Jour. Ind. Soc. Ag. Statistics
48(2), 1996 : 195–198

A Note on Ratio Estimator in Overlapping Clusters

S. J. Amdekar
G. B. Pant Univ. of Agri. and Tech., Pantnagar
(Received : January, 1994)

SUMMARY

The ratio method of estimation is used in case of overlapping clusters. From empirical study it is observed that overlapping cluster sampling with ratio estimator is generally better than simple random sampling with ratio estimator. Thus overlapping cluster sampling besides being operationally convenient is likely to be better from variance point of view also when used with Ratio estimator.

Key words : Overlapping clusters, Ratio estimator.

1. Introduction

Cluster sampling is adopted in practice due to its operational convenience even if list of all elements of the population is available. In survey with village as a sampling unit one of the practices is to select a random sample of some villages and form clusters by associating nearby villages (lying within a particular distance) to each of the selected village. This results in overlapping clusters in the sense that a particular village may belong to more than one cluster. Amdekar [1] has given an unbiased estimator of population mean for such situations.

In case of usual cluster sampling Zarkovic and Krane [4] have shown that correlation between auxiliary variable and study variable increases if cluster of units is taken as sampling unit and that the correlation increases with increase in cluster size. Subsequently, Mishro and Sukhatme [3] have shown that even if intraclass correlation is positive, cluster sampling in conjunction with ratio estimator is likely to be more efficient than simple random sampling in conjunction with ratio estimator. If cost is also considered the increase in efficiency may be even more. In light of these results in this note we study the use of ratio estimator in overlapping clusters.

2. The Estimator

Let there be N units in the population with y and x as study and auxiliary variables respectively. A SRS of n units is selected. With respect to each of the selected unit, a cluster of size M_i is formed by associating (M_i-1) nearby units. A reciprocal association scheme for cluster formation is assumed. That is, if unit U_j is a member of cluster of U_i then U_i also becomes member of cluster of $U_j (i \neq j = 1, 2, \ldots, N)$. Now, the ratio estimator of population mean is given by $\hat{\bar{Y}}_1 = (\bar{T}_n/\bar{W}_n)\bar{X}_N$ where $T_i = \sum_{j \in i} y_j/M_j$ and $W_i = \sum_{j \in i} x_j/M_j$, the sums being over those U_j which are members of U_i, $\bar{T}_n = \sum^{n} T_i/n$ and $\bar{W}_n = \sum^{n} W_i/n$. The estimator being biased, its MSE to the first order of approximation is given by

$$MSE(\hat{\bar{Y}}_1) = \frac{1}{n}(\sigma_T^2 - 2R\sigma_{TW} + R^2\sigma_W^2) = \frac{1}{nN}\sum^{N}(T_i - RW_i)^2$$

where σ_T^2, σ_{TW} and σ_W^2 have usual meaning and $R = \bar{T}_N/\bar{W}_N = \bar{Y}/\bar{X}$. On the other hand, if ratio method of estimation is used without formation of clusters the estimator and its MSE, based on an equivalent sample of $n\bar{M}$ units, $(\bar{M} = \sum^{N} M_i/N)$ are given respectively by $\hat{\bar{Y}}_2 = (\bar{y}/\bar{x})\bar{X}$ and

$$MSE(\hat{\bar{Y}}_2) = \frac{1}{n\bar{M}}\frac{1}{N}\sum(Y_i - RX_i)^2$$

Comparison of MSEs shows that $\hat{\bar{Y}}_1$ will be better than $\hat{\bar{Y}}_2$ if $\bar{M}\sum^{N}(T_i - RW_i)^2 < \sum^{N}(Y_i - RX_i)^2$ i.e. if $\bar{M}$ times variability in $(T_i - RW_i)$ is less than variability in $(Y_i - RX_i)$.

3. Empirical Study

For empirical study we consider two populations. The first population consists of 16 villages of a region (Lahiri [2]). The second population consists of 25 villages of a block of Nainital district (UP). Taking a number of different distances overlapping clusters were formed. Results are given in tables 1 and 2. While calculating cost efficiency we assume that for a given total cost we can have either n_o clusters of average size $\bar{M}$ or a sample of $n\bar{M}$ units without forming clusters. It is observed that correlation coefficient between T and W

is generally more than that between x and y. From the point of view of MSE the estimator $\hat{\bar{Y}}_1$ is better than estimator $\hat{\bar{Y}}_2$ except for a few cases. Further the cost efficiency of $\hat{\bar{Y}}_1$ is also generally, more than that of $\hat{\bar{Y}}_2$ as $n_o > n$. Thus, as in case of usual cluster sampling, the overlapping cluster sampling in conjunction with ratio estimator is likely to be more efficient than simple random sampling in conjunction with ratio estimator besides being operationally covenient.

Table 1. Performance of ratio estimator in overlapping clusters for Population I

x : Area of the village y : Population of the village Correlation coefficient between x and y = 0.5144				
Avg. cluster size	Corr. coeff. bet. T and W	MSE $(\hat{\bar{Y}}_1)$	MSE $(\hat{\bar{Y}}_2)$	Cost eff.
2.75	0.7165	5.93/n	6.00/n	$(1.01)n_o/n$
3.5	0.7285	3.66/n	4.71/n	$(1.29)n_o/n$
4.5	0.8631	1.56/n	3.67/n	$(2.35)n_o/n$
7.625	0.9051	0.79/n	2.16/n	$(2.73)n_o/n$

Table 2. Performance of ratio estimator in overlapping clusters for Population II

A.	x : Area y : No. of households in '81 Corr. coeff. between x and y = 0.6830			
Avg. cluster size	Corr. coeff. bet. T and W	MSE $(\hat{\bar{Y}}_1)$	MSE $(\hat{\bar{Y}}_2)$	Cost eff.
3.4	0.6814	2693/n	5085/n	$(1.88)n_o/n$
4.52	0.7078	1593/n	3825/n	$(2.40)n_o/n$
6.68	0.7147	1065/n	2588/n	$(2.43)n_o/n$
B.	x : Area y : No. of households in '71 Corr. coeff. between x and y = 0.8056			
3.4	0.8014	1621/n	1190/n	$(0.73)n_o/n$
4.52	0.7910	650/n	895/n	$(1.38)n_o/n$
6.68	0.8315	318/n	605/n	$(1.90)n_o/n$
C.	x : No. of households in '71 y : No. of households in '81 Corr. coeff. between x and y = 0.6287			
3.4	0.6466	1857/n	2061/n	$(1.11)n_o/n$
4.52	0.6700	1723/n	1558/n	$(0.90)n_o/n$
6.68	0.6848	852/n	1049/n	$(1.23)n_o/n$

ACKNOWLEDGEMENT

Author is grateful to the referee for suggesting the improvements in the presentation.

REFERENCES

[1] Amdekar S. J., 1985. An unbiased estimator in overlapping clusters. *Bull. Cal. Stat. Assoc.*, **34**, 231-232.

[2] Lahiri D. B., 1954. Technical paper on some aspects of the development of the sample design. *Sankhya*, **14**, 264-316.

[3] Mishro G. K. and Sukhatme B. V., 1972. Efficiency of cluster sampling in conjunction with ratio and regression methods of estimation. *J. Ind. Soc. Agri. Stat.*, **24**, 81-90.

[4] Zarkovic S. S. and Krane J., 1965. Some efficient ways of cluster sampling. Proc. of 35th session of Int. Stat. Instt., Belgrade.