EFFICIENCY OF DIFFERENT EXPERIMENTAL DESIGNS WITH SPECIAL REFERENCE TO INTRA-CLASS CORRELATIONS

By Y. S. Kapse

Institute of Plant Industry, Indore

1. General

In field experiments, the character of soil fertility is such that on dividing the field into blocks and further dividing the blocks into plots, the plot yields within a block are more strongly correlated than the yields of plots lying in different blocks. The variation within a block will reflect the natural variation of individual plots treated alike. Let this be σ^2; the variation between blocks, however, arises from larger difference in environmental conditions as well. Let the variation due to environmental conditions be ${\sigma_m}^2$. The sum $\sigma^2 + {\sigma_m}^2$ is the variation of plots taken at random from the entire field. The ratio of the two variances will be

$$\frac{{\sigma_m}^2}{\sigma^2 + {\sigma_m}^2} \tag{1.1}$$

This ratio is termed intra-class correlation.

In order to calculate intra-class correlation directly from a table of analysis of variance of a field experiment having 'k' individuals per group, the formula is

$$\frac{M_{\bar{x}} - M}{M_{\bar{x}} + (K-1)\, M} \tag{1.2}$$

where

$M_{\bar{x}}$ = mean square for group means and

M = mean square for individual plots.

With reference to split-plot designs, Cochran (1950) has shown that if σ^2 is the error variance in a randomised block design the expectation of the error variance of main treatments in a split-plot design, on sub-plot basis, will be

$$\sigma^2 [1 + \rho (k - 1)]$$

where k is the number of sub-plots per main plot, and the expectation of the sub-plot error variance will be

$$\sigma^2 (1 - \rho)$$

ρ being the intra-class correlation among sub-plots.

Similar considerations apply to other designs. In the confounded design as against one in simple randomised block, there is an additional statistical control of sub-blocks, so that a complete replicate is divided into two or more smaller blocks; intra-class correlation among plots within-sub-blocks is expected to be higher than that among plots in the whole replicate.

Expressions for calculating intra-class correlation will now be set up for the split-plot and confounded designs. For convenience, we may consider designs with dummy treatments as in a uniformity trial.

2. Intra-Class Correlation in Randomised Block Split-Plot Design

Consider a split-plot design having s sub-plots per main plot, p main plots per block and q blocks. The table of analysis of variance is then as follows:

Variance due to	D.F.	Mean square
Blocks ..	$q-1$	B
Main plots within blocks ..	$q(p-1)$	E_a
Sub-plots within main plots ..	$pq(s-1)$	E_b
Total ..	$pqs-1$	T

Following expression (1.2) above, the group (here main plot) mean square will be

$$B' = \frac{B(q-1) + E_a(p-1)q}{pq-1}$$

The mean square for individual plots (sub-plot here) will be the sub-plot error mean square E_b. Therefore, intra-class correlation for sub-plots within main plots in the randomised block split-plot design will be given by

$$r_1 = \frac{B' - E_b}{B' + (s-1) E_b} \quad (2.1)$$

s being the number of sub-plots per main plot. Had the design been randomised block instead of randomised block split-plot, the table of analysis of variance would have been as given below:

Variance due to	D.F.	Mean square
Blocks ..	$q-1$	B
Plots within blocks ...	$qps-q$	E'
Total ..	$qps-1$	T

The block mean square B will be the group mean square. The mean square for individual plots will, however, be obtained as

$$E' = \frac{q(p-1) E_a + pq(s-1) E_b}{q(ps-1)}$$

The intra-class correlation for the randomised block design without splitting will thus be obtained as

$$r_2 = \frac{B - E'}{B + (ps-1) E'} \quad (2.2)$$

where

p = number of main plots per block and

s = number of sub-plots per main plot.

It will be recalled that relative efficiency of split-plot design is given by the ratio of the error variance with and without splitting; therefore

$$Y = \frac{E'}{E_b} \quad (2.3)$$

The algebraic expressions for r_1 and r_2 and Y are functionally related among themselves along with the constants s, p and q which appear in the table of analysis of variance of split-plot design.

