



IMPROVED SAMPLING STRATEGIES FOR FINITE POPULATION USING AUXILIARY INFORMATION

Chandni Kumari and Ratan Kumar Thakur

Department of Applied Statistics, Bahasaheb Bhimrao Ambedkar University, (A Central University), Lucknow, 226025, India

ARTICLE INFO

Article History:

Received 13th January, 2020
 Received in revised form 11th February, 2020
 Accepted 8th March, 2020
 Published online 28th April, 2020

Key words:

Simple Random Sampling, Ratio type Estimator, Midzuno-Lahiri-Sen type Sampling Scheme.

ABSTRACT

This paper advocates the problem of estimating the finite population mean in sample surveys. We have suggested the generalized class of estimators under Midzuno (1950) - Lahiri - Sen (1952) type sampling scheme and its properties are studied up to the first degree of approximation. Further, we compare the proposed sampling strategy with the mean per unit estimator, classical ratio estimator, product estimator, linear regression estimator and the generalized ratio estimator (Walsh) under the simple random sampling without replacement. Also, we find out the unbiased estimate of the variance. On the basis of suitable range information, we give some concluding remarks related to proposed sampling strategy. An empirical study is given in support of the present study.

Copyright©2020 **Chandni Kumari and Ratan Kumar Thakur**. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Let the population consists of N units. Y_i and X_i denote the i^{th} characteristics of the population. The population mean of the study variable is denoted by \bar{Y} and population mean of the auxiliary variable which

is known is denoted by \bar{X} and it is given as

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i \quad \text{and} \quad \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \quad \text{respectively.}$$

The population variance of the study variable and the auxiliary

variable is given as $\sigma_y^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2$ and

$$\sigma_x^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2 \quad \text{respectively.}$$

Since the study variable and the auxiliary variable are correlated with each other and there exists a linear relationship between two variables. Thus ρ be the correlation coefficient between the variable under study and auxiliary variable which measures the degree of linear relationship between two variables, it is given as

$$\rho = \frac{Cov(Y, X)}{\sigma_y \sigma_x}$$

$$\text{where } Cov(Y, X) = E(Y - \bar{Y})(X - \bar{X})$$

$$V(Y) = E(Y - \bar{Y})^2$$

$$V(X) = E(X - \bar{X})^2$$

Also C_y and C_x be the coefficient of variation of 'Y' and 'X' respectively and it is given as

$$C_y = \frac{\sigma_y}{\bar{Y}} \quad \text{and} \quad C_x = \frac{\sigma_x}{\bar{X}}$$

where σ_y and σ_x are the standard deviation of the Y and X respectively.

Now let us take a random sample of size 'n'. y_i be the i^{th} characteristics of the study variable of the sample and x_i be the i^{th} characteristics of the auxiliary variable of the sample. Sample mean of study variable for estimating the population mean is denoted by \bar{y}_s and it is given as

$$\bar{y}_s = \frac{1}{n} \sum_{i=1}^n y_i$$

*Corresponding author: **Chandni Kumari**

Department of Applied Statistics, Bahasaheb Bhimrao Ambedkar University, (A Central University), Lucknow, 226025, India

The sample mean of auxiliary variable is denoted by \bar{x}_s and it

$$\text{is given as } \bar{x}_s = \frac{1}{n} \sum_{i=1}^n x_i$$

When the random sample s is selected by simple random sampling without replacement, the generalized ratio for estimating population mean \bar{Y} is

$$\bar{y}_{A\gamma} = \frac{\gamma \bar{y}(\alpha \bar{X} + \beta)}{(\alpha \bar{X} + \beta) + A\alpha(\bar{x} - \bar{X})}$$

where A is the characterizing scalar to be chosen(1952) suitably.

We now consider the generalized ratio estimator \bar{y}_A under Midzuno (1950) (1952)- Lahiri (1951)- Sen (1952) (1952) type sampling scheme and denote it by $\bar{y}_{A\gamma}$.

