JOURNAL OF NORTHEASTERN UNIVERSITY

Volume 26 Issue 01, 2023 ISSN: 1005-3026 https://dbdxxb.cn/ Original Research Paper

AN IMPROVED ESTIMATOR FOR ESTIMATION OF FINITE POPULATION MEAN USING AUXILIARY INFORMATION

Manish Kumar* and Peeyush Misra
Department of Statistics, D.A.V. (P.G.) College, Dehradun – 248001, Uttarakhand (India)
kumarmann03@gmailcom and dr.pmisra.dav@gmail.com
***Corresponding Author**

Abstract: In the recent time researchers have focused on offering estimators for unknown population mean utilising auxiliary information in the sampling field. Auxiliary information is very useful to increase accuracy of the estimator. It is usual practice to increase the precision of estimators by using auxiliary data. The bias & MSE formulation for the suggested estimator have been derived to the first order of approximation. A comparison technique was utilised to assess the effectiveness of the recommended estimator and it was found that it outperformed several of the estimators that were already in use in the literature.

Keywords: Auxiliary Information, Bias, Mean Squared Error and Efficiency.

1. Introduction

In the recent time sampling has become more popular because it takes less time and money in comparison to complete enumeration. There are several approaches to employ auxiliary data in the sampling theory to enhance the population mean estimator. Cochran developed the concept of auxiliary information in 1940 and methods for employing auxiliary information for assessments to increase precisions have made significant contributions to contemporary sampling research. Both the ratio estimation approach and the regression estimator are helpful when data on an auxiliary variable is available and there is a positive association between the main variable and the auxiliary variable. When comparing estimators, the one with the lowest mean square error is regarded as the most effective. Many statisticians, including L.N. Upadhyaya and H. P. Singh (1999), Kadilar and H. Cingi, have employed population mean estimation to enhance ratio product, and exponential estimators employing study auxiliary data (2004), Misra P (2016), Misra P(2018), Ahuja et al (2020) and Misra P(2021).

Let the variable of interest be Y and the auxiliary variable be X taking the values Y_i and X_i respectively for the i^{th} (i=1,2,…, N) unit of the population of size N

Further, let

$$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i, \overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \ \ and$$

$$\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^r\left(Y_i - \overline{Y}\right)^s .$$

Also

$$C_Y = \frac{\sigma_Y}{\overline{Y}}, C_X = \frac{\sigma_X}{\overline{X}}, \beta_2 = \frac{\mu_{04}}{\mu_{02}^2}, \gamma_1 = \frac{\mu_{03}}{\mu_{02}^{3/2}}$$

Submitted: 13/03/2023 **Accepted**: 09/04/2023

$\sigma_Y^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)^2$, $\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2$ and

$\sigma_{XY} = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)$ Also $\rho = \frac{\rho_{XY}}{\sigma_X \sigma_Y}, B = \frac{\sigma_{xy}}{\sigma_x^2}$

Also, let

$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2$, $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2$

$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)$, $b = \frac{S_{xy}}{S_x^2}$

where $y_1, y_2, \ldots, y_n$ are the observations on y and x_1, $x_2, \ldots, x_n$ are the observations on auxiliary variable x for a simple random sample of size n.

For estimating the population mean, An improved estimator is proposed as

$$\hat{\overline{y}} = \overline{y} + b\left(\overline{X} - \overline{x}\right) + k_1\left\{\exp\left(\frac{s_x^2}{C_X^2} - \overline{x}^2\right) - 1\right\} + k_2\left\{\exp\left(\frac{s_y^2}{C_Y^2} - \overline{y}^2\right) - 1\right\} \tag{1.1}$$

2. Bias and Mean Squared Error

Let, $\overline{y} = \overline{Y}\left(1 + e_0\right)$

$\overline{x} = \overline{X}\left(1 + e_1\right)$

$s_y^2 = \sigma_y^2\left(1 + e_2\right)$

$s_x^2 = \sigma_X^2\left(1 + e_3\right)$

$s_{xy} = \sigma_{XY}\left(1 + e_4\right)$

such that $E(e_0) = E(e_1) = E(e_2) = E(e_3) = E(e_4) = 0$ (2.1)

$$E\left(e_0^2\right) = \frac{C_Y^2}{n}, E\left(e_1^2\right) = \frac{C_X^2}{n} \tag{2.2}$$

