Yates, F. 1940. The recovery of inter□block information in balanced incomplete block designs. *Annals of Eugenics.* 10 (1), pp. 317-325.

THE RECOVERY OF INTER-BLOCK INFORMATION IN BALANCED INCOMPLETE BLOCK DESIGNS

BY F. YATES

1. INTRODUCTION

INCOMPLETE block and quasi-factorial designs of various kinds were first introduced by the author a few years ago (1936*a*, *b*). All these designs have the property that the number of varieties (or treatments) included in each block is smaller than the total number to be tested. There is consequently a gain in precision due to the use of smaller blocks, at the expense of loss of information on those varietal comparisons which are confounded with blocks. In the original papers only the complete elimination of inter-block differences was considered. These inter-block comparisons will, however, contain an appreciable amount of information, amounting in the limiting case, when the inter-block and intra-block comparisons are of equal accuracy, to a fraction $1-E$ of the total information, where E is the efficiency factor.

The recovery of this information has already been discussed for three-dimensional quasi-factorial designs (1939) and also for lattice (quasi-Latin) squares (1940). The present paper contains a similar discussion of balanced incomplete block designs. The case of two-dimensional quasi-factorial designs is to be dealt with in a publication by the Statistical Department of Iowa State College.

2. ESTIMATES OF THE VARIETAL DIFFERENCES

If v varieties are arranged in b blocks containing k varieties each, there being r replicates of each variety, the condition of balance will be fulfilled if each pair of varieties occurs together an equal number λ of times. A catalogue of possible designs and known solutions is given by Fisher & Yates (1938).

The following relations hold:
$$vr = kb,$$
$$(v-1)\lambda = (k-1)r.$$

The efficiency factor E, defined as the fraction of the total information contained in the intra-block comparisons, when inter-block and intra-block comparisons are of equal accuracy, is given by
$$E = \frac{1-1/k}{1-1/v} = \frac{v\lambda}{rk}.$$

Let V_s be the sum of all the yields of variety s, T_s the sum of all the block totals of blocks containing variety s, T'_s the sum of all the remaining block totals, and G the total yield of all plots. Then, as has been shown previously (1936*a*), the estimates of the varietal differences derived from the intra-block comparisons are obtained from the quantities
$$Q_s = V_s - T_s/k,$$

or, as is more convenient when $k > \frac{1}{2}v$,

$$Q'_s = V_s + T'_s/k.$$

The actual differences in units of the total yield of the r replicates are given by the differences of

$$Q_s/E \quad \text{or} \quad Q'_s/E,$$

the sum of the first set being zero, and the second set $rG/\lambda v$.

The error variance of these latter sets of quantities is r/Ew, where $1/w$ is the intra-block error variance.

The estimates of the varietal differences derived from the inter-block comparisons are similarly given by the differences of

$$rT_s/(r-\lambda) \quad \text{or} \quad rT'_s/(r-\lambda),$$

in units of the total yield of r replicates. The error variance of these sets of quantities is $kr^2/(r-\lambda)\,w'$, where $1/w'$ is the error variance of the inter-block comparisons, in units of a single plot.

If the weights w and w' are known, the most efficient estimates of the varietal differences will be given by the differences of the weighted means:

$$Y_s = \frac{\dfrac{Q_s}{E}\dfrac{Ew}{r} + \dfrac{rT_s}{r-\lambda}\dfrac{(r-\lambda)\,w'}{kr^2}}{\dfrac{Ew}{r} + \dfrac{(r-\lambda)\,w'}{kr^2}}.$$

The quantities Y_s may be termed the partially adjusted yields (totals of r replicates). If G is the total yield of all plots, Y_s may be written (after the addition of a quantity $\mu\,(k-1)\,G$) in the form

$$Y_s = V_s + \mu W_s,$$

where

$$W_s = (v-k)\,V_s - (v-1)\,T_s + (k-1)\,G,$$

and

$$\mu = \frac{w-w'}{wv(k-1)+w'(v-k)}.$$

The error variance of the Y's is

$$\frac{1}{\dfrac{Ew}{r} + \dfrac{(r-\lambda)\,w'}{kr^2}} = \frac{kr(v-1)}{wv(k-1)+w'(v-k)}.$$