The proposed Midzuno (1950)-Lahiri-Sen(1952) type sampling scheme for selecting a sample 's' of size n is given as selecting the first unit of the sample by probability proportional to $(\alpha \bar{X} + \beta) + A\alpha(x_i - \bar{X})$, where x_i is the size of first selected unit.

$$P(i) = \frac{(\alpha \bar{X} + \beta) + A\alpha(x_i - \bar{X})}{{}^N C_n(\alpha \bar{X} + \beta)} \quad \text{and selecting the}$$

remaining $(n-1)$ units in the sample from $(N-1)$ units in the population by simple random sampling without replacement. Thus,

$$P(s) = \sum_{i=1}^n \{ \text{probability of selecting } i \text{ th sample unit at first draw} \} \{ \text{probability of selecting } (n-1) \text{ units out of } (N-1) \text{ units by SRSWOR} \}$$

$$P(s) = \sum_{i=1}^n \frac{(\alpha \bar{X} + \beta) + A\alpha(x_i - \bar{X})}{{}^{N-1} C_{n-1}(\alpha \bar{X} + \beta)}$$

$$P(s) = \frac{(\alpha \bar{X} + \beta) + A\alpha(\bar{x} - \bar{X})}{{}^N C_n(\alpha \bar{X} + \beta)}$$

2 BIASEDNESS OF $(\bar{y}_{A\gamma})$

The formula for calculating biasness of $(\bar{y}_{A\gamma})$ is given as

$$\bar{y}_{A\gamma} = \frac{\gamma \bar{y}(\alpha \bar{X} + \beta)}{(\alpha \bar{X} + \beta) + A\alpha(\bar{x} - \bar{X})}$$

Taking expectation on both the sides,

$$E(\bar{y}_{A\gamma}) = E \left[\frac{\gamma \bar{y}(\alpha \bar{X} + \beta)}{(\alpha \bar{X} + \beta) + A\alpha(\bar{x} - \bar{X})} \right]$$

$$= \sum_{s=1}^{N C_n} \left[\frac{\gamma \bar{y}(\alpha \bar{X} + \beta)}{(\alpha \bar{X} + \beta) + A\alpha(\bar{x} - \bar{X})} \right] P(s)$$

$$= \sum_{s=1}^{N C_n} \left[\frac{\gamma \bar{y}(\alpha \bar{X} + \beta)}{(\alpha \bar{X} + \beta) + A\alpha(\bar{x} - \bar{X})} \right] \frac{\{(\alpha \bar{X} + \beta) + A\alpha(\bar{x} - \bar{X})\}}{{}^N C_n(\alpha \bar{X} + \beta)}$$

$$= \gamma \sum_{s=1}^{N C_n} \left[\frac{\bar{y}}{{}^N C_n} \right]$$

$$= \gamma E(\bar{y})$$

$$= \gamma \bar{Y}$$

$$\text{Bias}(\bar{y}_{A\gamma}) = E(\bar{y}_{A\gamma}) - \gamma \bar{Y}$$

Therefore, $(\bar{y}_{A\gamma})$ is not an unbiased estimator of population mean.

Mean Square Error Of $(\bar{y}_{A\gamma})$

$$MSE(\bar{y}_{A\gamma}) = E[\bar{y} - \bar{Y}]^2$$

$$= E \left[\frac{\gamma \bar{y}(\alpha \bar{X} + \beta)}{(\alpha \bar{X} + \beta) + A\alpha(\bar{x} - \bar{X})} - \bar{Y} \right]^2$$

$$= E \left[\frac{\gamma \bar{y}(\alpha \bar{X} + \beta) - \bar{Y}(\alpha \bar{X} + \beta) - \bar{Y}A\alpha(\bar{x} - \bar{X})}{(\alpha \bar{X} + \beta) + A\alpha(\bar{x} - \bar{X})} \right]^2$$

$$= \sum_{s=1}^{N C_n} \left[\frac{\gamma \bar{y}(\alpha \bar{X} + \beta) - \bar{Y}(\alpha \bar{X} + \beta) - \bar{Y}A\alpha(\bar{x} - \bar{X})}{(\alpha \bar{X} + \beta) + A\alpha(\bar{x} - \bar{X})} \right]^2 P(s)$$

