

# ON AUGMENTED DESIGNS<sup>1</sup>

by

W. T. Federer and D. Raghavarao<sup>2</sup>

Cornell University

BU-480-M

August 1973

## SUMMARY

When some treatments (checks) are replicated  $r$  times and other treatments (new treatments) are replicated once, an augmented design may be used. These designs may be minimum variance designs for estimating contrasts of check effects, of new variety effects, of new variety versus checks, or of all check and new varieties simultaneously. In this paper optimal augmented block and optimal augmented row-column designs for estimating contrasts of new treatments are presented.

## 1. INTRODUCTION

When new varieties or strains are developed in a plant improvement program, sufficient material is often not available for planting more than one experimental plot or unit of the new variety at a single location; and, in some cases, it may be undesirable to lay out more than one experimental unit for the treatment under consideration. In some plant breeding investigations even though one plot of a new variety is laid out at a single location, the new variety may be planted at a number of locations, with the standard or check varieties being replicated  $r$  times at each location. Federer [1956, 1960, 1961, 1963, 1972],

---

<sup>1</sup> This work was partially supported under an NIH Research Grant No. 5-R01-GM-05900

<sup>2</sup> On leave from Punjab Agricultural University (India).

Steel [1958], and Searle [1965] introduced a class of designs, called augmented designs, to handle this situation.

Experimental designs could be constructed for variance-optimality

- (i) among new variety yields,
- (ii) among check variety yields,
- (iii) between check and new variety yields, or
- (iv) among all check and new varieties simultaneously.

In this paper we concentrate on one-way and two-way elimination of heterogeneity designs which minimize the variance among the new varieties. These designs are in the class of augmented designs.

## 2. AUGMENTED DESIGNS - GENERAL THEORY

Suppose that  $v^*$  new varieties are to be tested and that sufficient seeds or plants are available to plant only single replicates of each variety. Furthermore, suppose that  $v$  other varieties, called standard or check varieties, are available in such quantities that  $r$  replications of each variety may be planted. The  $v + v^*$  varieties included in a particular experiment are then laid out in an appropriate experimental design for controlling the heterogeneity effects in the experimental area. Sufficient replications of the check varieties are included to have sufficient degrees of freedom for estimating the experimental error variance and for estimating the effects of the varieties and of the effects of the blocking used to control the heterogeneity.

The statistical analysis for experiment designs in which  $v$  check varieties have been replicated  $r$  times (or even  $r_i$  times for treatment  $i$ ) and in which  $v^*$  new varieties have been replicated once, may be carried out in the following two ways:

- (a) The trial on  $v + v^*$  varieties may be analyzed using standard methods for disproportionate numbers in the subclasses; then, contrasts among the check varieties, among the new varieties, and among the checks and new varieties may be made.
- (b) A statistical analysis is performed on the check variety yields only and blocking effects, a general mean effect, and check variety effects are estimated; an estimate of the experimental variance is obtained. Then, the estimated new variety means or effects are obtained and the varietal contrasts are made as in (a).

Though methods (a) and (b) might appear to result in different estimators for the effects and experimental error variance, it can be shown that this is not the case. Let  $\underline{y}$  be an  $n \times 1$  observational vector with

$$\left. \begin{aligned} E(\underline{y}) &= X_{11}\underline{\beta} \\ V(\underline{y}) &= \sigma^2 I_n \end{aligned} \right\}, \tag{2.1}$$

where  $E(\cdot)$  and  $V(\cdot)$  denote the expected value and the dispersion matrix respectively of the quantity inside the parentheses,  $I_n$  is the identity matrix of order  $n$ ,  $\underline{\beta}$  is a  $p \times 1$  column vector of unknown parameters,  $\sigma^2$  is an unknown scalar, and  $X_{11}$  is an  $n \times p$  matrix with known coefficients. Let  $\underline{z}$  be another  $m \times 1$  observational vector with

$$\left. \begin{aligned} E(\underline{z}) &= X_{21}\underline{\beta} + X_{22}\underline{\gamma} \\ V(\underline{z}) &= \sigma^2 I_m \end{aligned} \right\}, \tag{2.2}$$

where  $\underline{y}$  is a  $q \times 1$  column vector of another set of parameters and  $X_{21}$  and  $X_{22}$  are  $m \times p$  and  $m \times q$  matrices, respectively, of known coefficients. We assume that  $m = q$  and that the rank of  $X_{22}$  is equal to  $q$ .