3. The analysis of variance

The structure of the analysis of variance is a little complicated. The residual sum of squares for intra-block error may be calculated (*a*) by deducting the sum of squares for blocks (ignoring varieties) and for varieties (eliminating blocks) from the total sum of squares, or alternatively (*b*) by deducting the sum of squares for varieties (ignoring blocks) and for blocks (eliminating varieties).

As has previously been shown, the sum of squares for varieties (eliminating blocks) is derived from the sum of the squares of the quantities Q, with divisor rE.

The sum of squares for blocks (eliminating varieties) splits into two parts. The first, corresponding to $v-1$ degrees of freedom, is affected by varietal differences and is derived from the sum of the squares of the deviations of the quantities W_s, with divisor $rv(v-k)(k-1)$. The second, corresponding to $b-v$ degrees of freedom, is unaffected by varietal differences, and represents pure inter-block error. This latter sum of squares can best be computed by taking the difference of the total sum of squares for blocks (ignoring varieties) and the component of this sum of squares which is affected by varietal differences, this being given by the sum of squares of the deviations of T_s, with divisor $k(r-\lambda)$.

Table 1 shows these relations in tabular form. In this table dev^2 indicates the sum of the squares of the deviations, y the individual yields and B the block totals. By calculating both forms of the analysis a complete check is obtained, except for the total sum of squares and for the total sum of squares for blocks (ignoring varieties).

Table 1. *Structure of analysis of variance*

Method (*a*)	D.F.	S.S. (*a*)		S.S. (*b*)	Method (*b*)
Blocks (ignoring varieties):					Blocks (eliminating varieties):
Varietal component	$v-1$	$\frac{dev^2\, T}{k(r-\lambda)}$		$\frac{dev^2\, W}{rv(v-k)(k-1)}$	Varietal component
Remainder	$b-v$	†	$\rightarrow$	†	Remainder
Total	$b-1$	$\frac{dev^2\, B^*}{k}$		†	Total
Varieties (eliminating blocks)	$v-1$	$\frac{dev^2\, kQ}{k^2\, rE}$		$\frac{dev^2\, V}{r}$	Varieties (ignoring blocks)
Intra-block error	$rv-v-b+1$	†	$\longleftrightarrow$	†	Intra-block error
Total	$rv-1$	$dev^2\, y^*$	$\longleftrightarrow$	$dev^2\, y^*$	Total

* Requires checking. † Calculated by addition or subtraction.

4. Estimation of the relative weights

If the intra-block error variance is B, and the error variance of block totals is $k(kA+B)$, the expectations of the mean squares corresponding to the components of the sum of squares for varieties (eliminating blocks) are shown in Table 2.

Table 2. *Expectations of mean squares for blocks (eliminating varieties)*

	D.F.	Expectation
Varietal component	$v-1$	$EkA+B$
Remainder	$b-v$	$kA+B$
Total	$b-1$	$\frac{bk-v}{b-1}A+B$

The factor E in the first of the above expressions is derived as follows. If for any pair of varieties s and s' the coefficients of each plot yield in the difference $W_s - W_{s'}$ are written down, and summed by blocks, it will be found that $(r-\lambda)$ of these sums equal $v(k-1)$ and $(r-\lambda)$ equal $-v(k-1)$, the remainder being zero. Utilizing the divisor given in Table 1 (which is itself one-half the sum of the squares of these coefficients), we obtain as the coefficient of A

$$\frac{(r-\lambda)\,v^2(k-1)^2}{rv(v-k)\,(k-1)} = Ek.$$

As has been pointed out previously (Yates, 1939), it is sufficient, for the purpose of estimating w and w', to equate the expectation in terms of w and w' of the mean square for all the $b-1$ degrees of freedom for blocks (eliminating varieties) with the actual mean square M'', say. If the mean square for intra-block error is M, we obtain the equations

$$w = \frac{1}{M}, \quad w' = \frac{v(r-1)}{k(b-1)\,M''-(v-k)\,M}.$$

Since w may ordinarily be assumed to be greater than w' it will be sufficient, if M'' is less than M, to take w' as equal to w, i.e. to use the unadjusted yields as the final estimates.

