RESEARCH NOTES

A Remark on Sampling with Replacement

By H. S. Konijn

University of Sydney, Sydney, Australia

(Received July 1959)

Sukhatme and Narain (1952) discuss, among other matters, two methods of sampling in which n primary sampling units are selected with replacement (with probabilities proportionate to size) and the sub-units within the chosen primary units are sampled

(i) so that no sub-units can be included in the ultimate sample more than once, or

(ii) independently every time the particular primary unit occurs in the sample.

For details and notations we refer to the abovementioned paper. Their results can be readily extended to ratios and to the case of probabilities which are proportionate to an estimate of size.

Obviously the two methods do not differ when $n = 1$, so we shall assume that $n > 1$.

2. Unfortunately, the variance V_1 which the authors give for method I is not correct, because of the possibility that by using this method we may exhaust some primary sampling units. The purpose of this note is to correct this oversight. In particular we show that, nevertheless,

V_1 is always less than V_2; the difference is at most

$$\frac{n-1}{n}\sum_{i=1}^{N}\frac{p_i^2\sigma_i^2}{M_i},$$

and the lower bound is reached if and only if, for each i, m_i is less than or equal to M_i/n.

3. Method I differs from method II only in the way in which samples are drawn within primary units and so we need only compare

those parts of V_1 and V_2, which we shall call V_1' and V_2'. If the conditional expected values of the estimate of the mean arrived at by the two different methods, given that a particular sequence S of primary units is selected, are $V_1(S)$ and $V_2(S)$,

$$V_1' = V_1(S); \qquad V_2' = V_2(S).$$

Let $A(S)$ be the subset of elements s of S for which $\lambda_s \leq M_s/m_s$ and let

$$T_s = \frac{1}{n^2 M^2} \frac{M_s^2}{p_s^2} \frac{\sigma_s^2}{m_s}.$$

Then

$$V_1'(S) = \sum_{A(S)} \frac{M_s - m_s\lambda_s}{M_s} T_s\lambda_s, \quad V_2'(S) = \sum_{S} \frac{M_s - m_s}{M_s} T_s\lambda_s.$$

Now

$$V_2'(S) \geq \sum_{A(S)} \frac{M_s - m_s}{M_s} T_s\lambda_s \geq \sum_{A(S)} \frac{M_s - m_s\lambda_s}{M_s} T_s\lambda_s = V_1'(S),$$

and for some sequence S the first inequality holds if $m_s > M_s/n$ for some s is S and the second if $m_s \leq M_s/2$ for some s in S. Taking expectations on both sides, this gives that V_2' always exceeds V_1', since we excluded the case $n = 1$ from our discussion. When $m_s > M_s/n$ for some s in S,

$$\sum_{S} \frac{M_s - M_s\lambda_s}{M_s} T_s\lambda_s$$

includes some negative terms, so that it is less than $V_1'(S)$; otherwise it equals $V_1'(S)$. Therefore its mean is less than V_1' except when $m_i \leq M_i/n$ for $i = 1, \ldots, N$. Now

$$V_2' - \mathcal{E} \sum_{S} \frac{M_s - m_s\lambda_s}{M_s} T_s\lambda_s = \mathcal{E} \sum_{S} \frac{m_s}{M_s} T_s\lambda_s(\lambda_s - 1)$$

equals

$$\sum_{s_1=1}^{N} \cdots \sum_{s_n=1}^{N} p_{s_1} \cdots p_{s_n} \sum_{j=1}^{\nu} \frac{m_j}{M_j} T_j\lambda_j(\lambda_j - 1)$$

$$= \sum_{s_1=1}^{N} \cdots \sum_{s_n=1}^{N} p_{s_1} \cdots p_{s_n} \sum_{i=1}^{N} \frac{m_i}{M_i} T_i \lambda_i (\lambda_i - 1)$$

$$= \sum_{i=1}^{N} \frac{m_i}{M_i} T_i \sum_{s_1=1}^{N} \cdots \sum_{s_n=1}^{N} p_{s_1} \cdots p_{s_n} \lambda_i (\lambda_i - 1)$$

$$= \sum_{i=1}^{N} \frac{m_i}{M_i} T_i n (n-1) p_i^2 = \frac{n-1}{n} \sum_{i=1}^{N} \frac{p_i^2 \sigma_i^2}{M_i}.$$

