The Analysis of Latin Squares when Two or More Rows, Columns, or Treatments are Missing

By F. Yates and R. W. Hale

1. Introduction

In recent work comparing different rye-grass strains in Northern Ireland it was necessary to analyse a Latin-square trial of seven strains, in which the aftermath of two strains was not ready for cutting at the same time as the others. The aftermath yields had hence to be treated as a Latin-square trial with two treatments missing. Since an example of a trial with more than one treatment missing has not been described previously, it was thought worth while to record the method of analysis.

Opportunity has been taken to discuss also the case in which two or more rows or columns are missing. Although the analysis of an ordinary square with missing rows is complicated by lack of orthogonality, this complication may be avoided by the imposition of additional restrictions, analogous to those of incomplete randomized blocks. Such arrangements (known as Youden squares) provide valid and useful experimental designs.

2. Missing Treatments

Constants representing row, column, and treatment effects must be fitted by the method of least squares.

The system of constants given in a previous paper (Yates, 1936) dealing with cases in which one row, column, or treatment, or a row and column, or either and a treatment, are missing, may be adopted. If only the first s out of the p treatments exist, the constants will be :—

Mean :	m.
Rows :	$r_1,\ r_2,\ \ldots\ r_p;\ S(r) = 0.$
Columns :	$c_1,\ c_2,\ \ldots\ c_p;\ S(c) = 0.$
Treatments :	$t_1,\ t_2,\ \ldots\ t_s;\ S(t) = 0.$

If R_u, C_u, T_u are the totals of the u^{th} row, column, and treatment respectively, and G is the grand total, and if $S_{r_1}(c)$ represents the sum of the c's for those columns which have plots in row 1, etc., the normal equations are :—

$$\text{Mean}: \quad spm = G$$

$$\text{Rows}: \quad \begin{cases} sm + sr_1 + S_{r_1}(c) = R_1 \\ sm + sr_2 + S_{r_2}(c) = R_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots \end{cases}$$

$$\text{Columns}: \quad \begin{cases} sm + sc_1 + S_{c_1}(r) = C_1 \\ sm + sc_2 + S_{c_2}(r) = C_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots \end{cases}$$

$$\text{Treatments}: \quad \begin{cases} pm + pt_1 = T_1 \\ pm + pt_2 = T_2 \\ \ldots\ldots\ldots\ldots \end{cases}$$

The equations for m and for the t's are independent of the r's and c's, and can be solved immediately, m being the general mean, and the t's the deviations of the treatment means.

The equations for the r's and c's, however, are not independent. They may be written :—

$$\begin{array}{ll} sr_1 + S_{r_1}(c) = \mathrm{R}_1 - \bar{R} & sc_1 + S_{c_1}(r) = C_1 - \bar{C} \\ sr_2 + S_{r_2}(c) = R_2 - \bar{R} & sc_2 + S_{c_2}(r) = C_2 - \bar{C} \\ \ldots\ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots\ldots \end{array}$$

or, alternatively, if $S'_{r_1}(c)$ represents the sum of the c's for those columns in which plots are missing in row 1, etc.,

$$\begin{array}{ll} sr_1 - S'_{r_1}(c) = R_1 - \bar{R} & sc_1 - S'_{c_1}(r) = C_1 - \bar{C} \\ sr_2 - S'_{r_2}(c) = R_2 - \bar{R} & sc_2 - S'_{c_2}(r) = C_2 - \bar{C} \\ \ldots\ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots\ldots \end{array}$$

The latter form is more convenient if less than half the treatments are missing.

When only a few of the treatments are missing, the equations are most simply solved by successive approximation, as indicated in the example below. If a large proportion of the treatments are missing, the solution ceases to be of practical interest, since the number of degrees of freedom for error is small. With only two treatments there are no degrees of freedom for error.

When the values of the r's and the c's have been obtained, the reduction in the sum of squares due to the fitting of these constants may be computed in the ordinary manner from the formula

$$S\{r(R - \bar{R})\} + S\{c(C - \bar{C})\}$$

The remainder of the analysis follows the usual lines.

