ESTIMATION OF FINITE POPULATION MEAN USING LINEAR COMBINATION OF TRI-MEAN AND QUARTILE AVERAGE OF AUXILIARY VARIABLE UNDER PREDICTIVE MODELING APPROACH

Dharmendra K. Yadav, Department of Commercial Tax, Government of U.P., India
(dkumar.yadava@gmail.com)
Dinesh K. Sharma, University of Maryland, Eastern Shore, USA
(dksharma@umes.edu)

ABSTRACT

The present article addresses the problem of estimating the finite population mean under predictive modeling approach. Ratio type predictive estimator of finite population mean using Tri-Mean and Quartile Average of the auxiliary variable is proposed for this purpose. The asymptotic expressions of bias and MSE are also obtained. Theoretical efficiency comparison of the proposed estimator with Bahl and Tuteja estimator (1991) and Singh et al. estimator (2014) has also been made. Theoretical results are also supported by numerical illustration.

Key words: Tri Mean, Quartile Average, Bias, MSE, Predictive Modeling.

INTRODUCTION

With the increasing growth in the number and diverse uses of sample surveys worldwide, it is often desired to analyze and interpret the resulting voluminous data by swifter methods (Cochran (1940)). An essential requirement of a good survey is that a measure of precision is provided for each estimate computed from survey data collected based on the survey design. Like mean, median and mode, Tri-mean is also an essential measure of location. The Tri-mean takes not only the central tendency into account but also provides an idea about the distribution of data. Since it is not very much affected by outliers, Tri-mean is considered 'resistant' or 'robust.'

In the predictive approach, a model is specified for the population values and is used to predict the non-sampled values. Prediction theory for sample surveys (or model-based theory) can be considered as a general framework for statistical inferences on the character of a finite population. Srivastava (1983) has shown that if the usual product estimator is used as a predictor for the mean of the unobserved units of the population, the resulting estimator of the mean of the whole population is different from the customary (usual) product estimator. Agrawal and Roy (1999) and Nayak and Sahoo (2012) provided some predictive estimators for finite population variance. Sahoo and Panda (1999) developed the regression type estimator for a two-stage sampling procedure. Sahoo and Sahoo (2001) and Sahoo et al. (2009) introduced a class of estimators for the finite population mean availing information on two auxiliary variables in two-stage sampling. Yadav and Mishra (2015) also proposed an improved predictive estimator of finite population mean using a linear combination of Singh et al. (2014) estimators. Yadav et al. (2017) proposed an estimator of population variance by using a linear combination of Tri-mean and Inter-quartile range of the auxiliary variable. In the present article, we have proposed a new modified ratio-type predictive estimator of the population mean of the auxiliary variable by using a linear combination of Tri-mean and population semi interquartile average of the auxiliary variable. It is highly sensitive to outliers as its design structure is based on only extreme values of the data (for more details, see Ferrell (1953).

NOTATIONS

Let N= Population size, n = sample size,

Y= study variable, X = auxiliary variable

$\bar{X}$, $\bar{Y}$ = population means, $\bar{x}$, $\bar{y}$ = sample means, C_X , C_Y = Coefficient of varation

ρ = Correlation Coefficient,

Tri-mean: $TM = \frac{Q_1+2Q_2+Q_3}{4}$

Population Semi quartile average of X : $Q_a = \frac{Q_3+Q_1}{2}$

$\theta = \frac{\bar{X}}{\bar{X}+H}$, $H = TM + Q_a$

$f_1 = \frac{f}{1-f}$, $f = \frac{n}{N}$

SOME EXISTING ESTIMATORS IN LITERATURE

Some existing predictive estimators of population mean are given in the given table

Table 1: Review of some existing estimators

S.No	Author	Estimator	Bias	MSE
I	Bahl and Tuteja (1991)	$t_1 = \bar{y}\exp\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]$	$B(t_1) = \frac{1-f}{8n}\bar{Y}[3C_x^2 - 4C_{yx}]$	$MSE(t_1) = \lambda\bar{Y}^2[C_y^2 + \frac{C_x^2}{4} - C_{yx}]$
II	Singh et al.(2014)	$t_{Re} = \left[\frac{n}{N}\bar{y} + \left(\frac{N-n}{N}\right)\bar{y}\exp\left(\frac{\bar{X}_{\bar{s}} - \bar{x}}{\bar{X}_{\bar{s}} + \bar{x}}\right)\right]$	$Bias(t_{Re}) = \frac{\phi}{8}\bar{Y}C_x^2[3 - 4(C+f)]$	$MSE(t_{Re}) = \phi\bar{Y}^2\left[C_y^2 + \frac{C_x^2}{4}(1-4c)\right]$
III	Singh et al.(2014)	$t_{pe} = \left[\frac{n}{N}\bar{y} + \left(\frac{N-n}{N}\right)\bar{y}\exp\left(\frac{\bar{x} - \bar{X}_{\bar{s}}}{\bar{x} + \bar{X}_{\bar{s}}}\right)\right]$	$Bias(t_{pe}) = \frac{\phi}{8}\bar{Y}C_x^2\left[4C - \frac{1}{(1-f)}\right]$	$MSE(t_{pe}) = \phi\bar{Y}^2\left[C_y^2 + \frac{C_x^2}{4}(1+4c)\right]$

