Computational Approach to Generalized Ratio Type Estimator of Population Mean Under Two Phase Sampling

Subhash Kumar Yadav
Department of Mathematics and Statistics (A Centre of Excellence),
Dr. RML Avadh University,
Faizabad-224001, U.P., INDIA
Email: drskystats@gmail.com

S. S. Mishra*
Department of Mathematics and Statistics (A Centre of Excellence),
Dr. RML Avadh University,
Faizabad-224001, U.P., INDIA
Email: sant_x2003@yahoo.co.in

Vishwas Tiwari
Department of Statistics,
D.A-V. College, Kanpur,
U.P., INDIA
Email: vishwas.stats@gmail.com

Alok Kumar Shukla
Department of Statistics,
D.A-V. College, Kanpur, U.P., INDIA
Email: alokshukladav@gmail.com

Abstract-In the present draft, we propose the computational approach to generalized ratio type estimator of population mean of the main variable under study using auxiliary information. The expressions for the bias and mean square errors (MSE) have been obtained up to the first order of approximation. The minimum value of the MSE of the proposed estimator is also obtained for the optimum value of the characterizing scalar. A comparison has been made with the existing estimators of population mean in two phase sampling. A computing based on numerical example also carried out which shows improvement of proposed estimator over other estimators in two phase sampling as the proposed estimator has lesser mean squared error.

Key Words: *Computing, Two phase sampling, Auxiliary variable, Bias, MSE, Efficiency.*

*****

1. INTRODUCTION

The simple random sampling technique is the most appropriate sampling technique for the estimation of population parameters when the population units for the main characteristic under study are homogeneous. The auxiliary information provided by the auxiliary variable is sampling theory for the improved estimation of the population parameters of the main characteristic under study. The auxiliary information is in use since the development of the sampling theory itself and its application to different domains of the real world. The auxiliary variable is highly correlated (positively or negatively) with the main variable under study. The auxiliary information used at both the stages of designing and the estimation stages of the sampling. In the present manuscript we have used it at estimation stage for estimating the population mean of the main variable under study in two phase sampling.

Many authors including Upadhyaya and Singh (1999), Singh (2003), Singh and Tailor (2003), Singh *et.al* (2004, 2008, 2010), Singh *et al.* (2004), Kadilar and Cingi (2004, 2006), Tailor and Sharma (2009), Yan and Tian (2010), Yadav (2011), Pandey *et al.* (2011), Subramani and Kumarapandiyan (2012), Solanki *et al.* (2012), Onyeka (2012), Jeelani *et.al* (2013), Yadav and Kadilar (2013) etc. used auxiliary information for improved estimation of population mean of the study variable under simple random sampling and two phase sampling. Computational methods have been widely used in boiling down the numerical complexity including inventory populations in supply chain management etc for attaining efficient estimators of performance measures of inventory system vide for examples Mishra and Singh (2012, 2012, 2013), Mishra and Mishra (2013) and Yadav et al.(2014)

The fresh objective of this paper is to present the computational approach to generalized ratio type estimator of population mean of the main variable under study using auxiliary information. A comparison has been made with the existing estimators of population mean in two phase sampling. A computational algorithm has been developed to compute proposed estimator over other estimators in two phase sampling as the proposed estimator has lesser mean squared error.

2. STATISTICAL ANALYSIS

Let the main variable under study be denoted by y and the auxiliary variable by x respectively. When the main variable y under study is highly positively correlated with the auxiliary variable x and the line of regression of y on x passes through origin, the ratio type estimators are used to estimate the population parameters of the main variable under study and on the other hand the product type estimators are used to estimate the parameter under study when y and x are highly negatively correlated to each other otherwise regression method of estimation is used to estimate the population parameters of the main variable. In the present study we have considered the case of positive correlation and have used the ratio type estimators for the estimation of population mean in two phase sampling.

