

# An Improved Estimator of Population Mean using Information on Median of the Study Variable

<sup>1</sup>Dharmendra Kumar Yadav, Ravendra Kumar\*<sup>2</sup>, Prof Sheela Misra<sup>3</sup>, S.K. Yadav<sup>4</sup>

<sup>1</sup>Department of Statistics, University of Lucknow, Lucknow-226007, U.P., INDIA

<sup>2</sup>Department of Mathematics, Dr Ram Manohar Lohiya Government Degree College, Aonla, Bareilly, U.P., India.

<sup>3</sup>Department of Statistics University of Lucknow 226007

<sup>4</sup>Department of Mathematics and Statistics (A Centre of Excellence), Dr RML Avadh University, Faizabad-224001, U.P., INDIA

## ABSTRACT

The present paper advocates the estimation of population mean of the study variable by utilizing the information on median of the study variable. A generalized ratio type estimator has been proposed for this purpose. The expressions for the bias and mean squared error of the proposed estimator have been derived up to the first order of approximation. The optimum value of the characterizing scalar has also been obtained. The minimum value of the proposed estimator for this optimum value of the characterizing scalar is obtained. A theoretical efficiency comparison of the proposed estimator has been made with the mean per unit estimator, usual ratio of Cochran (1940) and usual regression estimator of Watson (1937), Bahl and Tuteja (1991) estimator, Kadilar (2016) and Subramani (2016) estimators. Theoretical results are supported by the numerical illustration and found that proposed estimator performs better than the existing estimators.

**Key words:** Study variable, Bias, Ratio estimator, Mean squared error, Simple random sampling, Efficiency.

## 1. INTRODUCTION:

In many real life situations we observed many cases where population mean of the study variable is not known but the population median of the main variable under study is known. For example if we ask for the weight or basic salary of a person, it is very hard to get the exact value but we get the information in terms of interval or the pay band. Here we can easily get the median of the study variable which can be used for improving the estimation procedure of population mean of study variable. In sampling theory use of auxiliary information is a very common practice for improving the precision of estimates. But the use of auxiliary information has a serious drawback in terms of increased survey cost for collection of this additional information. The use of median of study variable may be an important attempt in this direction as it increases the efficiency of estimator without any additional survey cost. In the present paper we have proposed an improved estimator of population mean of the study variable using median of the study variable.

Let us consider a finite population consisting of  $N$  distinct and identifiable units and let  $(x_i, y_i), i = 1, 2, \dots, n$  be a bivariate sample of size  $n$  taken from  $(X, Y)$  using a simple random sampling without replacement (*SRSWOR*) scheme. Let  $\bar{X}$  and  $\bar{Y}$  respectively be the population means of the auxiliary and the study variables, and let  $\bar{x}$  and  $\bar{y}$  be the corresponding sample means. In *SRSWOR*, It is well established fact that sample means  $\bar{x}$  and  $\bar{y}$  are unbiased estimators of population means of  $\bar{X}$  and  $\bar{Y}$  respectively.

To demonstrate the above problem in a more effective manner, let us consider an interesting example of mean estimation of study variable using median of study variable given by Subramani (2016). The table has been used with permission of the author.

**Example.1.** The problem is to estimate the average salary drawn by the faculty members (population mean) per month in an Indian university 800 faculty members are working in different categories and the basic salary drawn by different categories of the faculty members.

**Table2: Salary of University faculty members**

Category	Basic Salary in Indian Rupees (IRs) Per month*	Number of faculty members	Cumulative total
Senior Professor	56000+10000**	20	20
Professor - Grade I	43000+10000	40	60
Professor - Grade II	37400+10000	60	120
Associate Professor - Grade I	37400+10000	80	200
Associate Professor - Grade II	37400+9000	100	300
Assistant Professor - Grade I	15100+8000	110	410
Assistant Professor - Grade II	15100+7000	140	550
Assistant Professor - Grade III	15100+6000	250	800
Total		800	800

\*Actual salary depends on their experience in their designation and other allowances.

\*\*The Basic salary is the sum of the basic (the first value) and the academic grade pay (the second value), which will differentiate people with same designation but different grades.

The population median value will be assumed as IRs. 15100+8000 = IRs. 23100.