$$= \sum_{s=1}^{N C_n} \left[\frac{\gamma \bar{y}(\alpha \bar{X} + \beta) - \bar{Y}(\alpha \bar{X} + \beta) - \bar{Y}A\alpha(\bar{x} - \bar{X})}{(\alpha \bar{X} + \beta) + A\alpha(\bar{x} - \bar{X})} \right]^2 \frac{\{(\alpha \bar{X} + \beta) + A\alpha(\bar{x} - \bar{X})\}}{{}^N C_n(\alpha \bar{X} + \beta)}$$

$$= \sum_{s=1}^{N C_n} \left[\frac{\{ \gamma \bar{y}(\alpha \bar{X} + \beta) - \bar{Y}(\alpha \bar{X} + \beta) - \bar{Y}A\alpha(\bar{x} - \bar{X}) \}^2}{(\alpha \bar{X} + \beta) + A\alpha(\bar{x} - \bar{X})} \right] \frac{1}{{}^N C_n(\alpha \bar{X} + \beta)}$$

$$= \sum_{s=1}^{N C_n} \left[\frac{\{ \gamma \bar{y}(\alpha \bar{X} + \beta) - \bar{Y}(\alpha \bar{X} + \beta) - \bar{Y}A\alpha(\bar{x} - \bar{X}) \}^2}{\left\{ 1 + \frac{A\alpha(\bar{x} - \bar{X})}{(\alpha \bar{X} + \beta)} \right\}} \right] \frac{1}{{}^N C_n(\alpha \bar{X} + \beta)^2}$$

$$= \sum_{s=1}^{N C_n} \left[\{ \gamma \bar{y}(\alpha \bar{X} + \beta) - \bar{Y}(\alpha \bar{X} + \beta) - \bar{Y}A\alpha(\bar{x} - \bar{X}) \}^2 \left\{ 1 + \frac{A\alpha(\bar{x} - \bar{X})}{(\alpha \bar{X} + \beta)} \right\}^{-1} \right] \frac{1}{{}^N C_n(\alpha \bar{X} + \beta)^2}$$

$$= E \left[\{ \gamma \bar{y}(\alpha \bar{X} + \beta) - \bar{Y}(\alpha \bar{X} + \beta) - \bar{Y}A\alpha(\bar{x} - \bar{X}) \}^2 \left\{ 1 + \frac{A\alpha(\bar{x} - \bar{X})}{(\alpha \bar{X} + \beta)} \right\}^{-1} \right] \frac{1}{(\alpha \bar{X} + \beta)^2}$$

(1)

$$\text{Let } \varepsilon_0 = \frac{(\bar{y} - \bar{Y})}{\bar{Y}} \quad \text{and} \quad \varepsilon_1 = \frac{(\bar{x} - \bar{X})}{\bar{X}}$$

Such that $E(\varepsilon_0) = E(\varepsilon_1) = 0$

$$\text{and } V(\varepsilon_0) = E(\varepsilon_0^2) = \frac{(N-n)}{Nn} \frac{S_y^2}{\bar{Y}^2}$$

$$V(\varepsilon_1) = E(\varepsilon_1^2) = \frac{(N-n)}{Nn} \frac{S_x^2}{\bar{X}^2}$$

Thus (1) becomes,

$$= E \left[\{ \gamma \bar{Y}(1 + \varepsilon_0)(\alpha \bar{X} + \beta) - \bar{Y}(\alpha \bar{X} + \beta) - \bar{Y}A\alpha\varepsilon_1 \}^2 \left\{ 1 + \frac{A\alpha\bar{X}\varepsilon_1}{(\alpha \bar{X} + \beta)} \right\}^{-1} \right] \frac{1}{(\alpha \bar{X} + \beta)^2}$$

