and hence has two different characteristic roots; they are

$$e(A) = 1/e_i, \qquad i = 1, \cdots, v-1, \qquad \text{and } e_0(A);$$

$$e_i = 1/e(A) \text{ for } e_i \neq 0.$$

Therefore, C has roots $e_0 = 0$ and $e = 1/e(A)$, the latter of multiplicity $v - 1$.

Using the lemma of Section 1,

$$C = \frac{e}{v}\begin{pmatrix} v-1 & -1 & \cdots & -1 \\ -1 & v-1 & \cdots & -1 \\ \vdots & \vdots & & \vdots \\ -1 & -1 & \cdots & v-1 \end{pmatrix},$$

which says that $\lambda_{ij} = \lambda$, and hence our design is BIB.

THE EFFICIENCY FACTOR OF AN INCOMPLETE BLOCK DESIGN[1]

By Oscar Kempthorne

Statistical Laboratory, Ames, Iowa

1. Summary. It is shown that the efficiency factor of a design is r times the harmonic mean of the latent roots of the reduced intrablock normal equations exclusing the root which is always zero.

2. Properties of reduced normal equations. We consider an incomplete block design with the following properties:

(1) There are rt units arranged in b blocks of k units;

(2) Every one of the t treatments occurs r times;

(3) A treatment occurs once or not at all in a block. (This condition is easily seen to be unnecessary for all the results given below.)

Then, it is well known (see for example Kempthorne [1], pp. 541–543) that the reduced normal equations for intrablock estimates of the treatment effects τ_i are of the form

$$\left(rI - \frac{1}{k}\Lambda\right)(\hat{\tau}) = (Q),$$

where Λ_{ii} equals r for all i, $\Lambda_{ij}(= \Lambda_{ji})$ is the number of blocks which contain both treatments i and j, and Q_j is the total for treatment j adjusted for blocks. Also, the interblock estimates are given by the equation

$$(\Lambda)(\tilde{\tau}) = (R),$$

where R_j is equal to the total of blocks containing treatment j minus r/b times the grand total, and the condition $\sum \tilde{\tau}_j = 0$ is imposed for the nonestimable quantity $\sum \tau$.

Received July 5, 1955.

[1] Journal paper No. J-2819 of the Iowa Agricultural Experiment Station, Ames, Iowa, Project 890.

We first note that the matrix Λ is real and symmetric, so there exists an orthogonal matrix O,

$$OO' = I = O'O,$$

such that

$$O\Lambda O' = D,$$

where D is a diagonal matrix, with, say, $D_{jj} = d_j$, where the d_j's are the latent roots of Λ. It is also obvious that

$$O\left[rI - \frac{1}{k}\Lambda\right]O' = rI - \frac{1}{k}D$$

is a diagonal matrix for which the (jj)th element is $r - d_j/k$.

Now, let $O\tau = \rho$. Then we have

$$\left(rI - \frac{1}{k}D\right)\hat{\rho} = OQ, \qquad D\tilde{\rho} = OR.$$

We note immediately what was perhaps entirely obvious without the above elementary manipulations: that if $\hat{\rho}_j$ exists (i.e., if there is an intrablock estimate of ρ_j), there may or may not be an interblock estimate of ρ_j; *but* if $\hat{\rho}_j$ does not exist (i.e., if no intrablock estimate of ρ_j is given by the design), then $r - d_j/k$ is equal to zero, and hence d_j is not zero and an interblock estimate of ρ_j exists. Similarly, if $\tilde{\rho}_j$ does not exist, then $\hat{\rho}_j$ exists. Of course, in the case of many designs, $\hat{\rho}_j$ and $\tilde{\rho}_j$ both exist for some j.

It is also known that if W and W' denote respectively $1/\sigma^2$ and k/σ_b^2, where σ^2 is the variance within blocks and σ_b^2 is the variance between blocks of block totals, then the combined estimates of τ_j, say τ_j^*, are given by

$$\left[W\left(rI - \frac{1}{k}\Lambda\right) + \frac{W'}{k}\Lambda\right]\tau^* = WQ + \frac{W'}{k}R,$$

so that ρ^* is given by the equation

$$\left[W\left(rI - \frac{1}{k}D\right) + \frac{W'}{k}D\right]\rho^* = WOQ + \frac{W'}{k}OR,$$

or

$$\left[W\left(r - \frac{d_j}{k}\right) + \frac{W'}{k}d_j\right]\rho_j^* = \left[WOQ + \frac{W'}{k}OR\right]_j,$$

or

$$\left[Wr - \frac{d_j}{k}(W - W')\right]\rho_j^* = \left[WOQ + \frac{W'}{k}OR\right]_j.$$

This equation is, among other things, a mathematical presentation of the usual rules for analyzing quasifactorial designs in terms of effects and interactions of pseudo-factors. In that case, the estimates ρ_j^* can be written down by inspection of the design, and one merely has to use $\tau^* = O'\rho^*$ to get the combined estimates of the τ_j's.

