iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: IX Month of publication: September 2020

DOI: https://doi.org/10.22214/ijraset.2020.31389

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IX Sep 2020- Available at www.ijraset.com

An Improved Predictive Approach for Estimation of Population Mean

Shashi Bhushan[1], Pavitra Jaiswal[2], Shailja Pandey[1]
[1]Department of Mathematics and Statistics, Dr. Shakuntala Misra National Rehabilitation University, India
[2]Department of Applied Statistics, Babasaheb Bhimrao Ambedkar University, India

Abstract: In the present paper, we have proposed improved predictive approach for estimation of population mean. The proposed difference type approach works better in comparison with existing works by various authors. A comprehensive comparative study is carried out both theoretically and numerically to study the merit of the considered estimator.
Keywords: Predictive approach, Bias, Mean square error, Auxiliary information.

I. INTRODUCTION

Sample survey is a cost effective method of data collection and is used for drawing valid inference about population parameters. The main aim of sample survey is to get an efficient estimate of a population. Hence, in order to enhance the efficiency of estimators of parameters one can use additional information which is correlated with the information under study and about which the data is accessible before the initiation of the survey process known as auxiliary information.

The literature portrays a wide array of techniques for using auxiliary information in concern of product, ratio and regression methods for population mean estimation which most of the time leads to gain in terms of efficiency of the estimator. Basu (1971) encountered a prediction approach to estimate population mean by predicting or guessing the mean of unobserved units and combined the aforementioned mean with mean of sampled units of understudy population. Based on this approach, various decision-theorists might unwilling to make estimator choice. But to represent "heart of the matter" Basu (1971) adopted this prediction approach for estimating population mean [see Cassel et al. (1977, p.110)]. Further, Srivastava (1983) extended this approach by proposing a product estimator under predictive estimation approach for estimating finite population mean. Singh et al. (2014) developed the exponential based ratio and product type predictive estimators for finite population mean using auxiliary information. On the similar basis Yadav, Mishra and Kumar (2014) proposed improved exponential ratio and product based predictive estimators for prediction of finite population mean then after Yadav and Mishra (2015) developed an improved ratio cum product type predictive estimator for finite population mean using auxiliary information.

Motivated with the work based on prediction approach, we here proposed a Searls(1964) based regression type estimator as a predictor for the mean of the unobserved units of the population and is shown efficient of all previous estimators used to estimate the mean of a finite population through, an empirical study given in the last section of the paper, as the percentage relative efficiency of the proposed estimator with respect to usual estimator is maximum than all other estimators.

II. EXISTING ESTIMATORS AND METHOD

Let 'y' be the study variable which takes real value Y_i for the i^{th} unit $i = 1, 2, ..., N$ of the finite population U of size N. To estimate the population mean

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

with the help of prediction based approach. Let S comprises all possible samples obtained from the population U. For any given $s \in S$, let $\vartheta(s)$ denote the effective sample size of the target population U and $\tilde{s}$ denote the set of all those units of U which are not the part of S. So that, the mean per unit estimator $\bar{y}_s$ based on sampled units is defined as

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

Further, let $\bar{y}_{\tilde{s}}$ be the mean per unit estimator based on unsampled units is defined as

$$\bar{y}_{\tilde{s}} = \frac{1}{N - \vartheta(s)} \sum_{i \in \tilde{s}} y_i \quad (2.1)$$

According to Basu's (1971) model based theory of prediction approach, for given $s \in S$, we have,

$$\bar{Y} = \left[\frac{\vartheta(s)}{N} \bar{Y}_s + \frac{(N - \vartheta(s))}{N} \bar{Y}_{\tilde{s}} \right]$$

In the above represented population mean $\bar{Y}$, the sample mean $\bar{Y}_s$ is based on the observed y values based on the units of the known sample s and $\bar{Y}_{\tilde{s}}$ based on unobserved units in s. Thus, the statistician should attempt a prediction of the mean $\bar{Y}_{\tilde{s}}$ of the unobserved units of the population on the basis of observed units in s. For any given $s \in S$ using simple random sampling without replacement and effective sample size $\vartheta(s) = n$, the population mean $\bar{Y}$ is given by,

$$\bar{Y} = \frac{n}{N} \bar{Y}_s + \frac{N - n}{N} \bar{Y}_{\tilde{s}} \quad (2.2)$$

Considering T as a predictor of $\bar{Y}_{\tilde{s}}$, the estimator of population mean $\bar{Y}$ can be written as,

$$\tilde{y} = \frac{n}{N} \bar{y}_s + \frac{N - n}{N} T$$

It is seen that the use of auxiliary information for estimation of population mean by means of any methods like ratio, and product and regression always lead to the gain in terms of efficiency of the estimator. But, if no additional information is available, then a conspicuous choice of predictor T will be:

$$\tilde{y} = \bar{y}_{\tilde{s}}$$

Now, if x be the auxiliary variable correlated with study variable y and X_i being the value of x on the i^{th} unit $i = 1, 2, ..., N$ of the population U, let $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$ and further for x_i being the value of x on the i^{th} unit $i = 1, 2, ..., n$ of the sample S, let $\bar{x} = \frac{1}{n} \sum_{i=1}^{N} x_i$.

