EFFICIENCY OF STRATIFICATION IN SUB-SAMPLING DESIGNS FOR THE RATIO METHOD OF ESTIMATION

By V. K. Mokashi
Institute of Plant Industry, Indore

Sukhatme (1950 *a*, *b* and 1954) has worked out the efficiency of stratification in sub-sampling designs using the sample estimate. A ratio estimate may be used when supplementary information is available. Based on Sukhatme's approach, formulæ are developed in the present paper for determining the efficiency of stratification for the ratio method of estimation in sub-sampling designs.

Let,

N, be the number of units in the population.

N_j, the number of units in the jth stratum.

n, the number of units in the sample.

n_j, the number of units in the sample from the jth stratum.

M, the number of sub-units in each unit of sampling.

m, the number of sub-units sampled from each unit.

k, the number of strata in the population.

Also let,

y_{jil}, be the value of the character for the lth sub-unit in the ith unit of the jth stratum.

$\bar{y}_{ji(M)}$, be the true mean of the ith unit of the jth stratum so that

$$\bar{y}_{ji(M)} = \frac{1}{M} \sum_{l=1}^{M} y_{jil} \tag{1}$$

$\bar{y}_{ji(m)}$, be the corresponding sample mean so that

$$\bar{y}_{ji(m)} = \frac{1}{m} \sum_{l=1}^{m} y_{jil} \tag{2}$$

$\bar{y}_{j(N_j,M)}$, be the true mean of the jth stratum so that

$$\bar{y}_{j(N_j,M)} = \frac{1}{N_j} \sum_{i=1}^{N_j} \bar{y}_{ji(M)} \tag{3}$$

$\bar{y}_{nj}$ be the corresponding sample mean so that

$$\bar{y}_{nj} = \frac{1}{n_j} \sum_{i=1}^{nj} \bar{y}_{ji(m)} \tag{4}$$

$\bar{y}_{NM}$ be the population mean so that

$$\bar{y}_{NM} = \frac{1}{N} \sum_{j=1}^{k} N_j \bar{y}_{N_jM} \tag{5}$$

$\bar{y}_n$, be the corresponding sample mean so that

$$\bar{y}_n = \frac{1}{N} \sum_{j=1}^{k} N_j \bar{y}_{nj} \tag{6}$$

The expressions for another variate 'x' follow similarly.

The estimate of the ratio of y to x is given by

$$\bar{r} = \frac{\sum_{j=1}^{k} N_j \bar{y}_{nj}}{\sum_{j=1}^{k} N_j \bar{x}_{nj}} \tag{7}$$

and the variance of $\bar{r}$, is given by

$$V(\bar{r}) = \frac{\bar{y}^2_{NM}}{\bar{x}^2_{NM}} \left[\frac{V(\bar{y}_n)}{\bar{y}^2_{NM}} + \frac{V(\bar{x}_n)}{\bar{x}^2_{NM}} - \frac{2 \text{ cov. } (\bar{y}_n, \bar{x}_n)}{\bar{x}_{NM} \bar{y}_{NM}} \right] \tag{8}$$

The variance of $\bar{y}_n$ is given by Sukhatme (1954) as

$$V(\bar{y}_n) = \sum_{j=1}^{k} P_j^2 \left[\left(\frac{1}{n_j} - \frac{1}{N_j} \right) S^2_{jb} + \frac{1}{n_j} \left(\frac{1}{m} - \frac{1}{M} \right) \times S^2_{jw} \right] \tag{9}$$

where

P_j is the weight of the jth stratum,

and

$$S^2_{jb} = \frac{1}{N_j - 1} \sum_{i=1}^{Nj} (\bar{y}_{ji(M)} - \bar{y}_{NM})^2$$

$$\bar{S}^2_{jw} = \frac{1}{N_j} \sum_{i=1}^{Nj} \sum_{l=1}^{M} \frac{1}{N_j - 1} (y_{jil} - \bar{y}_{ji(M)})^2$$

Expressions for $V(\bar{x}_n)$ and cov. $(\bar{x}_n \bar{y}_n)$ follow similarly. The estimates for $V(\bar{y}_n)$ and $V(\bar{x}_n)$, without stratification, have been worked out by

Sukhatme (1954). In order to find out the efficiency of stratification for the ratio estimate, it is necessary to work out the formula for the variance of ratio without stratification and compare it with that given by (8) with stratification. We are therefore concerned with estimation of cov. $(\bar{y}_n \bar{x}_n)$ without stratification. Covariance without stratification is given by

$$\text{Cov. } (\bar{y}_n \bar{x}_n)_{U.S.} = \left(\frac{1}{n} - \frac{1}{N}\right)(M.P.)_b + \left(\frac{1}{m} - \frac{1}{M}\right)\frac{(M.P.)\,w}{n} \tag{10}$$

where '*U.S.*' stands for 'unstratified' and

$(M.P.)_b$ = mean products between 1st stage unit means in the population;

$$= \frac{1}{N-1}\sum^{N}(\bar{y}_{i(M)}\bar{x}_{i(M)} - \bar{x}_{NM}\bar{y}_{NM}) \tag{11}$$

and

$(M.P.)\,w$ = the mean product between second stage units within 1st stage units in the whole population.