## 2. REVIEW OF EXISTING ESTIMATORS

The sample mean is the most suitable estimator of population mean of the study variable, given by,

$$t_o = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (1)$$

It is an unbiased estimator and its variance, up to the first order of approximation, is given by

$$V(t_o) = \frac{1-f}{n} S_y^2 = \frac{1-f}{n} \bar{Y}^2 C_y^2 \quad (2)$$

where,  $C_y = \frac{S_y}{\bar{Y}}$ ,  $S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2 = \frac{1}{N} \sum_{i=1}^{N C_n} (\bar{y}_i - \bar{Y})^2$ ,  $f = \frac{n}{N}$ .

Watson (1937) first utilized the highly correlated auxiliary variable and proposed the usual linear regression estimator of population mean as,

$$t_1 = \bar{y} + b_{yx} (\bar{X} - \bar{x}) \quad (3)$$

Where  $b_{yx}$  is the regression coefficient of Y on X.

This estimator is also unbiased for population mean and its variance up to the first order of approximation, is given by,

$$V(t_1) = \frac{1-f}{n} \bar{Y}^2 C_y^2 (1 - \rho_{yx}^2). \quad (4)$$

Cochran (1940) made use of highly positively correlated auxiliary variable and proposed the following usual ratio estimator as,

$$t_2 = \bar{y} \frac{\bar{X}}{\bar{x}} \quad (5)$$

It is a biased estimator of population mean and the expressions for the bias and mean squared error for this estimator, up to the first order of approximation are given as,  $B(t_2) = \frac{1-f}{n} \bar{Y} [C_x^2 - C_{yx}]$  and

$$MSE(t_2) = \frac{1-f}{n} \bar{Y}^2 [C_y^2 + C_x^2 - 2C_{yx}], \quad (6)$$

where,  $C_x = \frac{S_x}{\bar{X}}$ ,  $S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2 = \frac{1}{N} \sum_{i=1}^N (\bar{x}_i - \bar{X})^2$ ,  $\rho_{yx} = \frac{Cov(x, y)}{S_x S_y}$ ,

$$Cov(x, y) = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X}), \text{ and } C_{yx} = \rho_{yx} C_y C_x.$$

Bahl and Tuteja (1991) proposed the following exponential ratio type estimator of population mean by making use of positively correlated auxiliary variable as,

$$t_3 = \bar{y} \exp \left[ \frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}} \right] \tag{7}$$

The above estimator is biased and the bias and the mean squared error of this estimator, up to the first order of approximation, are given respectively by,

$$B(t_3) = \frac{1-f}{8n} \bar{Y} [3C_x^2 - 4C_{yx}]$$

$$MSE(t_3) = \frac{1-f}{n} \bar{Y}^2 [C_y^2 + \frac{C_x^2}{4} - C_{yx}]. \tag{8}$$

Kadilar (2016), using positively correlated auxiliary variable proposed the following exponential type estimator of population mean as,

$$t_4 = \bar{y} \left( \frac{\bar{x}}{\bar{X}} \right)^\delta \exp \left( \frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}} \right) \tag{9}$$

where  $\delta$  is a characterizing scalar to be determined such that the MSE of above estimator is minimum. The bias and the mean squared error of the above estimator up to the first order of approximation respectively are,

$$B(t_4) = \frac{1-f}{n} \bar{Y} \left[ \left\{ \frac{\delta(\delta-1)}{2} + \frac{3}{8} \right\} C_x^2 + \left( \delta + \frac{1}{2} \right) C_{yx} \right]$$

$$MSE(t_4) = \frac{1-f}{n} \bar{Y}^2 \left[ C_y^2 + \left( \delta^2 + \delta + \frac{1}{4} \right) C_x^2 + (2\delta + 1) C_{yx} \right] \tag{10}$$

The optimum value of the characterizing scalar  $\delta$  which minimizes the mean squared error of  $t_6$  is,

$$\delta_{opt} = \left( \frac{1}{2} - \rho_{yx} C_y / C_x \right)$$

The minimum value of the mean squared error of above estimator is,

$$MSE_{min}(t_4) = \frac{1-f}{n} \bar{Y}^2 C_y^2 (1 - \rho_{yx}^2) \tag{11}$$

which is equal to the variance of the usual regression estimator of Watson (1937).

