Discussion Paper No. 7

BALANCED INCOMPLETE BLOCK DESIGNS

Peter Walker
Research Support Group

&

Salvatore Ceccarelli
Forage Improvement Program

The International Center for Agricultural Research in the Dry Areas
(ICARDA)

February,1982

This **Discussion Paper** is published by the International Center for Agricultural Research in the Dry Areas (ICARDA). For details of other ICARDA publications please write to:

Training and Communications Program
ICARDA, P.O.Box 5466, Aleppo, Syria

Preface

This Discussion Paper was written by Peter Walker, a biometrician with the Research Support Group at ICARDA, and Salvatore Ceccarelli, a plant breeder with the Forage Improvement Program.

This paper is intended for scientists involved in plant breeding and selection, and for other scientists and institutions in the Near East and North Africa region and elsewhere, interested in the practical application of experimental design.

CONTENTS

1. Introduction

The randomized complete block design is rightly seen as the norm to be aimed at in the designing of experiments. This is due to its simplicity, its capability of reducing error and the ease with which missing values can be handled. We use more complicated designs in agriculture only when:

(1) it is desired to remove further sources of variability (Latin Square, Graeco-Latin Square, Youden Square),

(2) such designs are demanded by the practical details of field operations (split-plots, criss-cross, etc.), and

(3) there are too many treatments for reasonably homogeneous replications to be laid out in the field, and each replication has to be divided into several smaller blocks (incomplete blocks, confounded factorials).

Under (3), 'confounding' is a well-known device used in cases where the treatments have a factorial structure. When faced with unstructured treatments, particularly with many cultivars of a crop, then we often resort to more general incomplete block designs (usually balanced). This Discussion Paper introduces these balanced, incomplete block designs in general terms.

Plant breeding and selection programs require the screening of large numbers of genotypes (lines, varieties, accessions, etc.), particularly in their early stages. When the amount of available seed does not permit more than one plot for each entry, the "augmented designs" (R.G.Petersen, 1980) probably provide the most suitable solution although other methods for adjusting the treatment observations are available.

If the amount of available seed allows a replicated trial, the "incomplete block designs" introduced by Yates (1936) can give better control of environmental heterogeneity better than do randomized block or latin square designs.

The incomplete block designs may be balanced or partially balanced, the main difference being that in the first case all the comparisons among pairs of treatments have equal precision, and in the second case they do not.

Lattices are probably the best known class of incomplete block designs and in this paper we will introduce the properties and the analysis of balanced lattices, deferring to a future paper the introduction to partially balanced lattices.

2. Design plan

The basic feature of incomplete block designs, either balanced or partially balanced, is that treatments are arranged in blocks, or groups, that are smaller than a complete replication.

When these blocks can be arranged in complete replications the designs are said to be resolvable.

Balanced designs can in theory be constructed for any number of varieties and of plots per incomplete block but in many cases the required number of blocks is too high to find practical application.

The designs can be arranged either in randomized blocks or in quasi-latin squares.

To illustrate the properties of balanced incomplete block designs the following notation will be used:

v = number of varieties

r = number of replications

b = total number of incomplete blocks

k = number of varieties per incomplete block

The following relations must hold:

$$v\ r = k\ b$$

and

$$\lambda\ (v - 1) = r\ (k - 1)$$

where $b \geqslant v$, and λ (an integer) is the number of times two varieties occur together within an incomplete block. λ is equal for all pairs of varieties if the design is balanced. Consider the following example:

Blocks	REP I	Blocks	REP II	Blocks	REP III	Blocks	REP IV
(1)	A B C	(4)	A D G	(7)	A E I	(10)	A H F
(2)	D E F	(5)	B E H	(8)	G B F	(11)	D B I
(3)	G H I	(6)	C F I	(9)	D H C	(12)	G E C

Note that:

$$v = 9$$
$$r = 4$$
$$b = 12$$
$$k = 3$$

so that:

$$9 \times 4 = 3 \times 12$$
$$\lambda \times 8 = 4 \times 2$$

and therefore

$$\lambda = 1$$

Every pair of varieties will be found to occur once, and only once, in the same block. For example variety A occupies the same block with varieties B and C in the first replication, with varieties D and G in the second, with varieties E and I in the third, with varieties H and F in the fourth.

This design in fact belongs to the so called balanced lattices for which the main restriction is that the number of varieties (v) must be an exact square while the number of varieties per incomplete block (k) is the corresponding square root. Balanced lattices exist for all values of (k) from 2 to 9, with the exception of 6.

For $k > 6$ the full balanced design is often too large to be practical. In such a case we can resort to laying down two (simple lattice) or three (triple lattice) replicates. These are examples of partially balanced incomplete blocks.

The term "lattice" is also used for certain other types of incomplete block design (partially balanced), i.e. rectangular lattice ($v = k\ (k-1)$) and cubic lattice ($v = k^3$).

