Jñānābha, Vol. 48 (1) (2018), 17-25

NEW EFFICIENT PREDICTIVE ESTIMATOR OF POPULATION MEAN USING MEDIAN OF STUDY VARIABLE

By
S. K. Yadav
Department of Mathematics and Statistics,
Babasaheb Bhimrao Ambedkar University,
Lucknow-226025, Uttar Pradesh, India
drskystats@gmail.com
S. S. Mishra
Department of Mathematics and Statistics (A Centre of Excellence)
Dr RML Avadh University, Faizabad-224001, Uttar Pradesh, India
profsantmishra@gmail.com
Surendra Kumar
Department of Mathematics,
Government Degree College Pihani, Hardoi-241406, Uttar Pradesh, India
surendra.kumar776@gmail.com
Ravendra Kumar
Department of Mathematics,
Dr. RML Government Degree College, Aonla, Bareilly, Uttar Pradesh, India
rkpmbly@gmail.com
(*Received: September* 09,2017 ; *Revised: April* 02,2018)

Abstract

In the present paper, a new generalized predictive estimator of population mean has been proposed which makes use of information on the median of the study variable. The proposed estimator is a ratio cum product type estimator. The properties of the proposed estimator such as bias and mean squared error have been studied up to the first order of approximation. The optimal value of the characterizing scalar which makes the ratio cum product estimator better is obtained. This optimum value is responsible for the minimum value of the proposed estimator and it has also been obtained. The proposed estimator is compared with the competing estimators and the conditions for its better performance have also been given. A numerical study is also made to verify the theoretical findings. Results shows that proposed estimator is better than the competing estimators of population mean.
2010 Mathematics Subject Classification. 62D05
Keywords and phrases. Predictive estimator, Study variable, Bias, Mean squared error, Efficiency.

1 Introduction

It is always good to calculate the parameters under consideration rather than to estimate them. Whenever the population is very large, it is very time taking and costly to get the information on every unit of the population and calculate the parameters under study. The best alternative of complete enumeration is the sampling. The thing which is to be cared of is that the sample should be well representative of the population under consideration. It is well established in sampling theory that the most suitable estimator of any parameter is the corresponding statistic. Thus the most appropriate estimator for population mean is the corresponding sample mean. The estimator of population mean is best if it has almost all desirable properties of good estimator that is unbiasedness, consistency, sufficiency and efficiency. The most important property is that the estimator should be most efficient equivalently have minimum variance that is its sampling distribution should be more and more concentrated around the true parameter under consideration. The biased estimator which has minimum mean squared error is always better than the unbiased estimator with larger variance. This is the case with the sample mean. Thus we search for even biased estimator but with least mean squared error.

The above purpose of searching more efficient estimator than sample mean of study variable is achieved through the use of auxiliary information. The auxiliary information is gathered from auxiliary variable. Auxiliary variable is the variable about which the interviewer has full information. The auxiliary variable is highly correlated may be positive or negative with the main variable under study. Ratio and product estimators are used respectively when auxiliary and study variables are highly positively and negatively correlated along with the condition that the line of regression, y on x passes through origin. In either case regression estimator is used for improved estimation of population mean. For detailed study of improved estimators of population mean using auxiliary information, the latest references can be made of Tailor and Sharma [8], Yan and Tian [12], Subramani [4], Subramani and Kumarapandiyan [5, 6], Yadav et al. [9, 10], Yadav et al. [11], and Abid et al. [1].

The auxiliary information is used at various stages of designing and estimation etc. and it is collected on some additional cost of the survey. This is the drawback with the auxiliary information. Thus it is best if we use the information on same known parameter of study variable and the efficiency of the estimator is increased. Here we use information on such parameter of study variable which does not require information on every unit of the population under consideration. It is median of the study variable which does not require information on all units of the population like other parameters do. There are so many examples in the literature, latest reference can be made of Subramani [7].

There are number of approaches in the literature which makes use of auxiliary information to construct more efficient estimators based on design and model based methods. When auxiliary information is used to construct more efficient estimators at estimation stage for model based method, it is known as predictive method of estimation. This method of estimation is based on super population models which is realization of super population random variables.