Let the finite population consists of N distinct and identifiable units under study. A random sample of size n is drawn using simple random sampling without replacement (SRSWOR) technique. Let $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ and $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ be the population means of study and the auxiliary variables and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ be the respective sample means. When $\bar{X}$ is not known, double sampling or two phase sampling is used to estimate the population mean of the study variable y. Under double sampling scheme the following procedure is used for the selection of the required sample,

(i) A large sample S' of size n' $(n' < N)$ is drawn from the population by SRSWOR and the observations are taken only on the auxiliary variable x to estimate the population mean $\bar{X}$ of the auxiliary variate.

(ii) Then the sample S of size n $(n < N)$ is drawn either from S' or directly from the population of size N to observe both the study variable and the auxiliary variable.

The appropriate estimator for estimating population mean is the sample mean,

$$t_0 = \bar{y} \qquad (1.1)$$

The variance of t_0, up to the first order of approximation is,

$$V(t_0) = f_1 C_y^2 \qquad (1.2)$$

Where,

$f_1 = \left(\frac{1}{n} - \frac{1}{N}\right)$, $C_y = \frac{S_y}{\bar{Y}}$ and

$$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2 .$$

Cochran (1940) proposed the classical ratio type estimator using auxiliary variable in simple random sampling as,

$$t_R = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right) \qquad (1.3)$$

The double sampling version of Cochran (1940) estimator is defined as,

$$t_R^d = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right) \qquad (1.4)$$

where $\bar{x}' = \frac{1}{n'}\sum_{i=1}^{n'} x_i$ is an unbiased estimator of population mean $\bar{X}$ of auxiliary variable based on the sample of size n'.

The bias and the mean square error of t_R^d, up to the first order of approximation respectively are,

$$B(t_R^d) = \bar{Y} f_3 [C_x^2 - \rho_{xy} C_x C_y] \qquad (1.5)$$

$$MSE(t_R^d) = \bar{Y}^2 [f_1 C_y^2 + f_3 (C_x^2 - 2\rho_{xy} C_x C_y)] \qquad (1.6)$$

where,

$f_3 = (f_1 - f_2) = \left(\frac{1}{n} - \frac{1}{n'}\right)$,

$f_2 = \left(\frac{1}{n'} - \frac{1}{N}\right)$ $C_x = \frac{S_x}{\bar{X}}$, $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$

and $\rho_{yx} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})(x_i - \bar{X})$.

Singh and Tailor (2003) proposed the following estimator of population mean of study variable in simple random sampling using correlation coefficient between x and y as,

$$t_{ST} = \bar{y}\left(\frac{\bar{X} + \rho_{yx}}{\bar{x} + \rho_{yx}}\right) \qquad (1.7)$$

Malik and Tailor (2013) suggested the double sampling version estimator of Singh and Tailor (2003) estimator as,

$$t_{ST}^d = \bar{y}\left(\frac{\bar{x}' + \rho_{yx}}{\bar{x} + \rho_{yx}}\right) \qquad (1.8)$$

The bias and the mean square error of t_{ST}^d, up to the first order of approximations respectively are,

$$B(t_{ST}^d) = \bar{Y} f_3 \theta [C_x^2 - \rho_{yx} C_x C_y] \qquad (1.9)$$

$$MSE(t_{ST}^d) = \bar{Y}^2 [f_1 C_y^2 + f_3 \theta (\theta C_x^2 - 2\rho_{yx} C_x C_y)] \qquad (1.10)$$

where, $\theta = \frac{\bar{X}}{\bar{X} + \rho_{yx}}$.

3. PROPOSED ESTIMATOR

Motivated by Malik and Tailor (2013) estimator, we propose the generalized estimator of population mean in two phase sampling as,

$$t = \bar{y}\left[\alpha + (1-\alpha)\left(\frac{\bar{x}' + \rho_{yx}}{\bar{x} + \rho_{yx}}\right)\right] \qquad (2.1)$$

where α is a characterizing scalar to be determined such that mean square error of t is minimum.