The two characteristics, $v = k^2$ and $\lambda = 1$, allow an easy determination of the size of the balanced experiment. For example if $v = 64$, k will be equal to $\sqrt{64} = 8$ varieties per incomplete block; in order to have $\lambda = 1$ the number of replications must be:

$$r = \frac{(v - 1)}{k - 1} = \frac{63}{7} = 9$$

and the total number of incomplete blocks will be:

$$b = \frac{v \times r}{k} = \frac{64 \times 9}{8} = 72$$

or more simply,

$$r \times k = 9 \times 8 = 72$$

As a consequence

$$v\ r = k\ b = 576$$

which is the total number of plots.

3. Advantages and disadvantages

The main advantage of lattice designs is that a large number of varieties can be compared with relatively small blocks (the incomplete blocks), with a gain in precision.

The disadvantages of lattice designs are:

a) the analysis is more complex especially if missing plots occur;

b) lattice designs are not available for all values of v, r and k;

c) the designs are more difficult to randomize.

The use of prepared arrangements (Clem and Federer, 1950; Cochran and Cox, 1957) and of a computer program greatly simplify the layout and the analysis of lattice experiments. Moreover, if a design for the required number of varieties is not available, the addition or subtraction of a few varieties is usually all that is required to obtain a lattice of the desired size.

4. Statistical analysis for balanced lattice

The statistical analysis can be illustrated by an experiment with 16 accessions of *Medicago rigidula* carried out in 1979-80 by the Forage Program.

In this case $v = 16$

$$k = \sqrt{16} = 4$$

In order to have $\lambda = 1$ $\qquad \lambda = \dfrac{r(k-1)}{(v-1)} = \dfrac{3r}{15}$

Therefore r must be 5

Since $v\,r = k\,b$ we have $\qquad b = \dfrac{vr}{k} = \dfrac{16 \times 5}{4} = \dfrac{80}{4} = 20$

The arrangement of the 16 varieties in 20 incomplete blocks, each containing 4 (k) varieties in 5 (r) replications, is shown in Table 1.

The steps in the analysis are as follows (algebraic formulae are generalized for k x k lattices with r = k + 1 replications):

a) Calculate the totals, the replication totals, the grand total G and the variety totals T (the last ones are reported in the second column of Table 2);

b) For each variety, calculate the total B_t for all blocks in which the variety appears. For example, for variety H, this is equal to

$$B_t = 412.0 + 121.4 + 423.8 + 206.1 + 146.8 = 1310.1$$

Check : The sum of the B_t values should be k times the sum of the T values. The B_t values are reported in the third column of Table 2.

c) For each variety calculate the quantity

$$W = kT - (k + 1)\ B_t + G$$

For example, for variety H this is equal to

$$W = 4 \times 401.9 - 5 \times 1310.1 + 4075.3 = -\ 867.6$$

The W values are reported in the fourth column of Table 2. The sum of W values must be zero.

d) Calculate the total sum of squares, the replications sum of squares and the varieties sum of squares in the usual way. These results are reported in Table 3. The sum of squares for blocks within replications, adjusted for varieties effect is:

$$\frac{W^2}{k^3\ (k + 1)} = \frac{(1224.3)^2 + (-366.4)^2 + \ldots\ldots + (-82.4)^2}{320} =$$

$$= 23386.35$$

Table 1: Experimental layout of a 4 x 4 balanced lattice.

block	REP I					block	REP II				
(1)	A	B	C	D		(5)	A	E	I	N	
	54.7	37.0	27.6	37.1	156.4		39.9	23.3	38.1	30.9	132.2
(2)	E	F	G	H		(6)	B	F	J	O	
	86.4	79.6	79.8	166.2	412.0		58.5	76.5	54.4	78.2	267.6
(3)	I	J	K	L		(7)	C	G	K	P	
	60.0	39.7	41.3	45.1	186.1		53.3	20.1	51.8	35.6	160.8
(4)	N	O	P	Q		(8)	D	H	M	Q	
	66.1	37.0	38.8	33.2	175.1		25.9	30.4	25.1	40.0	121.4
					929.6						682.0

block	REP III					block	REP IV				
(9)	A	F	K	Q		(13)	A	O	G	M	
	64.5	56.4	35.5	29.5	185.9		29.1	30.4	42.5	37.3	139.3
(10)	E	B	P	M		(14)	N	B	K	H	
	40.3	48.9	38.6	51.9	179.7		55.9	46.5	48.1	55.6	206.1
(11)	I	O	C	H		(15)	E	J	C	Q	
	146.6	82.6	67.4	127.2	423.8		73.4	39.9	64.2	40.3	217.8
(12)	N	J	G	D		(16)	I	F	P	D	
	37.9	24.1	54.9	33.7	150.6		34.6	49.0	36.6	44.2	164.4
					940.0						727.6

REP V

block					
(17)	A	J	P	H	
	49.8	37.4	37.1	22.5	146.8
(18)	I	B	G	Q	
	106.2	71.3	57.6	53.2	288.3
(19)	N	F	C	M	
	62.2	44.5	44.2	37.3	188.2
(20)	E	O	K	D	
	53.6	49.4	27.6	42.2	172.8
					796.1

G = 4075.3

Table 2: Steps in computation of adjusted varieties means.

