Vol.10.Issue.4.2022 (Oct-Dec)
©KY PUBLICATIONS

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

http://www.bomsr.com

Email:editorbomsr@gmail.com

INTERNATIONAL STANDARD SERIAL NUMBER 2348-0580

RESEARCH ARTICLE

NEW EFFICIENT ESTIMATORS FOR POPULATION MEAN USING DOUBLING SAMPLING SCHEME IN SAMPLE SURVEYS

Kamlesh Kumar[1], Sunit Kumar[2], Sanjay Kumar[3]
[1, 2]Department of Statistics, Central University of South Bihar, Gaya, Bihar, India.
[1]Email: kamalbhu03@gmail.com , [2]Email: sunit@cusb.ac.in
[3]Department of Statistics, Central University of Rajasthan, Ajmer, Rajasthan, India.
[3]Email: sanjay.kumar@curaj.ac.in
DOI:10.33329/bomsr.10.4.75

ABSTRACT

Many estimators for estimating the population mean using supporting (auxiliary) variable(s), have been constructed by several researchers. Further, some researchers have also constructed the estimators using information on supporting (auxiliary) attribute(s). In this research paper, we have built new efficient estimators by using the data from the supporting variable and the supporting attribute in double sampling. To compare the percent relative efficiency (PRE) of the built estimators, we have derived the expressions for bias and mean square error of the estimators in case of large sample approximation. Mathematical conditions under which the built estimators have more percent relative efficiency (PRE), are acquired. Further, comparision of PRE of the recommended estimators, has been done by using a numerical analysis with different sample sizes.

Keywords: Bias, double sampling, mean square error, supporting attribute, supporting variable.

1. INTRODUCTION

Estimation of population mean, ratio and product of two population means are generally required in many areas of sciences such as biological sciences, social sciences, environmental sciences, etc. These parameters can be estimated by using either population mean of supporting (auxiliary) variable or population proportion of supporting (auxiliary) attribute. By using population mean of supporting variable(s), Neyman (1938), Cochran (1940, 77), Murthy (1964), Kiregyera (1980, 84), Srivastava et al. (1990), Singh and Vishawakarma (2007) and Khare et al. (2013) built dissimilar types of estimators. Further, using proportion of supporting attribute(s), numerous research works have been completed by Naik and Gupta (1996), Jhajj et al. (2006), Shabbir and Gupta (2010), and Solanki and Singh (2013).

In the recent paper, we have suggested new efficient estimators in double sampling with the help of information on the supporting variable and the supporting attribute. This paper main goals to develop new efficient estimators for the population mean of the study variable that achieves more percent relative efficiency in comparison to the ratio, regression and exponential types estimators, suggested by Cochran (1977) and Singh and Vishawakarma (2007).

2. THE EARLIER EXITING ESTIMATORS IN DOUBLE SAMPLING

Let (y , x, ϕ) stand for the study variable, supporting variable, and supporting attribute and ($\bar{Y}$, $\bar{X}$, P) stand for the population means and population proportion of (y, x, ϕ) having the values (Y_l, X_l, ϕ_l), where l varies from 1 to M (population size). Let ϕ_l indicates 1 if the lth unit of population has the attribute and 0 if not.

Let $\bar{x}'$ denotes the mean of the first phase sample that can be acquired after drawing a large sample of size m' from the population of size M using the simple random sampling without replacement (SRSWOR) method. Again, let ($\bar{y}$, $\bar{x}$) denote the means of the second phase sample that can be acquired after drawing a second phase sample from first phase sample of size m' . If $\bar{X}$ is not available then $\bar{x}'$ can be used in place of $\bar{X}$. In this case, Cochran (1977) and Singh and Vishawakarma (2007) suggested the ratio, regression and exponential ratio type estimators which are given as follows:

$$\hat{\bar{Y}}_{C1} = \frac{\bar{y}}{\bar{x}}\bar{x}', \tag{2.1}$$

$$\hat{\bar{Y}}_{C2} = \bar{y} + \lambda_{yx}\left(\bar{x}' - \bar{x}\right) \tag{2.2}$$

and

$$\hat{\bar{Y}}_{SV} = \bar{y}\exp\left(\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right), \tag{2.3}$$

where $\lambda_{yx} = \hat{S}_{yx} / \hat{S}_x^2$, ($\hat{S}_{yx}, \hat{S}_x^2$) are defined on m units, and they are used to estimate (S_{yx}, S_x^2) which are defined on M units.

