USE OF THE ANALYSIS OF COVARIANCE IN TWO-STAGE SAMPLING

I. M. CHAKRAVARTI

Calcutta

Often we come across the problem of finding an estimate for the product of the means of two correlated variates x and y. For the unistage sampling this problem has been discussed by several authors, but little has been done for the case of a multistage sample. Some results for two-stage sampling which may prove useful in this connection are discussed below; these results can be extended to a greater number of stages without much difficulty.

For simplicity the "equal number" case is considered. Let the sampling plan be such that t first stage units are to be selected at random and p second stage or ultimate units are to be chosen at random within each first stage unit. The sampling in both the stages is further assumed to be simple.

Assuming the usual linear set up, we have

$y_{ij}=A+B_i+C_{ij}$; $E(B_i)=E(C_{ij})=0$, $V(B_i)=\sigma_1'^2$, $V(C_{ij})=\sigma_2'^2$

$x_{ij}=D+F_i+G_{ij}$; $E(F_i)=E(G_{ij})=0$, $V(F_i)=\sigma_1^2$, $V(G_{ij})=\sigma_2^2$

C's are mutually uncorrelated and uncorrelated to B's; B's are also mutually uncorrelated; similar is the case for F's and G's.

$E(B_iF_i)=\lambda_1$; $E(B_iF_j)=0$ for $i\neq j$

$E(C_{ij}G_{ij})=\lambda_2$; $E(C_{ij}G_{i'j'})=0$ for $i'\neq i$ and/or $j\neq j'$

In the two-satge analysis of co-variance, let us denote the covariance between groups by S_1 and the covariance within groups by S_2 and let

$$x_{i.}=\sum_{j=1}^{p} x_{ij}/p\ ; \qquad x_{..}=\sum_{i=1}^{t}\sum_{j=1}^{p} x_{ij}/pt$$

$y_{i.}$ and $y_{..}$ are similarly defined.

Then

$$S_1 = p \sum_{i=1}^{t} (y_{i.} - y_{..})(x_{i.} - x_{..})/(t-1)$$

$$S_2 = \sum_{i=1}^{t} \sum_{j=1}^{p} (y_{ij} - y_{i.})(x_{ij} - x_{i.})/t(p-1)$$

Then it can be easily shown that

$$E(S_2) = \lambda_2 \; ; \quad E(S_1) = \lambda_2 + p\lambda_1$$

$$\text{and Cov } (x_{..}, y_{..}) = \frac{\lambda_1}{t} + \frac{\lambda_2}{pt} = \lambda(\text{say})$$

So we can estimate λ_1, λ_2 and λ in terms of S_1 and S_2.

Application :

If we use $z_1 = x_{..} \times y_{..}$ (as is done in many cases) to estimate the product of the population means, AD we find

$$E(z_1) = \lambda + AD$$

For an alternative estimate,

$$z_2 = \sum_{i=1}^{t} \sum_{j=1}^{p} x_{ij} \, y_{ij}/pt$$

$$E(z_2) = AD + \lambda_1 + \lambda_2$$

So both the estimates z_1 and z_2 are biased but the bias can be estimated in both the cases in terms of S_1 and S_2, from the expression for $E(S_1)$ and $E(S_2)$ given above.

Now $$V(z_1) = E[(x_{..} - D)\{y_{..} - A - \theta(x_{..} - D)\} + \theta(x_{..} - D)^2 + D\{y_{..} - A - \theta(x_{..} - D)\} + D\theta(x_{..} - D) + A(x_{..} - D) - \lambda]^2$$

where $\theta = \lambda/(\sigma^2{}_1/t + \sigma_2{}^2/pt)$

On the assumption of the population having a normal bivariate distribution,

$$V(z_1) = A^2\left(\frac{\sigma_1{}^2}{t} + \frac{\sigma_2{}^2}{pt}\right) + D^2\left(\frac{\sigma_1{}'^2}{t} + \frac{\sigma_2{}'^2}{pt}\right) + \left(\frac{\sigma_1{}^2}{t} + \frac{\sigma_2{}^2}{pt}\right)\left(\frac{\sigma_1{}'^2}{t} + \frac{\sigma_2{}'^2}{pt}\right) + 2AD\lambda + \lambda^2 = V_1 \text{ (say)}$$

Again

$$V(z_2)=[A^2(\sigma_1^2+\sigma_2^2)+D^2(\sigma_1'^2+\sigma_2'^2)+(\sigma_1^2+\sigma_2^2)(\sigma_1'^2+\sigma_2'^2) +2AD(\lambda_1+\lambda_2)+(\lambda_1+\lambda_2)^2]/pt=V_2 \text{ (say)}$$

Then

$$V_2-V_1=\Big[(A^2\sigma_1^2+D^2\sigma_1'^2+2AD\lambda_1)\left(\frac{1}{pt}-\frac{1}{t}\right)\Big] +\frac{1}{pt}\Big[(\lambda_1+\lambda_2)^2-\frac{(\lambda_2+p\lambda_1)^2}{pt}\Big] +\frac{1}{pt}\Big[(\sigma_1^2+\sigma_2^2)(\sigma_1'^2+\sigma_2'^2)-(\sigma_2^2+p\sigma_1^2)(\sigma_1'^2+p\sigma_2'^2)/pt\Big]$$