To study the large sample properties of the proposed estimator, we have the following approximations as,

$\bar{y} = \bar{Y}(1+e_0)$, $\bar{x} = \bar{X}(1+e_1)$ and $\bar{x}' = \bar{X}(1+e_2)$

such that $E(e_0) = E(e_1) = E(e_2) = 0$ and

$E(e_0^2) = f_1 C_y^2$, $E(e_1^2) = f_1 C_x^2$, $E(e_2^2) = f_2 C_y^2$,

$E(e_0 e_1) = f_1 \rho_{yx} C_y C_x$, $E(e_0 e_2) = f_2 \rho_{yx} C_y C_x$,

$E(e_1 e_2) = f_2 C_x^2$.

Expressing the proposed estimator in terms of e_i's, we have

$$t = \bar{Y}(1+e_0)\left[\alpha + (1-\alpha)\left(\frac{\bar{X}(1+e_2)+\rho_{yx}}{\bar{X}(1+e_1)+\rho_{yx}}\right)\right]$$

$$= \bar{Y}(1+e_0)[\alpha + (1-\alpha)(1+\theta e_2)(1+\theta e_1)^{-1}]$$

$$= \bar{Y}(1+e_0)[\alpha + (1-\alpha)(1+\theta e_2)(1-\theta e_1 + \theta^2 e_1^2 - ...)$$

$$= \bar{Y}(1+e_0)[\alpha + (1-\alpha)(1-\theta e_1 + \theta e_2 + \theta^2 e_1^2 - \theta^2 e_1 e_2 + ...)]$$

$$= \bar{Y}(1+e_0)[1-(1-\alpha)(\theta e_1 - \theta e_2 - \theta^2 e_1^2 + \theta^2 e_1 e_2 + ...)]$$

$$= \bar{Y}(1+e_0)[1-(1-\alpha)(\theta e_1 - \theta e_2 - \theta^2 e_1^2 + \theta^2 e_1 e_2 + ...)]$$

$$= \bar{Y}(1+e_0)[1-\alpha_1(\theta e_1 - \theta e_2 - \theta^2 e_1^2 + \theta^2 e_1 e_2 + ...)] + \theta^2 e_1 e_2 + ...)]$$

where $\alpha_1 = (1-\alpha)$

$$= \bar{Y}[1+e_0 - \alpha_1(\theta e_1 - \theta e_2 - \theta^2 e_1^2 + \theta^2 e_1 e_2 + ...) - \alpha_1(\theta e_0 e_1 - \theta e_0 e_2)]$$

$$t - \bar{Y} = \bar{Y}[e_0 - \alpha_1(\theta e_1 - \theta e_2 - \theta^2 e_1^2 + \theta^2 e_1 e_2 + ...) - \alpha_1(\theta e_0 e_1 - \theta e_0 e_2)] \quad (2.2)$$

Taking expectations of above equation on both sides, we have bias up to the first order of approximation of t as,

$$B(t) = \bar{Y}\alpha_1[\theta^2 f_1 C_x^2 - \theta^2 f_2 C_x^2 - \theta(f_1 - f_2)\rho_{yx} C_y C_x]$$

$$= \bar{Y}\alpha_1[\theta^2 f_3 C_x^2 - \theta f_3 \rho_{yx} C_y C_x]$$

$$= \bar{Y}\alpha_1 \theta f_3[\theta C_x^2 - \rho_{yx} C_y C_x] \quad (2.3)$$

From (2.2) up to the first order of approximation, we have

$$t - \bar{Y} \approx \bar{Y}(e_0 - \alpha_1 \theta e_1 + \alpha_1 \theta e_2) \quad (2.4)$$