Then  $\underline{y}$  can be estimated either by estimating  $\underline{\beta}$  from (2.1) and substituting in (2.2), which is method (b), or by using the combined set of equations in (2.1) and (2.2) to estimate the varietal effects, which is method (a). By the method of least squares, the estimated vector  $\hat{\underline{\beta}}^{(1)}$  of  $\underline{\beta}$  from (2.1) is:

$$\hat{\underline{\beta}}^{(1)} = (X'_{11} X_{11})^{-1} X'_{11} \underline{y}, \quad (2.3)$$

where  $A^{-}$  denotes a generalized inverse of  $A$  (see Searle [1971], e.g.). Substituting the value  $\hat{\underline{\beta}}^{(1)}$  for  $\underline{\beta}$  in (2.2), we obtain the estimate  $\hat{\underline{y}}^{(1)}$  of  $\underline{y}$  as follows:

$$\hat{\underline{y}}^{(1)} = X_{22}^{-1} [\underline{z} - X_{21} (X'_{11} X_{11})^{-1} X'_{11} \underline{y}]. \quad (2.4)$$

Alternatively, from the combined set of equations (2.1) and (2.2), that is,

$$\begin{bmatrix} \underline{y} \\ \underline{z} \end{bmatrix} = \begin{bmatrix} X_{11} & 0_{n,q} \\ X_{21} & X_{22} \end{bmatrix} \begin{bmatrix} \underline{\beta} \\ \underline{y} \end{bmatrix}, \quad (2.5)$$

where  $0_{n,q}$  is the  $n \times q$  null matrix, the estimate  $(\hat{\underline{\beta}}^{(2)'} \hat{\underline{y}}^{(2)'})'$  of  $(\underline{\beta}' \underline{y}')$  is obtained by the method of least squares described below. Now,

$$\begin{bmatrix} X'_{11} X_{11} + X'_{21} X_{21} & X'_{21} X_{22} \\ X'_{22} X_{21} & X'_{22} X_{22} \end{bmatrix} \begin{bmatrix} \hat{\beta}^{(2)} \\ \hat{\gamma}^{(2)} \end{bmatrix} = \begin{bmatrix} X'_{11} \underline{y} + X'_{21} \underline{z} \\ X'_{22} \underline{z} \end{bmatrix} \quad (2.6)$$

Since the rank of  $X_{22}$  is  $q$ , there exists a  $q \times p$  matrix  $L$  such that

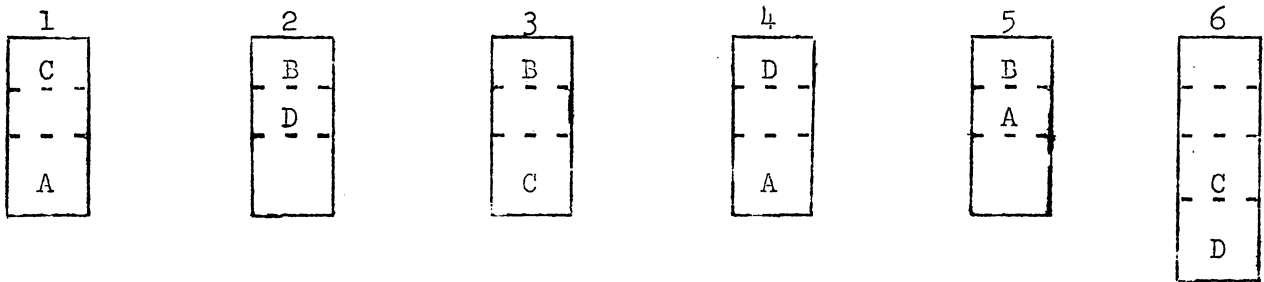
$$X_{21} = X_{22} L. \quad (2.7)$$

After substituting this value in (2.6) and eliminating  $\hat{\beta}^{(2)}$ , we obtain  $\hat{\gamma}^{(2)} = \hat{\gamma}^{(1)}$ . Thus, methods (a) and (b) as described previously lead to the same estimates of the varietal effects for both check and new varieties.

