

# EFFICIENT ESTIMATOR IN SUCCESSIVE SAMPLING USING POST-STRATIFICATION

M. Trivedi\* and D. Shukla\*\*

## ABSTRACT

It is often seen that a population having large number of elements remains unchanged in several occasions but the value of unit's changes. The sample surveys are also not limited to one-time inquiries. In this paper, an estimator has been introduced under successive surveys. This estimator is unbiased and efficient over Post-stratification estimator. In this paper, minimum variance of the optimum estimator has been derived and comparative study is incorporated.

**Key Words :** Post-stratification, Successive Occasions, Optimum, Estimator.

---

\*Department of Applied Mathematics ,Birla Institute of Technology, Mesra, Ranchi (JH)

\*\*Department of Statistics, Lucknow University, Lucknow(UP)

E-mail : [manish\\_trivedi1976@yahoo.com](mailto:manish_trivedi1976@yahoo.com) ( M. Trivedi – Corresponding Author )

## 1. INTRODUCTION:

Today, it is often seen that sample surveys are not limited to one time inquiries. If the value of study character of a finite population is subject to change over time, a survey carried out on a single occasion will provide information about the characteristics of the surveyed population for the given occasion only and can not give any information on the nature or the rate of change of the characteristic over all occasions or more recent occasion. Smith, T.M.F. (1991), Holt, D. and Smith, T.M.F. (1979), Jagers, P. Oden, A. and Trulsson, L. (1985) and Jagers P. (1986) have done a lot of inspired work in the area of post-stratification in sample survey. Further the theory of Post-stratification in sample survey was extended by Cochran, W.G. (1977) , Gupta, S.C. and Kapoor, V.K. (1977), Ravindra Singh and Sukatme, B.V. (1969) and (1973), Singh, D. and Chaudhary, F.S. (1984), Sukatme, P.V. Sukatme, B.V., Sukatme, S. and Asok, C. (1984). Data regarding changing properties of the populations of cities or counties, such as unemployment statistics, are collected regularly on a sample basis, to estimate the changes from one occasion to the next or to estimate the average over a certain period. An important aspect of continuous surveys is the structure of the sample on each occasion. To meet these requirements, successive sampling provides a strong tool for generating the reliable estimates at different occasions.

Theory of successive sampling appears to have started with the work of Jessen (1942). He was pioneer to utilize the entire information collected in the previous occasions. Further the theory of successive sampling was extended by Patterson (1950), Rao and Graham (1964), Gupta(1979), Das(1982), Chaturvedi and Tripathi and many others. Feng and Zou (1997) used the auxiliary information on both the occasions for estimating the current mean in successive sampling.

There are several types of procedures to adopt for estimating the population parameters: (i) the same sample may be used on each occasion (ii) a new sample may be taken on each occasion, (iii) a part of the sample may be retained while the remainder of the sample may be drawn afresh. Some conditions to consider are that (i) for estimating change from one occasion to the next, it may be best to retain the sample on each occasion, (ii) for estimating the mean on each occasion, it may be best to draw a fresh sample on each occasion, and (iii) if it is desire to estimate the mean on each occasion and also the change from one occasion to the next, it may be best to retain part of the sample and draw the remainder of the sample afresh.

## 2. NOTATIONS :

Let a population be of size  $N$ , that is sampled over two occasions. Assume that the size of the population remains unchanged so  $n_1 = n_2$ , but values of units change over occasions. Assume that we have two parts of the sample in second occasion. First part of the sample consists of retained  $n_2'$  units of first occasions and the second part has the  $n_2''$  units drawn afresh in second occasion where  $n_2 = n_2' + n_2''$ .

$\bar{y}_i$  - The population mean on the  $i$ -th occasion,  $i = 1, 2, \dots, n$ .

$S_i^2$  - The population mean square error for the  $i$ th occasion,  $i = 1, 2, \dots, n$ .

$\bar{Y}_1$  - The sample mean based on  $n_1$  units observed on the first occasion.

$\bar{y}_2'$  - The sample mean based on  $n_2'$  units observed on the second occasion and common with the first occasion.