Subramani (2016) used the population median of the study variable and proposed the following ratio estimator of population mean of the study variable,

$$t_5 = \bar{y} \left( \frac{M}{m} \right) \tag{12}$$

where  $M$  and  $m$  are the population and sample medians of study variable respectively.

It is a biased estimator and its bias and the mean squared error, up to the first order of approximation, are respectively given by,

$$B(t_5) = \frac{1-f}{n} \bar{Y} [C_m^2 - C_{ym} - \frac{Bias(m)}{M}] \text{ and}$$

$$MSE(t_5) = \frac{1-f}{n} \bar{Y}^2 [C_y^2 + R_5^2 C_m^2 - 2R_5 C_{ym}], \quad (13)$$

where

$$R_5 = \frac{\bar{Y}}{M}, \quad C_m = \frac{S_m}{M}, \quad S_m^2 = \frac{1}{N} \sum_{i=1}^{N} (m_i - M)^2, \quad S_{ym} = \frac{1}{N} \sum_{i=1}^{N} (\bar{y}_i - \bar{Y})(m_i - M) \text{ and}$$

$$C_{ym} = \frac{S_{ym}}{YM}.$$

Many authors have given various modified estimators of population mean using auxiliary variables. The latest references can be found in Subramani (2013), Subramani and Kumarapandiyam (2012, 2013), Yan and Tian (2010), Yadav and Kadilar (2013), Yadav et al. (2014, 2015), and Yadav et al. (2016).

### 3. PROPOSED ESTIMATOR

We propose the following ratio type estimators of population mean using known population median of study variable as,

$$t_R = \bar{y} \exp \left[ \frac{M}{M+a(m-M)} - 1 \right] \quad (14)$$

where  $a$  is a characterizing scalar to be determined such that the mean squared error of the proposed estimator  $t_R$  is minimum.

The following approximations have been made to study the properties of the proposed estimators as,

$$\bar{y} = \bar{Y}(1 + e_0) \text{ and } m = M(1 + e_1) \text{ such that } E(e_0) = 0, \quad E(e_1) = \frac{\bar{M} - M}{M} = \frac{Bias(m)}{M} \text{ and}$$

$$E(e_0^2) = \frac{1-f}{n} C_y^2, \quad E(e_1^2) = \frac{1-f}{n} C_m^2, \quad E(e_0 e_1) = \frac{1-f}{n} C_{ym},$$

$$\text{where, } \bar{M} = \frac{1}{n} \sum_{i=1}^n m_i$$

The proposed estimator  $t_R$  can be expressed in terms of  $e_i$ 's ( $i = 1, 2$ ) as,

$$t_R = \bar{Y} (1 + e_0) \exp \left[ \frac{M}{M + a\{M(1 + e_1) - M\}} - 1 \right]$$

$$t_R - \bar{Y} = \bar{Y} (e_0 - a e_1 - a e_0 e_1 + \frac{3}{2} a^2 e_1^2) \quad (15)$$

Taking expectation on both sides we get  $Bias(t_R)$

$$Bias(t_R) = \bar{Y} \left[ \frac{3}{2} a^2 C_m^2 - a C_{ym} - a \frac{Bias(m)}{M} \right]$$

Squaring (15) both sides and taking expectation we get the  $MSE(t_R)$  upto first order approximation,

$$E(t_R - \bar{Y})^2 = \bar{Y}^2 E[e_0 - a e_1]^2$$

$$= \bar{Y}^2 E[e_0^2 + a^2 e_1^2 - 2a e_0 e_1]$$

$$= \bar{Y}^2 [E(e_0^2) + a^2 E(e_1^2) - 2a E(e_0 e_1)]$$

$$= \bar{Y}^2 [\lambda C_y^2 + a^2 \lambda C_m^2 - 2a \lambda C_{ym}]$$

$$MSE(t_R) = \lambda \bar{Y}^2 [C_y^2 + a^2 C_m^2 - 2a C_{ym}] \quad (16)$$

which is minimum for,

$$a_{opt} = \frac{C_{ym}}{C_m^2}$$

and the minimum mean squared error of the proposed estimator  $t_R$  is,

$$MSE(t_R)_{min} = \lambda \bar{Y}^2 \left[ C_y^2 - \frac{C_{ym}^2}{C_m^2} \right]. \quad (17)$$

#### 4. EFFICIENCY COMPARISON

Under this section, a theoretical comparison of the proposed estimator has been made with the competing estimators of population mean. The conditions under which the proposed estimator performs better than the competing estimators have also been given.

