

Estimation of the Population Mean By Developing a New Estimator

*Sudhanshu Aggarwal¹, Shiv Shankar Soni², Ram Sahay Chaubey³

Author's Affiliation:

¹Assistant Professor, Department of Mathematics, National Post Graduate College Barhalganj, Gorakhpur-273402, Uttar Pradesh, India

E-mail: sudhanshu30187@gmail.com

²Assistant Professor, Department of Agricultural Statistics, National Post Graduate College Barhalganj, Gorakhpur-273402, Uttar Pradesh, India

E-mail: sonishivshankar@gmail.com

³Assistant Professor, Department of Agricultural Economics, National Post Graduate College Barhalganj, Gorakhpur-273402, Uttar Pradesh, India

E-mail: ramsahaychaubeynpg@gmail.com

***Corresponding Author: Sudhanshu Aggarwal**, Assistant Professor, Department of Mathematics, National Post Graduate College Barhalganj, Gorakhpur-273402, Uttar Pradesh, India

E-mail: sudhanshu30187@gmail.com

How to cite this article: Aggarwal S, Soni SS, Chaubey RS. (2022). Estimation of the Population Mean By Developing a New Estimator. *Bull. Pure Appl. Sci. Sect. E Math. Stat.* 41E(1), 11-15.

ABSTRACT

In the presented paper, authors propose a new estimator by combining two already exist ratio estimators and estimate the population mean in simple random sampling. Authors also determine the mean square error (MSE) of this estimator. In the last of this paper, they show that the presented estimator is more efficient than the existing ratio estimators theoretically and numerically.

KEYWORDS: Mean Square Error; Taylor Series Method; Ratio Estimator; Population Mean; Simple Random Sampling.

MSC2010: 62N02; 62G05; 62H12

1. INTRODUCTION

Sampling theory is an area of statistics which deals with the estimation of parameters of finite population. Parameter estimation is the requirement of almost all fields of investigation that includes agriculture, economics, industries, medical science, population studies, marketing and social science. In practice, parameters are unknown and estimated by taking a random sample using any sampling method suitable for the situation. There are various methods (simple random sampling, cluster sampling, double sampling, systematic sampling, stratified random sampling, two stage sampling, probability proportional to size sampling and multi-stage sampling) of selecting the samples from the population. Cochran [1] suggested ratio estimator for mean of the population. Srivastava [2] worked out a class of estimators of population mean using known population mean of the auxiliary variable while Srivastava and Jhaji [3] used known population mean as well as variance of the auxiliary variable and suggested a family of estimators for population mean.

Bahl and Tutaja [4] contributed for the estimation of population mean by suggesting ratio and product type exponential estimators. Singh and Tailor [5] discussed the problem of estimation of mean of the population by utilizing the coefficient of correlation. Kadilar and Cingi [6] suggested some ratio estimators in simple random sampling. Sisodia and Dwivedi [7] used coefficient of variation of auxiliary variable and presented a modified

ratio estimator. Kadilar et al. [8] used ranked set sampling and presented ratio estimator for the population mean.

The main aim of the presented paper is to develop a new estimator with its mean square error (MSE) by combining two already exist ratio estimators and estimating the population mean in simple random sampling.

2. NOTATIONS

$\bar{x}$ = Sample mean of the auxiliary variable

$\bar{y}$ = Sample mean of the study variable

$\bar{X}$ = Population mean of the auxiliary variable

$\bar{Y}$ = Population mean of the study variable

S_{xy} = Population covariance between auxiliary and study variables

S_{yx} = Population covariance between study and auxiliary variables

s_{xy} = Sample covariance between auxiliary and study variables

s_x^2 = Sample variance of auxiliary variable

S_x^2 = Population variance of auxiliary variable

S_y^2 = Population variance of study variable

$b = \frac{S_{xy}}{S_x^2}$ = Regression coefficient

N = Population size

n = Sample size

$R_1 = \frac{\bar{Y}}{\bar{X}}$ = Population ratio

C_x = Population coefficient of variation of auxiliary variable

$R_2 = \frac{\bar{Y}}{\bar{X} + C_x} = \text{Positive real constant}$

$f = \frac{n}{N}$ = Fraction population coefficient

$\rho = \left(\begin{array}{c} \text{Population correlation coefficient between} \\ \text{study and auxiliary variables} \end{array} \right)$