Further, Singh and Choudhury (2012) proposed exponential chain ratio and product type estimators for population mean in case of unknown population mean $\bar{X}$ of supporting variable x and known population mean $\bar{Z}$ of an additional supporting variable z, where z is probably cheaper and may be less correlated with the study variable y compared to the supporting variable x (i.e. $\rho_{yx} > \rho_{yz}$), which are given as:

$$\hat{\bar{Y}}_{SC1} = \bar{y}\exp\left(\frac{\bar{x}'\frac{\bar{Z}}{\bar{z}'} - \bar{x}}{\bar{x}'\frac{\bar{Z}}{\bar{z}'} + \bar{x}}\right), \tag{2.4}$$

and

$$\hat{\bar{Y}}_{SC2} = \bar{y}\exp\left(\frac{\bar{x} - \bar{x}'\frac{\bar{Z}}{\bar{z}'}}{\bar{x} + \bar{x}'\frac{\bar{Z}}{\bar{z}'}}\right), \tag{2.5}$$

where $\bar{z}' = \frac{1}{m'}\sum_{l=1}^{m'} Z_l$.

3. THE SUGGESTED ESTIMATORS

Motivated by Singh and Chaudhury (2012), we suggest the following new efficient estimators for the population mean in case of unknown population mean $\bar{X}$ of supporting variable x and known population proportion P of a supporting attribute ϕ, where ϕ is probably cheaper and may be less correlated with the study variable y compared to the supporting variable x (i.e. $\rho_{yx} > \rho_{y\phi}$).

$$\hat{\bar{Y}}_{K1} = \bar{y}\exp\left(\frac{\bar{x}'\frac{P}{p'} - \bar{x}}{\bar{x}'\frac{P}{p'} + \bar{x}}\right) \tag{3.1}$$

and

$$\hat{\bar{Y}}_{K2} = \bar{y}\exp\left(\frac{\bar{x} - \bar{x}'\frac{P}{p'}}{\bar{x} + \bar{x}'\frac{P}{p'}}\right), \tag{3.2}$$

where $(p',\ p)$ are the first and second phase sample proportions of supporting attribute ϕ based on $(m',\ m)$ units.

4. THE BIAS AND MEAN SQUARE ERROR (MSE)

We have obtained the following mathematical expressions for the estimators' biases and MSEs up to the $O(n^{-1})$.

$$B(\hat{\bar{Y}}_{K1}) = \bar{Y}\left[\frac{3}{8}\left(f_3C_x^2 + f_2C_\phi^2\right) + \frac{1}{2}\left(f_3C_{yx}C_x^2 - f_2C_{y\phi}C_\phi^2\right)\right], \tag{4.1}$$

$$B(\hat{\bar{Y}}_{K2}) = \bar{Y}\left[-\frac{1}{8}\left(f_3C_x^2 + f_2C_\phi^2\right) + \frac{1}{2}\left(f_3C_{yx}C_x^2 + f_2C_{y\phi}C_\phi^2\right)\right] \tag{4.2}$$

and

$$MSE\,(\hat{\bar{Y}}_{K1}) = \bar{Y}^2\left[f_1C_y^2 + f_3C_x^2\left(\frac{1}{4} - \rho_{yx}C_yC_x\right) + f_2C_\phi^2\left(\frac{1}{4} - \rho_{y\phi}C_yC_\phi\right)\right], \tag{4.3}$$

$$MSE\,(\hat{\hat{Y}}_{K2}) = \bar{Y}^2\left[\,f_1 C_y^2 + f_3 C_x^2\left(\frac{1}{4} + \rho_{yx} C_y C_x\right) + f_2 C_\phi^2\left(\frac{1}{4} + \rho_{y\phi} C_y C_\phi\right)\right], \tag{4.4}$$