From equation (17) and equation (2), we have,

$$V(t_0) - MSE_{min}(t_R) > 0 \text{ if } \frac{C_{ym}^2}{C_m^2} > 0, \text{ or if } C_{ym}^2 > 0$$

Thus the proposed estimator is better than the usual mean per unit estimator of population mean.

From equation (17) and equation (4), we have,

$$MSE(t_1) - MSE_{min}(t_d) > 0 \text{ if } \frac{C_{ym}^2}{C_m^2} - C_y^2 \rho_{yx}^2 > 0$$

Under the above condition, the proposed estimator is better than the usual regression estimator of Watson (1937).

From equation (17) and equation (6), we have,

$$MSE(t_2) - MSE_{min}(t_R) > 0 \text{ if } C_x^2 + \frac{C_{ym}^2}{C_m^2} > 2C_{yx}$$

Under the above condition, proposed estimators perform better than the usual ratio estimator given by Cochran (1940).

From equation (17) and equation (8), we have,

$$MSE(t_3) - MSE_{min}(t_R) > 0 \text{ if } \frac{C_x^2}{4} + \frac{C_{ym}^2}{C_m^2} > C_{yx}$$

Under the above condition, the proposed estimator performs better than Bahl and Tuteja (1991) ratio type estimator of population mean.

From equation (17) and equation (11), we have,

$$MSE(t_4) - MSE_{min}(t_R) > 0, \text{ if } \frac{C_{ym}^2}{C_m^2} - C_y^2 \rho_{yx}^2 > 0$$

Under the above condition, the proposed estimator is better than the Kadilar (2016) estimator. From equation (17) and equation (13), we have,

$$MSE(t_5) - MSE_{min}(t_R) > 0, \text{ if } R_5^2 C_m^2 + \frac{C_{ym}^2}{C_m^2} > 2R_5 C_{ym}$$

Under the above condition, the proposed estimator is better than the Subramani (2016) estimator of population mean using information on median of the study variable.

**5. NUMERICAL STUDY**

To judge the theoretical findings, we have considered the natural populations given in Subramani (2016). He has used three natural populations. The population 1 and 2 have been taken from Singh and Chaudhary (1986, page no. 177) and the population 3 has been taken from Mukhopadhyay (2005, page no. 96). In populations 1 and 2, the study variable is the estimate the area of cultivation under wheat in the year 1974, whereas the auxiliary variables are the cultivated areas under wheat in 1971 and 1973 respectively. In population 3, the study variable is the quantity of raw materials in lakhs of bales and the number of labourers as the auxiliary variable, in thousand for 20 jute mills. Tables 3 and 4 represent the parameter values along with constants, along with proposed estimator, variances and mean squared errors of existing and proposed estimator

**Table-3. Parameter values and constants for three natural populations**

Parameter	Population-1	Population-2	Population-3
$N$	34	34	20
$n$	5	5	5
${}^N C_n$	278256	278256	15504
$\bar{Y}$	856.4118	856.4118	41.5
$\bar{M}$	736.9811	736.9811	40.0552
$M$	767.5	767.5	40.5
$\bar{X}$	208.8824	199.4412	441.95
$R_7$	1.1158	1.1158	1.0247
$C_y^2$	0.125014	0.125014	0.008338
$C_x^2$	0.088563	0.096771	0.007845
$C_m^2$	0.100833	0.100833	0.006606
$C_{ym}$	0.07314	0.07314	0.005394
$C_{yx}$	0.047257	0.048981	0.005275
$\rho_{yx}$	0.4491	0.4453	0.6522

**Table-4. Mean squared error of various estimators**

Estimator	Popln-1	Popln-2	Popln-3
$t_0$	15640.97	15640.97	2.15
$t_1$	12486.75	12539.30	1.24
$t_2$	14895.27	15492.08	1.48
$t_3$	12498.01	12539.30	1.30
$t_4$	12486.75	12539.30	1.24
$t_5$	10926.53	10926.53	1.09
$t_R$	9002.22	9002.22	0.98

**6. RESULTS AND CONCLUSION**

From Table-4, it can be seen that the proposed estimator has minimum mean squared error among other competing estimators of population mean of study character. Thus proposed estimator is better than usual mean per unit estimator, Watson (1937) usual regression estimator, Cochran (1940) usual ratio estimator, Bahl and Tuteja (1991) exponential ratio type estimator, Kadilar (2016) estimator and Subramani (2016) estimator.