3. THE NEW PROPOSED ESTIMATOR

In this section, authors propose a new estimator by combining two already exist ratio estimators, which are given by $\left\{ \bar{y}_1 = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x}} \bar{X} \right\}$ and $\left\{ \bar{y}_2 = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x} + C_x} (\bar{X} + C_x) \right\}$, [6] as follows

$$\bar{y}_p = \left(\frac{\alpha}{\alpha + \beta} \right) \bar{y}_1 + \left(\frac{\beta}{\alpha + \beta} \right) \bar{y}_2 \quad (1)$$

where $\left(\frac{\alpha}{\alpha + \beta} \right)$ and $\left(\frac{\beta}{\alpha + \beta} \right)$ are weights.

The mathematical equations of Mean square error (MSE) of the estimators $\bar{y}_1$ and $\bar{y}_2$ were given by [6]

$$\text{MSE}(\bar{y}_1) = \left(\frac{1-f}{n} \right) [S_y^2(1 - \rho^2) + R_1^2 S_x^2] \quad (2)$$

$$\text{MSE}(\bar{y}_2) = \left(\frac{1-f}{n} \right) [S_y^2(1 - \rho^2) + R_2^2 S_x^2] \quad (3)$$

The Mean square error (MSE) of the estimator, which is given by equation (1), can be easily evaluated by Taylor series method using first degree approximation as

$$\text{MSE}(\bar{y}_p) \cong d \sum d' \quad (4)$$

where

d' = the transpose of the matrix d

$$d = \left[\begin{array}{cc} \left. \frac{\partial h(l, m)}{\partial l} \right|_{\bar{x}, \bar{y}} & \left. \frac{\partial h(l, m)}{\partial m} \right|_{\bar{x}, \bar{y}} \end{array} \right]$$

$$\sum = \left(\frac{1-f}{n} \right) \begin{bmatrix} S_y^2 & S_{yx} \\ S_{xy} & S_x^2 \end{bmatrix}$$

(see [9]), $h(\bar{x}, \bar{y}) = h(l, m) = \bar{y}_p$.

Using the above definition of d , the value of d for the proposed estimators is given by

$$d = \left[1 - \left(\frac{\alpha}{\alpha + \beta} \right) (R_1 + B) - \left(\frac{\beta}{\alpha + \beta} \right) (R_2 + B) \right]$$

$$\text{where } B = \frac{S_{xy}}{S_x^2} = \frac{\rho S_x S_y}{S_x^2} = \frac{\rho S_y}{S_x} \quad (5)$$

Here it is very important to point out that authors omit the difference $[E(b) - B]$ [10]
Using (4), the mathematical expression of MSE of our propose estimator is presented as

$$E(\bar{y}_p) = \left(\frac{1-f}{n} \right) \left[S_y^2 - 2 \left\{ \left(\frac{\alpha}{\alpha + \beta} \right) (R_1 + B) + \left(\frac{\beta}{\alpha + \beta} \right) (R_2 + B) \right\} S_{yx} \right. \\ \left. + \left\{ \left(\frac{\alpha}{\alpha + \beta} \right) (R_1 + B) + \left(\frac{\beta}{\alpha + \beta} \right) (R_2 + B) \right\}^2 S_x^2 \right] \quad (6)$$

The optimum value of weights $\left(\frac{\alpha}{\alpha + \beta} \right)$ and $\left(\frac{\beta}{\alpha + \beta} \right)$ to minimize equation (6) can easily determined and given by

$$\left(\frac{\alpha}{\alpha + \beta} \right)_{\text{optimum}} = \frac{R_2}{R_2 - R_1} \quad \left(\frac{\beta}{\alpha + \beta} \right)_{\text{optimum}} = \frac{R_1}{R_1 - R_2} \quad (7)$$