Reference

Sukhatme, P. V. and Narain, R. D. "Sampling with replacement," *Jour. Ind. Soc. Agri. Stat.*, 1952, **4**, 42.

A Problem in Reinforced Incomplete Block Design

By G. A. Kulkarni

Indian Council of Agricultural Research, New Delhi

(Received March 1959)

Reinforced incomplete block designs introduced by Das (1958) consist in augmenting an incomplete block design by including some treatments all to each of the blocks with a few more blocks, if necessary, each containing all the treatments. The normal equations in such designs are of the form as in the original design with some changes in the coefficients of treatment effects. If t_m $(m = 1, 2, \ldots v)$ represents the effect of the m-th treatment belonging to the former group and t_i that $(i = 1, 2, \ldots, a$ say) of the treatment belonging to the additional group, then it has been shown that $V(t_{m_1} - t_{m_2})$ is the same as that obtained in the case of the original incomplete block design excepting that the block size 'k' is to be replaced by $k + a$, where a is the number of additional treatments. Das further remarks that the $V(t_{m_1} - t_{m_2})$ obtained in the new design is less than the variance of the same difference in the original design or in other words the $V(t_{m_1} - t_{m_2})$ in the original design is a decreasing function of the block size. The general proof of this result could not be obtained, although Das did supply a proof in the case of two associate PBIB design, elsewhere. [*see* Giri (1957) and also Nair (1958)]. In the present note a general proof of the above result has been given.

Kempthorne (1956) obtained an interesting result that the efficiency factor of an incomplete block design is 'r' times the harmonic mean of the latent roots of the matrix of the reduced intra-block normal equations, the zero root being excluded. A few of the steps are repeated for convenience.

In an incomplete block design with v, b, r, k and λ_{ij} parameters, the reduced normal equations are:

$$\left(rI_v - \frac{1}{k} NN'\right)(\hat{\tau}_i) = Q_i \qquad i = 1,\ 2,\ \ldots\ v$$

where N is the incidence matrix (n_{ij}) such that

$$\begin{aligned} n_{ij} &= 1 \quad \text{if the } i\text{-th treatment occurs in the } j\text{-th block} \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

The incidence matrix N has v rows corresponding to the v treatments and b columns corresponding to the b blocks. The matrix NN' is a $(v \times v)$ matrix and I_v the identity matrix;

$$\begin{aligned} Q_i = {} & \text{(Total of the } i\text{-th treatment)} \\ & - 1/k \text{ (sum of the block totals containing the } i\text{-th treatment)} \end{aligned}$$

along with the additional restriction $\Sigma\, \hat{\tau}_i = 0$.

Since the matrix $NN' = \Lambda$ is real symmetric, there always exists an orthogonal matrix A such that

$$AA' = I_v = A'A$$

and

$$A \Lambda A' = D$$

where D is the diagonal matrix with say d_j as the diagonal elements.

Further,

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

is a diagonal matrix with the diagonal element $[r - (d_j/k)]$. These quantities are also the latent roots of the matrix

$$(c) = \left(rI_v - \frac{1}{k}\Lambda\right)$$

with one latent root zero. Putting $[r - (d_1/k)]$ as the zero root, Kempthorne has proved that the mean variance of differences between all pairs of treatments in this design is

$$V = \frac{2\sigma^2}{v-1} \sum_{j=2}^{v} \frac{1}{\left(r - \frac{d_j}{k}\right)}. \qquad (1)$$

With the help of this result the problem posed earlier can be proved as follows:

(i) The quantities d_j's do not contain k since these come from the matrix $\Lambda = NN'$ which does not contain k.

