Vol.2.Issue.4.2014

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

http://www.bomsr.com

ISSN:2348-0580

RESEARCH ARTICLE

IMPROVED RATIO TYPE ESTIMATOR OF POPULATION MEAN UNDER TWO PHASE SAMPLING

SUBHASH KUMAR YADAV[1*], SHEELA MISRA[2], SURENDRA KUMAR[3], ALOK KUMAR SHUKLA[4], VISHWAS TIWARI[5]

[1*]Department of Mathematics and Statistics (A Centre of Excellence), Dr. RML Avadh University, Faizabad, U.P., INDIA

[2]Department of Statistics, University of Lucknow, Lucknow, U.P., INDIA

[3]Department of Mathematics, Govt. Degree College, Pihani, U.P., INDIA

[4]Department of Statistics, D.A-V. College, Kanpur, U.P., INDIA

[5]Department of Statistics, D.A-V. College, Kanpur, U.P., INDIA

SUBHASH KUMAR YADAV

Article Info:
Article received :15/10/2014
Revised on:29/10/2014
Accepted on:31/10/2014

ABSTRACT

This manuscript deals with the estimation of population mean in two phase sampling using auxiliary information. In the present paper an improved estimator of population mean has been proposed under two phase sampling scheme. 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 constant (kappa). A comparison has been made with the existing estimators in two phase sampling. Finally am empirical study is also carried out which shows improvement of proposed estimator over other estimators in two phase sampling in the sense of having lesser mean squared error.

KEY WORDS: Two phase sampling, Auxiliary variable, Bias, MSE, Efficiency

INTRODUCTION

The auxiliary information is being used in sampling theory since the development of the sampling theory and its application to the applied areas of the society. It is well established among the statisticians and the researchers that the suitable use of auxiliary information improves the efficiency of the estimates of the parameters under consideration by increasing the precision of the estimates. The auxiliary variable which provides the auxiliary information is highly correlated (positively or negatively) with the main variable under study. The auxiliary information is used for different purposes in sampling theory. It is used for the purposes of stratification in stratified sampling, measures of sizes in PPS (Probability Proportional to Size) sampling etc. It is used at both the stages of designing and the estimation stages of the sampling. In the present draft we have used it at estimation stage for estimating the population mean of the main variable under study in two phase or double sampling.

Let y and x be the study and the auxiliary variables respectively. When the variable y under study and the auxiliary variable x is highly positively correlated 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 the product type estimators are used to estimate the parameter under study when y and x are highly negatively correlated to each other. When the regression line does not passes through origin or its neighbourhood, regression estimator is appropriate estimator for the estimation of population parameter of the main variable under study. 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.

2. MATERIAL AND METHODS

Let $U=(U_1,U_2,..........,U_N)$ be the finite population consisting of N distinct and identifiable units out of which a sample of size n is drawn with 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 This sampling technique the following procedure is used for the sample selection,

(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.

It is well known that to estimate any of the parameters the appropriate estimators are the corresponding statistics, therefore the appropriate estimator for estimating population mean is the sample mean given by,

$$t_0=\bar{y} \tag{2.1}$$

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

$$V(t_0)=f_1\bar{Y}^2C_y^2 \tag{2.2}$$

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) used the auxiliary information and proposed the classical ratio type estimator in simple random sampling as,

$$t_R=\bar{y}\left(\frac{\bar{X}}{\bar{x}}\right) \tag{2.3}$$

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

$$t_R^d=\bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right) \tag{2.4}$$

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,

$$B(t_R^d) = \bar{Y} f_3 [C_x^2 - \rho_{xy} C_x C_y] \tag{2.5}$$

$$MSE(t_R^d) = \bar{Y}^2 [f_1 C_y^2 + f_3 (C_x^2 - 2\rho_{xy} C_x C_y)] \tag{2.6}$$

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) utilized the correlation coefficient between x and y and proposed the following estimator of population mean in simple random sampling as,

$$t_{ST} = \bar{y}\left(\frac{\bar{X} + \rho_{yx}}{\bar{x} + \rho_{yx}}\right) \tag{2.7}$$

