RESEARCHERID
THOMSON REUTERS

[Dipika* *et al.*, 6(6): June, 2017]
IC™ Value: 3.00

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

AN EFFICIENT ESTIMATOR FOR ESTIMATING FINITEPOPULATION MEAN USING KNOWN MEDIAN OF THE STUDY VARIABLE

Dharmendra Kumar Yadav[1], Dipika[*2] & Sheela Misra[3]
[1,2&3]Department of Statistics , University of Lucknow, Lucknow-226007, U.P., INDIA

DOI: 10.5281/zenodo.817862

ABSTRACT
The present paper concerns with the estimation of population mean of the study variable by utilizing the known 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 ratioestimator of Cochran (1940), usual regression estimator of Watson (1937) ,Bahl and Tuteja estimator (1991), Kadilar (2016) and Subramani (2016) estimators. Through the numerical study, the theoretical findings are validated and it has been found that proposed estimatorperforms better than theexisting estimators.

KEYWORDS: Study variable,Bias, Ratio estimator, Mean squared error, Simple random sampling, Efficiency.

I. INTRODUCTION
In many practical situations it is seen that population mean of the study variable is not known but the population median of the study variable may be 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 utilized for improved estimation of population mean of study variable. It is a well-known fact in sampling theory that the use of auxiliary information will increase the efficiency of the estimator,but it is collected on additional cost of the survey. Therefore use of median of study variable can be a better option to improve the efficiency without increasing the cost of survey. 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 population of *N* distinct and identifiable units and let $(x_i, y_i), i = 1, 2, ..., n$ 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 problem in a more efficient way, 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 the permission of the author.

Example
The estimation of body mass index (BMI) of the 350 patients in a Hospital based on a small simple random sample without replacement has been considered.

Table1: Body mass index of 350 patients in a hospital

Category	BMI range – kg/m2	Number of patients	Cumulative total
Very severely underweight	less than 15	15	15
Severely underweight	from 15.0 to 16.0	35	50
Underweight	from 16.0 to 18.5	67	117
Normal (healthy weight)	from 18.5 to 25	92	209
Overweight	from 25 to 30	47	256
Obese Class I (Moderately obese)	from 30 to 35	52	308
Obese Class II (Severely obese)	from 35 to 40	27	335
Obese Class III (Very severely obese)	over 40	15	350
	Total	350	350

The median value will be between 18.5 and 25. So one can assume that the population median of the BMI is approximately 21.75.

II. REVIEW OF EXISTING ESTIMATORS

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

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

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

$$V(t_0) = \frac{1-f}{n} S_y^2 = \frac{1-f}{n}\bar{Y}^2 C_y^2 \qquad (2)$$

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

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

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

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

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

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

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

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

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

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

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

$Cov(x,y) = \frac{1}{N-1}\sum_{i-1}^{N}(Y_i - \bar{Y})(X_i - \bar{X})$, 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,

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

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,

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

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

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

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

where δ 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,

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

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

The optimum value of the characterizing scalar δ 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,

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

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,

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

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,

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

$$MSE(t_5) = \frac{1-f}{n}\bar{Y}^2[C_y^2 + R_5^2C_m^2 - 2R_5C_{ym}], \qquad (13)$$

where

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

$C_{ym} = \dfrac{S_{ym}}{\bar{Y} M}$.

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

III. PROPOSED ESTIMATOR

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

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

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

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

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

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

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

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

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

$$= \bar{Y}(1+e_0)\left[\alpha + (1-\alpha)\exp\left(\frac{-e_1}{2+e_1}\right)\right]$$

$$= \bar{Y}(1+e_0)\left[\alpha + (1-\alpha)\exp\left(\frac{-\theta}{1+\theta}\right)\right], \text{ where } \theta = \frac{e_1}{2}$$

$$\bar{Y}(1+e_0)\left[\alpha + (1-\alpha)\{1 + \theta(-1+\theta-\theta^2) + \frac{\theta^2(-1+\theta-\theta)^2}{2} + \}\right]$$

$$t - \bar{Y} = \bar{Y}\left[\frac{e_1^2}{4} + \frac{1}{2}e_1 + \frac{e_1^2}{8} + \alpha\frac{e_1^2}{4} + \alpha\frac{e_1}{2} - \alpha\frac{e_1^2}{8} + e_0 - \frac{e_0 e_1}{2} + \frac{\alpha e_0 e_1}{2}\right] \quad (15)$$

Taking expectation on both sides we get Bias(t)

$$E(t - \bar{Y}) = \bar{Y}\left[\frac{E(e_1^2)}{4} + \frac{1}{2}E(e_1) + \frac{E(e_1^2)}{8} + \alpha\frac{E(e_1^2)}{4} + \alpha\frac{E(e_1)}{2} - \alpha\frac{E(e_1^2)}{8} + E(e_0) - \frac{E(e_0 e_1)}{2} + \frac{\alpha E(e_0 e_1)}{2}\right]$$

$$Bias(t) = \bar{Y}\left[\frac{3\lambda C_m^2}{8} - \alpha\frac{\lambda C_m^2}{8} + \frac{\lambda C_{ym}}{2}(\alpha - 1) + \frac{Bias(m)}{2M}(\alpha + 1)\right]$$

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

$$MSE(t) = E(t - \bar{Y})^2 = E\left[\bar{Y}\left\{\alpha\frac{e_1}{2} - \frac{e_1}{2} + e_0\right\}\right]^2$$

$$= \bar{Y}^2 E\left[e_0 + \frac{e_1}{2}(\alpha - 1)\right]^2$$

$$= \bar{Y}^2 \left[E(e_0^2) + \frac{E(e_1^2)}{4}(\alpha-1)^2 + E(e_0e_1)(\alpha-1) \right]$$

$$= \bar{Y}^2 \left[\lambda C_y^2 + \frac{\lambda C_m^2}{4}(\alpha-1)^2 + (\alpha-1)\lambda C_{ym} \right]$$

$$MSE(t) = \lambda \bar{Y}^2 \left[C_y^2 + \frac{(\alpha-1)^2}{4} C_m^2 + (\alpha-1)C_{ym} \right] (16)$$

which is minimum for,

$$\alpha_{opt} = 1 - 2\frac{C_{ym}}{C_m^2}$$

and the minimum mean squared error of the proposed estimator t is,

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

IV. 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,

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

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

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

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

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,

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

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,

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

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,

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

Under the above condition, the proposed estimator is better than the Kadilar (2016) estimator.

From equation (17) and equation (13), we have,

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

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

V. 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
ρ_{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
$\boldsymbol{t_5}$	10926.53	10926.53	1.09
t	9002.22	9002.22	0.98

VI. 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 by the survey practitioners for improved estimation of population mean under simple random sampling scheme.

VII. REFERENCES

[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., KADILAR C.(2013): Improved Exponential Type ratio estimator of population variance. Rev. Colom. Estad.,36, 145-152

CITE AN ARTICLE

Yadav, D. K., D., & Misra, S. (2017). AN EFFICIENT ESTIMATOR FOR ESTIMATING FINITEPOPULATION MEAN USING KNOWN MEDIAN OF THE STUDY VARIABLE. *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY*, *6*(6), 503-509. doi:10.5281/zenodo.817862