$$\begin{aligned}
 &= \frac{1}{(\alpha\bar{X} + \beta)^2} E \left[\left\{ (\gamma-1)\bar{Y}(\alpha\bar{X} + \beta) + \gamma\epsilon_0\bar{Y}(\alpha\bar{X} + \beta) - \bar{Y}A\alpha\epsilon_1\bar{X} \right\}^2 \left\{ 1 - \frac{A\alpha\epsilon_1\bar{X}}{(\alpha\bar{X} + \beta)} + \frac{A^2\alpha^2\epsilon_1^2\bar{X}^2}{(\alpha\bar{X} + \beta)^2} - \dots \right\} \right] \\
 &= \frac{1}{(\alpha\bar{X} + \beta)^2} E \left[\left\{ (\gamma-1)^2\bar{Y}^2(\alpha\bar{X} + \beta)^2 + \gamma^2\epsilon_0^2\bar{Y}^2(\alpha\bar{X} + \beta)^2 + \bar{Y}^2A^2\alpha^2\epsilon_1^2\bar{X}^2 \right. \right. \\
 &\quad \left. \left. + 2(\gamma-1)\gamma\bar{Y}^2(\alpha\bar{X} + \beta)^2\epsilon_0 - 2(\gamma-1)\bar{Y}^2\bar{X}(\alpha\bar{X} + \beta)A\alpha\epsilon_1 - 2\gamma\bar{Y}^2\bar{X}(\alpha\bar{X} + \beta)A\alpha\epsilon_1\epsilon_0 \right\} \right. \\
 &\quad \left. \left\{ 1 - \frac{A\alpha\epsilon_1\bar{X}}{(\alpha\bar{X} + \beta)} + \left\{ \frac{A\alpha\epsilon_1\bar{X}}{(\alpha\bar{X} + \beta)} \right\}^2 - \dots \right\} \right]
 \end{aligned}$$

(ignoring higher power terms)

$$\begin{aligned}
 &= \frac{1}{(\alpha\bar{X} + \beta)^2} E \left[\left\{ (\gamma-1)^2\bar{Y}^2(\alpha\bar{X} + \beta)^2 + \bar{Y}^2(\gamma-1)^2A^2\alpha^2\epsilon_1^2\bar{X}^2 + \bar{Y}^2\gamma^2\epsilon_0^2(\alpha\bar{X} + \beta)^2 + \bar{Y}^2A^2\alpha^2\epsilon_1^2\bar{X}^2 \right. \right. \\
 &\quad \left. \left. - 2\gamma(\gamma-1)\bar{Y}^2\bar{X}(\alpha\bar{X} + \beta)A\alpha\epsilon_1\epsilon_0 + 2\bar{Y}^2(\gamma-1)\bar{X}^2A^2\alpha^2\epsilon_1^2 - 2\gamma\bar{Y}^2\bar{X}(\alpha\bar{X} + \beta)A\alpha\epsilon_1\epsilon_0 \right\} \right] \\
 &= \frac{1}{(\alpha\bar{X} + \beta)^2} E \left[\left\{ (\gamma-1)^2\bar{Y}^2(\alpha\bar{X} + \beta)^2 + \bar{Y}^2(\gamma-1)^2A^2\alpha^2\bar{X}^2E(\epsilon_1^2) + \bar{Y}^2\gamma^2(\alpha\bar{X} + \beta)^2E(\epsilon_0^2) \right. \right. \\
 &\quad \left. \left. + \bar{Y}^2\bar{X}^2A^2\alpha^2E(\epsilon_1^2) - 2\gamma(\gamma-1)\bar{Y}^2\bar{X}(\alpha\bar{X} + \beta)A\alpha E(\epsilon_0\epsilon_1) + 2\bar{Y}^2(\gamma-1)\bar{X}^2A^2\alpha^2E(\epsilon_1^2) \right. \right. \\
 &\quad \left. \left. - 2\gamma\bar{Y}^2\bar{X}(\alpha\bar{X} + \beta)A\alpha E(\epsilon_0\epsilon_1) \right\} \right] \\
 &= \frac{1}{(\alpha\bar{X} + \beta)^2} E \left[\left\{ (\gamma-1)^2\bar{Y}^2(\alpha\bar{X} + \beta)^2 + \bar{Y}^2(\gamma-1)^2A^2\alpha^2S_x^2 + \gamma^2(\alpha\bar{X} + \beta)^2S_y^2 + \bar{Y}^2A^2\alpha^2S_x^2 \right. \right. \\
 &\quad \left. \left. - 2\gamma(\gamma-1)\bar{Y}(\alpha\bar{X} + \beta)A\alpha S_{xy} + 2\bar{Y}^2(\gamma-1)A^2\alpha^2S_x^2 - 2\gamma\bar{Y}(\alpha\bar{X} + \beta)A\alpha S_{xy} \right\} \right] \\
 &= \frac{1}{(\alpha\bar{X} + \beta)^2} E \left[\left\{ (\gamma-1)^2\bar{Y}^2(\alpha\bar{X} + \beta)^2 + \bar{Y}^2\gamma^2A^2\alpha^2S_x^2 + \gamma^2(\alpha\bar{X} + \beta)^2S_y^2 - 2\gamma^2\bar{Y}A\alpha(\alpha\bar{X} + \beta)S_{xy} \right\} \right] \\
 &= \frac{1}{(\alpha\bar{X} + \beta)^2} E \left[\left\{ (\gamma-1)^2\bar{Y}^2(\alpha\bar{X} + \beta)^2 + \gamma^2\left\{ \bar{Y}^2A^2\alpha^2S_x^2 + (\alpha\bar{X} + \beta)^2S_y^2 - 2\bar{Y}A\alpha(\alpha\bar{X} + \beta)S_{xy} \right\} \right. \right. \\
 &\quad \left. \left. = \left[(\gamma-1)^2\bar{Y}^2 + \gamma^2 \left\{ S_y^2 + \frac{\bar{Y}^2A^2\alpha^2S_x^2}{(\alpha\bar{X} + \beta)^2} - \frac{2\bar{Y}A\alpha S_{xy}}{(\alpha\bar{X} + \beta)} \right\} \right] \right] \tag{2.1}
 \end{aligned}$$