(ii) Moreover, it can be proved* easily that the matrix NN' is positive semi-definite and hence its transformed diagonal matrix also contains the elements which are all non-negative.

Now differentiating (1) with respect to k we have

$$\frac{dV}{dk} = \frac{2\sigma^2}{v-1}\left[-\frac{d_2}{(rk-d_2)^2} - \frac{d_3}{(rk-d_3)^2} - \cdots - \frac{d_v}{(rk-d_v)^2}\right]$$

which is negative since each term is negative showing thereby that mean variance of the differences between all pairs of treatments is a decreasing function of the block-size k. That is, by adding a few extra treatments to an incomplete block design (v, b, r, k and λ_{ij}) so as to obtain a reinforced incomplete block design the mean variance of differences between all pairs of treatments belonging to the original design is decreased.

The author is grateful to Shri M. N. Das, I.C.A.R., for many helpful suggestions.

References

1. Das, M. N. .. "Reinforced incomplete block design," *Jour. Ind. Soc. Agri. Stat.*, 1958, **10**, 73–77.
2. Giri, N. C. .. "On reinforced partially balanced incomplete block design," *Ibid.*, 1957, **9**, 41–51.
3. Kempthorne, Oscar .. "Efficiency factor of an incomplete block design," *Ann. Math. Stat.*, 1956, **27**, 846–49.
4. Nair, K. R. .. "A note on reinforced incomplete block design," *Jour. Ind. Soc. Agri. Stat.*, 1958, **10**, 150–56.

* The fact that NN' is positive semi-definite can be proved as follows:

Let $X' NN' X$ be a quardatic form and let $X'N = (Y_1, Y_2 \cdots, Y_b)$ and therefore $X' NN' X = \Sigma y_i^2 \geqslant 0$. Hence NN' is positive semi-definite.

A Note on the Variance in Reinforced Incomplete Block Designs

By K. R. Shah*

Forest Research Institute, Dehra Dun

(Received April 1959)

1. Introduction

Das (1954) proposed a design in which, to every block of a BIB design (v, k, r, b, λ) a set of α new treatments is added. To the design, β complete blocks each containing all $(v+\alpha)$ treatments may also be added. A design thus obtained has come to be termed a reinforced design. The reinforcement in the above manner has been extended by Das (1958) to the design $(v, k, r, b, \lambda_{ij})$ where λ_{ij}'s are not all equal. This general case has been called "Reinforced Incomplete Block Design (RIBD)" by him.

Giri (1957) has given a proof due to Das that in the case of a reinforced two-associate PBIB design (with $\alpha > 0$, $\beta = 0$) the variances of differences between pairs of the original set of v treatments will be less than those in the corresponding PBIB design, assuming $\sigma_{k+\alpha} = \sigma_k$. Nair (1958) has given an alternative proof. The proof in the general case of RIBD is not yet given. The purpose of the present paper is to give it.

2. Proof

We have v treatments numbered 1, 2, ... v. They are applied to plots in b blocks, say $\beta_1, \ldots \beta_b$ of sizes $k_1, \ldots k_b$, respectively. Suppose that the i-th treatment is replicated r_i times. Let n_{ij} be the number of plots in the j-th block receiving the i-th treatment ($n_{ij} =$ 0 or 1), $T_i =$ Total yield of all plots receiving the i-th treatment, and $B_j =$ Total yield of all plots in the j-th block. It is well known that the normal equations giving intra-block estimates of treatment effects are

$$C\hat{\tau} = Q \tag{2.1}$$

where

$$Q = \{Q_1, \ldots Q_v\} \qquad Q_i = T_i - \sum_j \frac{n_{ij}B_j}{k_j} \tag{2.1}$$

*Now at the Indian Statistical Institute, Calcutta.

and

$$C = \text{diag} \quad (r_1, \ldots r_v) - N \,\text{diag}\left(\frac{1}{k_1}, \ldots \frac{1}{k_6}\right) N'. \tag{2.2}$$

where N is the incidence matrix (n_{ij}).