3. Example

The plan and yields of the experiment referred to in the introduction are given in Fig. 1, strains being indicated by letters and the yields being given in pounds of dry matter per plot.

FIGURE 1.

Plan and yields of experiment on rye-grass strains.

							Total
I 10·34	—	*M* 6·54	*S* 5·45	*N* 9·58	—	*W* 7·04	38·95
N 7·51	*W* 5·49	*I* 5·24	*M* 5·69	—	—	*S* 6·61	30·54
—	*I* 5·14	—	*W* 6·76	*S* 7·67	*N* 7·86	*M* 7·43	34·86
—	*M* 6·39	*W* 7·90	—	*I* 6·11	*S* 5·93	*N* 7·34	33·67
S 7·95	*N* 11·03	—	*I* 5·32	*W* 9·40	*M* 7·70	—	41·40
W 6·29	*S* 5·26	*N* 7·87	—	*M* 7·52	*I* 6·56	—	33·50
M 6·27	—	*S* 5·34	*N* 6·25	—	*W* 8·14	*I* 5·12	31·12
Total 38·36	33·31	32·89	29·47	40·28	36·19	33·54	244·04

	I.	*M.*	*S.*	*N.*	*W.*	
Strain means:	6·26	6·79	6·32	8·21	7·29	6·97257

The complete set of normal equations for the r's and c's is shown in Table I.

TABLE I

Normal Equations

$$5r_1 - c_2 - c_6 = +4{\cdot}087 \qquad 5c_1 - r_3 - r_4 = +3{\cdot}497$$
$$5r_2 - c_5 - c_6 = -4{\cdot}323 \qquad 5c_2 - r_1 - r_7 = -1{\cdot}553$$
$$5r_3 - c_1 - c_3 = -0{\cdot}003 \qquad 5c_3 - r_3 - r_5 = -1{\cdot}973$$
$$5r_4 - c_1 - c_4 = -1{\cdot}193 \qquad 5c_4 - r_4 - r_6 = -5{\cdot}393$$
$$5r_5 - c_3 - c_7 = +6{\cdot}537 \qquad 5c_5 - r_2 - r_7 = +5{\cdot}417$$
$$5r_6 - c_4 - c_7 = -1{\cdot}363 \qquad 5c_6 - r_1 - r_2 = +1{\cdot}327$$
$$5r_7 - c_2 - c_5 = -3{\cdot}743 \qquad 5c_7 - r_5 - r_6 = -1{\cdot}323$$

In solving equations of this type it is best to begin by substituting for the c's in the r equations, or vice-versa. This gives the equations

$$23r_1 - r_2 - r_7 = +20{\cdot}21$$
$$23r_2 - r_1 - r_7 = -14{\cdot}87$$
$$23r_3 - r_4 - r_5 = +1{\cdot}51$$
$$23r_4 - r_3 - r_6 = -7{\cdot}86$$
$$23r_5 - r_3 - r_6 = +29{\cdot}39$$
$$23r_6 - r_4 - r_5 = -13{\cdot}53$$
$$23r_7 - r_1 - r_2 = -14{\cdot}85$$

The equations may now be solved by successive approximation, starting with 1/23 of each of the numerical terms and using each equation in turn to obtain an improved approximation for that r. Thus the second approximation for r_1 is

$$\tfrac{1}{23}(+20{\cdot}21 - 0{\cdot}647 - 0{\cdot}646) = +0{\cdot}822$$

and, using this value, the second approximation for r_2 is

$$\tfrac{1}{23}(-14{\cdot}87 + 0{\cdot}822 - 0{\cdot}646) = -0{\cdot}639$$

In this manner the approximations shown in Table II are obtained. It is clear that the second approximation gives the results with all necessary accuracy.

TABLE II

Solution by Successive Approximation

	1st	2nd	3rd	Substitution	
r_1	+ 0·879	+ 0·822	+ 0·823	c_1	+ 0·648
r_2	− 0·647	− 0·639	− 0·638	c_2	− 0·274
r_3	+ 0·066	+ 0·106	+ 0·105	c_3	− 0·122
r_4	− 0·342	− 0·363	− 0·361	c_4	− 1·261
r_5	+ 1·278	+ 1·257	+ 1·259	c_5	+ 0·828
r_6	− 0·588	− 0·549	− 0·549	c_6	+ 0·302
r_7	− 0·646	− 0·638	− 0·638	c_7	− 0·123
	0·000	− 0·004	+ 0·001		− 0·002

The values of the c's may then be obtained directly by substitution.