Srivastava (1983) proposed following estimator under predictive approach

$$\tilde{y}_p = \bar{y}_s \left[\frac{n + (\bar{x}/\bar{X})(N - 2n)}{N - (\bar{x}/\bar{X})n} \right] \quad (2.4)$$

Singh et al. (2014) proposed following estimators as

$$\bar{y}_{Re} = \bar{y}_s \exp\left(\frac{\bar{X}_{\tilde{s}} - \bar{x}}{\bar{X}_{\tilde{s}} + \bar{x}} \right) \quad (2.5)$$

$$\bar{y}_{Pe} = \bar{y}_s \exp\left(\frac{\bar{x} - \bar{X}_{\tilde{s}}}{\bar{x} + \bar{X}_{\tilde{s}}} \right) \quad (2.6)$$

Similarly, Yadav, Mishra and Kumar (2014) proposed following estimators

$$\tau_{Re} = \kappa_1\left[\frac{n}{N}\bar{y}_s + \left(\frac{N-n}{N}\right)\bar{y}\exp\left(\frac{\bar{X}_{\tilde{s}} - \bar{x}}{\bar{X}_{\tilde{s}} + \bar{x}}\right)\right] \quad (2.7)$$

$$\tau_{Pe} = \kappa_2\left[\frac{n}{N}\bar{y}_s + \left(\frac{N-n}{N}\right)\bar{y}\exp\left(\frac{\bar{x} - \bar{X}_{\tilde{s}}}{\bar{x} + \bar{X}_{\tilde{s}}}\right)\right] \quad (2.8)$$

Yadav and Mishra (2015) also developed an improved predictive estimator for finite population mean using auxiliary information.

$$\tau = \alpha t_{Re} + (1-\alpha)\,t_{Pe} \quad (2.9)$$

where

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

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

The mean square errors of the above estimators are as follows

$$MSE(\tilde{y}_p) = \bar{Y}^2\theta(C_y^2 + C_x^2 + 2\rho C_y C_x) \quad (2.10)$$

$$MSE(\bar{y}_R) = \bar{Y}^2\theta[C_y^2 + C_x^2(1-2C)] \quad (2.11)$$

$$MSE(\bar{y}_{Re}) = \bar{Y}^2\theta[C_y^2 + \frac{C_x^2}{4}(1-4C)] \quad (2.12)$$

$$MSE(\bar{y}_{Pe}) = \bar{Y}^2\theta[C_y^2 + \frac{C_x^2}{4}(1+4C)] \quad (2.13)$$

$$MSE(\tau_{Re}) = \bar{Y}^2\left[1 - \frac{A_1^2}{B_1}\right] \quad (2.14)$$

$$MSE(\tau_{Pe}) = \bar{Y}^2\left[1 - \frac{A_2^2}{B_2}\right] \quad (2.15)$$

$$MSE(\tau) = \bar{Y}^2\theta C_y^2(1-\rho^2) \quad (2.16)$$

where,

$$A_1 = \left[1 - \frac{\theta C C_x^2}{2} + \theta\frac{C_x^2}{8}\frac{(3-4f)}{(1-f)}\right]$$

$$B_1 = \left[1 + \theta C_y^2 + \theta\frac{C_x^2}{2} - 2\theta C C_x^2 + \theta\frac{C_x^2}{4}\frac{(3-4f)}{(1-f)}\right]$$

$$A_2 = \left[1 + \frac{\theta C C_x^2}{2} - \theta\frac{C_x^2}{8}\frac{(1-4f)}{(1-f)}\right]$$

$$B_2 = \left[1 + \theta C_y^2 + \theta \frac{C_x^2}{2} + 2\theta C C_x^2 - \theta \frac{C_x^2}{4}\frac{(1-4f)}{(1-f)}\right]$$

$$\theta = \frac{1}{n} - \frac{1}{N},\ f = \frac{n}{N} \text{ and } C = \rho \frac{C_y}{C_x}$$

III. THE PROPOSED ESTIMATOR

The proposed regression type estimator in predictive estimation approach is given as follows:

$$t_1 = \alpha f \bar{y}_s + (1-f)\left\{\alpha \bar{y}_s + \beta\left(\bar{X}_{\tilde{s}} - \bar{x}\right)\right\} \quad (3.1)$$