ESTIMATION UNDER PREDICTIVE MODELLING APPROACH

Let Y_i $(i = 1,2,\ldots,N)$ be the real value taken by the variable under study from the finite population of U of size N. Here, the population parameter to be estimated is the population mean on the basis of observed values of y in an ordered sample of the finite population U of size N. Let S denote the collection of all possible samples from the finite population U. Let w(s) denote the effective sample size, for any given $s \in S$ and $\bar{s}$ denote the collection of all those units of U which are not in S.

We now denote:

$$\bar{y}_s = \frac{1}{w(s)}\sum_{i\in s} y_i$$

$$\bar{y}_{\bar{s}} = \frac{1}{N - w(s)}\sum_{i\in\bar{s}} y_i$$

We have,

$$\bar{Y} = \frac{w(s)}{N}\bar{y}_s + \frac{N - w(s)}{N}\bar{y}_{\bar{s}}$$

Basu (1971), asserted that in the representation of $\bar{Y}$ above the sample mean $\bar{y}_s$ being based on the observed y values on units in the samples is known, therefore the statistician should attempt a prediction of the mean $\bar{y}_{\bar{s}}$ of the unobserved units of the population on the basis of observed units in s.

For $s \in S$ under simple random sampling without replacement (SRSWOR) with sample size $w(s) = n$ and $\bar{y}_s = \bar{y}$, the population mean $\bar{Y}$ is given by

$$\overline{Y} = \frac{n}{N}\bar{y}_s + \frac{(N-n)}{N}\bar{y}_{\bar{s}} \quad (1)$$

In view of equation (1) above, an appropriate estimator of population mean $\overline{Y}$ is obtained as

$$t = \frac{n}{N}\bar{y} + \frac{(N-n)}{N}\,T$$

where T is taken as the predictor of $\bar{y}_{\bar{s}}$.

Let x_i $(i = 1,2,\dots,N)$ denote the i^{th}observation of the auxiliary variable x and $X_i(i = 1,2,\dots,N)$ be the values of x on the i^{th} unit of the population U. Auxiliary variable x is correlated with the variable under study y.
Let

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

and

$$\bar{x} = \frac{1}{N}\sum_{i\in s} x_i$$

$$t = \frac{n}{N}\bar{y} + \frac{(N-n)}{N}T \quad (2)$$

PROPOSED PREDICTIVE ESTIMATOR

Maqbool and Javaid (2017) proposed the following ratio type estimator for estimating population mean using linear combination of trimean and semi interquartile average of auxiliary variable X.

$$t_{sj} = \bar{y}\left[\frac{\overline{X}+(TM+Q_a)}{\bar{x}+(TM+Q_a)}\right] \quad (3)$$

where TM and Q_a are the tri-mean and quartile average respectively.

Motivated by Maqbool and Javaid (2017), we have used (3) as predictor for estimation of population mean under predictive modelling approach.

Therefore, on taking $T = \bar{y}\left[\frac{\overline{X}+(TM+Q_a)}{\bar{x}+(TM+Q_a)}\right]$ in (2), the proposed predictive estimator will be

$$t = \frac{n}{N}\bar{y} + \frac{(N-n)}{N}\bar{y}\left[\frac{\overline{X}_{\bar{s}} + (TM + Q_a)}{\bar{x} + (TM + Q_a)}\right]$$

Using the transformation $\overline{X}_{\bar{s}} = \frac{N\overline{X}-n\bar{x}}{N-n}$, we have

$$t = f\,\bar{y} + (1-f\,)\bar{y}\left[\frac{\frac{N\overline{X}-n\bar{x}}{N-n} + H}{\bar{x} + H}\right]$$

where $H = (TM + Q_a)$

$$t = f\,\bar{y} + (1-f\,)t_1 \quad (4)$$

where $t_1 = \bar{y}\left[\frac{\frac{N\overline{X}-n\bar{x}}{N-n}+H}{\bar{x}+H}\right]$