$$E\left(e_2^2\right) = \frac{\beta_2(y) - 1}{n} = \frac{A_Y}{n}, where\ A_Y = \beta_2(y) - 1 \tag{2.3}$$

$$E\left(e_0 e_1\right) = \frac{\rho C_X C_Y}{n} \tag{2.4}$$

$$E\left(e_0 e_2\right) = \frac{\mu_{03}}{n\sigma_Y^2 \overline{Y}} \text{ and } E\left(e_1 e_2\right) = \frac{\mu_{12}}{n\sigma_Y^2 \overline{X}} \tag{2.5}$$

$$E(e_1 e_3) = \frac{\mu_{30}}{n\sigma_X^2 \overline{X}} \tag{2.6}$$

$$E(e_3^2) = \frac{\beta_2(X)-1}{n} = \frac{A_X}{n}, where A_X = \beta_2(x)-1 \tag{2.7}$$

$$E(e_0 e_3) = \frac{\mu_{21}}{n\sigma_X^2 \overline{Y}} \tag{2.8}$$

$$E(e_2 e_3) = \frac{1}{n\sigma_X^2 \sigma_Y^2}(\mu_{22} - \mu_{20}\mu_{02}) = \frac{\delta - 1}{n}, \text{ where } \delta = \frac{\mu_{22}}{\mu_{20}\mu_{02}} \tag{2.9}$$

$$E(e_1 e_4) = \frac{\mu_{21}}{n\sigma_{XY} \overline{X}} \tag{2.10}$$

Now Expression (1.1) in terms of e_i's, we have

$$\hat{\bar{y}} = (\overline{Y} + e_0\overline{Y}) + \frac{\sigma_{XY}(1+e_4)}{\sigma_X^2(1+e_3)}(-\overline{X}e_1) + k_1\left\{\exp\left(\frac{\sigma_X^2(1+e_3)}{C_X^2} - \overline{X}^2(1+e_1)^2\right) - 1\right\}$$
$$+ k_2\left\{\exp\left(\frac{\sigma_Y^2(1+e_2)}{C_Y^2} - \overline{Y}^2(1+e_0)^2\right) - 1\right\}$$

$$\hat{\bar{y}} = (\overline{Y} + e_0\overline{Y}) - \frac{\sigma_{XY}}{\sigma_X^2}(1+e_4)(1 - e_3 + e_3^2)\overline{X}e_1 + k_1\left[\exp\left\{\overline{X}^2(1+e_3) - \overline{X}^2(1+e_1^2+2e_1)\right\} - 1\right]$$
$$+ k_2\left[\exp\left\{\overline{Y}^2(1+e_2) - \overline{Y}^2(1+e_0^2+2e_1)\right\} - 1\right]$$

$$\hat{\bar{y}} = (\overline{Y} + e_0\overline{Y}) - \frac{\sigma_{XY}}{\sigma_X^2}\overline{X}e_1(1 - e_3 + e_3^2 + e_4 - e_3e_4) + k_1\left[\exp\left\{\overline{X}^2(e_3 - 2e_1 - e_1^2)\right\} - 1\right]$$
$$+ k_2\left[\exp\left\{\overline{Y}^2(e_2 - 2e_0 - e_0^2)\right\} - 1\right]$$

Approximating it to the first order, we have

$$\hat{\bar{y}} = (\overline{Y} + e_0\overline{Y}) - \frac{\sigma_{XY}}{\sigma_X^2}\overline{X}(e_1 - e_1e_3 + e_1e_4) + k_1\left\{1 + \overline{X}^2(e_3 - 2e_1 - e_1^2) + \frac{\overline{X}^4}{2}(e_3^2 + 4e_1^2 - 4e_1e_3) - 1\right\}$$

$$+ k_2\left\{1 + \overline{Y}^2(e_2 - 2e_0 - e_0^2) + \frac{\overline{Y}^4}{2}(e_2^2 + 4e_0^2 - 4e_0e_2) - 1\right\}$$

or

$$\hat{\bar{y}} - \overline{Y} = e_0\overline{Y} - \frac{\sigma_{XY}}{\sigma_X^2}\overline{X}e_1 + k_1\overline{X}^2(e_3 - 2e_1) + k_2\overline{Y}^2(e_2 - 2e_0) + \frac{\sigma_{XY}}{\sigma_X^2}\overline{X}(e_1e_3 - e_1e_4)$$
$$- \frac{k_1\overline{X}^2}{2}\left\{2e_1^2 - \overline{X}^2(e_3^2 + 4e_1^2 - 4e_1e_3)\right\}$$

$$-\frac{k_2}{2}\bar{Y}^2\left\{2e_0^2-\bar{Y}^2\left(e_2^2+4e_0^2-4e_0e_2\right)\right\}$$