Variety	T	B_t	W	Adjusted totals $T + \mu W$	Adjusted means
A	238.0	760.6	1224.3	302.3	60.5
B	262.2	1098.1	- 366.4	243.0	48.6
C	256.7	1147.0	- 632.9	223.5	44.7
D	183.1	765.6	- 979.7	234.5	46.9
E	277.0	1114.5	- 389.2	256.6	51.3
F	306.0	1218.1	- 791.2	264.5	52.9
G	254.9	1151.0	- 660.1	220.2	44.0
H	401.9	1310.1	- 867.6	356.4	71.3
I	385.5	1194.8	- 356.7	366.8	73.4
H	195.5	968.9	12.8	196.2	39.2
K	204.3	911.7	334.0	221.8	44.4
M	196.7	814.7	788.6	238.1	47.6
N	253.0	852.2	826.3	296.4	59.3
O	277.6	1178.6	- 707.3	240.5	48.1
P	186.7	826.8	683.1	222.8	44.6
Q	196.2	988.5	- 82.4	191.9	38.4
	4075.3	16301.2	0.0	4075.4	

e) Calculate the adjustment factor

$$\mu = \frac{E_b - E_e}{k^2 E_b} = \frac{1599.09 - 255.02}{16 \times 1599.09} = 0.0525$$

where E_b and E_e are the mean squares for blocks and intra-block respectively. If E_b is less than E_e, μ is taken as zero and no adjustments are applied. The μ value is then used to calculate the adjusted variety totals as $T + \mu W$, the sum of these values (shown in the fifth column of Tab. 2) must be equal to G, apart from rounding errors.

f) For the t-test calculate the effective error mean square

$$E'_e = E_e \; (1 + k \mu) = 255.02 \; (1 + 4 \times 0.0525) = 308.57$$

This applies only when μ is different from zero. The least significant difference is then calculated in the usual way.

$$L.S.D. = t \times \sqrt{\frac{2 E'e}{r}}$$

using t for $(k - 1) \; (k^2 - 1)$ degrees of freedom.

Table 3: - ANOVA

S.V.	S.S.	D.F. General	D.F. k=4	M.S.
Replications	3402.69	k	4	850.67
Varieties	13070.71	$(k^2 - 1)$	15	871.38
Blocks (adjusted)	23386.35	$(k^2 - 1)$	15	1559.09
Intra-block error	11475.96	$(k - 1) \; (k^2 - 1)$	45	255.02
TOTAL	51335.71	$(k^3 + k^2 + 1)$	79	

g) The precision relative to randomized blocks can be estimated by pooling the mean squares for blocks (adjusted) and the intra-block error. The result is

$$(23386.35 + 11475.96) / 60 = 581.04 \text{ with } 60 \text{ d.f.}$$

and it is an unbiased estimated of the error variance that would have been present if the experiment had been arranged in randomized blocks. The relative precision is then:

$$\frac{581.04}{308.57} = 188 \%$$

5. <u>Missing values</u>

Since lattice designs are used in large experiments, missing data tend to be more common than with small experiments. Analysis of lattice experiments with incomplete data has been developed by Cornish (1943), and, as might be expected, computations are lengthy. A less accurate method is to compute missing data with the following "intra-block" formula:

$$x = \frac{k^2 T + k(k + 1) B - R + G - k T_b - k B_t}{k (k - 1)^2}$$

Where T, B and R are the totals for variety, block and replication which contain the missing value, and G is the grand total.

T_b = Total (over all replications) of all varieties that appear in the block with the missing value;

B_t = Total of all blocks in which the variety with the missing value appears.

Suppose that the observation 35.5 for variety K in block 9 of replication III is missing. The totals for variety K, for block 9 and for replication III would have been:

$$T = 168.8$$
$$B = 150.4$$
$$R = 904.5$$

and the grand total $G = 4039.8$

The other quantities needed are:

$$T_h = 238.0 + 306.0 + 168.8 + 196.2 = 909.0$$

$$B_t = 186.1 + 160.8 + 150.4 + 206.1 + 172.8 = 876.2$$

Hence

$$X = \frac{(16)(168.8)+(20)(150.4)-904.5 + 4039.8-(4)(909)-(4)(876.2)}{36}$$

$$= 47.3$$

The analysis of variance is now computed in the usual way except that one degree of freedom is lost for the total sum of squares and for the intra-block error.

If an appreciable number of plots are missing it will probably be necessary to analyse the results as a randomized block design (this is legitimate).

REFERENCES

Clem, M.A. and Federer, W.T. (1950). Random arrangements for lattice designs.
Iowa Agricultural Experiment Station Special Report No. 5.

Cochran, W.G. and Cox, G.M. (1957). Experimental designs. pp. 428 - 432.
John Wiley & Sons, Inc., New York.

Cornish, E.A. (1943). The recovery of intra-block information in quasifactorial designs with incomplete data. 1. Square, triple, and cubic lattices.
Australian Council for Science and Industry Research Bulletin 158.

Petersen, R.G. (1980). Augmented designs for International Yield Trials.
ICARDA Discussion Paper No. 3, July 1980.

Yates, F. (1936). A new method of arranging variety trials involving large numbers of varieties.
J. Agric. Sci., 26, 424 - 455.