The error mean squares from both methods can also be easily verified to be the same. The inferences drawn from both methods are thus identical. It is recommended that method (b) be used for the statistical analyses as it minimizes the algebra and the computations. Use of experiment designs with known statistical analyses further minimizes the algebraic and the numerical computations.

The randomization procedure (Federer, [1956, 1961]) is to follow the randomization procedure for the known design for the check varieties; then the new varieties are randomly allotted to the remaining empty experimental units. An example will suffice to clarify the procedure further. Suppose that  $v = 4$  check varieties (A,B,C,D) and  $v^* = 7$  new varieties (1,2,3,4,5,6,7) are to be laid out in an augmented balanced incomplete block design. Following the randomization procedure for the check varieties in a balanced incomplete block design for  $v = 4$ ,  $b = 6$ ,  $r = 4$ ,  $k = 2$ , and  $\lambda = 1$ , the experimental plan might be as follows:

block number



There are seven empty experimental units in the above lay-out. Now, randomly assign the numbers 1 to 7 to the new varieties and fill in the seven experimental units in numerical order. The result is the plan used for the experiment.

3. AUGMENTED DESIGNS ELIMINATING HETEROGENEITY  
IN ONE DIRECTION

Before proceeding to a discussion of augmented designs with one-way elimination of heterogeneity, the concept of a linked block design is needed. These designs are defined (cf. Youden [1951]) as follows: A linked block design (LB) is an arrangement of  $v$  treatments in  $b$  blocks each of size  $k$ , such that

- (i) every treatment occurs at most once in a block,
- (ii) every treatment occurs in  $r$  blocks, and
- (iii) every pair of blocks has exactly  $\mu$  varieties in common.

The constants  $v, k, r, b$ , and  $\mu$  are called the parameters of the linked block design. For further details, the reader is referred to the paper by Roy and Laha [1956].

Augmented designs eliminating heterogeneity in one direction are called augmented block designs. Since interest lies in estimating the block effects, and consequently the new variety effects, from that portion of the experiment on the check varieties, it would be desirable to select augmented block designs for which one or more optimality criteria (e.g., see Kiefer [1958]) for estimating block effects hold. The randomized complete block design and the incomplete block designs for which the variances of differences of block effects are all equal, that is the block effects are variance-balanced (also pairwise or combinatorially balanced in this case), (e.g., see Raghavarao [1971]), are such designs for which one or more of the optimality criteria hold. One class of incomplete block designs for which the above holds is the symmetrically balanced incomplete block designs for  $v = b$  and  $r = k$ . For these designs  $N'N = NN' = c_1 I_v + c_2 J_v$  where  $N$  is the treatment-block design matrix,  $c_1$  and  $c_2$  are constants,  $I_v$  is the identity matrix of order  $v$  and  $J_v$  is a square matrix of ones of order  $v$ . Since the block effects have minimum variance, the variance of a difference between two new varieties in different blocks will also be minimized.

Likewise if the dual of a balanced incomplete block design with parameters  $v', b', r', k'$  and  $\lambda'$  for which a treatment either occurs or does not occur in an incomplete block, is used, then the design is a linked block design and the block effects are pairwise and variance balanced. For example, the design matrix of the dual of the balanced incomplete block for  $v' = 4$ ,  $b' = 6$ ,  $r' = 3$ ,  $k' = 2$ , and  $\lambda' = 1$  is the following linked block design:

Treatment	Block			
	1	2	3	4
1	1	1	0	0
2	0	0	1	1
3	1	0	1	0
4	0	1	0	1
5	1	0	0	1
6	0	1	1	0

The variance-covariance matrix of the block effects is  $(\sigma^2 k' / \lambda' v') I_v$ , where  $\sigma^2$  is the intrablock error variance. The above design is variance optimal for estimating block effects. The parameters of the above linked block design are  $v = 6$ ,  $b = 4$ ,  $r = 2$ ,  $k = 3$ , and  $\mu = 1$ .