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

$$t_{ST}^d = \bar{y}\left(\frac{\bar{x}' + \rho_{yx}}{\bar{x} + \rho_{yx}}\right) \tag{2.8}$$

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

$$B(t_{ST}^d) = \bar{Y} f_3 \theta [C_x^2 - \rho_{yx} C_x C_y] \tag{2.9}$$

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

(2.10)

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

3. PROPOSED ESTIMATOR

Motivated by Malik and Tailor (2013) and Prasad (1989), we propose the following estimator of population mean in two phase sampling as,

$$t = \kappa\, \bar{y}\left(\frac{\bar{x}' + \rho_{yx}}{\bar{x} + \rho_{yx}}\right) \tag{3.1}$$

Where κ is a constant known as kappa to be determined such that the 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

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

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

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

$$= \kappa\bar{Y}(1+e_0)[(1+\theta e_2)(1-\theta e_1+\theta^2 e_1^2-...)]$$
$$= \kappa\bar{Y}(1+e_0)[1-\theta e_1+\theta e_2+\theta^2 e_1^2-\theta^2 e_1 e_2+...]$$
$$= \kappa\bar{Y}[1+e_0-\theta e_1+\theta e_2+\theta^2 e_1^2-\theta^2 e_1 e_2-\theta e_0 e_1+\theta e_0 e_2+...] \tag{3.2}$$

Subtracting $\bar{Y}$ on both sides of (2.2) and simplifying, we get

$$t-\bar{Y} = (\kappa-1)\bar{Y}+\kappa\bar{Y}[e_0-\theta e_1+\theta e_2+\theta^2 e_1^2-\theta^2 e_1 e_2-\theta e_0 e_1+\theta e_0 e_2+...] \tag{3.3}$$

Taking expectations on both the sides of above equation, we get the bias of t, up to the first order of approximation after putting the values of different expectations as,

$$B(t) = (\kappa-1)\bar{Y}+\kappa\bar{Y}[\theta^2 f_1 C_x^2-\theta^2 f_2 C_x^2-\theta f_1\rho_{yx}C_y C_x+\theta f_2\rho_{yx}C_y C_x]$$
$$= (\kappa-1)\bar{Y}+\kappa\bar{Y}[\theta^2 f_3 C_x^2-\theta f_3\rho_{yx}C_y C_x] \tag{3.4}$$

Squaring equation (2.3) on both sides and taking expectations, we have MSE of t up to the first order of approximation as,

$$MSE(t) = \kappa^2\bar{Y}^2 E[1+e_0-\theta e_1+\theta e_2+\theta^2 e_1^2-\theta^2 e_1 e_2-\theta e_0 e_1+\theta e_0 e_2+...]^2+\bar{Y}^2$$
$$-2\kappa\bar{Y}^2 E[1+e_0-\theta e_1+\theta e_2+\theta^2 e_1^2-\theta^2 e_1 e_2-\theta e_0 e_1+\theta e_0 e_2+...]$$
$$= \kappa^2\bar{Y}^2 E[1+e_0^2-\theta^2 e_1^2+\theta^2 e_2^2+\theta^2 e_1^2-2e_0-2\theta e_1+2\theta e_2+2\theta^2 e_1^2-2\theta^2 e_1 e_2-2\theta e_0 e_1+2\theta e_0 e_2$$
$$-2\theta e_0 e_1+2\theta e_0 e_2-2\theta^2 e_1 e_2]+\bar{Y}^2$$
$$-2\kappa\bar{Y}^2 E[1+e_0-\theta e_1+\theta e_2+\theta^2 e_1^2-\theta^2 e_1 e_2-\theta e_0 e_1+\theta e_0 e_2]$$
$$= \bar{Y}^2[(\kappa-1)^2+\kappa^2 E\{2e_0-2\theta e_1+2\theta e_2+e_0^2+3\theta^2 e_1^2+\theta^2 e_2^2-4\theta^2 e_1 e_2-4\theta e_0 e_1+4\theta e_0 e_2\}$$
$$-2\kappa E[e_0-\theta e_1+\theta e_2+\theta^2 e_1^2-\theta^2 e_1 e_2-\theta e_0 e_1+\theta e_0 e_2]$$
$$= \bar{Y}^2[(\kappa-1)^2+\kappa^2\{2E(e_0)-2\theta E(e_1)+2\theta E(e_2)+E(e_0^2)+3\theta^2 E(e_1^2)+\theta^2 E(e_2^2)-4\theta^2 E(e_1 e_2)$$
$$-4\theta E(e_0 e_1)+4\theta E(e_0 e_2)\}$$
$$-2\kappa[E(e_0)-\theta E(e_1)+\theta E(e_2)+\theta^2 E(e_1^2)-\theta^2 E(e_1 e_2)-\theta E(e_0 e_1)+\theta E(e_0 e_2)]$$