Since M'' is frequently based on a somewhat small number of degrees of freedom, there is of course some inaccuracy in the estimated weights. The effect of this inaccuracy on the accuracy of the weighted estimates has been investigated in various extreme cases (Yates, 1939, 1940; and Cochran, unpublished material). The results obtained are summarized in Table 3.

Table 3. *Loss of information due to inaccuracies of weighting*

(*a*) Particulars of cases investigated

Case	Type of design	Replications	Degrees of freedom		Expectation of block M.S.		Efficiency factor	Reference to literature
			Blocks	Error	Actual	Unconfounded		
1	5×5 lattice	2	8	16	$\frac{5}{2}A+B$	$5A+B$	0·75	Cochran (unpublished)
2	4×4 triple lattice	3	9	21	$\frac{8}{3}A+B$	$4A+B$	0·769	Cochran (unpublished)
3	$3\times3\times3$ lattice	3	24	28	$2A+B$	$3A+B$	0·591	Yates (1939)
4	5×5 lattice squares	3	12	24	$\frac{10}{3}A+B$	$5A+B$	0·667	Yates (1940)

(*b*) Percentage losses of information for various values of w/w'

w/w'	1	2	4	6	8	12
Case 1	2·21	3·07	4·54	4·37	3·91	—
2	1·73	3·00	3·73	3·19	—	—
3	1·71	2·68*	2·54*	—	—	—
4	2·52	4·04	4·02	3·14	2·53	1·81

* These values are approximate only, being calculated on the assumption that the 24 degrees of freedom for blocks are homogeneous, with mean square expectation $2A+B$.

The actual loss of efficiency depends not only on the numbers of degrees of freedom for M'' and M, but also on the efficiency factor. From the cases already investigated, however, it may be concluded that this source of loss is of little importance in cases likely to occur in practice.

5. Modification when groups of blocks form complete replications

In certain cases the structure of the design is such that blocks fall into groups containing one or more complete replications of all the varieties. When this is so it is clearly advisable in agricultural trials to arrange such groups of blocks in compact large blocks on the ground, since the variation affecting the inter-block varietal estimates will thereby be reduced. Allowance for this must be made in the analysis of variance given above by eliminating complete replications (or groups of replications) from the remainder component of blocks. If there are c such large blocks (containing r/c replications each), the expectations given in Table 2 will require modification as in Table 4.

Table 4. *Expectations of mean squares when groups of blocks contain complete replications*

	D.F.	Expectation
Groups of blocks	$c-1$	—
Varietal component	$v-1$	$EkA+B$
Remainder	$b-v-c+1$	$kA+B$
Varietal component + remainder	$b-c$	$\frac{bk-v-k(c-1)}{b-c}A+B$

The formula for w' will also require modification, being in fact

$$w' = \frac{v(r-1)-k(c-1)}{k(b-c)\,M''-(v-k)\,M}.$$

In the common case in which each large block contains a single replication, $c = r$, and the expectation of the mean square for the $b-r$ degrees of freedom is $\frac{k(r-1)}{r}A+B$, the formula for w' being

$$w' = \frac{r-1}{rM''-M}.$$

6. Simplification when $b = v$, and when $v = k^2$

When $b = v$ the analysis of variance reduces to the simplified form given in Table 5.

