring treatment effects need be solved for the complete analysis. The general method of writing and solving equations in the two-way classification has been given by the author (Das, 1954) elsewhere.

3. Special Cases

The whole process of analysis in the general case will not be easy. But algebraic solutions of all the sets of equations are available in two special cases, *viz.*, (1) when the plots are each missing in a separate row, column and a treatment, and (2) when a row, column or a treatment is partly but not completely missing.

Case I.—The equations in Case I can be written as:

$$rc_i - (r-1)P_i + \sum_{i'\neq i}' P_{i'} = Q_i^{C}$$

$$rt_j - (r-1)P_i + \sum_{i'\neq i}' P_{i'} = Q_j^{T}$$

$$rc_a + \Sigma' P_i = Q_a^{C}$$

$$rt_b + \Sigma' P_i = Q_b^{T}$$

where P_i corresponds to the missing plot with column and treatment effects c_i and t_j respectively.

Thus

$$P_i = \frac{c_i + t_j}{r-1}.$$

Adding the first two equations

$$\{(r-2)(r-1) - 2\} P_i + 2\Sigma' P = Q_i^{C} + Q_j^{T}$$

Again adding such equations for all the missing plots

$$\Sigma' P_i = \frac{\Sigma Q_i^{C} + \Sigma Q_j^{T}}{(r-1)(r-2) + 2(x-1)}.$$

Thus

$$c_i = \frac{Q_i^{C}}{r} + \frac{Q_i^{C} + Q_j^{T} - (r-1)\Sigma' P}{r(r-3)}$$

$$t_j = \frac{Q_j^{T}}{r} + \frac{Q_i^{C} + Q_j^{T} - (r-1)\Sigma' P}{r(r-3)}$$

$$c_a = \frac{Q_a^{C} - \Sigma' P}{r}$$

$$t_b = \frac{Q_b^{T} - \Sigma' P}{r}.$$

From these equations it is evident that variance

$$(t_j - t_{j'}) = \frac{2\sigma^2}{r} \cdot \frac{r-2}{r-3}$$

where t_j and $t_{j'}$ are both affected treatments.

Again, variance

$$(t_b - t_j) = \frac{\sigma^2}{r}\left\{2 + \frac{r(r-3) + 2(x-1)}{(r-3)\{r(r-3) + 2x\}}\right\}$$

where t_b is an unaffected treatment.

The error sum of squares can be obtained from the following relation, *viz.*, Error S.S. = Total S.S. $- (\Sigma c\, Q^c + \Sigma t\, Q^T) -$ Row S.S.

with d.f. equal to $(r-1)(r-2) - x$.

To get the adjusted treatment sum of squares it is necessary to solve the set of equations containing column effects only to be obtained by eliminating row effects and ignoring treatment effects. The solution of this set of equations is easily seen to be:

$$C_i' = \frac{(r-1)}{r(r-2)} \left\{ Q_i^C - \frac{\Sigma Q_i^C}{r(r-2)+x} \right\}$$

$$C_a' = \frac{Q_a^C}{r} - \frac{\Sigma Q_i^C}{r\{r(r-2)+x\}}.$$

The adjusted treatment S.S. can now be obtained from

$$\Sigma c\, Q^C + \Sigma t\, Q^T - \Sigma c'\, Q^C.$$

Case II.—In the second case if all the plots are missing in the ith column, say, the equations turn out to be

$$rc_i - (r-1)\Sigma' P = Q_i^C$$

$$rt_j - rP_j + \Sigma' P = Q_j^T$$

where P_j stands for the plot where the jth treatment is missing.

$$rc_a + \Sigma' P = Q_a^C$$

$$rt_b + \Sigma' P = Q_b^T$$

By adding the first two equations

$$rP_j - \Sigma' P = \frac{Q_i^C + Q_j^T}{r-2}$$

i.e.,

$$\Sigma' P = \frac{xQ_i^C + \Sigma Q_j^T}{(r-2)(r-x)}$$

and

$$c_i = \frac{Q_i^C + (r-1)\Sigma' P}{r}$$

$$t_j = \frac{Q_j^T}{r} + \frac{Q_i^C + Q_j^T}{r(r-2)}$$

$$c_a = \frac{Q_a^C - \Sigma' P}{r}$$

$$t_b = \frac{Q_b^T - \Sigma' P}{r}.$$

The variance of the difference between any two affected treatments is evidently $\frac{2(r-1)}{r(r-2)}\sigma^2$.