Under this approach, the prior information on the population is formalized and it is being utilized for the prediction of values of the population which are not sampled. Thus the finite population quantities such as mean and other parameters of the main variable under consideration are estimated. The benefit of this approach is that the predictive estimation theory makes use of various estimators as predictors for population parameters of the unobserved units of the population. Using this predictor, the estimators of the population parameters for the whole population are constructed.

In this paper a new efficient predictive estimator of population mean using median of study variable has been constructed and its sampling properties have been studied up to the approximation of order one.

Let the population under consideration is constituted of N distinct and identifiable units $U_1, U_2, ..., U_N$ and the main and auxiliary variables respectively are Y and X. Let (Y, X) be the bivariate distribution of the random variables Y and X. A random sample of size n from this bivariate distribution is derived using simple random sampling without replacement technique. Here the parameter to be estimated is the population mean of the study variable denoted by

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

Further let S be the set of all possible ${}^{N}C_n$ sample of size n from the population under consideration of size N. Let s be the element of the set S and $\vartheta(s)$ represents the effective sample size having distinct elements in S and $\bar{S}$ be the set of all units of the population which are not in S.

Now we have the following notations:

$$\bar{Y}_s = \frac{1}{\vartheta(s)}\sum_{i\epsilon s} y_i,$$

$$\bar{Y}_{\bar{s}} = \frac{1}{[N - \vartheta(s)]}\sum_{i\epsilon \bar{s}} y_i,$$

Thus for a given sample $s\epsilon S$, we can write population mean of study variable as,

$$\bar{Y} = \frac{\vartheta(s)}{N}\bar{Y}_s + \frac{[N - \vartheta(s)]}{N}\bar{Y}_{\bar{s}}, \tag{1.1}$$

The sample mean from the sample of size n that is $\vartheta(s) = n$ drawn form the population of size N using simple random sampling technique is given by,

$$\bar{y} = \frac{1}{n}\sum_{i\epsilon s} y_i,$$

or equivalently

$$\bar{Y}_s = \bar{y}.$$

Thus equation (1.1) can be rewritten as,

$$\bar{Y} = \frac{n}{N}\bar{Y}_s + \frac{N - n}{N}\bar{Y}_{\bar{s}}. \tag{1.2}$$

Now the most suitable estimator of population mean $\bar{Y}$ in the view of equation is,

$$t = \frac{n}{N}\bar{y} + \frac{N - n}{N}T, \tag{1.3}$$

where T is considered as the predictor of $\bar{Y}_{\bar{s}}$.

Srivastava [3] proposed the following usual and ratio estimators of population mean of unobserved units of the population, known as predictor of $\bar{Y}_{\bar{s}}$ as, mean per unit estimator,

$$T = \bar{y} = \frac{1}{n}\sum_{i\epsilon\bar{s}} y_i,$$

usual ratio estimator,

$$T = \bar{y}_{\bar{R}} = \bar{X}_{\bar{s}}\left\{\frac{\bar{y}}{\bar{x}}\right\},$$

where,

$$\bar{x} = \frac{1}{n}\sum_{i\epsilon s} x_i, \bar{X}_{\bar{s}} = \frac{1}{N-n}\sum_{i\epsilon\bar{s}} x_i = \frac{N\bar{X}-n\bar{x}}{N-n}, \bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i.$$

The above estimators have been used as predictive estimators of $\bar{Y}_{\bar{s}}$ by Srivastava [3]. Using above estimators as predictive estimators, the estimators of population mean $\bar{Y}$ in equation (1.3) may be written respectively as,
the sample mean or mean per unit estimator,

$$\bar{y} = \frac{1}{n}\sum_{i\epsilon\bar{s}} y_i,$$

the usual ratio estimator,

$$\bar{y}_R = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right).$$

However in case of product method of estimation of population mean $\bar{Y}_{\bar{s}}$, Srivastava [3] has shown that using usual product predictive estimator $T = \bar{y}_{\bar{P}} = \bar{y}\left(\frac{\bar{x}}{\bar{X}_{\bar{s}}}\right)$, the estimator t in equation (1.3) does not results in usual product estimator $\bar{y}_P = \bar{y}\left(\frac{\bar{x}}{\bar{X}}\right)$ of population mean $\bar{Y}$ of study variable.