3. The efficiency factor of a design. The efficiency factor (EF) of an incomplete block design is defined to be

$$\text{EF} = \frac{\text{mean variance of treatment differences in complete block design}}{\text{mean variance of intrablock estimates of treatment differences in incomplete block design}} \times \frac{\text{within-block variance in incomplete block design}}{\text{within-block variance in complete block design}},$$

Alternatively, the efficiency factor may be defined as the first factor of the above right-hand side, assuming the within-block variances are the same in the two designs. The first factor on the right-hand side above gives the relative efficiency of the design. The efficiency factor does not depend on the actual within-block variances and is purely a property of the design.

We now proceed to get the relationship of the efficiency factor of the design to the d_j, which are the latent roots of the matrix Λ. We have the identity,

$$\frac{1}{t(t-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{t} (x_i - x_j)^2 = \frac{2}{(t-1)} \sum_{i=1}^{t} (x_i - \bar{x})^2,$$

Hence,

$$\sum_{\substack{i,j \\ i \neq j}} (\hat{\tau}_i - \hat{\tau}_j)^2 = 2t \sum_j \hat{\tau}_j^2,$$

when the condition $\sum_j \hat{\tau}_j = 0$ is imposed to give unique estimates. But,

$$\sum_j \hat{\tau}_j^2 = \hat{\tau}'\hat{\tau} = \hat{\rho}'\hat{\rho} = \sum_j \hat{\rho}_j^2.$$

Also,

$$V(\hat{\tau}_i - \hat{\tau}_j) = E[(\hat{\tau}_i - \hat{\tau}_j)^2 / \tau_i - \tau_j = 0].$$

One of the roots of Λ is zero regardless of the design, within the restrictions given above, and this corresponds to the fact that the reduced normal equations for treatments can be summed to give the equation

$$0(\sum \hat{\tau}_j) = 0,$$

so that $\sum \tau_j$ is not estimable. We shall suppose that the root of $(rI - (1/k)\Lambda)$, which is zero regardless of the design, is the root $r - d_1 / k$. It will then follow that $(OQ)_1$ is identically zero. We take the characteristic vector corresponding to this root to have all its elements equal. The imposition of the condition $\sum \hat{\tau}_j = 0$ will then cause the imposition of the condition $\hat{\rho}_1 = 0$. Hence, we have

$$E[\sum \hat{\tau}_j^2 / \text{all } \tau_j = 0] = E[\sum \hat{\rho}_j^2 / \text{all } \tau_j = 0],$$

and ρ_1 is always zero. Also,

$$V(\hat{\rho}_j) = \sigma_1^2 \Big/ \left(r - \frac{1}{k} d_j\right), \qquad j = 2, 3, \cdots, t,$$

where σ_1^2 is the within-block variance for the incomplete block design, so

$$\frac{1}{t(t-1)} E\left[\sum_{\substack{i,j \\ i \neq j}} (\hat{\tau}_i - \hat{\tau}_j)^2 \Big/ \text{all } \tau_j = 0\right] = \frac{2\sigma_1^2}{t-1} \sum_{j=2}^{t} \frac{1}{\left(r - \frac{d_j}{k}\right)},$$

which is the mean variance of a treatment difference.

For the complete block design, the mean variance of a treatment difference is $2\sigma_2^2/r$, where σ_2^2 is the variance within blocks for the complete block design.

Hence, we have the final result. The efficiency factor (EF) of an incomplete block design is equal to r times the harmonic mean of the latent roots of the matrix of coefficients of the reduced normal equations for the intrablock estimates, excluding the always-present zero root, whose characteristic vector consists of the same number repeated t times.

It may be of interest to record the view point that while the efficiency factor is a reasonable criterion of the loss due to confounding by blocking, from some points of view the generalized variance would be better. This, of course, corresponds in a certain sense to the geometric mean of the latent roots.

4. Notes on the Result. The result is interesting to the author and appears to be worth recording in the literature. It was obtained in a search for a proof of a theorem that the design with the highest efficiency factor is a balanced incomplete block design if such a design exists. To the author's knowledge, this theorem is yet to be proved.

REFERENCE

[1] OSCAR KEMPTHORNE, *The Design and Analysis of Experiments*, John Wiley and Sons, New York, 1952.

THE NULL DISTRIBUTION OF THE DIFFERENCE BETWEEN THE TWO LARGEST SAMPLE VALUES[1]

BY J. ST-PIERRE AND A. ZINGER

University of Montreal, Canada

1. Introduction. A decision procedure to select the population with the largest mean, proposed by Bose and St-Pierre [1], involves the auxiliary statistic $u = x_{(0)} - x_{(1)}$, where $x_{(0)}$ and $x_{(1)}$ are respectively the largest and second largest

Received July 20, 1955.

[1] Work done under the sponsorship of the National Research Council of Canada.