After putting optimum values of weights in equation (6), we obtain the expression of minimum value of MSE of the present estimator as

$$\text{Min MSE}(\bar{y}_p) = \left(\frac{1-f}{n} \right) \left[S_y^2 - 2 \left\{ \left(\frac{R_2}{R_2 - R_1} \right) (R_1 + B) + \left(\frac{R_1}{R_1 - R_2} \right) (R_2 + B) \right\} S_{yx} \right. \\ \left. + \left\{ \left(\frac{R_2}{R_2 - R_1} \right) (R_1 + B) + \left(\frac{R_1}{R_1 - R_2} \right) (R_2 + B) \right\}^2 S_x^2 \right] \\ \Rightarrow \text{Min MSE}(\bar{y}_p) = \left(\frac{1-f}{n} \right) \left[S_y^2 - 2 \left\{ \left(\frac{R_2 R_1 + B R_2 - R_1 R_2 - B R_1}{R_2 - R_1} \right) \right\} S_{yx} \right. \\ \left. + \left\{ \left(\frac{R_2 R_1 + B R_2 - R_1 R_2 - B R_1}{R_2 - R_1} \right) \right\}^2 S_x^2 \right] \\ \Rightarrow \text{Min MSE}(\bar{y}_p) = \left(\frac{1-f}{n} \right) [S_y^2 - 2 B S_{yx} + B^2 S_x^2] \quad (8)$$

Using equation (5) in equation (8), we have

$$\text{Min MSE}(\bar{y}_p) = \left(\frac{1-f}{n} \right) \left[S_y^2 - 2 \left(\frac{\rho S_y}{S_x} \right) \rho S_x S_y + \left(\frac{\rho S_y}{S_x} \right)^2 S_x^2 \right] \\ \Rightarrow \text{Min MSE}(\bar{y}_p) = \left(\frac{1-f}{n} \right) [S_y^2 - 2 \left(\frac{\rho S_y}{S_x} \right) \rho S_x S_y + \rho^2 S_y^2] \\ \Rightarrow \text{Min MSE}(\bar{y}_p) = \left(\frac{1-f}{n} \right) [S_y^2 - \rho^2 S_y^2] \\ \Rightarrow \text{Min MSE}(\bar{y}_p) = \left(\frac{1-f}{n} \right) S_y^2 [1 - \rho^2] \quad (9)$$

4. COMPARISONS OF EFFICIENCY

In this section of the paper, we shall compare the efficiency of proposed estimator $\bar{y}_p$ with the already established estimators $\bar{y}_1$ and $\bar{y}_2$ by the help of their MSE equations, which are given by equations (9), (2) and (3) respectively as follows

$$\text{MSE}(\bar{y}_p) - \text{MSE}(\bar{y}_1) < 0 \\ \Rightarrow \left(\frac{1-f}{n} \right) S_y^2 [1 - \rho^2] - \left(\frac{1-f}{n} \right) [S_y^2 (1 - \rho^2) + R_1^2 S_x^2] < 0$$

$$\begin{aligned} \Rightarrow -R_1^2 S_x^2 &< 0 \\ \Rightarrow R_1^2 S_x^2 &> 0 \end{aligned} \quad (10)$$

Since equation (10) is always true so it implies that the estimator of our study i.e. $\bar{y}_p$ is more efficient than the already established estimator $\bar{y}_1$.

Again $MSE(\bar{y}_p) - MSE(\bar{y}_2) < 0$

$$\begin{aligned} \Rightarrow \left(\frac{1-f}{n}\right) S_y^2 [1 - \rho^2] - \left(\frac{1-f}{n}\right) [S_y^2 (1 - \rho^2) + R_2^2 S_x^2] &< 0 \\ \Rightarrow -R_2^2 S_x^2 &< 0 \\ \Rightarrow R_2^2 S_x^2 &> 0 \end{aligned} \quad (11)$$

Since equation (11) is always true so it implies that the estimator of our study i.e. $\bar{y}_p$ is more efficient than the already established estimator $\bar{y}_2$.