The variance of $(t_b - t_j) = \frac{\sigma^2}{r}\left\{2 + \frac{r-x+1}{(r-2)(r-x)}\right\}$.

When the jth treatment is missing partly, the corresponding results can be obtained by interchanging c_i's and t_j's in the equations. Thus the variance of $(t_b - t_j)$ in this case becomes

$$\frac{\sigma^2}{r}\left\{2 + \frac{rx}{(r-2)(r-x)}\right\}.$$

The error S.S. in the case of one affected column can be obtained from the following, *i.e.*,

$$\text{Error S.S.} = \text{Total S.S.} - \text{Row S.S.} - \Sigma c\, Q^{\text{C}} - \Sigma t\, Q^{\text{T}}.$$

The further solution required for obtaining the adjusted treatment S.S. is

$$C_i' = \frac{Q_i^{\text{C}}}{r-x}.$$

$$C_a' = \frac{Q_a^{\text{C}}}{r} - \frac{xQ_i^{\text{C}}}{r(r-1)(r-x)}.$$

The adjusted treatment sum of squares is then given by

$$\Sigma c\, Q^{\text{C}} + \Sigma t\, Q^{\text{T}} - \Sigma c' Q^{\text{C}}.$$

In the case of one affected treatment, the error sum of squares can be obtained by writing c_i's for t_j's and *vice versa* in the equations and substituting in the above expression for error sum of squares. But the adjusted treatment sum of squares will be given by

$$\Sigma c\, Q^{\text{C}} + \Sigma t\, Q^{\text{T}} - \Sigma c' Q^{\text{C}}$$

where

$$C_i' = \frac{(r-1)}{r(r-2)}\left\{Q_i^{\text{C}} - \frac{\Sigma Q_i^{\text{C}}}{r(r-2)+x}\right\}$$

$$C_a' = \frac{Q_a^{\text{C}}}{r} - \frac{\Sigma Q_i^{\text{C}}}{r\{r(r-2)+x\}}.$$

All the results in the case of one affected row can be obtained from those of one affected column by treating the rows as columns and proceeding as above.

The cases of one and two plots missing in whatever manner, come out as particular cases of the results derived above. The variances of the treatment differences obtained by Yates (1933) and Pearce (1953) in these cases, agree with those derived in the paper.

4. Accuracy of the Approximate Variance Suggested by Yates

In the general case of an $r \times r$ Latin Square the approximate variance of the difference between two affected treatments, as obtained from Yates' method (1933) is equal to $2\sigma^2/r - 5/3$ when the plots are missing as in the Special Case I; and when all the missing plots are in the same row or column, as in Special Case II, it is $2\sigma^2/r - 4/3$. The corresponding actual variances have been found to be $\frac{2\sigma^2}{r} \cdot \frac{r-2}{r-3}$ and $\frac{2\sigma^2}{r} \cdot \frac{r-1}{r-2}$ respectively. In the former case the approximate value is equal to the exact one when $r = 5$. It is less than the actual by $\sigma^2/7$ when $r = 4$. For values of r greater than 5 the approximate value is greater than the actual, the maximum difference being $\sigma^2/56$ corresponding to $r = 7$.

In the second case the two variances become equal when $r = 4$. For greater values of r the approximate variance is greater, the maximum difference being $\sigma^2/82 \cdot 5$ for $r = 5$.

The approximate variance of the difference between an affected and an unaffected treatment is equal to

$$\sigma^2 \left\{ \frac{1}{r-1} + \frac{1}{r - 2/3} \right\},$$

if more than one treatment is missing, as in Special Cases I or II. The corresponding actual variances are

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

in Case I and

$$\frac{\sigma^2}{r} \left\{ 2 + \frac{r - x + 1}{(r-2)(r-x)} \right\}$$

in Case II. Evidently the difference between the two estimates is that whereas the variance as obtained from Yates' approximate method is independent of the number of the affected treatments, the actual variance is a function of x.