(2.11)

`

The bias in the first degree of approximation is produced by taking expectation on both sides of (2.11),

$$Bias\left(\hat{\bar{y}}\right)=E\left(\hat{\bar{y}}-\bar{Y}\right)=\frac{\sigma_{XY}}{\sigma_X^2}\bar{X}\left\{\frac{\mu_{30}}{n\sigma_X^2\bar{X}}-\frac{\mu_{21}}{n\sigma_{XY}\bar{X}}\right\}-\frac{k_1}{2}\bar{X}^2\left\{2\frac{C_X^2}{n}-\bar{X}^2\left(\frac{A_X}{n}+4\frac{C_X^2}{n}-4\frac{\mu_{30}}{n\sigma_X^2\bar{X}}\right)\right\}$$
$$-\frac{k_2}{2}\bar{Y}^2\left\{2\frac{C_Y^2}{n}-\bar{Y}^2\left(\frac{A_Y}{n}+4\frac{C_Y^2}{n}-4\frac{\mu_{03}}{n\sigma_Y^2\bar{Y}}\right)\right\}$$

(2.12)

Now squaring (2.11) and approximating to the first order, we have

$$\left(\hat{\bar{y}}-\bar{Y}\right)^2=e_0{}^2\bar{Y}^2-\frac{\sigma_{XY}^2}{\sigma_X^4}\bar{X}^2e_1{}^2-\frac{2\overline{XY}\sigma_{XY}}{\sigma_X^2}e_0e_1+k_1^2\bar{X}^4\left(e_3^2+4e_1^2-4e_1e_3\right)$$
$$+k_2^2\bar{Y}^4\left(e_2^2+4e_0^2-4e_0e_2\right)+2k_1\bar{X}^2\left\{\bar{Y}\left(e_0e_3-2e_0e_1\right)-\frac{\sigma_{XY}}{\sigma_X^2}\bar{X}\left(e_1e_3-2e_1^2\right)\right\}$$
$$+2k_2\bar{Y}^2\left\{\bar{Y}\left(e_0e_2-2e_0^2\right)-\frac{\sigma_{XY}}{\sigma_X^2}\bar{X}\left(e_1e_2-2e_0e_1\right)\right\}$$
$$+2k_1k_2\bar{X}^2\bar{Y}^2\left(e_2e_3-2e_0e_3-2e_1e_2+4e_0e_1\right)$$

(2.13)

The mse to the first degree of approximation is provided by taking expectation on both sides of (2.11) and using values of the expectations given from (2.1) to (2.10), we have

$$MSE\left(\hat{\bar{y}}\right)=E\left(\hat{\bar{y}}-\bar{Y}\right)^2$$
$$=\left(\bar{Y}^2\frac{C_Y^2}{n}+\frac{\sigma_{XY}^2}{\sigma_X^4}\bar{X}^2\frac{C_X^2}{n}-\frac{2\overline{XY}\sigma_{XY}}{\sigma_X^2}\frac{\rho C_XC_Y}{n}\right)+k_1^2\bar{X}^4\left(\frac{A_X}{n}+4\frac{C_X^2}{n}-4\frac{\mu_{30}}{n\sigma_X^2\bar{X}}\right)$$
$$+k_2^2\bar{Y}^4\left(\frac{A_Y}{n}+4\frac{C_Y^2}{n}-4\frac{\mu_{03}}{n\sigma_Y^2\bar{Y}}\right)+2k_1\bar{X}^2\left\{\bar{Y}\left(\frac{\mu_{21}}{n\sigma_X^2\bar{Y}}-2\rho\frac{C_XC_Y}{n}\right)\right.$$
$$\left.-\frac{\sigma_{XY}}{\sigma_X^2}\bar{X}\left(\frac{\mu_{30}}{n\sigma_X^2\bar{X}}-2\frac{C_X^2}{n}\right)\right\}+2k_2\bar{Y}^2\left\{\bar{Y}\left(\frac{\mu_{03}}{n\sigma_Y^2\bar{Y}}-2\frac{C_Y^2}{n}\right)\right\}-\frac{\sigma_{XY}}{\sigma_X^2}\bar{X}\left(\frac{\mu_{12}}{n\sigma_Y^2\bar{X}}-2\rho\frac{C_XC_Y}{n}\right)$$
$$+2k_1k_2\bar{X}^2\bar{Y}^2\left(\frac{\delta-1}{n}-2\frac{\mu_{21}}{n\sigma_X^2\bar{Y}}-2\frac{\mu_{12}}{n\sigma_Y^2\bar{X}}+4\rho\frac{C_XC_Y}{n}\right)$$