Table 5. *Analysis of variance when $b = v$*

	D.F.	S.S.
Blocks (eliminating varieties)	$v-1$	$\frac{dev^2\,W}{rv(v-k)\,(k-1)}$
Varieties (ignoring blocks)	$v-1$	$\frac{dev^2\,V}{r}$
Intra-block error	$(k-2)\,v+1$	†
Total	$kv-1$	$dev^2\,y$

There is little point in tabulating the Q's, though they will provide a general check, as before, if this is desired.

A similar simplification is possible in the series of designs $v = k^2$, $r = k+1$, $b = k(k+1)$ (balanced lattices), where the remaining k degrees of freedom for blocks correspond to the contrasts of complete replications.

7. First example

An example of a dummy trial of nine treatments (e.g. dietary treatments) superimposed on the scores of eighteen litters of four rats in a discrimination test is given by Fisher & Yates (1938). Here $v = 9$, $r = 8$, $k = 4$, $b = 18$, $\lambda = 3$, $E = 27/32$.

The individual scores have been given in the publication referred to. Table 6 shows the values of V, T, $4Q$ and W for the nine treatments *a–i*. The analysis of variance is shown in Table 7, which corresponds in arrangement to Table 1.

Table 6. *Calculation of adjusted scores in discrimination test*

	V	T	$4Q = 4V - T$	$W = 5V - 8T + 3G$	$Y = V + \mu W$
a	43·9	152·2	+23·4	+ 65·1	45·7
b	39·1	156·4	0	+ 7·5	39·3
c	41·3	169·6	− 4·4	− 87·1	38·9
d	43·6	151·7	+22·7	+ 67·6	45·4
e	41·7	159·2	+ 7·6	− 1·9	41·6
f	35·6	162·0	−19·6	− 54·8	34·1
g	28·6	138·3	−23·9	+ 99·8	31·3
h	42·8	172·5	− 1·3	−102·8	40·0
i	37·8	155·7	− 4·5	+ 6·6	38·0
	354·4	1417·6	0	0	354·3
Divisor	8	4.5 = 20	$4^2.8.27/32$ = 108	8.9.5.3 = 1080	

Table 7. *Analysis of variance, discrimination test*

	D.F.	S.S. (*a*)	S.S. (*b*)	M.S. (*b*)
Blocks:				
Varietal component	8	41·4684	37·0634	4·6329
Remainder	9	138·2011	138·2011	15·3557
Total	17	179·6695	175·2645	10·3097
Varieties	8	19·6044	24·0094	—
Error	46	119·4506	119·4506	2·5968
Total	71	318·7245	318·7245	—

From the results of the analysis of variance we obtain

$$w = \frac{1}{2\cdot5968} = 0\cdot3851, \quad w' = \frac{63}{68 \times 10\cdot3097 - 5 \times 2\cdot5968} = 0\cdot0916,$$

$$\mu = \frac{0\cdot3851 - 0\cdot0916}{27 \times 0\cdot3851 + 5 \times 0\cdot0916} = \frac{0\cdot2935}{10\cdot8557} = 0\cdot02704.$$

The final adjusted scores in terms of the total scores of eight rats are given in the last column of Table 6. The standard error of these scores is

$$\sqrt{\frac{256}{10\cdot8557}} = \sqrt{23\cdot58} = 4\cdot86.$$

The standard error of the completely adjusted scores (which are equal to Q/E) is

$$\sqrt{(8\times2\cdot5968\times27/32)} = \sqrt{24\cdot62} = 4\cdot96.$$

Thus the gain in information from the recovery of the inter-block information is $24\cdot62/23\cdot58-1$ or $4\cdot4$ % (excluding losses due to inaccuracy of weighting). If inter-litter and intra-litter comparisons had been of equal accuracy, the gain would have been $18\cdot5$ %.

8. Second example

Table 8 gives the arrangement and yields of a tomato trial of 21 varieties arranged in twenty-one blocks of five plots. (I am indebted to the Statistical Department of Iowa State College for the data of this example.)

