# ON COLLAPSING OF STRATA

BY G. R. SETH
*Institute of Agricultural Research Statistics, New Delhi*

When the stratification is deep, only one unit may be sampled from each stratum. In this event, a device to find the estimate of the variance of the customary estimate is to be found. The device suggested by Cochran is the collapsing of the strata in pairs, the two strata to be collapsed to form one strata should have common mean and same number of units. Let $L$ be the number of strata. Let $N_i$, $\mu_i$ and $\sigma_i^2$ denote number of units in the $i$th stratum mean and variance of the character $Y$ under study. The stratified simple random sample estimate is given by—

$$\sum_{i=1}^{N} \frac{N_i}{N} y_i \qquad \ldots(1)$$

and its variance is given by

$$\frac{1}{N^2} \sum_{i=1}^{N} N_i^2 \sigma_i^2 \qquad \ldots(2)$$

The estimate (2) as proposed by Cochran is given by

$$\sum_{i=1}^{L/2} \frac{N_i^2}{N^2} (y_{i1}-y_{i2})^2,$$

where

$$N_{i1}=N_{i2}=N_i \qquad \ldots(3)$$

where $i_1$th, $i_2$th strata are collapsed to form the $i$th stratum and $L$ is assumed to be an even integer.

The assumption of equality of units in the pair of strata to be collapsed together with the equality of strata means in the pair is rather restrictive in the practical utilisation of this technique for estimating the variance of the estimate of population mean with one unit per stratum.

The assumption of equality of units in the strata to be collapsed can be relaxed if we adopt the following estimate of (2):

$$\frac{1}{N^2}\sum_{i=1}^{L/2}(N_{i1}^2\, y_{i1}-N_{i2}^2\, y_{i2})(y_{i1}-y_{i2}) \qquad \ldots(4)$$

The expected value of (4) in repeated samples is given by

$$\frac{1}{N^2}\sum_{i=1}^{L} N_i^2\,\sigma_i^2+\frac{1}{N^2}\sum_{i=1}^{L/2}(N_{i1}^2\,\mu_{i1}-N_{i2}^2\,\mu_{i2})(\mu_{i1}-\mu_{i2}) \qquad \ldots(5)$$

It can be easily seen that estimate (4) is always positive and is thus usable and it reduces to estimate (3) if $N_{i1}=N_{i2}=N_i$. Further the bias in (4) is given by

$$\frac{1}{N^2}\sum_{i=1}^{L/2}(N_{i1}^2\,\mu_{i1}-N_{i2}^2\,\mu_{i2})(\mu_{i1}-\mu_{i2}) \qquad \ldots(6)$$

and it is negligible when the means of the pair collapsed are nearly equal or when the strata are formed in such a way that means of original strata are inversely proportional to the square of the number of units or at least when they are so pairwise.

This provides an idea for an efficient collapsing of strata in pairs provided one has the knowledge about the formation of strata. For example, if strata were formed in such a way that strata means are proportional to some power '$p$' of the number of units in the stratum or at least so in pairs.

That is, let

$$N_{i_1}^{p}\,\mu_{i1}=N_{i_2}^{p}\,\mu_{i2} \qquad \ldots(7)$$

in pairs, where $p$ is any integer. Then we consider

$$\frac{1}{N^2}\sum_{i=1}^{L/2}\left(N_{i_1}^{p}\,y_{i1}-N_{i_2}^{p}\,y_{i2}\right)\left(N_{i_1}^{2-p}\,y_{i1}-N_{i_2}^{2-p}\,y_{i2}\right) \qquad \ldots(8)$$

as an estimate of (2).

The expected value of (8) is given by

$$\frac{1}{N^2}\sum_{i=1}^{L/2}\left(N_{i_1}^2\,\sigma_{i_1}^2+N_{i_2}^2\,\sigma_{i_2}^2\right)+\frac{1}{N^2}\sum_{i=1}^{L/2}\left(N_{i_1}^{p}\,\mu_{i1}-N_{i_2}^{p}\,\mu_{i2}\right)(N_{i1}\,\mu_{i1}-N_{i2}\,\mu_{i2})$$

The bias in (8) will be small as $N_{i1}^{p}\,\mu_{i1}-N_{i2}^{p}\,\mu_{i2}$ are small.

The one interesting case is when $p=1$, *i.e.*, the strata are formed in-pairs such that total value of the character $Y$ is constant over pairs.

The treatment, when $L$ is odd, is to first collapse in pairs as usual and attach the last one to the pair which is most similar to it so far as property (7) is concerned. In that case the estimate (8) remains the same except for the contribution from the three strata collapsed together. The contribution from the latter will be

$$\frac{1}{2}\Big[\Big(N_{i_1}^{p}\,y_{i1}-N_{i_2}^{p}\,y_{i2}\Big)\Big(N_{i_1}^{2-p}y_{i1}-N_{i_2}^{2-p}y_{i2}\Big)$$
$$+\Big(N_{i_2}^{p}\,y_{i2}-N_{i_3}^{p}\,y_{i3}\Big)\Big(N_{i_2}^{2-p}y_{i2}-N_{i_3}^{2-p}y_{i3}\Big)$$
$$+\Big(N_{i_1}^{p}\,y_{i1}-N_{i_3}^{p}\,y_{i3}\Big)\Big(N_{i_1}^{2-p}y_{i1}-N_{i_3}^{2-p}y_{i3}\Big)\Big]$$

The proposed estimate (8) is, of course, subject to the same objection as estimate (3), that is, that it over-estimates the variance and is likely to be less efficient compared to the design where the strata are reduced to half and two units per stratum are selected to form the sample.

## SUMMARY

An estimate of the variance of customary stratified simple random estimate based on one sample unit per stratum is suggested which is of more general application than the one suggested by Cochran.

## REFERENCE

Cochran, W. G. Sampling Technique, 1962, pp. 105-106.