$\bar{y}_2''$  - The sample mean based on  $n_2''$  units drawn afresh on the second occasion.

$\bar{x}_1'$  - The sample mean based on  $n_2'$  units common to both the occasions and observed on the first occasion.

$W_i$  - the proportion  $N_i/N$

$n_2''$  - the sample size of the sample drawn afresh on the second occasion.

$n_1$  - sample units observed on the first occasion.

$n_2'$  - units observed on the second occasion and common with the first occasion.

## 3. ESTIMATION STRATEGY :

- (i) We assumed the population of size  $N$  remains unchanged over both occasions.
- (ii) We have  $n_2$  units constitutes the sample on the first occasion of which  $n_2'$  are retained on the second occasion while  $n_2'' = n_2 - n_2'$  are drawn afresh on the second occasion from  $(N - n_1)$  units.
- (iii) Now we have post stratified the  $n_2''$  units in to  $k$  strata, which are drawn afresh in second occasion. While the  $n_2'$  units, which retained from first occasion remain unchanged.

#### 4. THE ESTIMATOR :

For estimating  $\bar{Y}_2$  based on successive sampling using post-stratification scheme, we propose an estimator  $\bar{y}_{pss}$ , such that

$$\bar{y}_{pss} = \phi \bar{y}_{2s}'' + (1 - \phi) \bar{y}_{12}' \quad \dots \quad (4.1)$$

where,  $\phi$  is a constant. The motivation of taking this constant is taken from Agrawal, M.C. and Panda, K.B. (1993) and (1995) and the form of equation (4.1) is important for various applications, as to be mentioned in section 5. This appears to belong to the class of estimators known as “composite estimators.” See, section 5 in Henry, Strudler, and Chen (2007), and also FCSM (1993).

$\bar{y}_{2s}''$  is the mean based on post stratified  $n_2''$  units  $\left( \sum_{i=1}^k W_i \bar{y}_{2i}'' \right)$

$\bar{y}_{12}'$  is the mean based on the matched sample  $n_2'$ , which is termed as

$$\bar{y}_{12}' = \bar{y}_2' + \beta_{21} (\bar{y}_1' - \bar{x}_1')$$

Where  $\beta_{21}$  the regression coefficient of the variate of the second occasion on the variate of the first occasion is assumed to be known. In addition,  $\bar{y}_2', \bar{y}_1'$  and  $\bar{x}_1'$  are defined before. Then the proposed estimator  $\bar{y}_{pss}$  can be written as

$$\bar{y}_{pss} = \phi \left( \sum_{i=1}^k W_i \bar{y}_{2i}'' \right) + (1 - \phi) \left[ \bar{y}_2' + \beta_{21} (\bar{y}_1' - \bar{x}_1') \right] \quad \dots \quad (4.2)$$

**THEOREM 4.1** The estimator  $\bar{y}_{pss}$  is unbiased for  $\bar{Y}_2$

**Proof:** Clearly  $\bar{y}_1'$  is an unbiased estimator of  $\bar{Y}_1$  with variance given by

$$V(\bar{y}_1') = \left( \frac{1}{n_1} - \frac{1}{N} \right) S_1^2.$$

To estimate the mean  $\bar{Y}_2$  on the second occasion we have two estimators, one based on  $n_2''$  the sample drawn afresh on the second occasion and then after post stratified in to k-strata's. In addition, the other based on  $n_2'$  the sample common to both the occasion.

The estimator  $\bar{y}_{2s}''$  based on  $n_2''$  is an unbiased estimator of  $\bar{Y}_2$  such that

$$E \left[ \bar{y}_{2s}'' \right] = E \left[ \sum_{i=1}^k W_i \bar{y}_{2i}'' \right] = \sum_{i=1}^k W_i \left( \bar{y}_{2i}'' \right) = \sum_{i=1}^k W_i \bar{y}_{2i} = \bar{Y}_2 \quad \dots \quad (4.1.1)$$