Squaring on both sides and taking expectations on both sides, we get the mean square error of t up to the first order of approximation as,

$$MSE(t) = \bar{Y}^2 E(e_0^2 + \alpha_1^2 \theta^2 e_1^2 + \alpha_1^2 \theta^2 e_2^2 - 2\alpha_1^2 \theta^2 e_1 e_2 - 2\alpha_1 \theta e_0 e_1 + 2\alpha_1 \theta e_0 e_2)$$

$$= \bar{Y}^2[E(e_0^2) + \alpha_1^2 \theta^2 E(e_1^2) + \alpha_1^2 \theta^2 E(e_2^2) - 2\alpha_1^2 \theta^2 E(e_1 e_2) - 2\alpha_1 \theta E(e_0 e_1) + 2\alpha_1 \theta E(e_0 e_2)]$$

which is minimum for,

$$\alpha_1 = \frac{1}{\theta}\frac{[E(e_0 e_1) - E(e_0 e_2)]}{[E(e_1^2) + E(e_2^2) - 2E(e_1 e_2)]} = \frac{1}{\theta}\frac{A}{B}$$

where,

$$A = E(e_0 e_1) - E(e_0 e_2) = (f_1 - f_2)\rho_{yx} C_y C_x = f_3 \rho_{yx} C_y C_x$$

$$B = E(e_1^2) + E(e_2^2) - 2E(e_1 e_2) = f_3 C_x^2$$

The minimum mean square error of t the optimum value of α_1 is,

$$MSE_{\min}(t) = \bar{Y}^2\left[f_1 C_y^2 - \frac{A^2}{B}\right] \quad (2.5)$$

4. EFFICIENCY COMPARISON

From (1.2) and (2.5), we have

$$V(t_0) - MSE_{\min}(t) = \bar{Y}^2 f_3 \rho_{yx}^2 C_y^2 > 0 \quad (3.1)$$

From (1.6) and (2.5), we have

$$MSE(t_R^d) - MSE_{\min}(t) = \bar{Y}^2 f_3 (C_x - \rho_{yx} C_y)^2 > 0 \quad (3.2)$$

From (1.10) and (2.5), we have

$$MSE(t_{ST}^d) - MSE_{\min}(t) = \bar{Y}^2 f_3 (\theta C_x - \rho_{yx} C_y)^2 > 0 \quad (3.3)$$

5. COMPUTING ALGORITHM AND NUMERICAL ILLUSTRATION

The following algorithm has been developed to compute the estimator and its efficiency.

i. Begin
ii. Data input
iii. Compute sample mean of first inventory population
iv. Compute sample mean of second inventory population
v. Compute sample and population means of auxiliary inventory population
vi. Compute estimator for ratio of two inventory populations
vii. Compute biases for all estimators
viii. Compute MSE
ix. Compute efficiency (Percentage Relative Efficiency-PRE)
x. If PRE is greater than previous ones