From a closer scrutiny it is found that in Case I the approximate value is greater or less than the actual according as $r \geqslant$ or < 6 whatever x may be.

In Case II the approximate variance is less than the actual when $r = 4$ or 5 whatever x may be. When $r = 6$, it is greater or less than the actual according as $x <$ or $\geqslant 3$. For values of $r > 6$ and up to 12, the approximate value is greater or less than the actual according as $x <$ or $\geqslant r - 2$.

When several replicates (say, x) of the *same* treatment are missing, the approximate variance is equal to

$$\sigma^2 \left\{\frac{1}{r - x} + \frac{1}{r - 2x/3}\right\},$$

while the actual value is

$$\frac{\sigma^2}{r} \left\{2 + \frac{rx}{(r-2)\,(r-x)}\right\}.$$

In this case also when $r = 4$ or 5, the approximate value is less than the actual whatever x may be. It is greater or less than the actual according as (1) $x <$ or $\geqslant r - 2$ when $r = 6$ or 7, and (2) $x <$ or $= r - 1$ for greater values of r, except that the two variances become equal when $r = 6$ and $x = 3$ and also when $r = 8$ and $x = 6$.

Generally the approximate variance is considerably less than the actual when r is less than 6. For greater values of r the approximate value tends to be larger than the actual unless x is very nearabout r.

5. Example

In an yield trial of promising strains of Tur (*Cajanus indicus*) crop in Central Provinces, a 6×6 Latin Square design was adopted. The data were complete, but, for the sake of illustration four observations each from a separate row, column and treatment, have been omitted and the remaining data analysed according to the method described in the paper. The following table gives the layout plan and the observations.

TABLE I

Layout plan and observations (lb. per plot)

Rows	Columns						Total with No. of observations	Average
	×(1)	6·9(5)	8·9(3)	7·4(6)	11·6(2)	8·1(4)	42·9(5)	8·58
	3·9(6)	8·0(4)	5·0(2)	10·7(1)	×(3)	6·0(5)	33·6(5)	6·72
	4·6(2)	4·6(1)	×(5)	7·3(3)	8·1(4)	6·1(6)	30·7(5)	6·14
	5·1(3)	3·3(6)	8·3(4)	7·4(2)	6·5(5)	5·1(1)	35·7(6)	5·95
	2·2(5)	7·6(3)	5·2(1)	×(4)	7·1(6)	6·0(2)	28·1(5)	5·62
	4·4(4)	5·4(2)	6·0(6)	8·2(5)	8·1(1)	6·0(3)	38·1(6)	6·35
Column totals with No. of observations	20·2(5)	35·8(6)	33·4(5)	41·0(5)	41·4(5)	37·3(6)	209·1(32)	39·36
Row adjusted column totals (Q^c)	−10·58	−3·56	0·18	7·26	8·76	−2·06		
Treatment totals with No. of observations	(1) 33·7(5)	(2) 40·0(6)	(3) 34·9(5)	(4) 36·9(5)	(5) 29·8(5)	(6) 33·8(6)	209·1(32)	
Row adjusted treatment totals (Q^T)	2·92	0·64	2·26	3·16	−3·42	−5·56		

Here,

$$\Sigma Q_i^C = 5\cdot 62, \qquad \Sigma Q_j^T = 4\cdot 92$$

$$\Sigma' P = \frac{5\cdot 62 + 4\cdot 92}{26} = \cdot 4054$$

$$c_1 = \frac{Q_1^c}{r} + \frac{Q_1^c - Q_1^T - (r-1)\,\Sigma' P}{r\,(r-3)} = -2\cdot 3015$$

$$c_2 = \frac{Q_2^c - \Sigma' P}{r} = -0\cdot 6609$$

$$c_3 = \frac{Q_3^c}{r} + \frac{Q_3^c + Q_5^T - (r-1)\,\Sigma' P}{r\,(r-3)} = -0\cdot 2626$$