We are using the following approximations

$$\bar{y} = \overline{Y}(1 + e_0), \bar{x} = \overline{X}(1 + e_1), E(e_0) = E(e_1) = 0, E(e_0^2) = \lambda C_y^2, E(e_1^2) = \lambda C_x^2, E(e_0e_1) = \lambda C_{yx}$$

Now,

$$t_1 = \bar{y}\left[\frac{\frac{N\bar{X}-n\bar{x}}{N-n}+H}{\bar{x}+H}\right]$$

$$= \bar{Y}(1+e_0)\left[\frac{\frac{N\bar{X}-n\bar{X}(1+e_1)}{N-n}+H}{\bar{X}(1+e_1)+H}\right]$$

$$= \bar{Y}(1+e_0)\left[\frac{(\bar{X}+H)-\left(\frac{f}{1-f}\right)\bar{X}e_1}{(\bar{X}+H)+\bar{x}e_1}\right], \text{ where } \left(\frac{f}{1-f}\right)=\frac{n}{N-n}$$

$$= \bar{Y}(1+e_0)\left[\frac{1-f_1\theta e_1}{1+\bar{X}e_1}\right], \quad \text{where} \quad f_1 = \left(\frac{f}{1-f}\right), \theta = \left(\frac{\bar{X}}{\bar{X}+H}\right)$$

$$= \bar{Y}(1+e_0)(1-f_1\theta e_1)(1+\bar{X}e_1)^{-1}$$

$$t_1 = \bar{Y}(1+e_0)[1-\bar{X}e_1+\bar{X}^2e_1^2-f_1\theta e_1+f_1\theta\bar{X}e_1^2] \tag{5}$$

On substituting the value of t_1 from (5) to (4), we have

$$t-\bar{Y} = \bar{Y}[e_0-\bar{X}e_1-f_1\theta e_1-f\bar{X}e_1+ff_1\theta e_1+\bar{X}^2e_1^2+f_1\theta\bar{X}e_1^2-f\bar{X}e_1^2+ff_1\theta\bar{X}e_1^2-\bar{X}e_0e_1-f_1\theta e_0e_1-f\bar{X}e_0e_1+ff_1\theta e_0e_1] \tag{6}$$

taking expectation both sides in (6), we get the Bias(t),

$$\text{Bias}(t) = \lambda\bar{Y}\,[2\bar{X}^2C_X^2+f_1\theta\bar{X}C_X^2-f\bar{X}C_X^2+ff_1\theta\bar{X}C_X^2-\bar{X}C_{YX}-f_1\theta C_{YX}-f\bar{X}C_{YX}+ff_1C_{YX}]$$

Now squaring (6) and taking expectation on both sides we get the MSE (t) up to the first order of approximations.

$$\text{MSE}(t) = E(t-\bar{Y})^2$$

$$\text{MSE}(t) = \lambda\bar{Y}^2[C_Y^2+\{\bar{X}^2(1+f)^2+2\bar{X}(f^2f_1\theta+ff_1\theta+\theta ff_1+\theta f_1)+\theta f_1^2+2\theta^2ff_1^2\}C_X^2-2C_{YX}(\bar{X}+\theta f_1+f\bar{X}+ff_1\theta)] \tag{7}$$

THEORETICAL EFFICIENCY COMPARISON

In this section, we have made a theoretical efficiency comparison of the proposed estimator with Bahl and Tuteja (1991) estimator and Singh et al (2014) estimators.

1. The proposed estimator will be more efficient than the Bahl and Tuteja (1991) estimator, if

$$\text{MSE}(t) < MSE\ (t_1)$$

$$[\{\bar{X}^2(1+f)^2+2\bar{X}(f^2f_1\theta+ff_1\theta+\theta ff_1+\theta f_1)+\theta f_1^2+2\theta^2ff_1^2\}C_X^2-2C_{YX}(\bar{X}+\theta f_1+f\bar{X}+ff_1\theta)] < \frac{C_x^2}{4}-C_{yx}$$

2. The proposed estimator will be more efficient than the Singh et al.(2014) estimator, if

(a) $\text{MSE}(t) < MSE\ (t_{Re})$

$$\lambda[C_Y^2+\{\overline{X}^2(1+f)^2+2\overline{X}(f^2f_1\theta+ff_1\theta+\theta ff_1+\theta f_1)+\theta f_1^2+2\theta^2 ff_1^2\}C_X^2-2C_{YX}(\overline{X}+\theta f_1+f\overline{X}+ff_1\theta)] < \phi\left[C_y^2+\frac{C_x^2}{4}(1-4c)\right]$$