In this particular case the equations may also be solved directly without difficulty, since they split into two independent groups.

The sum of squares attributable to rows and columns may now be calculated by multiplying each r and c by the numerical term in the corresponding normal equation, and summing the products. Thus:—

$$+ 4{\cdot}087 \times 0{\cdot}823 + 4{\cdot}323 \times 0{\cdot}638 - 0{\cdot}003 \times 0{\cdot}105 + \ldots + 3{\cdot}497 \times 0{\cdot}648 + \ldots = 32{\cdot}70$$

The total sum of squares and the sum of strains are calculated in the ordinary way. The complete analysis of variance is shown in Table III.

TABLE III

Analysis of Variance

	D.F.	S.S.	M.S.	Variance Ratio
Rows and columns ...	12	32·70	2·725	1·97 (5% pt., 2·34)
Strains	4	18·13	4·532	3·28 (5% pt., 2·93)
Error	18	24·86	1·381	
Total	34	75·69		

It thus appears that there are significant differences amongst the strains. The smallest significant difference between strains is 1·32 lb. per plot, so that strain *N* yielded significantly more than all others except *W*. The standard error of a single plot is 16·8 per cent. of the mean, so that the experiment is of a low order of accuracy.

It might be thought that the above procedure is somewhat

elaborate, and that some approximate procedure might suffice. It is, of course, possible to eliminate either rows or columns by the ordinary methods, the experiment being then treated as if it were one in randomized blocks. If inspection of the results indicates that only one of these components is contributing any additional variation to the results, this procedure will be reasonably satisfactory, but when both components appear on inspection to be equally variable, as in the present example, the exact procedure must be followed. Even if the major part of the variation is confined to one component, there may still be some appreciable variation attributable to the second component. Again, if no additional variation is contributed by either component, and the larger component only is eliminated, an under-estimate of the residual variance will result. When dealing with important material, therefore, the exact procedure should always be followed.

It may be noted that in the present example the elimination of rows only would have given a residual mean square of 1·603, and a variance ratio, which just reaches significance, of 2·83.

4. Missing Rows or Columns

When a number of rows or columns are missing, the least-square solution is clearly similar to that already obtained. If rows are missing, for instance, columns and treatments will be non-orthogonal.

In this case the final estimates for the treatment differences will be given by the values of the treatment constants, and not by the ordinary treatment means. If it is desired to make a general test of significance on these differences, the sum of squares accounted for by columns only (ignoring treatments) must be deducted from the sum of squares accounted for by both columns and treatments.

There is, moreover, an additional complication, in that the standard errors of the treatment differences are no longer obtainable directly from the residual mean square. To make exact tests of significance we must evaluate the reciprocal matrix, as when testing the significance of the difference of the regression coefficients of a partial regression. This procedure, however, is somewhat laborious, and will not normally be worth while. Provided that only a small proportion of the rows are missing, the approximate procedure indicated below is likely to give reasonable results.

In order to illustrate the additional computations we will test the significance and evaluate the standard errors of the row constants in the above example.

The sum of squares due to columns only (ignoring rows) is derived directly from the column totals, the sum of the squares of their

deviations being divided by 5. This gives 16·09, and we thus obtain the analysis of Table IV.

TABLE IV

Significance of Row Differences

	D.F.	S.S.	M.S.
Rows and columns	12	32·70	
Columns (ignoring rows)	6	16·09	
Rows	6	16·61	2·77

The mean square, 2·77, when tested against the residual mean square (Table III), gives a variance ratio of 2·01, which is not significant. Had the rows been tested as if they were independent, we should have obtained a mean square of 3·18, and a variance ratio of 2·30.