Table 8. *Arrangement and yields of a tomato variety trial*

Block ...	1		2		3		4		5		6		7	
	s	22·25	*m*	32·00	*e*	51·75	*r*	45·75	*g*	49·25	*j*	59·00	*b*	61·25
	b	51·50	*l*	44·00	*i*	58·50	*c*	37·25	*a*	33·75	*u*	72·50	*p*	47·75
	o	41·56	*j*	52·50	*k*	29·75	*k*	17·50	*f*	45·75	*o*	49·50	*l*	45·00
	g	36·75	*a*	50·75	*n*	70·75	*q*	26·75	*h*	55·25	*e*	46·50	*t*	35·00
	k	21·00	*k*	32·25	*t*	56·00	*h*	37·25	*i*	62·80	*h*	78·00	*h*	53·00
Total		173·06		211·50		266·75		164·50		246·80		305·50		242·00

Block ...	8		9		10		11		12		13		14	
	e	31·25	*u*	51·00	*k*	24·75	*n*	55·25	*d*	36·50	*o*	38·75	*s*	28·00
	c	35·25	*r*	49·00	*f*	47·25	*o*	37·75	*t*	43·50	*t*	42·25	*q*	40·50
	a	40·50	*a*	40·50	*u*	50·50	*a*	39·50	*q*	35·25	*c*	42·50	*l*	50·25
	b	58·50	*t*	47·75	*p*	58·75	*q*	46·75	*g*	44·00	*f*	50·25	*f*	62·50
	d	45·50	*s*	38·50	*d*	51·25	*p*	48·25	*j*	51·75	*m*	30·75	*e*	41·00
Total		211·00		226·75		232·50		227·50		211·00		204·50		222·25

Block ...	15		16		17		18		19		20		21	
	i	65·00	*n*	57·50	*g*	67·00	*f*	74·00	*b*	67·00	*m*	44·75	*j*	74·25
	d	47·50	*g*	55·00	*r*	70·50	*b*	68·25	*u*	56·75	*d*	62·00	*p*	68·50
	o	49·75	*u*	55·50	*m*	46·00	*r*	86·00	*i*	66·00	*n*	76·75	*c*	46·25
	l	51·00	*c*	38·75	*e*	43·00	*n*	93·25	*m*	31·75	*s*	46·75	*s*	50·25
	r	64·50	*l*	51·25	*p*	64·50	*j*	98·12	*q*	49·00	*h*	82·25	*i*	65·50
Total		277·75		258·00		291·00		419·62		270·50		312·50		304·75

Here $v = 21$, $r = 5$, $k = 5$, $b = 21$, $\lambda = 1$, $E = 21/25$. The values of V, T, W, and the adjusted yields are shown in Table 9, and the analysis of variance in Table 10.*

* Labour would have been saved had the yields been rounded off to 1 decimal place before analysis. The fact that three, and only three, of the yields are not exact quarters may also point to the existence of certain errors of transcription.

Table 9. *Calculation of the adjusted yields, tomato trial*

	V	T	$W = 16V - 20T + 4G$	$Y = V + \mu W$
a	205·00	1123·55	+ 1927·92	225·57
b	306·50	1316·18	− 300·68	303·29
c	200·00	1142·75	+ 1463·92	215·62
d	242·75	1244·75	+ 107·92	243·90
e	213·50	1296·50	− 1395·08	198·62
f	279·75	1325·67	− 918·48	269·95
g	252·00	1179·86	+ 1553·72	268·58
h	305·75	1271·30	+ 584·92	311·99
i	317·80	1366·55	− 1127·28	305·77
j	335·62	1452·37	− 2558·56	308·32
k	125·25	1048·31	+ 2156·72	148·26
l	241·50	1211·50	+ 752·92	249·53
m	185·25	1290·00	− 1717·08	166·93
n	353·50	1484·37	− 2912·48	322·43
o	217·31	1188·31	+ 829·68	226·17
p	287·75	1297·75	− 232·08	285·27
q	198·25	1095·75	+ 2375·92	223·60
r	315·75	1379·62	− 1421·48	300·59
s	185·75	1239·31	− 695·28	178·33
t	224·50	1151·00	+ 1690·92	242·54
u	286·25	1293·25	− 166·08	284·48
	5279·73	26398·65	0	5279·74
Divisor		5	5·21 . 16·4 = 6720	