$$= \frac{1}{N(M-1)}\sum_{i=1}^{N}\sum_{l=1}^{M}(y_{il}x_{il} - \bar{y}_{i(M)}\bar{x}_{i(M)}) \tag{12}$$

We want an estimate of $(M.P.)_b$. We have from (11)

$$\begin{aligned}(N-1)(M.P.)_b &= \sum^{N}(\bar{y}_{i(M)}\bar{x}_{i(M)} - \bar{x}_{NM}\bar{y}_{NM})\\ &= \Sigma\Sigma(\bar{x}_{ji(M)}\bar{y}_{ji(M)} - \bar{x}_{NM}\bar{y}_{NM})\\ &= \Sigma\Sigma(\bar{x}_{ji(M)}\bar{y}_{ji(M)} - \bar{x}_{N_j}\bar{y}_{N_j} + \bar{x}_{N_j}\bar{y}_{N_j} - \bar{x}_{NM}\bar{y}_{NM})\\ &= \Sigma(N_j-1)(M.P.)_{jb} + \sum^{k} N_j(\bar{x}_{N_j}\bar{y}_{N_j} - \bar{x}_{NM}\bar{y}_{NM})\\ &= \Sigma(N_j-1)(M.P.)_{jb} + \Sigma N_j\bar{x}_{N_j}\bar{y}_{N_j} - N\bar{x}_{NM}\bar{y}_{NM}.\end{aligned} \tag{13}$$

We want to estimate the second and third terms in (13). Now, we have

$$\begin{aligned}\text{Cov. } (\bar{y}_{n_j}\bar{x}_{n_j}) &= E\,\bar{y}_{n_j}\bar{x}_{n_j} - \bar{x}_{N_j}\bar{y}_{N_j}\\ &= \left(\frac{1}{n} - \frac{1}{N_j}\right)(M.P.)_{jb} + \frac{\left(\frac{1}{m} - \frac{1}{M}\right)(M.P.)_{jw}}{n_j}\end{aligned}$$

or

$$\text{Est. } \bar{x}_{N_j}\bar{y}_{N_j} = \bar{y}_{n_j}\bar{x}_{n_j} - \left(\frac{1}{n_j} - \frac{1}{N_j}\right)(M.P.)_{jb}$$

$$- \left(\frac{1}{m} - \frac{1}{M}\right) \frac{(M.P.)_{jw}}{n_j}$$

or

$$\text{Est. } \overset{k}{\Sigma} N_j \bar{x}_{Nj} \bar{y}_{Nj} = \Sigma N_j \bar{x}_{nj} \bar{y}_{nj} - \Sigma N_j \left(\frac{1}{n_j} - \frac{1}{N_j}\right)(M.P.)_{jb}$$

$$- \left(\frac{1}{m} - \frac{1}{M}\right) \Sigma \frac{N_j}{n_j} (M.P.)_{jw} \qquad (14)$$

Now

$$\text{Cov. } (\bar{x}_n\bar{y}_n) = E\,\Sigma\,(P_j\bar{x}_{nj})\,\Sigma\,(P_j\bar{y}_{nj}) - \bar{x}_{NM}\,\bar{y}_{NM}$$

Therefore

$$\text{Est. } (N\,\bar{x}_{NM}\,\bar{y}_{NM}) = N\Sigma\,(P_j\bar{x}_{nj})\,\Sigma\,(P_j\,\bar{y}_{nj}) - N \text{ Est. cov. } (\bar{x}_n\bar{y}_n)$$

or

$$\text{Est. } N\bar{x}_{NM}\,\bar{y}_{NM} = N\,\bar{x}_n\bar{y}_n - N\,\Sigma\,P_j^2 \left[\left(\frac{1}{n_j} - \frac{1}{N_j}\right)(M.P.)_{jb}\right.$$

$$\left. + \frac{1}{n_j}\left(\frac{1}{m} - \frac{1}{M}\right)(M.P.)_{jw}\right] \qquad (15)$$

Subtract (15) from (14), and we get

$$\text{Est.}\left[\overset{k}{\Sigma} N_j\bar{x}_{Nj}\bar{y}_{Nj} - N\bar{x}_{NM}\bar{y}_{NM}\right] = \overset{k}{\Sigma} N_j\,(\bar{x}_{nj}\bar{y}_{nj} - \bar{x}_n\bar{y}_n)$$

$$- \left(\frac{1}{m} - \frac{1}{M}\right)\left[\overset{k}{\Sigma}\left(\frac{N_j}{n_j} - \frac{N}{N_j} P_j^2\right)(M.P.)_{jw}\right]$$

$$- \overset{k}{\Sigma} N_j\left(\frac{1}{n_j} - \frac{1}{N_j}\right)\left(1 - \frac{N}{N_j} P_j^2\right)(M.P.)_{jb} \qquad (16)$$