We shall consider only connected designs so that rank of C is $v - 1$. If we consider the design $(v, k, r, b, \lambda_{ij})$,

$$C = rI - \frac{NN'}{k}.$$

rk is a characteristic root of $NN' = \Lambda$; the corresponding *vector* is $E_{v1}/\sqrt{v}$. Let $d_2, \ldots d_v$ be the other roots and $L_2, \ldots L_v$ be the set of corresponding vectors so chosen that $L = (E_{v1}/\sqrt{v}, d_2, \ldots d_v)$ is an orthogonal matrix. $L' \Lambda L = \text{diag.} \; (rk, d_2, \ldots d_v)$.

When the reinforcement in the above manner is applied, *i.e.*, when α new treatments are added to every block and β complete blocks are added the following further matrices are required for writing the normal equations:

$$\left.\begin{aligned} C_{ii} &= r + \beta - \frac{r}{k+\alpha} - \frac{\beta}{v+\alpha} \qquad (i = 1, 2, \ldots v) \\ C_{ii'} &= -\frac{\lambda_{ii'}}{k+\alpha} - \frac{\beta}{v+\alpha} \qquad (i, i' = 1, 2, \ldots v) \\ C_{ii'} &= -\frac{r}{k+\alpha} - \frac{\beta}{v+\alpha} \\ &\qquad (i = 1, 2, \ldots v, i' = v+1, \ldots v+\alpha) \end{aligned}\right\} \tag{2.3}$$

If

$$C = \begin{pmatrix} C_{11} & C_{12} \\ & \\ C_{21} & C_{22} \end{pmatrix}, \quad Q_{(1)} = \begin{pmatrix} Q_1 \\ \cdot \\ \cdot \\ \cdot \\ Q_v \end{pmatrix}, \quad Q_{(2)} = \begin{pmatrix} Q_{v+1} \\ \cdot \\ \cdot \\ \cdot \\ Q_{v+a} \end{pmatrix},$$

$$\tau_{(1)} = \begin{pmatrix} \tau_1 \\ \cdot \\ \cdot \\ \cdot \\ \tau_v \end{pmatrix}, \quad \tau_{(2)} = \begin{pmatrix} \tau_{v+1} \\ \cdot \\ \cdot \\ \cdot \\ \tau_{v+a} \end{pmatrix}$$

the adjusted normal equations are

$$\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \begin{pmatrix} \hat{\tau}_{(1)} \\ \hat{\tau}_{(2)} \end{pmatrix} = \begin{pmatrix} Q_{(1)} \\ Q_{(2)} \end{pmatrix}.$$

By (2.3)

$$\left.\begin{aligned} C_{11} &= (r + \beta) I_v - \frac{NN'}{k + \alpha} - \frac{\beta E_{v,v}}{v + \alpha} \\ C_{12} &= - \frac{r}{k + \alpha} E_{v,\alpha} - \frac{\beta}{v + \alpha} E_{v,\alpha} \end{aligned}\right\}. \tag{2.4}$$

Put $L'\hat{\tau}_{(1)} = \theta$. Since L is an orthogonal matrix, $L\theta = \hat{\tau}_{(1)}$. θ_1 is not estimable. $\theta_2, \ldots \theta_v$ are all estimable. The first v normal equations are

$$C_{11}\hat{\tau}_{(1)} + C_{12}\hat{\tau}_{(2)} = Q_{(1)}.$$

Putting $\hat{\tau}_{(1)} = L\theta$ we get

$$C_{11}L\theta + C_{12}\hat{\tau}_{(2)} = Q_{(1)}.$$

$$\therefore \quad L'C_{11}L\theta + L'C_{12}\hat{\tau}_{(2)} = L'Q_{(1)}. \tag{2.5}$$