xi. Find efficient estimator
xii. Data output
xiii. End

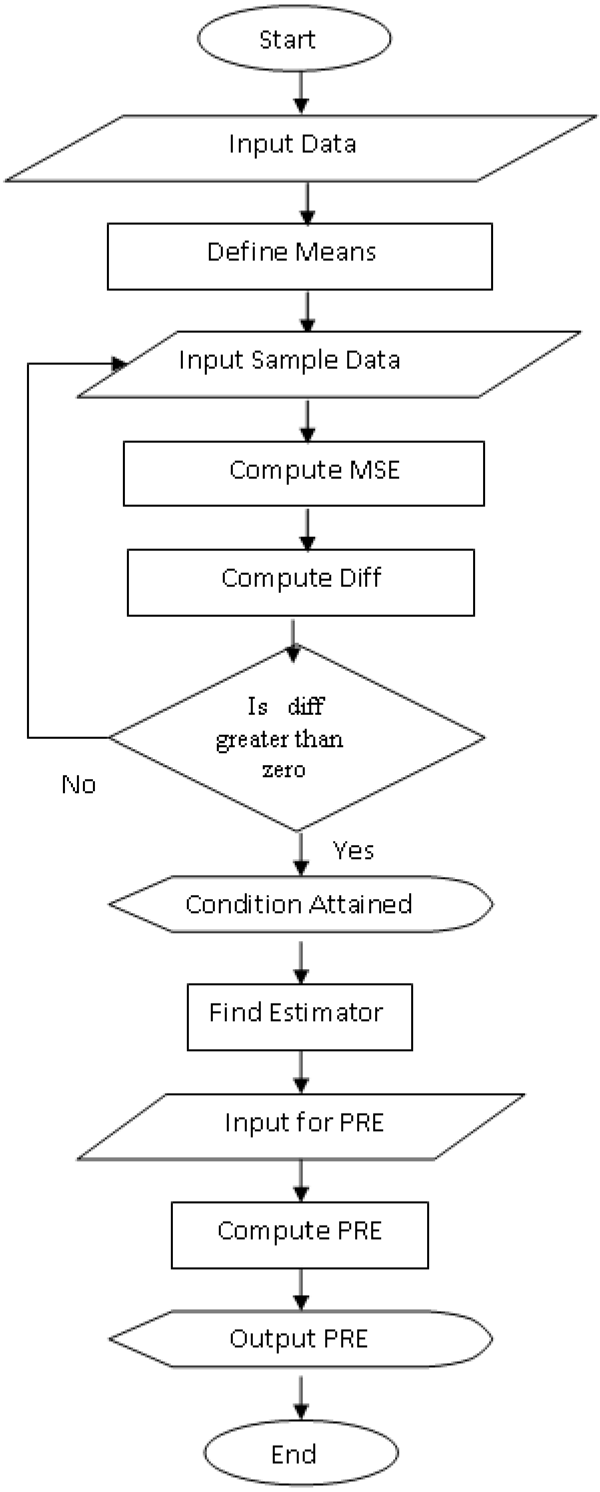

Figure 5.1: Tabular form of the algorithm (Computing flow chart)

6. EMPIRICAL EXAMPLE

To examine the performances of the proposed and the existing estimators of population mean in two phase random sampling, we have considered two populations given below,

Population-I [Source: Das, 1988]

Y: the number of agricultural labours for 1971,

X: the number of agricultural labours for 1961,

$\bar{Y} = 39.068$, $\bar{X} = 25.111$, $N = 278$, $n = 60$, $n' = 180$,

$C_y = 1.4451$, $C_x = 1.6198$, $\rho_{yx} = 0.7213$.

Population-II [Source: Cochran, 1977]

Y: the number of persons per block,

X: the number of rooms per block,

$\bar{Y} = 101.10$, $\bar{X} = 58.80$, $N = 20$, $n = 8$, $n' = 12$,

$C_y = 0.14450$, $C_x = 0.12810$, $\rho_{yx} = 0.6500$.

Table-1: Percentage Relative Efficiency (PRE) of t_0, t_R^d, t_{ST}^d and t with respect to t_0.

Estimator	PRE(., $t_0 = \bar{y}$)	
	Population-I	**Population-II**
t_0	100.00	100.00
t_R^d	142.11	117.65
t_{ST}^d	150.00	125.00
t	**178.60**	**130.67**

7. RESULTS AND CONCLUSION

In the present manuscript we have proposed a computational approach to generalized exponential type ratio estimator for population mean using auxiliary information. The large sample properties have been studied up to the first order of approximation. A comparison has been made with the existing estimators of population mean in two phase random sampling. From the strong theoretical recommendation and the results in table-1, we infer that the proposed estimator t is better than the sample mean, classical ratio estimator and the Malik and Tailor (2013) estimator as it has lesser mean square error. Therefore the proposed estimator should be preferred for the estimation of population mean in two phase random sampling.

REFERENCES

[1] Cochran W.G. (1977): *Sampling Techniques.* Third U.S. Edition. Wiley Eastern Limited. A. 325.

[2] Das A.K. (1988): Contribution to the Theory of Sampling Strategies Based on Auxiliary Information. Ph.D. Thesis, BCKV, West Bengal, India.

[3] Jeelani, M.I., Maqbool, S. and Mir, S.A. (2013): Modified Ratio Estimators of Population Mean Using Linear Combination of Co-efficient of Skewness and Quartile Deviation. International Journal of Modern Mathematical Sciences, 6, 3, 174-183.

[4] Kadilar, C. and Cingi, H. (2004): Ratio estimators in simple random sampling, Applied Mathematics and Computation 151, 893-902.