Thus, from equations (10) and (11), it is finalized that the estimator of our study i.e. $\bar{y}_p$ is more efficient than the already established estimators $\bar{y}_1$ and $\bar{y}_2$.

5. NUMERICAL ILLUSTRATION

For numerical purpose, we have considered the following population data set given by [6] (see Table: 1)

Table 1: Data Statistics

$N = 106, n = 20$	$\rho = 0.86$	$C_x = 2.10$	$S_{yx} = 568176176.10$
$\bar{X} = 2212.59$	$R_1 = 0.0807$	$R_2 = 0.0807$	$S_x = 57460.61$
$\bar{Y} = 27421.70$	$S_y = 11551.53$		

Table 2: MSE Values of Estimators

ESTIMATOR	Value of MSE
$(\bar{y}_1)$	2318722.45
$(\bar{y}_2)$	2318589.19
$(\bar{y}_p)$	1446719.34

Table: 2 presents the values of MSE for the proposed estimator $\bar{y}_p$ with already established estimators $\bar{y}_1$ and $\bar{y}_2$ using their MSE equations and Table: 1. From the Table: 2, it is clear that the estimator of our study i.e. $\bar{y}_p$ is more efficient than the already established estimators $\bar{y}_1$ and $\bar{y}_2$.

6. CONCLUSION

In the presented paper, authors successfully developed a new estimator $\bar{y}_p$ by combining already established two estimators $\bar{y}_1$ and $\bar{y}_2$. Authors also derived the MSE equation of this new estimator and showed that this estimator is more efficient than the previously established estimators $\bar{y}_1$ and $\bar{y}_2$. Further, authors used the data of previous study and proved numerically using this data that the estimator of our study i.e. $\bar{y}_p$ is more efficient than the previously established estimators $\bar{y}_1$ and $\bar{y}_2$. In future, we can developed more efficient new estimators using combination techniques as discuss in this paper.

7. DATA AVAILABILITY

Authors of this paper confirm that the datasets that are used in this paper are available from the author upon request.

8. ACKNOWLEDGMENT

The authors are thankful to the referees and to the Editor-in-Chief for their critical comments which provided them an opportunity to improve this paper.

9. COMPLIANCE WITH ETHICS REQUIREMENTS

This paper does not contain any studies with human or animal subjects.

10. CONFLICTS OF INTEREST

There are no conflicts of interest between the authors.

11. FUNDING STATEMENT

This research has no funding.

REFERENCES

1. Cochran, W.G. (1940) The estimation of yields of cereal experiments by sampling for ratio of grain to total produce, *J. Agri. Sci.*, 30, 262-275.
2. Srivastava, S.K. (1980) A class of estimators using auxiliary information in sample surveys, *Canad. J. Statist.*, 8, 253-254.
3. Srivastava, S.K. and Jhaggi, H.S. (1981) A class of estimators of the population mean in survey sampling using auxiliary information, *Biometrika*, 68(1), 341-343.
4. Bahl, S. and Tuteja, R.K. (1991) Ratio and product type exponential estimators, *Info. & Opti. Sci.*, 12, 159-163.
5. Singh, H.P. and Tailor, R. (2003) Use of known correlation coefficient in estimating the finite population mean, *Statist. In Transi.*, 6, 555-560.
6. Kadilar, C. and Cingi, H. (2004) Ratio estimators in simple random sampling, *App. Math. and Comp.*, 151, 893-902.
7. Sisodia, B.V.S. and Dwivedi, V.K. (1981) A modified ratio estimator using coefficient of variation of auxiliary variable, *J. Indian Soc. Agri. Stat.*, 33, 13-18.
8. Kadilar, C., Unyazici, Y. and Cingi, H. (2009) Ratio estimator for the population mean using ranked set sampling, *Statistical Papers*, 50(2), 301-309.
9. Wolter, K.M. (2005) *Introduction to variance estimation*, 2ed, Springer, New York, USA.
10. Cochran, W.G. (1953) *Sampling techniques*, 1ed, Asian Publishing House, Bombay.