Thus in the light of product predictive estimator $T = \bar{y}_{\bar{P}} = \bar{y}\left(\frac{\bar{x}}{\bar{X}_{\bar{s}}}\right)$, the estimator of population mean $\bar{Y}$ is,

$$t = \frac{n\bar{X} + (N-2n)\,\bar{x}}{N\bar{X} - n\bar{x}} = t_P. \tag{1.4}$$

Upto the approximation of degree one, the biases and mean squared errors of the estimators $\bar{y}_R, \bar{y}_P$ and t_p respectively are,

$$B(\bar{y}_R) = \theta\bar{Y}C_x^2(1-C),$$

$$B(\bar{y}_P) = \theta\bar{Y}CC_x^2,$$

$$B(t_P) = \theta\bar{Y}C_x^2[C + f(1-f)^{-1}],$$

$$MSE(\bar{y}_R) = \theta\bar{Y}^2[C_y^2 + C_x^2(1-2C)], \tag{1.5}$$

$$MSE(\bar{y}_P) = MSE(t_P) = \theta\bar{Y}^2[C_y^2 + C_x^2(1+2C)], \tag{1.6}$$

where,

$$\theta = \frac{(1-f)}{n}, f = (n/N), C_y^2 = (S_y^2/\bar{Y}^2), C_x^2 = (S_x^2/\bar{X}^2), C = \rho(C_y/C_x), \rho = (S_{yx}/S_yS_x),$$

$$S_y^2 = (N-1)^{-1}\sum_{i=1}^{N}(y_i-\bar{Y})^2, S_x^2 = (N-1)^{-1}\sum_{i=1}^{N}(x_i-\bar{X})^2, S_{yx} = (N-1)^{-1}\sum_{i=1}^{N}(y_i-\bar{Y})(x_i-\bar{X}).$$

Singh et al. [2] proposed the following ratio and product type estimators of population mean $\bar{Y}$ under predictive modeling approach respectively as,

$$t_{Re} = \left[\frac{n}{N}\bar{y} + \left(\frac{N-n}{N}\right)\bar{y}exp\left(\frac{\bar{X}_{\bar{s}}-\bar{x}}{\bar{X}_{\bar{s}}+\bar{x}}\right)\right] = \left[\frac{n}{N}\bar{y} + \left(\frac{N-n}{N}\right)\bar{y}exp\left(\frac{N(\bar{X}-\bar{x})}{N(\bar{X}-\bar{x})-2n\bar{x}}\right)\right] \quad (1.7)$$

$$t_{Pe} = \left[\frac{n}{N}\bar{y} + \left(\frac{N-n}{N}\right)\bar{y}exp\left(\frac{\bar{x}-\bar{X}_{\bar{s}}}{\bar{x}+\bar{X}_{\bar{s}}}\right)\right] = \left[\frac{n}{N}\bar{y} + \left(\frac{N-n}{N}\right)\bar{y}exp\left(\frac{N(\bar{x}-\bar{X})}{N\bar{X}+(N-2n)\bar{x}}\right)\right] \quad (1.8)$$

Upto the approximation of degree one, the biases and mean squared errors of the estimators t_{Re} and t_{Pe} respectively are,

$$B(t_{Re}) = \frac{\theta}{8}\bar{Y}C_x^2\left[3-4(C+f)\right],$$

$$B(t_{Pe}) = \frac{\theta}{8}\bar{Y}C_x^2\left[4C - \frac{1}{(1-f)}\right],$$

$$MSE(t_{Re}) = \theta\bar{Y}^2\left[C_y^2 + \frac{C_x^2}{4}(1-4C)\right], \quad (1.9)$$

$$MSE(t_{Pe}) = \theta\bar{Y}^2\left[C_y^2 + \frac{C_x^2}{4}(1+4C)\right]. \quad (1.10)$$

2 Proposed Estimator

Motivated by Singh et al. [2] and Subramani [7], we have proposed the following ratio and product estimators of population mean under predictive modeling approach respectively as,

$$t_{RM} = \left[\frac{n}{N}\bar{y} + \left(\frac{N-n}{N}\right)\bar{y}exp\left(\frac{N(M-m)}{N(M-m)-2nm}\right)\right], \quad (2.1)$$

$$t_{PM} = \left[\frac{n}{N}\bar{y} + \left(\frac{N-n}{N}\right)\bar{y}exp\left(\frac{N(m-M)}{NM+(N-2n)m}\right)\right]. \quad (2.2)$$