Now, the estimator based on the matched sample  $n_2'$  is also unbiased estimator for  $\bar{Y}_2$  such that

$$\begin{aligned} E \left[ \bar{y}_{12}' \right] &= E \left[ \bar{y}_2' + \beta_{21} (\bar{y}_1' - \bar{x}_1') \right] \\ &= \left( \bar{y}_2' \right) + \beta_{21} \left\{ E(\bar{y}_1') - E(\bar{x}_1') \right\} \\ &= \bar{Y}_2 + \beta_{21} \left\{ \bar{Y}_1 - \bar{Y}_1 \right\} = \bar{Y}_2 \quad \dots \quad (4.1.2) \end{aligned}$$

Hence, the estimator  $\bar{y}_{pss}$  is unbiased estimator for  $\bar{y}_2$ , such as,

$$\begin{aligned} E(\bar{y}_{pss}) &= \phi E[\bar{y}_{2s}^{\prime\prime}] + (1-\phi) E(\bar{y}_{12}^{\prime}) \\ &= \phi \bar{Y}_2 + (1-\phi) \bar{Y}_2 = \bar{Y}_2 \quad \dots \quad (4.1.3) \end{aligned}$$

From equation no. (4.1.1) & (4.1.2).

**THEOREM 4.2** The variance of the estimator  $\bar{y}_{pss}$  is given by

$$\begin{aligned} V(\bar{y}_{pss}) &= \phi^2 \left[ \left( \frac{1}{n} - \frac{1}{N} \right) \sum_{i=1}^k W_i S_{2i}^2 + \frac{(N-n)}{(N-1)n^2} \sum_{i=1}^k (1-W_i) S_{2i}^2 \right] \\ &+ (1-\phi)^2 \left[ \left( \frac{1}{n_2} - \frac{1}{n_1} \right) S_2^2 (1-\rho_{2i}^2) + \left( \frac{1}{n_1} - \frac{1}{N} \right) S_2^2 \right] - 2\phi(1-\phi) \frac{S_{2i}^2}{N} \end{aligned}$$

**Proof:** We have the estimator

$$\bar{y}_{pss} = \phi \bar{y}_{2s}^{\prime\prime} + (1-\phi) \bar{y}_{12}^{\prime}$$

Then

$$V(\bar{y}_{pss}) = \phi^2 \text{Var}(\bar{y}_{2s}^{\prime\prime}) + (1-\phi)^2 \text{Var}(\bar{y}_{12}^{\prime}) + 2\phi(1-\phi) \text{Cov}(\bar{y}_{2s}^{\prime\prime}, \bar{y}_{12}^{\prime}) \dots \quad (4.2.1)$$

The estimator  $\bar{y}_{2s}^{\prime\prime}$  is a stratified mean then

$$V(\bar{y}_{2s}^{\prime\prime}) = V \left[ \sum_{i=1}^k W_i \bar{y}_{2i}^{\prime\prime} \right] = \left( \frac{1}{n} - \frac{1}{N} \right) \sum_{i=1}^k W_i S_{2i}^2 + \frac{(N-n)}{(N-1)n^2} \sum_{i=1}^k (1-W_i) S_{2i}^2 \dots \quad (4.2.2)$$

In addition, the estimator  $\bar{y}_{12}^{\prime}$  is a regressed estimator, which has the variance.

$$\begin{aligned} V(\bar{y}_{12}^{\prime}) &= V \left[ \bar{y}_2 + \beta_{21} (\bar{y}_1 - \bar{x}_1) \right] \\ &= V(\bar{y}_2) + \beta_{21}^2 \left[ V(\bar{y}_1) - V(\bar{x}_1) \right] \\ &= \left( \frac{1}{n_2} - \frac{1}{N} \right) S_2^2 + \rho_{21}^2 \frac{S_2^2}{S_1^2} \left[ \left( \frac{1}{n_1} - \frac{1}{N} \right) S_1^2 - \left( \frac{1}{n_2} - \frac{1}{N} \right) S_1^2 \right] \\ &= \left( \frac{1}{n_2} - \frac{1}{n_1} \right) S_2^2 + \left( \frac{1}{n_1} - \frac{1}{N} \right) S_2^2 + \rho_{21}^2 \left( \frac{1}{n_1} - \frac{1}{n_2} \right) S_2^2 \\ &= \left( \frac{1}{n_2} - \frac{1}{n_1} \right) S_2^2 (1-\rho_{21}^2) + \left( \frac{1}{n_1} - \frac{1}{N} \right) S_2^2 \quad \dots \quad (4.2.3) \end{aligned}$$