In terms of the total mean square

$$B' = \frac{T(qps - 1) - E_b (s - 1) pq}{pq - 1}$$

Substituting in (2.1) we get

$$r_1 = \frac{\frac{T}{E_b} - 1}{\frac{T}{E_b} - a_1}$$

where

$$a_1 = \frac{s - 1}{qps - 1} \qquad (2.4)$$

r_1 is expressed here in terms of the variance ratio and a constant a_1. Similarly

$$B = \frac{T(qps - 1) - E'(qps - q)}{q - 1}$$

Substituting in (2.2) we get

$$r_2 = \frac{\frac{T}{E'} - 1}{\frac{T}{E'} - a_2}$$

where

$$a_2 = \frac{ps - 1}{qps - 1} \qquad (2.5)$$

from (2.4) and (2.5) we derive the relation

$$Y = \frac{(1 - a_1 r_1)(1 - r_2)}{(1 - a_2 r_2)(1 - r_1)} \qquad (2.6)$$

Thus the relative efficiency of split-plots is expressed in terms of r_1, r_2 and the physical constants of the split-plot design.

In order to verify the results obtained in expression (2.6) the yield data of a uniformity trial on cotton conducted at the Institute of Plant Industry, Indore, in 1933 (Hutchinson and Panse, 1935) was utilised. There were in all 1,280 ultimate units, square in shape, the dimensions of each being 4′ 8″ by 4′ 8″ or 1/2000 acre in area. With a view to make a study of the efficiency of split-plot designs, 5-unit totals, 10-unit totals, 20-unit totals and 40-unit totals were computed to give varying plot size, *viz.*, 0·0025 acre, 0·0050 acre, 0·01 acre and 0·02 acre respectively. In all 20 randomised block split-plot arrangements were superposed on the data; the number of sub-plots per main plot were varied from 2 to 16, the numbers of main plots per block from 2 to 16 and the number of replications from 2 to 32. For lack of space details of these arrangements are not given in this article.

Values of r_1 and r_2 were calculated for all the split-plot arrangements superposed; these values and the difference $(r_1 - r_2)$ are given along with the relative efficiency in Table I.

TABLE I

Relative efficiency of split-plots (Y) *in randomised block split-plot designs superposed on the uniformity trial on cotton, intra-class correlation coefficients* r_1 *and* r_2 *and difference* $r_1 - r_2$

No.	$Y \times 100$	r_1	r_2	$r_1 - r_2$
1	118·9	0·5733	0·5065	+0·0668
2	104·5	0·5733	0·5598	+0·0135
3	122·2	0·5866	0·5065	+0·0801
4	113·8	0·5599	0·5065	+0·0534
5	449·0	0·8069	0·1539	+0·6530
6	463·5	0·8069	0·1037	+0·7032
7	294·8	0·7121	0·1539	+0·5582
8	96·8	0·1037	0·1539	−0·0502
9	329·9	0·7318	0·1180	+0·6138
10	152·8	0·7519	0·6319	+0·1200
11	201·4	0·5599	0·1282	+0·4317
12	294·1	0·7051	0·1317	+0·5734
13	293·6	0·7051	0·1523	+0·5528
14	353·4	0·7519	0·1282	+0·6237
15	352·7	0·7519	0·1523	+0·5996
16	99·8	0·1317	0·1523	−0·0206
17	229·4	0·6156	0·1330	+0·4826
18	230·8	0·6319	0·1523	+0·4796
19	369·0	0·7318	−0·0148	+0·7466
20	111·4	0·1180	−0·0178	+0·1358

Equation (2.6) was used to get the estimated value of the relative efficiency of split-plots for a randomised block split-plot design superposed, using the values of r_1 and r_2 from Table I.

Figure 1 exhibits the relationship between relative efficiency of split-plots and difference $r_1 - r_2$ for the split-plot designs superposed. The trend is that as $r_1 - r_2$ increases, relative efficiency increases and there appears to exist a strong correlation between the two. The correlation coefficient obtained was $+0 \cdot 9413$.

3. INTRA-CLASS CORRELATION IN A CONFOUNDED DESIGN

Consider a confounded design having k plots per sub-block, h sub-blocks per replicate and q replicates. The analysis of variance of the confounded design or a partially confounded experiment with dummy treatments will be

Variance due to	D.F.	Mean square
Sub-blocks ..	$qh-1$	B
Plots within sub-blocks ..	$qh(k-1)$	E
Total ..	$qkh-1$	T

In confounded designs, the sub-block will be considered as the group so that the sub-block mean square will be the group mean square.