To study the sampling properties of the proposed estimators, following approximations have been made as,
$\bar{y} = \bar{Y}(1+e_0)$ and $m = M(1+e_1)$ such that $E(e_0) = 0, E(e_1) = \frac{\bar{M}-M}{M} = \frac{Bais(m)}{M}$ and $E(e_0^2) = \frac{1-f}{n}C_y^2, E(e_1^2) = \frac{1-f}{n}C_m^2, E(e_0e_1) = \frac{1-f}{n}C_{ym}$,
where,

$$\bar{M} = \frac{1}{n}\sum_{i=1}^{n}m_i.$$

In the light of above approximations and results, the biases and mean squared errors of above estimators t_{RM} and t_{PM} up to the first order of approximation respectively are,

$$B(t_{RM}) = -\frac{1}{2f}\bar{Y}\left[\frac{B(m)}{M} + \frac{(1-f)}{n}\left\{C_m^2 + C_{ym}\right\}\right],$$

$$B(t_{PM}) = \frac{1}{2(1-f)}\bar{Y}\left[\frac{B(m)}{M} + \frac{(1-f)}{n}\left\{\frac{C_m^2}{2(1-f)} + C_{ym}\right\}\right],$$

$$MSE(t_{RM}) = \theta\bar{Y}^2\left[C_y^2 + C_m^2/4 - C_{ym}\right], \quad (2.3)$$

$$MSE(t_{PM}) = \theta\bar{Y}^2\left[C_y^2 + C_m^2/4 + C_{ym}\right]. \quad (2.4)$$

3 Efficiency Comparison

The proposed estimator in equation (2.3) is better than mean per unit estimator if,

$$V(t_0) - MSE(t_{RM}) = \theta\bar{Y}^2\left[C_m^2/4 - C_{ym}\right] > 0,$$

or,

$$C_{ym} > C_m^2/4. \quad (3.1)$$

Thus the proposed estimator is better than mean per unit estimator under the above condition. From equations (1.5) and (2.3), we can see that the proposed ratio estimator performs better than the usual ratio estimator of population mean if,

$$V(t_R) - MSE(t_{RM}) = \theta\bar{Y}^2\left[C_x^2 - 2C_{yx} - C_m^2/4 + C_{ym}\right] > 0,$$

or,

$$\left(C_x^2 - 2C_{yx}\right) > \left(C_m^2/4 - C_{ym}\right). \quad (3.2)$$

When above condition is fulfilled, the proposed estimator is better than the usual ratio estimator.

The proposed estimator in euation (2.3) is better than usual product estimator if,

$$V(t_P) - MSE(t_{RM}) = \theta\bar{Y}^2\left[C_x^2 + 2C_{yx} - C_m^2/4 - C_{ym}\right] > 0,$$

or,

$$\left(C_x^2 + 2C_{yx}\right) > \left(C_m^2/4 + C_{ym}\right). \quad (3.3)$$

Under the above condition, proposed estimator is better than the usual product estimator of population mean.

From equations (1.9) and (2.3), we can see that the proposed ratio estimator performs better than the exponential ratio type estimator of population mean if,

$$V(t_{Re}) - MSE(t_{RM}) = \theta\bar{Y}^2\left[C_x^2/4 - C_{yx} - C_m^2/4 + C_{ym}\right] > 0,$$

or,

$$\left(C_x^2 - C_m^2\right) > 4\left(C_{yx} - C_{ym}\right). \quad (3.4)$$

When above condition is fulfilled, proposed estimator is better than the usual exponential ratio estimator of population mean.

The proposed estimator in equation ((2.3)) is better than usual product exponential type estimator if,

$$V(t_{Pe}) - MSE(t_{RM}) = \theta\bar{Y}^2\left[C_x^2/4 + C_{yx} - C_m^2/4 - C_{ym}\right] > 0,$$

or,

$$\left(C_x^2 - C_m^2\right) > 4\left(C_{yx} - C_{ym}\right). \tag{3.5}$$

Under the above condition, proposed estimator is better than the usual exponential product estimator of population.

Note: similar comparison can be made for proposed product type estimator of population mean and the conditions under which it performs better than the competing estimators of population mean.