or

$$MSE\left(\hat{\bar{y}}\right)=\left(\bar{Y}^2\frac{C_Y^2}{n}+\frac{\sigma_{XY}^2}{\sigma_X^2 n}-2\rho^2\frac{\sigma_Y^2}{n}\right)+k_1^2\delta_{11}+k_2^2\delta_{22}+2k_1\delta_{10}+2k_2\delta_{02}+2k_1k_2\delta_{12}$$

(2.14)

where $\delta_{11}=\bar{X}^4\left(\frac{A_X}{n}+4\frac{C_X^2}{n}-4\frac{\mu_{30}}{n\sigma_X^2\bar{X}}\right)$

$$\delta_{22}=\bar{Y}^4\left(\frac{A_Y}{n}+4\frac{C_Y^2}{n}-4\frac{\mu_{03}}{n\sigma_Y^2\bar{Y}}\right)$$

$$\delta_{10}=\bar{X}^2\left\{\bar{Y}\left(\frac{\mu_{21}}{n\sigma_X^2\bar{Y}}-2\frac{\rho C_X C_Y}{n}\right)-\frac{\sigma_{XY}}{\sigma_X^2}\bar{X}\left(\frac{\mu_{30}}{n\sigma_X^2\bar{X}}-2\frac{C_X^2}{n}\right)\right\}$$

$$\delta_{02}=\bar{Y}^2\left\{\bar{Y}\left(\frac{\mu_{03}}{n\sigma_Y^2\bar{Y}}-2\frac{C_Y{}^2}{n}\right)-\frac{\sigma_{XY}}{\sigma_X^2}\bar{X}\left(\frac{\mu_{12}}{n\sigma_Y^2\bar{X}}-2\rho\frac{C_X C_Y}{n}\right)\right\}$$

$$\delta_{12}=\bar{X}^2\bar{Y}^2\left(\frac{\delta-1}{n}-2\frac{\mu_{21}}{n\sigma_X^2\bar{Y}}-2\frac{\mu_{12}}{n\sigma_Y^2\bar{X}}+4\frac{\rho C_X C_Y}{n}\right)$$

On solving these two normal equations for k_1 &$k_2$the optimum values of k_1& $k_2$are given by

$$k_1=\frac{\delta_{22}\delta_{10}-\delta_{02}\delta_{12}}{\delta_{12}^2-\delta_{11}\delta_{22}}$$

(2.15)

$$k_2=\frac{\delta_{11}\delta_{02}-\delta_{12}\delta_{10}}{\delta_{12}^2-\delta_{11}\delta_{22}}$$

(2.16)

For these optimum values of k_1 and $k_2$the minimum mean squared error of $\hat{\bar{y}}_{ME}$is given by

$$MSE\left(\hat{\bar{y}}\right)_{\min}=\left(\bar{Y}^2\frac{C_Y^2}{n}+\frac{\sigma_{XY}^2}{\sigma_X^2 n}-2\rho^2\frac{\sigma_Y^2}{n}\right)-\frac{\left(\delta_{11}\delta_{02}^2+\delta_{22}\delta_{10}^2-2\delta_{02}\delta_{10}\delta_{12}\right)}{\left(\delta_{11}\delta_{22}-\delta_{12}^2\right)}$$

(2.17)

3. Theoretical Comparison

i) The mse of the usual unbiased estimator $\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y$ of population mean is given by

$$MSE\left(\bar{y}\right)=\bar{y}^2\frac{C_Y^2}{n}$$

(3.1)

Hence MSE of the proposed estimator will be more efficient if

$$MSE\left(\bar{y}\right)-MSE\left(\hat{\bar{y}}\right)_{\min}>0$$

or

$h < \rho^2 \dfrac{\sigma_Y^2}{n}$

(3.2)

or

h <0

(3.3)