(b) $\mathrm{MSE}(t) < MSE\ (t_{pe})$

$$\lambda[C_Y^2+\{\overline{X}^2(1+f)^2+2\overline{X}(f^2f_1\theta+ff_1\theta+\theta ff_1+\theta f_1)+\theta f_1^2+2\theta^2 ff_1^2\}C_X^2-2C_{YX}(\overline{X}+\theta f_1+f\overline{X}+ff_1\theta)] < MSE(t_{pe})=\phi\overline{Y}^2\left[C_y^2+\frac{C_x^2}{4}(1+4c)\right]$$

NUMERICAL ILLUSTRATION

The performance of the proposed estimator is assessed with that of existing estimators. We use the data of Murthy (1967) page 228 in which fixed capital is denoted by X (auxiliary variable) and output of 80 factories are denoted by Y(study variable). We apply the proposed and existing estimators to this data set and the data statistics are given below:

N = 80, n = 20, f = .25

$\overline{Y} = 51.8264, \overline{X} = 11.2624$

$C_Y = .3542, C_X = .7507$

$\rho = .9413, C_{YX} = \rho C_Y C_X = .2501$

$f_1 = .33, Q_1 = 9.318,,\quad Q_2 = 7.5750, Q_3 = 16.975$

$Q_D = 5.9125, Q_a = 11.6625, Q_R = 11.82$

T.M.= 9.318, H = 20.9805, $\lambda = .0375, \theta = .349$

Table: 2 : Values of Mean Squared Error (MSE) of different Estimators

S.No	Estimators	Mean Squared Error
1	Bahl and Tuteja (1991) Estimator	$\mathrm{MSE}(t_2) = 2{,}70{,}383.86$
2	Singh et al (2014) Estimator	(i) $\mathrm{MSE}(t_{Re}) = 9590029.95$ (ii) $\mathrm{MSE}(t_{pe}) = 9591799.89$
3	Proposed Estimator (t)	MSE(t) = 19,886.72

CONCLUSION

In the present article, a modified ratio-type predictive estimator has been proposed by using a linear combination of Tri-Mean and Quartile Average of auxiliary variable. Use of Tri-Mean and Quartile Average is beneficial for estimation of population mean because it is highly sensitive to outliers as its design structure is based on only extreme values of the data. From Table 2, it is seen that our proposed estimator has lesser MSE than that of competing estimators. Thus, proposed estimator performs better than the existing estimators. Therefore, proposed estimator is recommended to survey practitioners for estimation of population mean under predictive modeling approach, especially in the situations where outliers are present in the data.

REFERENCES

Agrawal, M. C., Roy, D.C. (1999). Efficient estimators of population variance with regression-type and ratio-type predictor-inputs. *Metron*, 57(3), pp. 4–13

Basu, D. (1971). *An Essay on the Logical Foundation of Survey Sampling*, Part I. Foundations of Statistical Inference, eds. GODAMBE,V. P., SPROTT, D. A. Toronto: Holt, Rinehart and Winston, 1971, pp. 203–242.

Bahl, S. and Tuteja, R.K. (1991). Ratio and product type exponential estimator, *Information and Optimization Sciences*, XII (I), 159-163.

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.

Ferrell, E. B. (1953). Control charts using Mid-ranges and Medians. *Industrial Quality control*, 9(5), 30-34.

Maqbool S., Javaid S. (2017).Variance Estimation Using Linear Combination of Tri-mean and Quartile Average, *American Journal of Biological and Environmental Statistics.* 3(1), 5-9

Murthy, M. N. (1967). *Sampling theory and methods*. Calcutta Statistical Publishing House, India

Nayak, R., Sahoo, L (2012). Some alternative predictive estimators of population variance. *Revista Colombiana de Estadistica*, 35(3), pp. 509–521.

Sahoo, L. N., Das, B. C., Sahoo, J. (2009). A Class of Predictive Estimators in Two stage sampling. *Journal of the Indian Society of Agricultural Statistics*, 63(2), pp. 175–180.

Sahoo, L. N., Panda P. (1999). A predictive regression-type estimator in two-stage sampling. *Journal of the Indian Society of Agricultural Statistics*, 52(3), pp. 303–308.

Sahoo, L. N., Sahoo, R. K. (2001). Predictive estimation of finite population mean in two phase sampling using two auxiliary variables. *Journal of the Indian Society of Agricultural Statistics*, 54(2), pp. 258–264.

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

Srivastava, S. K. (1983). Predictive estimation of finite population mean using product estimator. *Metrika*, 30, pp. 93–99.

Yadav, S. K., Mishra, S. S (2015). Developing Improved Predictive Estimator for Finite Population Mean Using Auxiliary Information, *Statistika*, 95(1).

Yadav, S.K., Misra, S., Mishra, S.S., Khanal, S.P. (2017). Variance Estimator using Tri-mean and Inter Quartile Range of Auxiliary Variable. *Nepalese Journal of Statistics*, 83-91.