### 3.1 Augmented block design when check varieties are in a randomized complete

block design. Let the  $v$  check varieties be randomly allotted to the experimental units within each of the  $r$  blocks which are complete for the check varieties and let the yield of the  $i$ -th check variety in the  $j$ -th block be  $y_{ij}$ . Let the  $v^*$  new varieties be randomly arranged such that  $l_j$  new varieties occur in the  $j$ -th block, let  $\sum_{j=1}^r l_j = v^*$ , and let  $z_{ij}$  be the yield of the  $i$ -th new variety in the  $j$ -th block in which it occurs. Let us, for simplicity, assume that the block effects are nonrandom effects. Note that for the  $z_{ij}$  yields  $j = 1, 2, \dots, r$  and  $i = 1, 2, \dots, v^*$ . Using the check variety yields alone, run the statistical analysis for a randomized complete block design, and let EMS be the estimated error mean square obtained from this analysis. Then for any pair of those new varieties, say  $i$  and  $i'$ , occurring together in the same block, the difference in their effects is estimated by  $z_{ij} - z_{i'j}$  with a standard error of  $\sqrt{2 \text{ EMS}}$ ; the difference between two new varieties,  $i$  and  $i'$ , occurring in blocks  $j$  and  $j'$  respectively is estimated by  $z_{ij} - z_{i'j} - (\bar{y}_{.j} - \bar{y}_{.j'})$  with a standard error of  $\sqrt{2(r+1) \text{ EMS}/r}$ , where  $\bar{y}_{.j}$  is the mean of the  $j$ -th block computed on check variety yields only.

### 3.2 Augmented block design when the check varieties are in a linked block design.

Let the  $v$  check varieties be arranged in an LB design with parameters  $v, b, r, k$ , and  $\mu$  and let  $y_{ij}$  be the yield of the  $i$ -th check variety in the  $j$ -th block for  $i = 1, 2, \dots, v$  and  $j = 1, 2, \dots, b$ . Let  $B_j$  be the  $j$ -th block total, let  $T_i$  be the  $i$ -th check variety total, and let  $G$  be the total of the  $vr = bk$  check variety yields. Let  $P_i$  be the difference between the block total  $Y_{.j}$  and  $1/r$  times the



sum of the totals of treatments appearing in the  $j$ -th block. The  $j$ -th block effect is computed as  $rP_j/\mu b$  and an analysis of variance of the check variety yields is given in Table 3.1.

Table 3.1  
Analysis of Variance for  
Check Variety Yields From An LB Design

Source of variation	df	Sum of Squares	Mean Square
Correction for the mean	1	$G^2/rv$	
Check varieties (ignoring block effects)	$v-1$	$v \sum_{i=1}^v T_i^2/r - G^2/rv = S_1$	
Block effects (adjusted for check variety effects)	$b-1$	$r \sum_{j=1}^b P_j^2/\mu b = S_2$	
Intrablock error	$bk-v-b+1$	by subtraction	EMS
Total	$bk$	$\sum_{i,j} y_{ij}^2$	
Check varieties (adjusted for block effects)	$v-1$	$S_1 + S_2 - \sum_{j=1}^b B_j^2 + G^2/rv$	

Let  $v^*$  of the new varieties be arranged such that  $l_j$  of them occur in the  $j$ -th incomplete block, and  $\sum_{j=1}^b l_j = v^*$ . Let  $z_{ij}$  be the yield of the  $i$ -th new variety in the  $j$ -th block. For nonrandom block effects the difference between the  $i$ -th and  $i'$ -th new varieties in the  $j$ -th block is estimated by  $z_{ij} - z_{i'j}$  with a standard error of  $\sqrt{2 \text{ EMS}}$ ; the difference between the  $i$ -th and  $i'$ -th new varieties appearing in block  $j$  and  $j'$  respectively, is estimated as  $z_{ij} - z_{i'j} - \frac{r}{\mu b}(P_j - P_{j'})$  with a standard error of  $\sqrt{2(r + \mu b) \text{ EMS}/\mu b}$ .

4. AUGMENTED DESIGNS ELIMINATING HETEROGENEITY  
IN TWO DIRECTIONS

A Youden design as originated by Youden [1937] may be defined as follows:

Definition 4.1. A Youden design is an arrangement of  $v$  symbols in a  $k \times v$  rectangular array such that exactly one of the symbols appears in the  $vk$  cells and

- (i) every symbol appears in each of the  $k$  rows,
- (ii) every symbol occurs at most once in a column,
- (iii) every symbol occurs in exactly  $r$  columns, and
- (iv) every pair of symbols occurs together in exactly  $\lambda$  columns.

