Relation Between Efficiency of Incomplete Block Designs and the Intra-Class Correlations Associated with Incomplete and Complete Blocks

By K. R. Nair

Forest Research Institute, Dehra Dun

One of the earliest studies of the factors influencing the efficiency of lattice designs was by Goulden (1937). Using mainly uniformity trial data he tried to assess the influence of factors like size and shape of plots, size and shape of blocks and the degree of soil heterogeneity.

Goulden took the intra-class correlation (r_i) of the incomplete blocks as a general measure of soil heterogeneity. His findings with regard to the influence of this factor on efficiency are quoted below:

> "There is a certain degree of correlation between this measure and efficiency ratio in that we must have fairly high values of r_i before high efficiency ratios can be obtained, but in certain cases of high values of r_i the efficiency ratio is low. This arises from the fact that in these cases the intra-class correlation (r_c) for the complete blocks is also high * * * An arbitrary constant which seems from a preliminary examination of the data to be reasonably well correlated with the efficiency ratio is $(r_i - r_c^2)$."

The efficiency ratio referred to by Goulden is the efficiency without recovery of inter-block information. In the case of square lattice designs for k^2 varieties with m mutually orthogonal groupings into k incomplete blocks of k plots each, the expression for this efficiency is

$$\frac{(m-1)\left(k+\frac{w}{w'}\right)}{\{(k+1)(m-1)+m\}}, \qquad (1)$$

where $\frac{1}{w}$ and $\frac{1}{w'}$ are the intra- and inter-block variances for blocks of k plots.

The corresponding expression for the efficiency with recovery of inter-block information can be shown to be

$$\frac{\left(k+\frac{w}{w'}\right)}{\left[(k-m+1)+\frac{m^2 w}{\{(m-1)w+w'\}}\right]} \qquad (2)$$

using the method developed by Nair (1944) for resolvable type of partially balanced incomplete block designs of which the square lattice design is a special case (Bose and Nair, 1939).

By putting $m = 2$, 3 and $(k + 1)$ respectively in (1) and (2) we get the expressions for double, triple and balanced lattices respectively.

To obtain values of (1) and (2) from the field data for a square lattice design with specified values of k and m, we will have to estimate the value of $\frac{w}{w'}$. It becomes particularly simple when uniformity trial data, without superimposing dummy varieties on the various plots, are used for the purpose of this estimation.

We shall discuss this method of estimation in the general case of partially balanced incomplete block designs of the resolvable type where v, the number of varieties, is a multiple of k, the number of plots in the incomplete block. We may then write $v = nk$, where n is a positive integer like v and k.

Let E_k and E_v be the intra-block variances for the incomplete and the complete blocks of the design. Expectation of E_k is $\frac{1}{w}$ and of E_v is $\left\{\frac{n(k-1)}{w} + \frac{(n-1)}{w'}\right\} \Big/ (nk-1)$ so that as estimate of $\frac{w}{w'}$ we may use the quantity

$$\frac{(nk-1)}{(n-1)} \cdot \frac{E_v}{E_k} - \frac{n(k-1)}{(n-1)}. \tag{3}$$

Let c_k and c_v be the intra-class correlations for the incomplete and complete blocks of the design. They correspond to r_i and r_o of Goulden's paper. Since the letter 'r' stands for the number of replications of each variety, I have used the letter 'c' to represent correlation coefficient.

Using the method given by Fisher for estimating intra-class correlation from a table of analysis of variance into *between* and *within* classes, it can be shown that

$$\frac{E_v}{E_k} = \frac{(1-c_v)\ (1-a_k c_k)}{(1-c_k)\ (1-a_v c_v)} \tag{4}$$

where

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

The right-hand side of (4) can be expanded as an infinite series, namely,

$$1 + \frac{k}{(nkr-1)}\{(nr-1)\ c_k - n(r-1)\ c_v\} + \text{higher degree terms in } c_k \text{ and } c_v. \tag{5}$$

By substituting the right-hand side of (4) for E_v/E_k in (3), the estimate of w/w' can be expanded as the infinite series

$$1 + \frac{k(nk-1)}{(n-1)(nkr-1)}\{(nr-1)\ c_k - n\ (r-1)\ c_v\} + \text{higher degree terms in } c_k \text{ and } c_v. \tag{6}$$

The corresponding expressions for square and cubic lattices follow, by putting $n = k$ and k^2 respectively. In the case of the square lattice, r will be multiple of 2, 3 or $(k + 1)$ according as the design is a double, triple or balanced lattice. For cubic lattice r will be a multiple of 3.

By putting $n = k$ and then substituting (6) for w/w' in (1) and (2) we can express the efficiency with or without recovery of inter-block information of square lattice designs in terms of the intra-class correlations associated with incomplete and complete blocks. In general, for any partially balanced incomplete block design, the expression for efficiency with or without recovery of inter-block information involves w/w' besides the parameters of the design. When the design is of the resolvable type, by substituting (6) for w/w' in these expressions we obtain the relationship between efficiency and the intra-class correlations c_k and c_v.

The numerical example given in Table I of Goulden's paper where he has analysed the variance of 810 observations of a uniformity trial as a balanced lattice design for 81 varieties in blocks of 9 plots, and the intra-class correlations calculated for this design in Table II of the paper may be used as a handy illustration of the relationship given by equation (4). From Goulden's Table I we get $E_k = 11912$ and $E_v = 15103$. Table II gives $c_k = 0\cdot4850$ and $c_v = 0\cdot3678$. Remembering that $k = 9$, $v = 81$ and $r = 10$ for this data, we can easily see that each side of equation (4) has the value $1\cdot268$.

In sum, since a mathematical relationship exists between efficiency and intra-class correlation, as shown above, it is unnecessary, and indeed undesirable, to try to establish, by the statistical method, a correlation between efficiency of a design and the two intra-class correlation coefficients c_k and c_v. Although the exact mathematical relationship is complicated, it could be expanded as an infinite series of which terms beyond the second degree in c_k and c_v are likely to be of negligible magnitude. In other words it is preferable to use convenient mathematical approximations to this infinite series than to seek regression relationships between efficiency and arbitrary 'constants' like $c_k - c_v^2$.

REFERENCES

Bose, R. C., and Nair, K. R. "Partially balanced incomplete block designs," *Sankhya*, 1939, **4**, 337–372.

Fisher, R. A. .. *Statistical Methods for Research Workers*, 10th Edition, Oliver and Boyd, Edinburgh, 1946.

Goulden, C. H. .. "Efficiency in field trials of pseudo-factorial and incomplete randomized block methods," *Canadian Journal of Research*, 1937, C, **15**, 231–241.

Nair, K. R. .. "The recovery of inter-block information in incomplete block designs," *Sankhya*, 1944, **6**, 383–390.