Putting the values of different expectations, we have MSE of t up to the first order of approximation as,

$$MSE(t) = \bar{Y}^2[(\kappa-1)^2+\kappa^2\{f_1 C_y^2+3f_3\theta^2 C_x^2-4f_3\theta\rho_{yx}C_y C_x\}-2\kappa\{f_3\theta^2 C_x^2-f_3\theta\rho_{yx}C_y C_x\}]$$

(3.5)

Which is minimum for,

$$\kappa = \frac{[1+f_3\theta^2 C_x^2-f_3\theta\rho_{yx}C_y C_x]}{[1+f_1 C_y^2+3f_3\theta^2 C_x^2-4f_3\theta\rho_{yx}C_y C_x]} = \frac{A}{B}$$

Where,

$$A = [1+f_3\theta^2 C_x^2-f_3\theta\rho_{yx}C_y C_x]$$
$$B = [1+f_1 C_y^2+3f_3\theta^2 C_x^2-4f_3\theta\rho_{yx}C_y C_x]$$

The minimum mean square error of t up to the first order of approximation is,

$$MSE_{\min}(t) = \bar{Y}^2\left[1-\frac{A^2}{B}\right] \tag{3.6}$$

4. EFFICIENCY COMPARISON

From (2.2) and (3.6), we have

$V(t_0)-MSE_{\min}(t)>0$ if,

$$A^2 > B(1-f_1 C_y^2) \tag{4.1}$$

From (2.6) and (3.6), we have

$MSE(t_R^d) - MSE_{\min}(t) > 0$ if,

$$A^2 > B[(1 - f_1 C_y^2) - f_3(C_x^2 - 2\rho_{yx} C_x C_y)] \qquad (4.2)$$

From (2.10) and (3.6), we have

$MSE(t_{ST}^d) - MSE_{\min}(t) > 0$ if,

$$A^2 > B[(1 - f_1 C_y^2) - f_3\theta(\theta C_x^2 - 2\rho_{yx} C_x C_y)] \qquad (4.3)$$

5. EMPIRICAL EXAMPLE

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

Population-I [Source: Das, 1988]

Y: the number of agricultural labours for 1971,

X: the number of agricultural labours for 1961,

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

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

Population-II [Source: Cochran, 1977]

Y: the number of persons per block,

X: the number of rooms per block,

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

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

6. RESULTS AND CONCLUSION

From the theoretical discussions in section-3 and the results in table-1, we see that the proposed estimator has lesser mean squared error and thus highest percentage relative efficiency. Therefore 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 under two phase random sampling technique. Therefore the proposed estimator should be preferred for the estimation of population mean in two phase random sampling.

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	**166.48**	**132.86**

REFERENCES

[1]. Cochran W.G. Sampling Techniques. Third U.S. Edition. Wiley Eastern Limited, 1977.

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

[3]. Jeelani, M.I., Maqbool, S. and Mir, S.A. 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, 2013, 174-183.

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

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

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

[7]. Onyeka A.C. Estimation of population mean in poststratified sampling using known value of some population parameter(s). Statistics in Transition-New Series 13, 2012, 65-78.

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

[9]. Solanki R.S, Singh, H.P, Rathour, A. 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.

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

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

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