The reciprocal matrix may now be evaluated. If all the constants fitted were independent, we should require the solution of fourteen auxiliary sets of equations, each set being obtained by replacing one of the numerical terms of the normal equations by 1 and the remainder by 0. The constants, however, are not all independent, being governed by the relations

$$S(r) = 0 \text{ and } S(c) = 0$$

To form the first set of equations, therefore, we replace the coefficient of the first row equation by $+\frac{6}{7}$, those of the remaining row equations by $-\frac{1}{7}$, and those of the column equations by zero, as proved in the Appendix. The other sets are formed similarly.

Any set may be solved by substitution and successive approximation in the same way as the original equations. Considerable symmetry exists, and only two such solutions are necessary, the second being expedited by the relations $c_{rs} = c_{sr}$. The full reciprocal matrix (with all values multiplied by 7) is given in Table V.

The estimate of the standard error of the difference of any two r's may now be immediately evaluated from the formula :—

$$V(r_s - r_t) = (c_{ss} + c_{tt} - 2c_{st})s^2$$

Thus

$$V(r_2 - r_6) = \tfrac{1}{7}(1{\cdot}290 + 1{\cdot}290 + 2 \times 0{\cdot}238)1{\cdot}381 = 0{\cdot}603$$

and the standard error is $\pm$ 0·776.

It will be noted that all the diagonal terms of the matrix have the same value, $\frac{1}{7}$(1·290), and that the other terms involving rows only range between $-\frac{1}{7}$(0·169) and $-\frac{1}{7}$(0·238). Thus $V(r_s - r_t)$, which

TABLE V

The Reciprocal matrix ($7c_{rs}$)

r_1	r_2	r_3	r_4	r_5	r_6	r_7	c_1	c_2	c_3	c_4	c_5	c_6	c_7
+ 1·290	− 0·169	− 0·238	− 0·238	− 0·238	− 0·238	− 0·169	− 0·095	+ 0·224	− 0·095	− 0·095	− 0·068	+ 0·224	− 0·095
	+ 1·290	− 0·238	− 0·238	− 0·238	− 0·238	− 0·169	− 0·095	− 0·068	− 0·095	− 0·095	+ 0·224	+ 0·224	− 0·095
		+ 1·290	− 0·171	− 0·171	− 0·232	− 0·238	+ 0·224	− 0·095	+ 0·224	− 0·081	− 0·095	− 0·095	− 0·081
			+ 1·290	− 0·232	− 0·171	− 0·238	+ 0·224	− 0·095	− 0·081	+ 0·224	− 0·095	− 0·095	− 0·081
				+ 1·290	− 0·171	− 0·238	− 0·081	− 0·095	+ 0·224	− 0·081	− 0·095	− 0·095	+ 0·224
					+ 1·290	− 0·238	− 0·081	− 0·095	− 0·081	+ 0·224	− 0·095	− 0·095	+ 0·224
						+ 1·290	− 0·095	+ 0·224	− 0·095	− 0·095	+ 0·224	− 0·068	− 0·095
							+ 1·290	− 0·238	− 0·171	− 0·171	− 0·238	− 0·238	− 0·232
								+ 1·290	− 0·238	− 0·238	− 0·169	− 0·169	− 0·238
									+ 1·290	− 0·232	− 0·238	− 0·238	− 0·171
										+ 1·290	− 0·238	− 0·238	− 0·171
											+ 1·290	− 0·169	− 0·238
												+ 1·290	− 0·238
													+ 1·290

would have the value $\frac{2}{5}s^2$ if there were complete orthogonality, has limits

$$\frac{2}{4{\cdot}80}s^2 \text{ and } \frac{2}{4{\cdot}58}s^2$$

In other words, the loss of information due to non-orthogonality ranges from 4 to 8 per cent. for the various comparisons. If there were complete balance, as in the Youden square, the loss of information would be the same as in balanced incomplete blocks, *i.e.*

$$1 - E = 1 - \frac{1 - 1/5}{1 - 1/7} = \frac{1}{15}$$

where E is the efficiency factor.