[5] Kadilar, C. and Cingi, H. (2006): An improvement in estimating the population mean by using the correlation co-efficient, Hacettepe Journal of Mathematics and Statistics Volume 35 (1), 103-109.

[6] Malik, K.A. and Tailor, R. (2013). Ratio Type Estimator of Population Mean in Double Sampling, International Journal of Advanced Mathematics and Statistics, 1, 1, 34-39.

[7] Mishra S S and Singh P K (2012): Computational approach to an inventory model with ramp- type demand and linear deterioration, International Journal of Operations Research, p.p. 337-357, Vol. 15, No. 3.

[8] Mishra S S and Singh P K(2012): Computing of total optimal cost of an EOQ model with quadratic deterioration and occurrence of shortages,International journal of Management Science and Engineering Management, World Academic Press, ISSN 1750-9653, England, UK 7(4): 243- 252 http://www.ijmsem.org,

[9] Mishra S S and Mishra P P(2012): Phase Wise Supply Chain Model of EOQ with Normal Life Time for Queued Customers: A Computational Approach, American Journal of Operations (http://www.SciRP.org/journal/ajor).

[10] Mishra S S and Singh P K(2013): Partial Backlogging EOQ Model for Queued Customers with Power demand and Quadratic Deterioration: Computational Approach, American Journal of Operations Research, Vol.3. Issue 2, pp. 13-27. Most read paper on Scientific and Academic Press(USA)

[11] Onyeka AC (2012) Estimation of population mean in post stratified sampling using known value of some population parameter(s). Statistics in Transition-New Series 13:65-78

[12] Pandey H, Yadav S.K., Shukla, A.K. (2011). An improved general class of estimators estimating population mean using auxiliary information. International Journal of Statistics and Systems 6:1-7

[13] Solanki R.S, Singh, H.P, Rathour, A. (2012) An alternative estimator for estimating the finite population mean using auxiliary information in sample surveys. International Scholarly Research Network: Probability and Statistics, 2012:1-14.

[14] Singh,G.N.(2003): On the improvement of product method of estimation in sample surveys. JISAS, 56, 267-275.

[15] Singh, H.P. and Tailor, R. (2003): Use of known correlation co-efficient in estimating the finite population means, Statistics in Transition 6 (4), 555-560.

[16] Singh, H.P., Tailor, R., Tailor, R. and Kakran, M.S. (2004): An improved estimator of population mean using power transformation, Journal of the Indian Society of Agricultural Statistics 58(2), 223-230.

[17] Subramani, J. and Kumarapandiyan, G. (2012): Estimation of Population Mean Using Co-Efficient of Variation and Median of an Auxiliary Variable. International Journal of Probability and Statistics 1(4): 111-118.

[18] Tailor, R., Chouhan, S. and Kim, J-M. (2014). Ratio and Product Type Exponential Estimators of Population Mean in Double Sampling for Stratification, Communications for Statistical Applications and Methods, 21, 1, 1-9.

[19] Upadhyaya, L.N. and Singh, H.P. (1999): Use of transformed auxiliary variable in estimating the finite population mean, Biometrical Journal 41 (5), 627-636.

[20] Yadav, S.K. (2011) Efficient estimators for population variance using auxiliary information. Global Journal of Mathematical Sciences: Theory and Practical 3:369-376.

[21] Yadav, S.K. and Kadilar, C. (2013): Improved class of ratio and product estimators, Applied Mathematics and Computation, 219, 10726–10731.

[22] Yan, Z. and Tian, B. (2010): Ratio method to the mean estimation using co-efficient of skewness of auxiliary variable, ICICA , Part II, CCIS 106, pp. 103–110.

[23] [23] Yadav S K, Mishra S S and Shukla A (2014): Improved ratio estimators for population mean based on median using linear combination of population mean and median of an auxiliary variable, American Journal of Operations Research, Scientific and Academic Publishing, Vol. 4, Issue 2, pp 21-27, 2014.