The mean square for individual plots will be the error mean square E; the intra-class correlation according to expression (1.2) will be given by

$$r_1 = \frac{B - E}{B + (k-1) E} \tag{3.1}$$

Intra-class correlation in the corresponding randomised block design may be calculated from the following analysis of variance:

Variance due to	D.F.	Mean square
Replicates ..	$q-1$	R
Plots within replicates ..	$qhk-q$	E'
Total ..	$qhk-1$	T

The complete replication without restriction of sub-block will be the group; so that the replicate mean square R will be the group mean square.

The mean square for individual plots will be the error mean square E' when the design is considered as a randomised block one.

From expression (1.2) the intra-class correlation in the corresponding randomised block design will be

$$r_2 = \frac{R - E'}{R + (hk-1) E'} \tag{3.2}$$

hk being the number of plots per replicate. The relative efficiency of sub-blocks in a confounded design is given by the ratio of the error

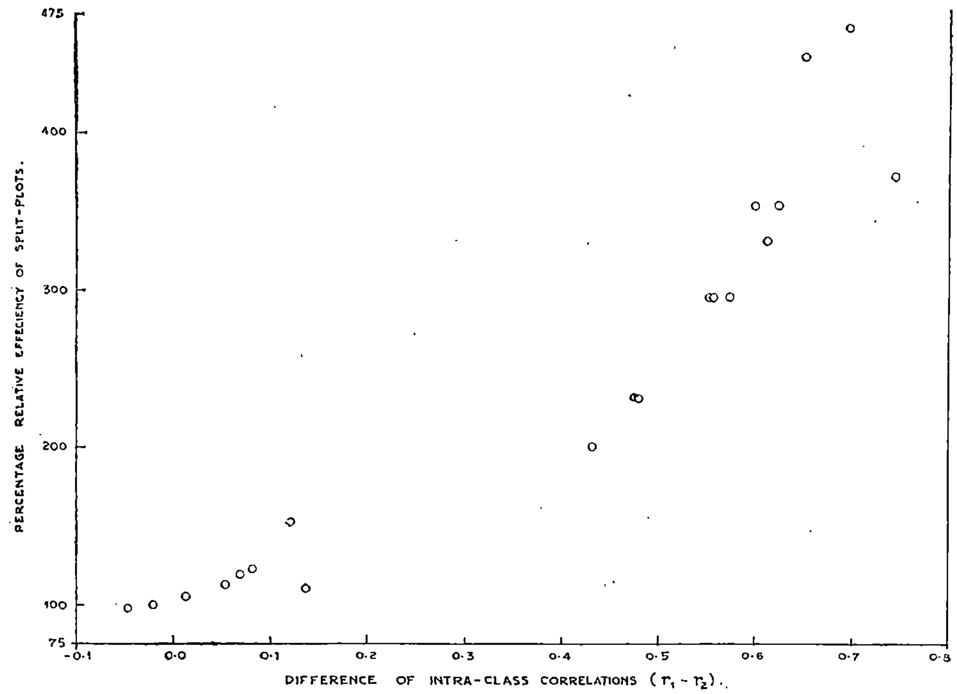

FIG. 1. Showing relationship between percentage relative efficiency of split-plots and difference of intra-class correlations ($r_1 - r_2$) in case of randomised block split-plot designs.