Table 10. *Analysis of variance, tomato trial*

	D.F.	S.S.	M.S.
Blocks (eliminating varieties)	20	7105·99	355·30
Varieties (ignoring blocks)	20	14222·31	—
Error	64	2363·20	36·92
Total	104	23691·50	—

From Table 10 we have

$$w = 0{\cdot}02709, \quad w' = \frac{84}{100 \times 355{\cdot}30 - 16 \times 36{\cdot}92} = 0{\cdot}00240,$$

$$\mu = \frac{w - w'}{84w + 16w} = \frac{0{\cdot}02469}{2{\cdot}314} = 0{\cdot}01067.$$

The standard error of the adjusted yields is $\sqrt{500/2{\cdot}314} = \sqrt{216{\cdot}1} = 14{\cdot}70$. The standard error of the fully adjusted yields would be $\sqrt{219{\cdot}8}$, so that the gain in information from the use of the inter-block information is trivial. If the inter-block and intra-block comparisons were of equal accuracy the gain would be 19·1 %, less losses due to inaccuracies of weighting.

9. General remarks

In both the examples given the gain in information due to the recovery of the inter-block information is small. Cases will arise, however, in which the chosen blocks do not account for much of the general variability, and in such cases the recovery of the inter-block information will lead to an appreciable increase in efficiency. Since this recovery involves little additional work, and the resulting gain cannot in any case be assessed until the analysis of variance (on the lines set out in this paper) is performed, it would appear best to follow this method of analysis in all cases.

In agricultural experiments, however, the gains from the use of inter-block information will not in general be so great as in similar quasi-factorial (lattice) designs, since complete replications cannot (except in special cases) be arranged in compact groups of blocks. For this reason also, cases will arise in which the use of ordinary randomized blocks will be more efficient than the use of incomplete blocks, whereas lattice designs can never be less efficient than ordinary randomized blocks. Nor is it at all easy, except in data from uniformity trials, to determine exactly what is the efficiency of an incomplete block design, relative to an arrangement in ordinary randomized blocks on the same land.

It will be remembered that lattice designs can be analysed as if they were arrangements in ordinary randomized blocks, the errors of the unadjusted yields being correctly estimated by this process. This property does not hold for incomplete block designs (except those which can be arranged in complete replications) and the full analysis must therefore always be performed.

For these reasons incomplete block designs which cannot be arranged in complete replications are likely to be of less value in agriculture than ordinary lattice designs. Their greatest use is likely to be found in dealing with experimental material in which the block size is definitely determined by the nature of the material. A further use is in co-operative experiments in which each centre can only undertake a limited number of treatments. Here the use of balanced incomplete blocks (each centre forming a block) is frequently much preferable to the common practice of assigning a standard treatment (or control) to each centre.

10. Summary

The recovery of inter-block information in incomplete block designs is discussed, and the method of computation is illustrated by examples.

REFERENCES

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

F. Yates (1936*a*). "Incomplete randomized blocks." *Ann. Eugen., Lond.*, **7**, 121–40.

—— (1936*b*). "A new method of arranging variety trials involving a large number of varieties." *J. Agric. Sci.* **26**, 424–55.

—— (1939). "The recovery of inter-block information in variety trials arranged in three dimensional lattices." *Ann. Eugen., Lond.*, **9**, 136–56.

—— (1940). "Lattice squares." *J. Agric. Sci.* **30**, 672–87.