Mean square errors of the other exiting estimators up to $O(n^{-1})$ are derived as:

$$MSE\,(\bar{y}) = \bar{Y}^2 f_1 C_y^2\,, \tag{4.5}$$

$$MSE\,(\hat{\hat{Y}}_{C1}) = \bar{Y}^2\left[f_1 C_y^2 + f_3\left(C_x^2 - 2\rho_{yx} C_y C_x\right)\right], \tag{4.6}$$

$$MSE\,(\hat{\hat{Y}}_{C2}) = \bar{Y}^2 C_y^2\left[f_1 - f_3 \rho_{yx}^2\right], \tag{4.7}$$

$$MSE\,(\hat{\hat{Y}}_{SV}) = \bar{Y}^2\left[\,f_1 C_y^2 + f_3\left(\frac{C_x^2}{4} - \rho_{yx} C_y C_x\right)\right], \tag{4.8}$$

where $f_1 = \left(\frac{1}{m} - \frac{1}{M}\right)$, $f_2 = \left(\frac{1}{m'} - \frac{1}{M}\right)$, $f_3 = \left(\frac{1}{m} - \frac{1}{m'}\right)$, $C_y = \frac{S_y}{\bar{Y}}$, $C_x = \frac{S_x}{\bar{X}}$, $C_\phi = \frac{S_\phi}{P}$, $S_y^2 = \frac{1}{M-1}\sum_{l=1}^{M}(Y_l - \bar{Y})^2$, $S_x^2 = \frac{1}{M-1}\sum_{l=1}^{M}(X_l - \bar{X})^2$, $S_\phi^2 = \frac{1}{M-1}\sum_{l=1}^{M}(\phi_l - P)^2$ and ρ_{yx} denotes the correlation coefficient between the study variable y and supporting variable x, $\rho_{y\phi}$ denotes point bi-serial correlation coefficient between y and ϕ, $\rho_{x\phi}$ denotes the point bi-serial correlation coefficient between x and ϕ respectively.

5. COMPARISONS

In this section, we have obtained mathematical conditions under which the suggested estimators have lower MSEs than the existing estimators.

Comparing the estimator $\hat{\hat{Y}}_{K1}$ with $(\bar{y},\ \hat{\hat{Y}}_{C1},\ \hat{\hat{Y}}_{C2},\ \hat{\hat{Y}}_{SV})$, we get

$MSE(\hat{\hat{Y}}_{K1}) < MSE(\bar{y})$

$$\text{If } \left[\left(\frac{1}{4} - \rho_{yx} C_y C_x\right)\right] < 0 \text{ and } \left[\left(\frac{1}{4} - \rho_{y\phi} C_y C_\phi\right)\right] < 0 \tag{5.1}$$

$MSE(\hat{\hat{Y}}_{K1}) < MSE(\hat{\hat{Y}}_{C1})$

$$\text{If } \left[\left(-\frac{3}{4}C_x^2 + \rho_{yx} C_y C_x (2 - C_x^2)\right)\right] < 0 \text{ and } \left[\left(\frac{1}{4} - \rho_{y\phi} C_y C_\phi\right)\right] < 0 \tag{5.2}$$

$MSE(\hat{\hat{Y}}_{K1}) < MSE(\hat{\hat{Y}}_{C2})$

$$\text{If } \left[\left(\frac{1}{4} - \rho_{yx} C_y C_x + \rho_{yx}^2 \frac{C_y^2}{C_x^2}\right)\right] < 0 \text{ and } \left[\left(\frac{1}{4} - \rho_{y\phi} C_y C_\phi\right)\right] < 0 \tag{5.3}$$