$v = b$ ,  $k = r$ , and  $\lambda$  are called the parameters of the Youden design. These designs are used to remove heterogeneity from two directions or sources; their variance optimality for estimating the  $v$  treatment effects has been determined by Kiefer [1958]. Since treatments and columns are orthogonal to rows and since the columns and treatments form a symmetrical BIB design, the Youden design is also variance optimal for estimating the column effects.

From a Youden design we may interchange the roles of the  $k$  rows and the  $v$  treatments to obtain a  $v \times v$  square array where the  $k$  check varieties are each replicated  $v$  times.  $v^* = v(v - k)$  new varieties are included in the remaining empty cells. To illustrate, take the Youden design with parameters  $v = 7 = b$ ,  $k = 3 = r$ , and  $\lambda = 1$  given below:

column

row	1	2	3	4	5	6	7
A	1	2	3	4	5	6	7
B	2	3	4	5	6	7	1
C	4	5	6	7	1	2	3

Interchange the roles of the rows and the treatments to obtain the following  $7 \times 7$  square for  $k = 3$  check varieties (A,B,C):

	column						
row	1	2	3	4	5	6	7
1	A				C		B
2	B	A				C	
3		B	A				C
4	C		B	A			
5		C		B	A		
6			C		B	A	
7				C		B	A

The  $49 - 21 = 28$  empty cells of the above plan may be filled with 28 new varieties 1,2,...,28 each with one replicate as follows:

	column						
row	1	2	3	4	5	6	7
1	A	1	2	3	C	4	B
2	B	A	5	6	7	C	8
3	9	B	A	10	11	12	C
4	C	13	B	A	14	15	16
5	17	C	18	B	A	19	20
6	21	22	C	23	B	A	24
7	25	26	27	C	28	B	A

Let  $y_{hij}$  be the yield of the  $i$ -th check variety in the  $h$ -th row and  $j$ -th column, and let  $z_{hij}$  be the yield of the  $i$ -th new variety in the  $h$ -th row and  $j$ -th column. Let  $y_{h..}$  be the  $h$ -th row total from check variety yields,  $y_{..j}$  be the  $j$ -th column total from check variety yields,  $y_{.i}$  be the  $i$ -th check variety total yield,  $y_{...}$  be the grand total of the check variety yields,  $h, j = 1, 2, \dots, v$ , and  $i = 1, 2, \dots, v$  for  $y_{hij}$  yields and  $i = v + 1, v + 2, \dots, v + v^*$  for the  $z_{hij}$  yields. Let the adjusted row totals be obtained as  $R_h = y_{h..} - (\text{sum of column totals in which a check variety appeared in row } h)/k$  and let the adjusted column totals be obtained as  $C_j = y_{..j} - (\text{sum of row totals in which a check variety appeared in column } j)/k$ . An analysis of variance on the check variety yields is given in Table 4.1. Note that the check variety effects are orthogonal to both row and column effects.

Table 4.1  
 Analysis of Variance of Check  
 Variety Yields When The Check Varieties Are  
 The Rows of a Youden Design

Source of variation	df	Sum of Squares	Mean Square
Correction for the mean	1	$y_{...}^2 / kv$	
Columns (ignoring row effects)	$v-1$	$\sum_{j=1}^v y_{..j}^2 / k - y_{...}^2 / kv$	
Rows (adjusted for column effects)	$v-1$	$k \sum_{h=1}^v R_h^2 / \lambda v$	
Check varieties	$k-1$	$\sum_{i=1}^k y_{.i}^2 / v - y_{...}^2 / kv$	
Error	$vk-2v$ $-k+2$	subtraction	EMS
Total	$vk$	$\sum_{h,i,j} y_{hij}^2$	-