$$c_4 = \frac{Q_4^c}{r} + \frac{Q_4^c + Q_4^T - (r-1)\,\Sigma' P}{r\,(r-3)} = 1\cdot 6762$$

$$c_5 = \frac{Q_5^c}{r} + \frac{Q_5^c + Q_3^T - (r-1)\,\Sigma' P}{r\,(r-3)} = 1\cdot 9596$$

$$c_6 = \frac{Q_6^c - \Sigma' P}{r} = -0\cdot 4109.$$

Similarly

$$t_1 = -0\cdot0515 \qquad c_1' = -2\cdot2460$$
$$t_2 = 0\cdot0391 \qquad c_2' = -0\cdot6271$$
$$t_3 = 0\cdot8763 \qquad c_3' = -0\cdot0043$$
$$t_4 = 0\cdot9929 \qquad c_4' = 1\cdot4707$$
$$t_5 = -0\cdot8626 \qquad c_5' = 1\cdot7832$$
$$t_6 = -0\cdot9942 \qquad c_6' = -0\cdot3771.$$

Now

$$\Sigma cQ^{C} + \Sigma t\, Q^{T} = 56\cdot8372 + 13\cdot4705 = 70\cdot3077$$

and

$$\Sigma c' Q^{C} = 53\cdot0693.$$

Both the expressions can be obtained directly also without actually solving the equations.

Thus

$$\Sigma c\, Q^{C} + \Sigma t\, Q^{T} = \frac{\Sigma Q^{c2} + \Sigma Q^{T2}}{r} + \frac{\Sigma\,(Q_i^{c} + Q_j^{T})^2}{r\,(r-3)} - \frac{2\,\{\Sigma Q_i^{c} + \Sigma Q_j^{T}\}^2}{r\,(r-3)\,\{(r-1)\,(r-2) + 2\,(x-1)\}}$$
$$= 70\cdot3026$$

and

$$\Sigma c' Q^{c} = \frac{\Sigma Q^{c2}}{r} + \frac{\Sigma Q_i^{c2}}{r\,(r-2)} - \frac{(\Sigma Q_i^{c})^2}{r\,(r-2)\,\{r\,(r-2) + x\}}$$
$$= 53\cdot0671.$$

Total S.S. $= 1493\cdot1100$

Row S.S. $= 1394\cdot6440$

Error S.S. $= 1493\cdot1100 - 1394\cdot6440 - 70\cdot3077$
$= 28\cdot1583$ on 16 d.f.

Adjusted treatment S.S. $= 70\cdot3077 - 53\cdot0693$
$= 17\cdot2384$

Variance $(t_j - t_{j'}) = 4/9\,\sigma^2$.

Variance $(t_h - t_j) = 5/13\,\sigma^2$.

The corresponding approximate variances from Yates' method come out to be $6/13\,\sigma^2$ and $31/80\,\sigma^2$ respectively. As the approximate variances are greater than the actual ones, tests based on the variances obtained from Yates' method may lead to less number of significant cases than is expected.

6. Summary

The general method of analysis of incomplete Latin Squares when any number of plots are missing in any manner provided there is at least one column or row and one treatment unaffected, has been indicated. The exact method of analysis including the finding of standard errors for all possible treatment differences, has been deduced in two special cases, *viz.*, (1) when the missing plots are each in a separate row, column and treatment, and (2) when the missing plots belong to the same row, column or treatment and the row, etc., are not completely missing. Cases of one and two plots missing in whatever manner come out as particular cases of the above. Accuracy of the approximate variances suggested by Yates has been examined in these cases. The method of analysis in the first case has been illustrated by means of an example.

References

Das, M. N. .. *J. Ind. Soc. Agric. Stat.*, 1954, **6**, 58–76.

Pearce, S. C. .. *C.A.B. Technical Communication*, 1953, p. 23.

Yates, F. .. *Emp. J. Exp. Agric.*, 1933, **1**, 129–42.

——— .. *J. Agric. Sci.*, 1936, **26**, 301–15.