Now from (13) we get

$$(M.P.)_b = \frac{1}{N-1}\,\Sigma\,N_j\,(\bar{x}_{nj}\bar{y}_{nj} - \bar{x}_n\bar{y}_n)$$

$$+ \frac{1}{N-1}\overset{k}{\Sigma}\left[N_j - \frac{N_j}{n_j} + \frac{N}{n_j}P_j^2 - \frac{NP_j^2}{N_j}\right](M.P.)_{jb}$$

$$- \frac{1}{N-1}\left(\frac{1}{m} - \frac{1}{M}\right)\left[\overset{k}{\Sigma}\frac{N_j}{n_j}\left(1 - \frac{N}{N_j}P_j^2\right)(M.P.)_{jw}\right] \qquad (17)$$

Substituting the value of $(M.P.)_b$ as obtained from (17) in (10), we get the value of covariance of $\bar{x}_n, \bar{y}_n$ without stratification.

If $P_j = N_j/N$, Sukhatme (1954) has estimated the difference between variances without stratification and with stratification as,

$$\text{Est. } [V(\bar{y}_n)_{U.S.} - V(\bar{y}_n)_S] = \frac{N-n}{n(N-1)} \sum^k P_j (\bar{y}_{nj} - \bar{y}_{ns})^2$$

$$+ \sum^k \left[\frac{P_j}{n} - \frac{P_j^2}{n_j} - \frac{N-n}{N-1} \cdot \frac{1}{n} \cdot \frac{P_j(1-P_j)}{n_j}\right] S_{jb}^2 \quad (18)$$

Similar expression follows for x. Estimated difference between covariance $(\bar{y}_n \bar{x}_n)$ without stratification and with stratification can be expressed as,

$$\text{Est. } [\text{cov. } (\bar{y}_n \bar{x}_n)_{U.S.} - \text{cov. } (\bar{y}_n \bar{x}_n)_S] = \frac{N-n}{n(N-1)} \sum^k P_j (\bar{x}_{nj} \bar{y}_{nj} - \bar{x}_n \bar{y}_n)$$

$$+ \sum^k \left[\frac{P_j}{n} - \frac{P_j^2}{n_j} - \frac{N-n}{N-1} \cdot \frac{1}{n} \cdot \frac{P_j(1-P_j)}{n_j}\right] (M.P.)_{jb} \quad (19)$$

The estimated difference between the variances of a ratio without stratification and with stratification can be expressed as

$$V(\bar{r})_{U.S.} - V(\bar{r})_S$$

$$= \frac{\bar{y}_{NM}^2}{\bar{x}_{NM}^2} \left\{ \frac{\frac{N-n}{n(N-1)} \sum^k P_j (\bar{y}_{nj} - \bar{y}_{ns})^2 + \sum^k \left[\frac{P_j}{n} - \frac{P_j^2}{n_j} - \frac{N-n}{N-1} \cdot \frac{1}{n} \cdot \frac{P_j(1-P_j)}{n_j}\right] S_{jby}^2}{\bar{y}_{NM}^2} \right.$$

$$+ \frac{\frac{N-n}{n(N-1)} \sum^k P_j (\bar{x}_{nj} - \bar{x}_{ns})^2 + \sum^k \left[\frac{P_j}{n} - \frac{P_j^2}{n_j} - \frac{N-n}{N-1} \cdot \frac{1}{n} \cdot \frac{P_j(1-P_j)}{n_j}\right] S_{jbx}^2}{\bar{x}_{NM}^2}$$

$$\left. - 2 \, \frac{\frac{N-n}{n(N-1)} \sum^k P_j (\bar{x}_{nj} \bar{y}_{nj} - \bar{x}_{ns} \bar{y}_{ns}) + \sum^k \left[\frac{P_j}{n} - \frac{P_j^2}{n_j} - \frac{N-n}{N-1} \cdot \frac{1}{n} \cdot \frac{P_j(1-P_j)}{n_j}\right] (M.P.)_{jb}}{\bar{x}_{NM} \bar{y}_{NM}} \right\} \quad (20)$$

SUMMARY

In two-stage stratified random sampling supplementary information is sometimes available on the sub-units and can be used to estimate that ratio of the two characters y and x. Following Sukhatme (1950 *a*, *b* and 1954), who has worked out the efficiency of stratification for a single variate, formulæ for determining the efficiency of stratification

for the ratio method of estimation in two-stage sampling from each stratum, have been derived and are given in the present article.

ACKNOWLEDGEMENTS

The author is thankful to Dr. V. G. Panse, for intiating him into the problem.

REFERENCES

1. Sukhatme, P. V. .. "Sample Surveys in Agriculture," *Presidential Address, Ind. Sci. Cong. Poona*, 1950 *a*.

2. ——— .. "Efficiency of sub-sampling designs in yield surveys" *Jour. Ind. Soc. Agri. Stat.*, 1950 *b*, **2**, 212.

3. ——— .. *Sampling Theory of Surveys with Applications*, Ind. Soc. Agric. Stat., New Delhi and Iowa State College Press, Ames, Iowa, U.S.A., 1954.