Assuming nonrandom row and column effects, the difference in yields between new varieties  $i$  and  $i'$  occurring in the same row and in columns  $j$  and  $j'$  for  $j \neq j'$  is  $z_{hij} - z_{hij'} - \frac{k}{\lambda v} (C_j - C_{j'})$  with an estimated standard error of  $\sqrt{2(\lambda v + k)EMS/\lambda v}$ . If new varieties  $i$  and  $i'$  occur in the same column and in rows  $h$  and  $h'$ , the difference in yield is estimated by  $z_{hij} - z_{h'ij} - \frac{k}{\lambda v} (R_h - R_{h'})$  with an estimated standard error of  $\sqrt{2(\lambda v + k)EMS/\lambda v}$ . If new varieties  $i$  and  $i'$  occur in rows  $h$  and  $h'$  and columns  $j$  and  $j'$ , respectively, for  $h' \neq h$  and  $j' \neq j$ , the difference between their yields is estimated by  $z_{hij} - z_{h'i'j'} - \frac{k}{\lambda v} (R_h - R_{h'} + C_j - C_{j'})$  with an estimated standard error of  $\sqrt{2[\lambda v + 2k + n_{ij'} + n_{i'j}]EMS/\lambda v}$ . When new varieties  $i$  and  $i'$  occur in cells  $(i,j)$  and  $(i',j')$  for  $i \neq i'$ ,  $j \neq j'$ , where  $n_{ij}$  is equal to one if  $n_{i'j}$  and/or  $n_{ij'}$  contains a check variety and is equal to zero otherwise.

It should be noted that the above design is variance optimal for estimating the variances between new treatments and for eliminating heterogeneity in two directions. Also, the design is variance optimal for check varieties since these effects are orthogonal to both row and column effects.

In the event that the experimental area and material is not suitable for a design of the above type, then a suitable two-way  $k$  row  $\times$   $v$  column design may be used to generate a  $v \times v$  square to accommodate the new varieties under investigation. Some of the designs given by Agrawal (1966a, 1966b, 1966c), Federer [1972], and Hedayat and Raghavarao [1973], may suffice for this purpose for the check varieties with the empty cells being filled with the new varieties. For simplicity of analysis the generalized row-column design should be chosen to have the rows and columns in as nearly a variance balanced arrangement as is possible.

REFERENCES

- Agrawal, H. L. [1966a]. "A method of construction of three factor balanced designs." Jour. Indian Statist. Assoc., 4, 10-13.
- Agrawal, H. L. [1966b]. "Some methods of construction of designs for two-way elimination of heterogeneity - I". Jour. Am. Statist. Assoc., 61, 1153-1171.
- Agrawal, H. L. [1966c]. "Some systematic methods of construction of designs for two-way elimination of heterogeneity". Calcutta Statist. Assoc. Bull., 15, 93-108.
- Federer, W. T. [1956]. "Augmented (or hoonuiaka) designs". Hawaiian Planters' Record, 55, 191-208.
- Federer, W. T. [1960]. Augmented design with two-, three-, and higher-way elimination of heterogeneity. BU-329-M in the Biometrics Unit Mimeo Series, Cornell University. (Also see Abstract in Biometrics 17:166)
- Federer, W. T. [1961]. "Augmented designs with one-way elimination of heterogeneity". Biometrics, 17, 447-473.
- Federer, W. T. [1963]. Procedures and designs useful for screening material in selection and allocation, with a bibliography. Biometrics, 19, 553-587.
- Federer, W. T. [1972]. Construction of classes of experimental designs using transversals in Latin squares and Hedayat's sum composition method. Chapter 7 in Statistical Papers in Honor of George W. Snedecor (edited by T. A. Bancroft), Iowa State University Press, Ames, Iowa.
- Hedayat, A. and Raghavarao, D. [1973]. 3-way BIB designs. BU-477-M in the Biometrics Unit Mimeo Series, Cornell University.
- Kiefer, J. [1958]. "On the nonrandomized optimality and randomized nonoptimality of symmetrical designs". Ann Math. Statist., 29, 675-699.
- Raghavarao, D. [1971]. Constructions and Combinatorial Problems in Design of Experiments. John Wiley and Sons, Inc., New York, London, Sydney, Toronto.
- Roy, J. and Laha, R. G. [1956]. "Classification and analysis of Linked Block designs." Sankhyā, 17, 115-132.
- Searle, S. R. [1965]. Computing formulae for analyzing augmented random complete block designs. BU-207-M in the Biometrics Unit Mimeo Series, Cornell University.
- Steel, R. G. D. [1958]. A class of augmented designs. Math. Res. Center Tech. Summary Report No. 56, University of Wisconsin.
- Youden, W. J. [1937]. "Use of incomplete block replications in estimating tobacco-mosaic virus". Contributions from Boyce Thompson Institute, 9, 317-326.
- Youden, W. J. [1951]. "Linked blocks: a new class of incomplete block designs" (abstract). Biometrics, 7, 124.