$MSE(\hat{\hat{Y}}_{K1}) < MSE(\hat{\hat{Y}}_{SV})$

If $\left[\left\{\rho_{yx}C_yC_x(1-C_x^2)\right\}\right]<0$ and $\left[\left(\frac{1}{4}-\rho_{y\phi}C_yC_\phi\right)\right]<0$ (5.4)

Comparing the estimator $\hat{\hat{\bar{Y}}}_{K2}$ with $(\bar{y},\ \hat{\hat{\bar{Y}}}_{C1},\ \hat{\hat{\bar{Y}}}_{C2},\ \hat{\hat{\bar{Y}}}_{SV})$, we get

$MSE(\hat{\hat{\bar{Y}}}_{K2})<MSE(\bar{y})$

If $\left[\left(\frac{1}{4}+\rho_{yx}C_yC_x\right)\right]<0$ and $\left[\left(\frac{1}{4}+\rho_{y\phi}C_yC_\phi\right)\right]<0$ (5.5)

$MSE(\hat{\hat{\bar{Y}}}_{K2})<MSE(\hat{\hat{\bar{Y}}}_{C1})$

If $\left[\left(-\frac{3}{4}C_x^2+\rho_{yx}C_yC_x(2+C_x^2)\right)\right]<0$ and $\left[\left(\frac{1}{4}+\rho_{y\phi}C_yC_\phi\right)\right]<0$ (5.6)

$MSE(\hat{\hat{\bar{Y}}}_{K2})<MSE(\hat{\hat{\bar{Y}}}_{C2})$

If $\left[\left(\frac{1}{4}+\rho_{yx}C_yC_x+\rho_{yx}^2\frac{C_y^2}{C_x^2}\right)\right]<0$ and $\left[\left(\frac{1}{4}+\rho_{y\phi}C_yC_\phi\right)\right]<0$ (5.7)

$MSE(\hat{\hat{\bar{Y}}}_{K2})<MSE(\hat{\hat{\bar{Y}}}_{SV})$

If $\left[\left\{\rho_{yx}C_yC_x(1+C_x^2)\right\}\right]<0$ and $\left[\left(\frac{1}{4}+\rho_{y\phi}C_yC_\phi\right)\right]<0$ (5.8)

6. A NUMERICAL STUDY

To validate our theoretical outcomes, we have used a dataset of 96 villages from the rural region, Singur, Hooghly, West Bengal, which was taken from the District Census Handbook 1981 released by the Government of India. Here, y, x, ϕ denote the number of labourers working in agriculture, the village's total area, and houses that are occupied but fewer than or equal to 250, respectively.

Below are the parameter values for the dataset mentioned above:

$\bar{Y}$ =162.67, $\bar{X}$ =164.01, $P=0.6146,\ C_y=1.431$, $C_x=0.800,\ C_\phi=0.796,\ \rho_{yx}=0.779,\ \rho_{y\phi}=0.502$.

The percent relative efficiencies (PRE) of the estimators $\hat{\hat{\bar{Y}}}_i : i=1$, 2, 3, 4 with respect to $\bar{y}$, are obtained as:

PRE = $\frac{MSE(\bar{y})}{MSE(\hat{\hat{\bar{Y}}}_i)}$ X 100, where $\hat{\hat{\bar{Y}}}_1=\hat{\hat{\bar{Y}}}_{C1}$, $\hat{\hat{\bar{Y}}}_2=\hat{\hat{\bar{Y}}}_{C2}$, $\hat{\hat{\bar{Y}}}_3=\hat{\hat{\bar{Y}}}_{SV}$ and $\hat{\hat{\bar{Y}}}_4=\hat{\hat{\bar{Y}}}_{K1}$.