$-\dfrac{\sigma_{XY}^2}{\sigma_X^2 n} + 2\rho^2 \dfrac{\sigma_Y^2}{n} Y + h > 0$ or $\overline{Y}^2 h > 0$

where $h = \dfrac{\left(\delta_{11}\delta_{02}^2 + \delta_{22}\delta_{10}^2 - 2\delta_{02}\delta_{10}\delta_{12}\right)}{\left(\delta_{11}\delta_{22} - \delta_{12}^2\right)}$ (3.4)

Hence under the condition (3.4), MSE of the suggested estimator will be more effective than the usual unbiased estimator of population mean.

4. Empirical Study

We examine the effectiveness of the suggested estimator taken into account in this paper using a set of known population data. The population set is described as follows. The authors Ye K., Myers S.L., Myers R.H., and Walpole R.E. provided the information used in the empirical investigation (2011, Page 502). The oxygen intake in volume per unit body weight per unit time is a key indicator of aerobic fitness. Thirty-one people participated in an experiment to measure their oxygen consumption (y) while running 1.5 miles (x). The following values have been calculated as the necessary values.

$n = 31,$

$\overline{X} = 10.58613,$

$\overline{Y} = 47.37581$

$\sigma_X^2 = 1.86282,$

$\sigma_Y^2 = 27.46392,$

$C_y = 0.11062,$

$C_X = 0.3379,$

$\rho_{XY} = 0 - 0.86219,$

$\beta_{2Y} = 1.9026,$

$\beta_{2X} = 3.34559,$

MSE’s of estimators

$MSE\left(\overline{y}_m\right) = 0.88593$

$MSE\left(\hat{\overline{y}}\right)_{\min} = 0.50314$

5. Conclusion

In this section the concept of mean squared error criteria was used to test the performance of the estimators and the performance of the suggested estimator was established theoretically and empirically. The mean per unit unbiased estimator compared with the suggested estimator and it is shown that the proposed estimator outperforms in terms of MSE. The (PRE) of the suggested estimator under over the mean per unit estimators is 176%, and hence the proposed estimator exhibits its improved efficiency.

Acknowledgement

The authors are sincerely thankful to the anonymous referees and the editor in chief for their valuable ideas regarding improvement of the paper.

References

[1] Ahuja T.K. and Misra P. (2020): A generalized double sampling estimator of population mean using auxiliary information in survey sampling, International Journal of Agricultural and Statistical Sciences,16(2),751-756.

[2] C. Kadilar and H. Cingi (2004): Ratio Estimators in Simple Random Sampling. Applied Mathematics and Computation, 151: 893 – 902.

[3] C. Kadilar and H. Cingi(2006) : An Improvement in Estimating the Population Mean By Using the Correlation Coefficient, Hacettepe Journal of Mathematics and Statistics, 35,103 – 109.

[4] H. P. Singh, and R. Tailor (2005): An Improved Estimation of Population Mean using Power Transformation". Journal of the Indian Society of Agricultural Statistics, 8 (2), 223–230, 2005.

[5] L.N. Upadhyaya, and H. P. Singh (1999): Use of Transformed Auxiliary Variable in Estimating the Finite Population Mean , Biometrical Journal. 41,627 – 636.

[6] M. Subzar, S. Maqbool, T.A. Raja, and M. Abid (2018) : Ratio Estimators for Estimating Population Mean in Simple Random Sampling Using Auxiliary Information, Applied Mathematics and Information Sciences Letters, 6(3), 126 – 130.

[7] Misra P., Tiwari N. and Ahuja T.K. (2021). An enhanced two phase sampling ratio estimator for estimating population mean, Journal of Scientific Research 65(3), 177-183.

[8] Misra P. (2016): Generalized class of estimators for population mean using information on auxiliary variable, International Review of pure and Applied Mathematics, Volume 12, Number 2, 139-151.

[9] Misra P. (2018): Improved double sampling estimator of population mean using auxiliary information, International Journal of Agriculture and Statistics Sciences, Volume 14, Number 1, 181-186.

[10] Walpole R.E. , Myers R.H., Myers S.L. and Ye K. page no -502 (2011) : probability and Statistics for Engineers and Scientist Ninth Edition.

[11] W.G. Cochran (1940): Sampling Techniques, 3rd Edition, Wiley Eastern Limited, India