An approximate method of evaluating the standard errors is as follows. We have, for instance :—

$$r_2 - r_6 = \tfrac{1}{5}(-4{\cdot}32 + 1{\cdot}36 + c_5 + c_6 - c_4 - c_7)$$

The variances of the two numerical terms are both $5s^2$, and if the c's were orthogonal with the r's, their variances would each be $\frac{1}{5}s^2$, since each would then be the mean of 5 plot yields. Hence, approximately,

$$V(r_2 - r_6) = \tfrac{1}{25}(5 + 5 + \tfrac{1}{5} + \tfrac{1}{5} + \tfrac{1}{5} + \tfrac{1}{5})s^2 = 0{\cdot}432s^2$$

Similarly, since r_1 and r_2 have a c in common,

$$V(r_1 - r_2) = \tfrac{1}{25}(5 + 5 + \tfrac{1}{5} + \tfrac{1}{5})s^2 = 0{\cdot}416s^2$$

The approximations to the true values, 0·437 and 0·417, given by the reciprocal matrix are very close.

5. THE YOUDEN SQUARE

It was shown in the previous paper (1936) that when only one row, column, or treatment of a Latin square was missing, the normal equations could be made orthogonal without difficulty, and it was pointed out that squares with one row or column missing provide valid experimental arrangements, which may on occasion be of practical use. Youden, by imposing additional restrictions on balanced incomplete block designs, constructed what were in effect Latin squares with several missing rows, the internal relations being such that the normal equations could still be reduced to orthogonality.

Youden utilized these arrangements in virus experimental work, in which he was using, as a measure of virulence, the number of local lesions produced on the leaves of plants when inoculated with different strains or concentrations of virus. By taking the plants as the columns and the leaf positions on the plant as the rows of an incom-

plete square, both these sources of variation in susceptibility were eliminated from the experimental comparisons.

In order that the normal equations shall be capable of being made orthogonal, the groups of treatments forming the columns must satisfy the conditions required for balance in incomplete blocks—*i.e.*, every treatment must fall in a column together with every other treatment the same number of times. The following arrangement for seven treatments in three rows and seven columns, for example, satisfies this condition :—

$$\begin{array}{ccccccc} b & f & e & a & d & g & c \\ f & c & a & b & g & e & d \\ e & a & d & g & f & c & b \end{array}$$

In general any arrangement appropriate to balanced incomplete blocks, in which the number of treatments or varieties v is equal to the number of blocks b, would appear to be capable of being set out in the form of a Youden square. A table of such arrangements is given in *Statistical Tables* (1938).

Since rows are orthogonal with columns and treatments, the procedure of analysis is exactly the same as for balanced incomplete block arrangements, except that a component for rows is also included in the analysis of variance, the sum of squares for this component being computed in the ordinary manner from the row totals. The procedure for incomplete blocks is described in the introduction to the above tables.

It may easily be shown that the estimate of error is unbiased, even with correlated material, provided that the rows and columns are randomized amongst themselves. The proof follows the same lines as that previously given (1936) for the case of a single missing row. Youden squares are therefore valid experimental arrangements.

Youden squares are not likely to be of very frequent value in agricultural field experiments, since in variety trials involving a large number of varieties, in which the incomplete block type of arrangement is most likely to be of use, it will generally be more profitable to arrange the plots in compact blocks rather than to set them out in line to form the columns of an incomplete square. Occasionally, however, the possibility of imposing the additional restrictions may be of value, as, for instance, in an irrigated area, in which it may be feasible to arrange that each block is bounded by an irrigation channel. The additional restrictions would then enable the effect of position in relation to irrigation channels to be completely eliminated.

It should be noted that the additional restrictions do not affect

the efficiency factor, so that if any reduction of variance results the efficiency of the experiment will be increased.

6. SUMMARY

Methods of analysing a Latin square with two or more missing treatments, rows, or columns are described, and illustrated by an example.

Attention is drawn to a special type of incomplete square, introduced by Youden, which is capable of simple analysis. Youden squares provide valid experimental arrangements, which are likely to be of value in biological experiments, and occasionally in variety trials.

We are indebted to Mr. P. A. Linehan of the Seed Testing and Plant Disease Research Division of the Ministry of Agriculture for Northern Ireland for the data on which this note is based.