Table 1: Percent Relative efficiency (PRE) of the suggested estimators with respect to the unbiased estimator $\bar{y}$

Estimators	($M = 96,\ m' = 70,\ m = 25$)	($M = 96,\ m' = 65,\ m = 20$)
	PRE	PRE
$\bar{y}$	100.00	100.00
$\hat{\hat{Y}}_{C1}$	355.25	360.85
$\hat{\hat{Y}}_{C2}$	367.28	373.34
$\hat{\hat{Y}}_{SV}$	201.02	202.26
$\hat{\hat{Y}}_{K1}$	387.24	393.17

7. DISCUSSION AND CONCLUSION

In tables 1, different first phase and second phase samples has been considered to check the percent relative efficiency (PRE) of the proposed estimator $\hat{\hat{Y}}_{K1}$ with respect to the unbiased estimators estimator ($\bar{y}$). For $m' = 70,\ m = 25,$ the suggested estimator $\hat{\hat{Y}}_{K1}$ have more percent relative efficiency in comparision to the estimators $\bar{y},\ \hat{\hat{Y}}_{C1},\ \hat{\hat{Y}}_{C2},\ \hat{\hat{Y}}_{SV}$. Further, after changing the sizes of the first phase and second phase samples, the suggested estimator again have more percent relative efficiency in comparision to the earliar existing estimators. Hence, we can conclude that the suggested estimator work better than the other exiting estimators, in situation when population mean of additional supporting variable is not known but population proportion is known. So, we recommend the proposed estimors for estimating the population mean when population mean of the additional supporting variable is not known.

8. REFERENCES

[1]. Cochran, W.G., 1940, The estimation of the yields of the cereal experiments by sampling for the ratio of grain to total produce. *J. Agri. Sci*. **30**, 262-275.

[2]. Cochran, W.G., 1977, Sampling Techniques, 3rd edition, John Wiley, New York.

[3]. Jhajj, H.S., Sharma, M.K., Grover, L.K., 2006, A family of estimators of population mean using information on auxiliary attribute. *Pak. J. Stat*. **22**, 43–50.

[4]. Khare, B.B., Srivastava, U., Kumar, K., 2013, A generalized chain ratio in regression estimator for population mean using two auxiliary characters in sample survey. *J. Sci. Res. BHU, Varanasi* **57,** 147-153.

[5]. Kiregyera, B., 1980, A chain ratio-type estimator in finite population in double sampling using two auxiliary variables. *Metrika* **27**, 217-223.

[6]. Kiregyera, B., 1984, Regression type estimators using two auxiliary variables and the model of double sampling from a finite population. *Metrika* 31, 215-226.

[7]. Murthy, M.N., 1964, Product method of estimation. *Ind. J. Stat. A* **26**, 69-74.

[8]. Naik, V.D., Gupta, P.C., 1996, A note on estimation of mean with known population proportion of an auxiliary character. *J. Ind. Soc. Agri. Stat*. **48(2**), 151-158.

[9]. Neyman J., 1938, Contributions to the theory of sampling human population. *J. Amer. Stat. Assoc*. **33**, 101-116.

[10]. Shabbir, J., Gupta, S., 2010, Estimation of the finite population mean in two phase sampling when auxiliary variables are attributes. *Hacett. J. Math. Stat*. **39**, 121–129.

[11]. Singh, B.K., Chaudhary, S., 2012, Exponential chain ratio and product type estimators for finite population mean under double sampling scheme. *Glob. J. Sci. and Front. Res. Maths and Dec. Sci*., **12(6),**13-24.

[12]. Singh, H.P., Vishwakarma, G.K., 2007, Modified exponential ratio and product estimators for finite population mean in double sampling. *Austrian Journal of Statistics*, **36(3)**, 217–225.

[13]. Solanki, R.S., Singh, H.P., 2013, Improved estimation of population mean using population proportion of an auxiliary character. *Chil. J. Stat*. **4(1),** 3-17.

[14]. Srivastava S. Rani, Khare, B.B., Srivastava, S.R., 1990, A generalised chain ratio estimator for mean of finite population. *J. Ind. Soc. Agri. Stat.* **1,** 108-117.