4 Numerical Study

To verify the theoretical findings of the proposed and competing estimators, the following numerical study is carried out. For this study we have used the population given by Subramani [3]. The following parameters of the population for both study and auxiliary variable have been calculated and are presented in Table 1, Table 2, and Table 3 represents the biases, mean squared errors and efficiencies of proposed and competing estimators respectively.

Note: All the calculations in Table 1, Table 2 and Table 3 representing parameters of two natural populations, variance and mean squared errors of various estimators and percentage relative efficiencies of various estimator along with proposed one of population mean have been calculated using MS-EXCEL.

Table 1: Various Parameters of three natural populations

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

*Since both the populations are positively correlated so these are not applicable on product type estimators.

Table 2: Variance / Mean squared error of the existing and proposed estimators

Estimator	Population-1	Population-2
t_0	15640.97	2.15
$\bar{y}_R$	14895.27	1.48
$\bar{y}_P, t_P$	38547.52*	6.91*
t_{Re}	12498.01	1.30
t_{Pe}	24324.14*	4.02*
t_{RM}	9644.07	1.19
t_{PM}	27946.29*	3.93*

Table 3: PRE of the proposed and competing estimators with respect to t_0

Estimator	Population-1	Population-2
t_0	100.00	100.00
$\bar{y}_R$	105.00	145.27
$\bar{y}_P, t_P$	40.57	31.11
t_{Re}	125.15	165.38
t_{Pe}	64.30	53.48
t_{RM}	162.18	180.67
t_{PM}	55.96	54.71

5 Results and Conclusion

The present manuscript deals with the estimation of population mean using median of study variable under predictive modeling approach. The bias and MSE of proposed estimator are obtained up to the approximation of degree one. The proposed estimator is compared with other competing estimators of population mean. A numerical study is also done to verify the theoretical findings. From Table 2 and Table 3, we see that the proposed estimator is most efficient among all estimators of population mean under simple random sampling scheme as it has least mean squared errors and highest percentage relative efficiencies for both the populations.

Acknowledgement

The authors are very much thankful to the editors of Jnanabha and anonymous referee for critically examining the manuscript and giving excellent valuable comments which made significant improvement in the earlier draft.

References

[1] M. Abid, N. Abbas, R.A.K. Sherwani and H. Z. Nazir, Improved Ratio Estimators for the population mean using non-conventional measure of dispersion, *Pakistan Journal of Statistics and Operations Research*, **XII (2)** (2016), 353-367.

[2] H. P. Singh, R. S. Solanki and A. K. Singh, Predictive Estimation of Finite Population Mean Using Exponential Estimators, *Statistika. Statistics and Economy Journal*, **94 (1)**(2014), 41-53.

[3] S. K. Srivastava, Predictive estimation of finite population mean using product estimator, *Metrika*,**30** (1983), 93-99.

[4] J. Subramani, Generalized modified ratio estimator of finite population mean, *Journal of Modern Applied Statistical Methods*, **12 (2)** (2013), 121-155.

[5] J. Subramani and G. Kumarapandian, Estimation of population mean using coefficient of variation and median of an auxiliary variable, *International Journal of Probability and Statistics*,**1 (4)** (2012), 111-118.

[6] J. Subramani and G. Kumarapandian, Estimation of population mean using deciles of an auxiliary variable, *Statistics in Transition-New Series*, **14(1)** (2013), 75-88.

[7] J. Subramani, A new median based ratio estimator for estimation of the finite population mean, *Statistics in Transition New Series*, **17, 4** (2016), 1-14.

[8] R.Tailor and B. Sharma, 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)** (2009), 15-24.

[9] S.K. Yadav, S.S. Mishra and A.K. Shukla, 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**(2014), 21-27.

[10] S.K. Yadav, S.S. Mishra and A.K. Shukla, Estimation Approach to Ratio of Two Inventory Population Means in Stratified Random Sampling, *American Journal of Operational Research*, **5, 4** (2015), 96-101

[11] S.K. Yadav, S. Misra, S.S. Mishra and N. Chutiman, Improved ratio estimators of population mean in Adaptive Cluster Sampling, *Journal of Statistics Applications and Probability Letter*, **3, 1** (2016), 1-6.

[12] Z. Yan and B. Tian, Ratio method to the mean estimation using coefficient of skewness of auxiliary variable. *ICICA, PartII, CCIS*, **106** (2010), 103-110.