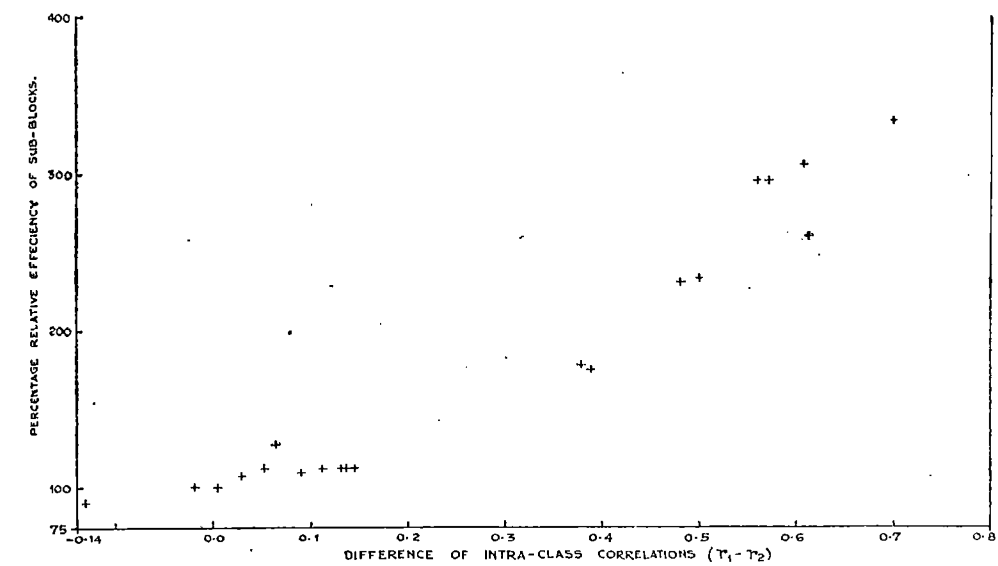

FIG. 2. Showing relationship between percentage relative efficiency of sub-blocks and difference of intra-class correlations ($r_1 - r_2$) in case of confounded designs.

mean squares as

$$Y = \frac{E'}{E} \tag{3.3}$$

A little algebra will bring out the relationship between r_1, r_2 and Y in case of the confounded designs also; we have

$$r_1 = \frac{B - E}{B + (k-1)E} = \frac{\frac{T}{E} - 1}{\frac{T}{E} - a_1}$$

where

$$a_1 = \frac{k-1}{qhk-1} \tag{3.4}$$

r_1 is here expressed in terms of variance ratio and a constant.

The intra-class correlation in the corresponding randomised block design is

$$r_2 = \frac{R - E'}{R + (hk-1)E'} = \frac{\frac{T}{E'} - 1}{\frac{T}{E'} - a_2}$$

where

$$a_2 = \frac{hk-1}{qhk-1} \tag{3.5}$$

Equations (3.4) and (3.5) therefore give

$$Y = \frac{(1 - a_1 r_1)(1 - r_2)}{(1 - a_2 r_2)(1 - r_1)} \tag{3.6}$$

It is interesting to note that equations (2.6) and (3.6) are similar expressing relative efficiency in terms of r_1 and r_2 for the randomised block split-plot and confounded designs respectively; the constants a_1 and a_2 in the two equations are however made up of different physical factors relevant to the particular design of the experiment.

In order to verify the relationship as was done in case of the split-plot designs, data of the uniformity trial on cotton described earlier, was utilised by superposing confounded designs, having different plot size, number of plots per sub-block, block size, and so on. Intra-class correlation values r_1 in the confounded designs, intra-class correlation values r_2 in the corresponding randomised block design and the differences $r_1 - r_2$ were computed for the different dummy experiments. In all 20 confounded arrangements were superposed;

the relative efficiency value and the intra-class correlations r_1, r_2 and difference $(r_1 - r_2)$ are given in Table II.

TABLE II

Relative efficiency of sub-blocks (Y) *in confounded designs superposed on the uniformity trial on cotton, intra-class correlation coefficient r_1, r_2 and difference $r_1 - r_2$*

Sl. No.	$r \times 100$	r_1	r_2	$r_1 - r_2$
1	107·4	0·5866	0·5599	0·0267
2	113·8	0·5599	0·5065	0·0534
3	175·7	0·5065	0·1187	0·3878
4	177·0	0·5065	0·1282	0·3783
5	304·3	0·7121	0·1037	0·6084
6	294·8	0·7120	0·1539	0·5581
7	333·8	0·7121	0·0105	0·7016
8	109·7	0·1037	0·0105	0·0932
9	113·2	0·1539	0·0105	0·1434
10	89·4	−0·0148	0·1180	−0·1328
11	111·3	−0·1180	−0·0178	0·1358
12	100·4	−0·0148	−0·0178	0·0030
13	127·2	0·7051	0·6319	0·0632
14	294·1	0·7051	0·1317	0·5734
15	231·3	0·6319	0·1317	0·5002
16	99·8	0·1317	0·1523	−0·0206
17	230·8	0·6319	0·1523	0·4796
18	112·4	0·1523	0·0203	0·1320
19	112·2	0·1317	0·0203	0·1114
20	259·5	0·6319	0·0203	0·6116

Figure 2 is a graph plotted between relative efficiency of sub-blocks and the corresponding difference of intra-class correlations $r_1 - r_2$, in case of the confounded arrangements superposed. Again the trend was that relative efficiency of sub-blocks increased as $(r_1 - r_2)$ increased. The correlation coefficient obtained between relative efficiency of sub-blocks and $r_1 - r_2$ was 0·9913. It may, therefore, be concluded that the functional relationship as obtained in equation (3.6) between relative efficiency of sub-blocks and the intra-class correlations r_1 and r_2 in case of a confounded experiment is not too far from a linear relationship.