Therefore it is recommended that the proposed estimator may be used for improved estimation of population mean under simple random sampling scheme.

## REFERENCE

- 1) ABID, M., ABBAS, N. SHERWANI, R.A.K. AND NAZIR, H. Z. (2016). Improved Ratio Estimators for the population mean using non-conventional measure of dispersion. *Pakistan Journal of Statistics and Operations Research*, XII (2), 353-367.
- 2) BAHL, S. AND TUTEJA, R.K. (1991). Ratio and product type exponential estimator, *Information and Optimization Sciences*, XII (I), 159-163.
- 3) COCHRAN, W. G. (1940). The Estimation of the Yields of the Cereal Experiments by Sampling for the Ratio of Grain to Total Produce. *The Journal of Agric. Science*, 30, 262-275.
- 4) JERAJUDDIN, M. AND KISHUN, J. (2016). Modified Ratio Estimators for Population Mean Using Size of the Sample, Selected From Population, *IJSRSET*, 2, 2, 10-16.
- 5) KADILAR, G.O. (2016). A New Exponential Type Estimator for the Population Mean in Simple Random Sampling, *Journal of Modern Applied Statistical Methods*, 15, 2, 207-214.
- 6) REDDY, V.N. (1974): On a transformed ratio method of estimation. *Sankhya*, C, 36(1), 59-70.
- 7) SRIVASTAVA, S.K. (1967): An estimator using auxiliary information in sample surveys. *Cal. Statist. Assoc. Bull.*, 16, 62-63.
- 8) SUBRAMANI, J. (2013). Generalized modified ratio estimator of finite population mean, *Journal of Modern Applied Statistical Methods*, 12 (2), 121-155.
- 9) SUBRAMANI, J., KUMARAPANDIYAN, G. (2012). Estimation of population mean using coefficient of variation and median of an auxiliary variable, *International Journal of Probability and Statistics*, 1 (4), 111-118.
- 10) SUBRAMANI, J., KUMARAPANDIYAN, G. (2013). A new modified ratio estimator of population mean when median of the auxiliary variable is known, *Pakistan Journal of Statistics and Operation Research*, Vol. 9 (2), 137-145.
- 11) SUBRAMANI, J. (2016). A new median based ratio estimator for estimation of the finite population mean, *Statistics in Transition New Series*, 17, 4, 1-14.
- 12) TAILOR, R., SHARMA, B. (2009). A modified ratio-cum-product estimator of finite population mean using known coefficient of variation and coefficient of kurtosis, *Statistics in Transition-New Series*, 10 (1), 15-24.
- 13) WATSON, D.J. (1937). The estimation of leaf area in field crops, *The Journal of Agricultural Science*, 27, 3, 474-483.
- 14) YADAV, S.K; MISHRA, S.S. AND SHUKLA, A.K. (2014). Improved Ratio Estimators for Population Mean Based on Median Using Linear Combination of Population Mean and Median of an Auxiliary Variable. *American Journal of Operational Research*, 4, 2, 21-27.
- 15) YADAV, S.K; MISHRA, S.S. AND SHUKLA, A.K. (2015). Estimation Approach to Ratio of Two Inventory Population Means in Stratified Random Sampling, *American Journal of Operational Research*, 5, 4, 96-101.
- 16) YADAV, S. K; GUPTA, SAT; MISHRA, S. S. AND SHUKLA, A. K. (2016). Modified Ratio and Product Estimators for Estimating Population Mean in Two-Phase Sampling, *American Journal of Operational Research*, 6, 3, 61-68.
- 17) YADAV S.K. AND KADILAR, C. (2013). Improved Class of Ratio and Product Estimators, *Applied Mathematics and Computation*, 219, 10726-10731.
- 18)