APPENDIX

The Evaluation of the Reciprocal Matrix when Redundant Constants or Regression Coefficients are Introduced into Least-Square Solutions

If a redundant constant is introduced, either by accident or design, the k normal equations

$$\begin{aligned} b_1Sx_1^2 + b_2Sx_1x_2 + b_3Sx_1x_3 + \ldots &= Sx_1y \\ b_1Sx_1x_2 + b_2Sx_2^2 + Sx_2x_3 \quad + \ldots &= Sx_2y \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \end{aligned}$$

will be indeterminate, reducing to $0 = 0$. If this is the case, λ's may be chosen such that on multiplying each equation by a λ, and summing, the coefficients of all the terms are identically zero—*i.e.* :—

$$\lambda_1Sx_1^2 + \lambda_2Sx_1x_2 + \ldots\ldots + \lambda_kSx_1x_k = 0 \quad . \quad . \quad (1)$$

etc.

In order to obtain a solution, any one of the b's, say b_1, may be replaced by any desired linear function of the remaining b's. Suppose the function is given by

$$\mu_1b_1 + \mu_2b_2 + \mu_3b_3 + \ldots\ldots = \mu_0 \quad . \quad . \quad . \quad (2)$$

If, instead of making the substitution directly in the normal equations, it is made in the original regression equation

$$Y = b_1x_1 + b_2x_2 + b_3x_3 + \ldots$$

and new normal equations formed, these will be found to be

$$b_2S(x_2 - \mu_2x_1/\mu_1)^2 + b_3S(x_2 - \mu_2x_1/\mu_1)(x_3 - \mu_3x_1/\mu_1) + \ldots = S(x_2 - \mu_2x_1/\mu_1)(y - \mu_0x_1/\mu_1)$$

$$b_2S(x_2 - \mu_2x_1/\mu_1)(x_3 - \mu_3x_1/\mu_1) + b_3S(x_3 - \mu_3x_1/\mu_1)^2 + \ldots = S(x_3 - \mu_3x_1/\mu_1)(y - \mu_0x_1/\mu_1)$$

. . . . . .

The b's so found will, of course, be identical with the b's derived from the original normal equations and equation (2).

The standard errors of $b_2, \ldots b_k$, or any linear function of them, may now be found in the ordinary manner by evaluating the reciprocal matrix of this second set of normal equations. We have, for instance :—

$$\left.\begin{aligned} c_{22}S(x_2 - \mu_2x_1/\mu_1)^2 + c_{23}S(x_2 - \mu_2x_1/\mu_1)(x_3 - \mu_3x_1/\mu_1) + \ldots &= 1 \\ c_{22}S(x_2 - \mu_2x_1/\mu_1)(x_3 - \mu_3x_1/\mu_1) + c_{23}S(x_3 - \mu_3x_1/\mu_1)^2 + \ldots &= 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \end{aligned}\right\} \quad (3)$$

The first equation reduces to :—

$$c_{22}Sx_2^2 + c_{23}Sx_2x_3 + \ldots - Sx_1x_2(\mu_2c_{22} + \mu_3c_{23} + \ldots)/\mu_1 - \mu_2(c_{22}Sx_1x_2 + c_{23}Sx_1x_3 + \ldots)/\mu_1 + \mu_2\ Sx_1^2(\mu_2c_{22} + \mu_3c_{23} + \ldots)/\mu_1^2 = 1$$

Putting $c_{12} = -(\mu_2c_2^2 + \mu_3c_{22} + \ldots)/\mu_1$ and

$$c_{12}Sx_1^2 + c_{22}Sx_1x_2 + c_{23}Sx_1x_3 + \ldots = -\mu_1d \quad . \quad . \quad (4)$$

we obtain for the remaining equations :—

$$\left.\begin{aligned} c_{12}Sx_1x_2 + c_{22}Sx_2^2 + c_{23}Sx_2x_3 + \ldots &= 1 - \mu_2d \\ c_{12}Sx_1x_3 + c_{22}Sx_2x_3 + c_{23}Sx_3^2 + \ldots &= -\mu_3d \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \end{aligned}\right\} \quad . \quad (5)$$