which is the mean square error of the estimator under the proposed sampling strategy.

Differentiating (2.1) with respect to constants and equating to zero, we get the optimum values of these constant by solving the two normal equations i.e.

$$\left[(\gamma-1)\bar{Y}^2 + \gamma \left\{ S_y^2 + \frac{\bar{Y}^2A^2\alpha^2S_x^2}{(\alpha\bar{X} + \beta)^2} - \frac{2\bar{Y}A\alpha}{(\alpha\bar{X} + \beta)} S_{xy} \right\} \right] = 0 \tag{2.2}$$

$$= \left\{ 2\gamma^2\bar{Y}^2A\alpha^2S_x^2 - 2\bar{Y}\gamma^2\alpha(\alpha\bar{X} + \beta)S_{xy} \right\} = 0 \tag{2.3}$$

From (2.2) we get

$$\gamma = \frac{\bar{Y}^2}{\bar{Y}^2 + \left\{ S_y^2 + \frac{\bar{Y}^2A^2\alpha^2S_x^2}{(\alpha\bar{X} + \beta)^2} - \frac{2\bar{Y}A\alpha}{(\alpha\bar{X} + \beta)} S_{xy} \right\}} \tag{2.4}$$

From (2.3) we get

$$A_{opt} = \frac{S_{xy}}{\bar{Y}\alpha S_x} \tag{2.5}$$

Substituting (2.5) in (2.4), we get

$$\gamma_{opt} = \frac{\bar{Y}^2}{\bar{Y}^2 + \left\{ S_y^2 (1 - \rho^2) \right\}} \tag{2.6}$$

Now minimum mean square error under the proposed sampling strategy is obtained by putting (2.5) and (2.6) in (2.1) we get

$$MSE(\bar{y}_{Ay})_{opt} = \frac{\bar{Y}^2S_y^2(1 - \rho^2)}{\bar{Y}^2 + \left\{ S_y^2(1 - \rho^2) \right\}} \tag{2.7}$$

Efficiency Comparison

Comparison with mean per unit estimator

$$MSE(\bar{y}_{Ay}) < MSE(\bar{y})_{wor} \tag{3.1.1}$$

Comparison with ratio estimator

$$MSE(\bar{y}_{Ay}) < MSE(\bar{y})_{ratio} \tag{3.2.1}$$

Comparison with product estimator

$$MSE(\bar{y}_{Ay}) < MSE(\bar{y})_{pro} \tag{3.3.1}$$

Comparison with linear regression estimator

$$MSE(\bar{y}_{Ay})_{opt} < MSE(\bar{y})_{lr} \tag{3.4.1}$$

Comparison with generalized ratio estimator

$$MSE(\bar{y}_{Ay})_{opt} < MSE(\bar{y}_A)_{opt} \tag{3.5.1}$$

Empirical study

Population 1 is taken from the book ‘Singh & Chaudhary’ from page number 107. Population 2 is taken from the book ‘Singh & Chaudhary Singh’ from page number 176. Population 3 is taken from the book ‘P. V. Sukhatme’ from page number 51. Population 4 is taken from the book ‘Singh & Chaudhary’ from page number 141.