It is easy to see that

$$L'C_{11}L = (r + \beta) I_v - \frac{1}{k + \alpha} \text{ diag. } (rk, d_2, \ldots d_v)$$

$$- \frac{\beta}{v + \alpha} \text{ diag. } (v, 0 \ldots 0)$$

and

$$L'C_{12} = - \left(\frac{r}{k + \alpha} + \frac{\beta}{v + \alpha}\right) \begin{pmatrix} \sqrt{v}, & \ldots & \sqrt{v_1}. \\ 0 & \ldots & 0 \\ 0 & \ldots & 0 \end{pmatrix}.$$

Hence (2.5) gives us

$$\left(r + \beta - \frac{d_i}{k + \alpha}\right) \theta_i = L_i'Q_{(1)}. \qquad (i = 2, 3, \ldots v)$$

$$V(Q_{(1)}) = C_{11}\sigma^2{}_{k+\alpha}$$

$$\therefore \quad V(L'Q_{(1)}) = (L'C_{11}L)\,\sigma^2{}_{k+\alpha}.$$

Hence

$$V(L_iQ_{(1)}) = V\left[\left\{r + \beta - \frac{d_i}{k + \alpha}\right\} \theta_i\right]$$

$$= \left\{r + \beta - \frac{d_i}{k + \alpha}\right\} \sigma^2{}_{k+\alpha}.$$

$$\therefore \quad V(\theta_i) = \frac{\sigma^2_{k+\alpha}}{r + \beta - \dfrac{d_i}{k + \alpha}}. \tag{2.6}$$

Also it follows that

$$\text{Cov.}\,(\theta_i, \theta_j) = 0 \qquad (i \neq j).$$

For the corresponding design with $\alpha = 0$, $\beta = 0$

$$V(\theta_i) = \frac{\sigma_k^2}{r - \dfrac{d_i}{k}}. \tag{2.7}$$

Since NN' is positive semi-definite $d_i \geqslant 0$ $(i = 2, \ldots v)$ when $\sigma_k^2 = \sigma^2_{k+\alpha}$ it can be easily seen that R.H.S. in (2.6) is less than R.H.S. in (2.7) for $i = 2, 3, \ldots v$.

Now consider any elementary treatment contrast $\hat{\tau}_1 - \hat{\tau}_2$, say. It must be a linear function of $\theta_2, \ldots \theta_v$.

Let $\hat{\tau}_1 - \hat{\tau}_2 = l_2\theta_2 + \ldots + l_v\theta_v$.

$$V(\hat{\tau}_1 - \hat{\tau}_2) = \sum_{i=2}^{v} l_i^2\, V(\theta_i). \tag{2.8}$$

Hence $V(\hat{\tau}_1 - \hat{\tau}_2)$ is less for the reinforced design than for the design with $\alpha = 0$, $\beta = 0$. Since this is true for any elementary treatment contrast $\hat{\tau}_i - \hat{\tau}_j$ $(i, j = 1, 2, \ldots v)$ our result is proved.

3. Acknowledgement

My thanks are due to Dr. K. R. Nair for calling my attention to this problem and for giving me general guidance.

4. References

Das, M. N. .. "Missing plots and a randomised block design with balanced incompleteness," *Jour. Ind. Soc. Agri. Stat.*, 1954, **6**, 58–76.

——— .. "On reinforced incomplete block designs," *Ibid.*, 1958, **10**, 73–77.

Giri, N. C. .. "On reinforced partially balanced incomplete block designs," *Ibid.*, 1957, **9**, 41–52.

Kempthorne, O. .. "Efficiency factor of an incomplete block design," *Ann. Math. Stat.*, 1956, **27**, 846–49.

Nair, K. R. .. "A note on reinforced incomplete block designs," *Jour. Ind. Soc. Agri. Stat.* 1958,, **10**, 150–56.