4. INTRA-CLASS CORRELATION IN CASE OF AN INCOMPLETE BLOCK DESIGN

While considering the efficiency of designs commonly employed in agricultural experimentation, some remarks may be made regarding quasi-factorial designs, these designs are frequently used by plant-breeders in order to compare a large number of strains in a single experiment. Arrangements in randomised blocks comprising all the strains will usually be ineffective in eliminating soil fertility differences. For such trials, quasi-factorial designs make possible the use of smaller blocks consisting of only a few plots. The Lattice square may be cited as an example; if k^2 varieties are to be tested, the lattice square experiment will consist of k small blocks, in each of which k varieties are randomised. These designs are essentially confounded designs.

Goulden (1937) while studying the efficiency of pseudo-factorial and incomplete block designs pointed out an empirical relationship between efficiency and $(r_i - r_c)^2$ where

r_i = intra-class correlation of the incomplete block,

r_c = intra-class correlation for the complete block.

Using Fisher's method for estimating intra-class correlation from a table of analysis of variance into between and within classes, Nair (1952) has shown that

$$\frac{E_v}{E_k} = \frac{(1 - c_v)(1 - a_k c_k)}{(1 - c_k)(1 - a_v c_v)} \qquad (4.1)$$

where

c_k = intra-class correlation for the incomplete block design,

c_v = intra-class correlation for the complete block,

and

$$\frac{a_k}{(k-1)} = \frac{a_v}{(v-1)} = \frac{1}{(v_r - 1)}$$

v is the total number of varieties to be tested, being randomised in n small blocks of k plots each, r being used here for the number of replicates. E_k and E_v are the intra-block variances for the incomplete and complete blocks of the design, so that the ratio E_v/E_k is easily recognised as the efficiency ratio.

It should be noted that the expression (4.1) for the incomplete block design resembles closely the expressions (2.6) and (3.6) for the randomised block split-plot and the confounded designs.

5. Summary

The statistical control of sub-blocks in a confounded design and of main plot in a split-plot design brings into play a variation in the intra-class correlation between plots. If r_1 stands for the intra-class correlation coefficient corresponding to the error component of the given design whether split-plot or confounded and r_2 for similar intra-class correlation in the randomised block design, algebraic expression

were found for the relationship between relative efficiency of the design on one hand and on the other, the coefficients r_1 and r_2 together with certain other factors such as number of sub-plots per main plot, number of main plots per block and number of blocks in case of split-plot designs, and number of plots per sub-block, number of sub-blocks per replicate and number of replications in case of confounded designs.

Though the algebraic relationship between relative efficiency and r_1 and r_2 was complex, it was established empirically from actual data that the difference $(r_1 - r_2)$ was strongly correlated with relative efficiency. In split-plot designs, the correlation between relative efficiency and the difference $(r_1 - r_2)$ was 0·9413 while for confounded designs the correlation was as high as 0·9913.

Acknowledgement

The present investigations were carried out under the guidance of Dr. V. G. Panse, Ph.D. (London), F.N.I., F.A.Sc. (formerly Director, Institute of Plant Industry, Indore), Statistical Adviser, Indian Council of Agricultural Research, New Delhi. I wish to express my deep gratitude for the keen interest that he took in guiding me and for granting permission to make use of the data of the uniformity trial on cotton.

The work was done by the author (1952) in partial fulfilment of a thesis for the M.Sc. Degree in Statistics at the Bombay University.

References

1. Cochran, W. G. and Gertrude, M. Cox — *Experimental Designs*, 1950; Chapters II & VII, John Wiley & Sons, Inc., New York.
2. Goulden, C. H. .. "Efficiency in field trials of Pseudo-factorial and incomplete randomised block methods," *Canadian Journal of Research*, 1937, C, **15**, 231–241.
3. Hutchinson, J. B. and Panse, V. G. — "Size, Shape and Arrangement of Plots in Cotton Trials," *Ind. Jour. Agric. Sci.*, 1935, **5**, Part V, 545–554.
4. Kapse, Y. S. .. "A statistical study of the Efficiency of field experimental designs with special reference to agronomic trials," *M.Sc. Thesis*, 1952, Bombay University.
5. Nair, K. R. .. "Relation between Efficiency of Incomplete Block Designs and the Intra-class correlations associated with incomplete and complete Blocks," *Jour. Ind. Soc. Agric. Stat.*, 1952, **4**, No. 2, 149–152.