$$\text{Where } \beta_{21} = \frac{\rho_{21} S_2}{S_1}$$

After that one can easily get

$$\text{Cov}(\bar{y}_{2s}^{\prime\prime}, \bar{y}_{12}^{\prime}) = -S_{2i}^2 / N \quad \dots \quad (4.2.4)$$

Now by putting the values from (4.2.2), (4.2.3) and (4.2.4) in equation (4.2.1) one gets

$$\begin{aligned}
V(\bar{y}_{pss}) &= \phi^2 \left[ \left( \frac{1}{n} - \frac{1}{N} \right) \sum_{i=1}^k W_i S_{2i}^2 + \frac{(N-n)}{(N-1)n^2} \sum_{i=1}^k (1-W_i) S_{2i}^2 \right] \\
&+ (1-\phi)^2 \left[ \left( \frac{1}{n_2} - \frac{1}{n_1} \right) S_2^2 (1-\rho_{21}^2) + \left( \frac{1}{n_1} - \frac{1}{N} \right) S_2^2 \right] - 2\phi(1-\phi) \sum_{i=1}^k \delta_{2i}^2 / N
\end{aligned}$$

## 5 OPTIMUM CHOICE

For getting the optimum value of,  $\phi$  we have to differentiate  $\text{Var}(\bar{y}_{pss})$  expression with respect to  $\phi$  and then equate it to zero. Then we can easily get

$$\phi_{\text{Opt}} = \frac{\text{Var}(\bar{y}_{12}^{\prime}) - \text{Cov}(\bar{y}_{21}^{\prime}, \bar{y}_{12}^{\prime})}{\text{Var}(\bar{y}_{12}^{\prime}) + \text{Var}(\bar{y}_{2S}^{\prime}) - \text{Cov}(\bar{y}_{2S}^{\prime}, \bar{y}_{12}^{\prime})} \quad \dots \quad (5.1)$$

By putting this value in variance expression one can get

$$V(\bar{y}_{pss})_{\text{opt}} = \frac{V(\bar{y}_{2S}^{\prime})V(\bar{y}_{12}^{\prime}) - [\text{Cov}(\bar{y}_{2S}^{\prime}, \bar{y}_{12}^{\prime})]^2}{V(\bar{y}_{12}^{\prime}) + V(\bar{y}_{2S}^{\prime}) - 2\text{Cov}(\bar{y}_{2S}^{\prime}, \bar{y}_{12}^{\prime})} \quad \dots \quad (5.2)$$

The derived equation (5.2) is the expression of the optimum variance. It has recently been noted that under a different context, Granger and Newbold (1986), using Bates and Granger (1969), arrived at a variance estimate of this same form. In that case it was used for the combination of forecasts for time series. However, it may be used for any combination of estimators of the form (4.1).

## 6. COMPARISON:

We see the variance expression

$$V(\bar{y}_{pss}) = \phi^2 \left[ \text{Var}(\bar{y}_{2S}^{\prime}) \right] + (1-\phi)^2 \left[ \text{Var}(\bar{y}_{12}^{\prime}) \right] + 2\phi(1-\phi) \text{Cov}(\bar{y}_{2S}^{\prime}, \bar{y}_{12}^{\prime})$$

$$\text{And } V(\bar{y}_s) = \phi^2 \left[ \text{Var}(\bar{y}) \right] + (1-\phi)^2 \left[ \text{Var}(\bar{y}_{12}^{\prime}) \right] + 2\phi(1-\phi) \text{Cov}(\bar{y}, \bar{y}_{12}^{\prime})$$

Where  $V(\bar{y}_s)$  is the variance of general systematic sampling scheme. Obviously, one can say easily  $\text{Var}(\bar{y}_{pss}) \leq \text{Var}(\bar{y}_s)$

$$\text{Because } \text{Var}(\bar{y}_{2S}^{\prime}) \leq \text{Var}(\bar{y})$$

Where  $\text{Var}(\bar{y}_{2S}^{\prime})$  and  $\text{Var}(\bar{y})$  are the variances of sample mean of post stratified sample and sample mean of simple random sample scheme.