The following inequality relations connect σ's and λ's,

(i) $(\sigma_1^2+\sigma_2^2)(\sigma_1'^2+\sigma_2'^2)\geqslant(\lambda_1+\lambda_2)^2$

(ii) $(\sigma_1^2/t+\sigma_2^2/pt)(\sigma_1'^2/t+\sigma_2'^2/pt)\geqslant(\lambda_1/t+\lambda_2/pt)^2$

and obviously $(A^2\sigma_1^2+D^2\sigma_1'^2+2AD\lambda_1\lambda_2)\geqslant 0$

Whether V_2-V_1 will be positive or negative will, however, depend upon the values of the parameters ; so partcular situations will call for use of the particular estimates, z_1 and z_2.

A NOTE ON THE RESOLVABILITY OF BALANCED INCOMPLETE BLOCK DESIGNS

PURNENDU MOHON ROY
Calcutta

1. A balanced incomplete block design with parameters, v, b, r, k, λ is said to be resolvable if the blocks of a design are separable into r sets of n blocks each (implying $b=nr$) in such a way that each variety occurs only once among the n blocks of a set forming a complete replication (obviously $v=nk$). Bose (1942) has proved a number of relations obtaining among the parameters of a resolvable design. The important ones are as follows :—

(A) $b \geqslant v+r-1$

(B) If $b=v+r-1$, k^2 must be divisible by v

and so the design is affine resolvable ; and conversely.

The object of this note is to show that these relations remain true in a wider class of designs than the class of resolvable designs.

2. **Condition** (A). We will prove the following proposition.

Whenever v of a balanced incomplete block design with arithmetically consistent parameters : v, b, r, k, λ is divisible by k, the inequality $b \geqslant v+r-1$ holds.

Proof. Note that
$$bk=vr \tag{1}$$
$$\lambda(v-1)=r(k-1) \tag{2}$$

If v is divisible by k, $v=nk$ where n is an integer greater than one. From (1), we have $b=nr$. From (2) it follows

$$r=\frac{\lambda(v-1)}{k-1}=\lambda\,\frac{(nk-1)}{k-1}=n\lambda+\frac{n-1}{k-1}\,\lambda.$$

Since r is an integer, $(n-1)\lambda/(k-1)$ must be an integer, g (say)

$$\therefore\ r=n\lambda+g\ ;\ \frac{k-1}{n-1}=\frac{\lambda}{g}\ ;\ n\lambda+g=kg+\lambda \tag{3}$$

Now $b >, =$ or $< v+r-1$ according as

$$n(n\lambda+g) >, = \text{or} < nk+n\lambda+g-1$$

$$\text{or } n\lambda+g >, = \text{ or } < \frac{nk-1}{n-1} = k+\frac{k-1}{n-1}$$

$$= k+\frac{\lambda}{g} \quad \text{from (3)}$$

$$= \frac{kg+\lambda}{g}$$

$$= \frac{n\lambda+g}{g}$$

$\therefore b > v+r-1$ unless $g=1$ when $b=v+r-1$.

Thus the balanced incomplete block designs for which v is divisible by k belongs to the series :

(P) $v=nk$, $b=n(n\lambda+g)$, $r=n\lambda+g$, k,λ where g is an integer given by $\frac{n-1}{k-1}\lambda$. and the sub-class of these degins for which $b=v+r-1$ or $g=1$ belongs to the series :

(R) $v=n(n-1)\lambda+n$, $b=n(n\lambda+1)$, $r=n\lambda+1$, $k=(n-1)\lambda+1$, λ.

It may be noted that for any design of (R) b, r and λ do not have any common factor.

3. **Condition** (B). Instead of assuming the design with parameters $v=nk$, $b=nr$, r, k, λ to be resolvable, let us suppose that there is one set of n blocks containing a complete replication of the varieties and let these blocks be $B_1, B_2, \ldots, B_n$. The other blocks $B_{n+1}, \ldots\ldots$ may or may not be separable into one or more replications. Then consider l_j to be the number of varieties common to the block B_i for any fixed i $(1 \leqslant i \leqslant n)$ and the block $B_j (n+1 \leqslant j \leqslant b)$. Proceding in the same way as Bose (1942), we have

$$\text{Mean } (l_j) = k/n = k^2/v \tag{4}$$

$$\text{Variance } (l_j) = \frac{k(v-k)(b-v-r+1)}{n^2(r-1)(v-1)}$$

Thus it is to be observed that if there is at least one complete replication of all varieties in n blocks and $b=v+r-1$ then any block belonging to the replication has equal number of varieties (namely k^2/v) in common with each of the blocks of the solution not in the replication. The converse of this result is also true.

Thus any balanced incomplete block design belonging to the series (R) cannot have a single set of blocks providing a complete replication of all the varieties unless λ is of the form $nt+1$ *i.e.* unless the design belongs to the series D (Bose 1942) :

(D) $v=n^2(n-1)t+n^2, \quad b=n(n^2t+n+1),$
$r=n^2t+n+1, \quad k=n(n-1)t+n, \quad \lambda=nt+1$

REFERENCE

Bose, R. C. (1942) *Sankhya*, 6, 105.