Table 1 Summary of the data

Population 1		Population 2		Population 3		Population 4	
N	12	N	20	N	25	N	22
n	5	n	8	n	12	n	8
\bar{X}	27.33	\bar{X}	641.05	\bar{X}	927.36	\bar{X}	22.62
\bar{Y}	550.33	\bar{Y}	816.45	\bar{Y}	735.8	\bar{Y}	1467.55
σ_y^2	429.59	σ_y^2	562.21	σ_y^2	785	σ_y^2	2503.23
σ_x^2	19.92	σ_x^2	517.11	σ_x^2	568.97	σ_x^2	32.29
ρ	0.93	ρ	0.89	ρ	0.96	ρ	0.90

Table 2 Mean square error (M) of various estimators

Estimator	Pop 1	Pop 2	Pop 3	Pop 4
$M(\bar{y})$	21530.53	23706.39	26703.08	498447.4
$M(\bar{y}_R)$	2876.249	6621.88	5929.65	94906.49
$M(\bar{y}_P)$	77726.36	105852.8	65139.19	1600009
$M(\bar{y}_{lr})$	2876.07	4788.30	1892.50	92776.46
$M_{min}(\bar{y}_{Ay})$	2659.59	4369.77	1751.23	60183.83

CONCLUSION

For all the population, the minimum mean square error of the proposed sampling strategy is less than the mean square error of the mean per unit estimator, ratio estimator, product estimator and the linear regression estimator. Thus the proposed sampling strategy is the best among mean per unit, ratio, product and regression estimator.

References

1. Bhushan, S. and Kumari, C. (2018). An Unbiased Family of Sampling Strategies for Estimation of Population Variance, *International Journal of Statistics and Systems*, 13 (1), 41-52
2. Bhushan, S. (2013). Improved sampling strategies in finite population. Scholars Press, Germany.
3. Bhushan S. (2012). Some efficient sampling strategies based on ratio type estimator, *Electronic Journal of Applied Statistical Analysis*, 5(1), 74-88.
4. Bhushan S., Masaldan R. N. and Gupta P. K. (2011). Improved sampling strategies based on modified ratio estimator , *International Journal of Agricultural and Statistical Sciences*, 7(1), 63-75.
5. Bhushan S. and Katara, S. (2010). On the class of unbiased sampling strategies , *Journal of Reliability and Statistical Studies*, 3(2), 93-101.
6. Bhushan S., Singh, R. K. and Katara, S. (2009). Improved estimation under Midzuno-Lahiri-Sen-type sampling scheme, *Journal of Reliability and Statistical Studies*, 2(2), 59-66.
7. Lahiri D. B. (1951). A method of sample selection providing unbiased ratio estimates, *Bull. Int. Stat. Inst.*, 3, p. 133-140.
8. Midzuno H. (1952). On the sampling system with probability proportional to the sum of the sizes, *Ann. Ints. Stat. Math.*, 3, p. 99-107.
9. Murthy M. N. (1967). *Sampling theory and method* , Statistical Publishing Society, Calcutta, India.
10. Sen A. R. (1952). Present status of probability sampling and its use in the estimation of characteristics, *Econometrika*, p. 20-103
11. Walsh J. E. (1970). Generalization of ratio estimator for population total, *Sankhya*, A, 32, p. 99-106.
12. Das, A. K., and Tripathi, T. P (1978). Use of auxiliary information in estimating the finite population variance. *Sankhya*, 40(C), 139-148.

How to cite this article:

Chandni Kumari and Ratan Kumar Thakur (2020) 'Improved Sampling Strategies for Finite Population Using Auxiliary Information', *International Journal of Current Advanced Research*, 09(04), pp. 21932-21935.
DOI: <http://dx.doi.org/10.24327/ijcar.2020.21935.4319>