## REFERENCES

1. Agrawal, M.C. and Panda, K.B. (1993): An efficient estimator in Post stratification, *METRON*, Vol. 51, 3-4, 179-187.
2. Agrawal, M.C. and Panda, K.B. (1995): An efficient estimator in Post stratification, *METRON*, LIII, 3-4, 107-115.
3. Bates, J.M., and Granger, C.W.J. [1969], The combination of forecasts, *Oper. Res. Q.* **20**, 451-468.
4. Cochran, W.G. (1977): Sampling techniques, John Weily and Sons Inc. New York.
5. Das, A. K. (1982) : Estimation of Population ratio on two occasions. *Jour. Ind.Soc. Ag. Statist.*, 34, 1-9.
6. Feng, S. and Zou, G. (1997) : Sample Rotation Method with auxiliary Variable. *Commun. Statist. Theo-Meth.*,26, 6, 1497-1509.
7. FCSM (1993): Indirect Estimators in Federal Programs. (US) Federal Committee on Survey Methodology, Working Paper #21, <http://www.fcsm.gov/>.
8. Gupta, S.C. and Kapoor, V.K. (1977): Fundamentals of Applied Statistics: Sultanchand and sons, New Delhi.
9. Gupta, P.C. (1979) : Sampling on Two Successive Occassions. *Jour. Statist. Res.*,13,7-16.
10. Granger, C.W.J. and Paul Newbold (1986): The Combination of Forecasts, *Forecasting Economic Time Series*, 2<sup>nd</sup> edition, Academic Press, pp. 266 – 276.
11. Henry, K., Strudler, M., and Chen, W. (2007): An Empirical Evaluation of Various Direct, Synthetic, and Traditional Composite Small Area Estimators, American Statistical Association, Survey Research Methods Section, proceedings from August 2007 Joint Statistical Meetings, page 3122. (On CD.)
12. Holt, D. and Smith, T.M.F. (1979): Post-stratification, *J.R. Stat. Soc.*, A, 142, 33-36.
13. Jagers, P. Oden, A. and Trulsson, L. (1985): Post-stratification and ratio estimator, *Int. Stat. Rev.*, 53, 221-238.
14. Jagers P. (1986) : Post-stratification against bias in sampling, *Int. Stat. Rev.*, 54, 159-167.
15. Jessen, R. J. (1942): Statistical investigation of a sample survey for obtaining form facts. *Iowa Agr. Expt. Sta. Res. Bull.*, 304.
16. Patterson, H.D. (1950) : Sampling on successive occasions with partial replacement of units. *Jour. Royal Statist. Assoc., Scr. B*, 12,241-255.
17. Rao, J. N.K. AND Graham, J. E. (1964) : Rotation Design for sampling on repeated occasions. *Jour. Amer. Statist. Assoc.*, 59, 492-509.
18. Ravindra Singh and Sukatme, B.V. (1969): Optimum stratification. *Ann. Inst. Stat. Math.*, 21, 515-528.
19. Ravindra Singh and Sukatme, B.V. (1973): Optimum stratification with ratio and regression methods of estimation. *Ann. Inst. Stat. Math*, 25, 627-633.
20. Smith, T.M.F. (1991): Post stratification, the statistician, 40, 315-323.
21. Singh, D. and Chaudhary, F.S. (1984), Theory and analysis of sample survey design: New Age International (P) Lon (formally) Wiley Eastern Ltd. New Delhi.
22. Sukatme, P.V. Sukatme, B.V., Sukatme, S. and Asok, C. (1984): Sampling Theory of Survey with Applications, Iowa State University Press, Indian Society of Agricultural Statistics, New Delhi.