Equations (4) and (5) together form equations for the c's which are of the usual form except for the numerical terms. From equation (1) these numerical terms must satisfy the relation

$$\lambda_1(-\mu_1d) + \lambda_2(1 - \mu_2d) + \lambda_3(-\mu_3d) + \ldots = 0 \quad . \quad (6)$$

which provides an equation for determining d. We may then solve the equations by successive approximation, starting from any set of approximations which satisfy the equation

$$\mu_1c_{12} + \mu_2c_{22} + \mu_3c_{23} + \ldots = 0 \quad . \quad . \quad . \quad (7)$$

It is immediately obvious, from the symmetry of the solution, that the c's obtained in this manner will be applicable to tests of significance involving the eliminated coefficient b_1, as well as the other coefficients.

If there is more than one redundant constant, similar equations

will hold. With two redundant constants, for instance, we may take

$$\mu_1 b_1 + \mu_2 b_2 + \mu_3 b_3 + \mu_4 b_4 + \ldots\ldots = \mu_0$$
$$\mu'_1 b_1 + \mu'_2 b_2 + \mu'_3 b_3 + \mu'_4 b_4 + \ldots\ldots = \mu'_0$$

The numerical terms of the auxiliary normal equations must then be adjusted by quantities

$$-\mu_1 d - \mu'_1 d', -\mu_2 d - \mu'_2 d', \ldots\ldots\ldots$$

d and d' being chosen so as to satisfy two equations similar to (6).

In the special case above we have :—

$$r_1 + r_2 + \ldots\ldots + r_7 = 0$$
$$c_1 + c_2 + \ldots\ldots + c_7 = 0$$

and consequently $-d$ must be added to each of the row equations, and $-d'$ to each of the column equations. If the first row equation is the one with a numerical term of unity, we have $1 - 7d = 0$, and $-7d' = 0$. Hence the adjusted numerical terms of the first seven equations of Table I are

$$\tfrac{6}{7}, -\tfrac{1}{7}, -\tfrac{1}{7}, \ldots -\tfrac{1}{7},$$

and those of the remaining seven are zero.

If the normal equations are such that a direct solution by the ordinary methods, instead of by successive approximation, appears desirable, equations (3) must be used instead of equations (4) and (5), since the latter contain a redundant constant.

The formation of the coefficients of equations (3) can be performed in two steps. The first consists of multiplying the terms of the first column of the original matrix by $-\mu_2/\mu_1$, $-\mu_3/\mu_1$, . . . in turn, adding the products so obtained to the corresponding terms of second, third, . . . columns of the matrix respectively. The second consists of performing the same operation on the rows of this new matrix.

The same result would be reached by eliminating c_{12} from equations (4) and (5) by means of equation (7), and using equation (4) to restore the diagonal symmetry of equations (5).

Elimination of certain of the regression constants from the normal equations by substitution from some of the equations into the others presents no new features. Thus in the case above, where the column equations were used to eliminate the c's from the row equations, the equations for determining c_{11}, c_{12}, . . . c_{17}, are obtained from the equations containing the r's only by substituting the numerical terms

$$\tfrac{30}{7}, -\tfrac{5}{7}, -\tfrac{5}{7}, \ldots -\tfrac{5}{7}.$$

Although one of the r's (or, alternatively, m) is redundant, these equations are not indeterminate, since m has been eliminated, and consequently they could be solved directly without further elimination. The solution only differs from the usual reciprocal matrix solution in the numerical terms, and can be reduced to the usual form by adding $\frac{5}{147}$ to each of the c's.

It would also be possible to use the relationship $S(r) = 0$ to eliminate one of the r's, but if diagonal symmetry is restored the resulting equations are more complex, so that this course would not be profitable.

References

Yates, F., "Incomplete Latin Squares," *Journ. Agric. Sci.* (1936), Vol. XXVI, Part II, pp. 301–315.

Fisher, R. A., and Yates, F., *Statistical Tables for Biological, Agricultural and Medical Research.* Oliver and Boyd, Edinburgh. 1938.