Expressing above in terms of e_i using the following transformations,

$$\bar{y}_s = (1+e_0)\bar{Y},\ \bar{x} = (1+e_1)\bar{X}$$

we get,

$$t_1 = \alpha f(1+e_0)\bar{Y} + (1-f)\left[\alpha(1+e_0)\bar{Y} + \beta\left\{\frac{(N\bar{X} - n\bar{x})}{(N-n)} - (1+e_1)\bar{X}\right\}\right]$$

$$\text{here, } \bar{X}_{\tilde{s}} = \frac{1}{(N-n)}\sum_{i \in \tilde{s}} x_i = \frac{(N\bar{X} - n\bar{x})}{(N-n)}$$

$$t_1 = \alpha(1+e_0)\bar{Y} - \beta\bar{X}e_1 \quad (3.2)$$

IV. THE BIAS AND MEAN SQUARE ERROR

For the proposed estimator under prediction approach, the bias is obtained by subtracting $\bar{Y}$ on both sides of (3.2) we get,

$$\left(t_1 - \bar{Y}\right) = \bar{Y}\left(\alpha + \alpha e_0 - 1\right) - \beta\bar{X}e_1 \quad (4.1)$$

Now taking expectation on both sides of (4.1), we get

$$E\left(t_1 - \bar{Y}\right) = \bar{Y}\left\{\alpha + \alpha E(e_0) - 1\right\} - \beta\bar{X}E(e_1)$$

Using results,

$$E(e_0) = E(e_1) = 0$$

And up to first degree of approximation:

$$E\left(e_0^2\right) = \theta C_y^2,\ E\left(e_1^2\right) = \theta C_x^2 \text{ and } E\left(e_0 e_1\right) = \theta C C_x^2$$

we get the bias of the proposed estimator as,

$$\text{Bias}(t_1) = \bar{Y}(\alpha - 1) \quad (4.2)$$

Squaring both sides of equation (4.1)

$$\left(t_1 - \bar{Y}\right)^2 = \bar{Y}^2\left(\alpha + \alpha e_0 - 1\right)^2 + \beta^2\bar{X}^2 e_1^2 - 2\beta\bar{X}e_1\bar{Y}\left(\alpha + \alpha e_0 - 1\right)$$

$$= \bar{Y}^2\left\{(\alpha-1)^2 + \alpha^2 e_0^2 + 2\alpha e_0(\alpha - 1)\right\} + \beta^2\bar{X}^2 e_1^2 - 2\bar{X}\bar{Y}\beta e_1(\alpha-1) - 2\bar{X}\bar{Y}\beta e_1 \alpha e_0$$

Now taking expectations on both sides of above we get the MSE of the estimator as,

$$E\left(t_1 - \bar{Y}\right)^2 = E\left[\bar{Y}^2\left\{\left(\alpha-1\right)^2 + \alpha^2 e_0^2 + 2\alpha e_0\left(\alpha-1\right)\right\} + \beta^2\bar{X}^2 e_1^2 - 2\bar{X}\bar{Y}\beta e_1\left(\alpha-1\right) - 2\bar{X}\bar{Y}\beta e_1\alpha e_0\right]$$

$$MSE\left(t_1\right) = \bar{Y}^2\left\{\left(\alpha-1\right)^2 + \alpha^2\theta C_y^2\right\} + \beta^2\bar{X}^2\theta C_x^2 - 2\bar{X}\bar{Y}\beta\alpha\rho\theta C_y C_x \quad (4.3)$$

For the optimum value of MSE of estimator, differentiating with respect to constant α and β, we get

$$\alpha_{opt} = \frac{1}{1+\theta C_y^2\left(1-\rho^2\right)}$$

The optimum value of β is given by

$$\beta_{opt} = \frac{\bar{Y}\rho C_y}{\bar{X}C_x}.\frac{1}{1+\theta C_y^2\left(1-\rho^2\right)}$$

The minimum MSE for the optimum value of α and β is given by

$$MSE_{min}\left(t_1\right) = \frac{\bar{Y}^2\theta C_y^2\left(1-\rho^2\right)}{1+\theta C_y^2\left(1-\rho^2\right)}$$

From, the expression (2.16), we get

$$MSE_{min}\left(t_1\right) = \frac{MSE(\tau)}{1+\frac{MSE(\tau)}{\bar{Y}^2}}$$

V. COMPARITIVE STUDY

On comparing the mean square errors of the estimator τ with the proposed estimator, we get

$$MSE_{min}\left(t_1\right) < MSE\left(\tau\right)$$

If $\frac{1}{1+\frac{MSE\left(\tau\right)}{\bar{Y}^2}} < 1$

$$\Rightarrow MSE\left(\tau\right) > 0$$

which is true always. Hence, the proposed estimator is always better than the estimator proposed by Yadav and Mishra (2015). Further, since the estimator τ is better than the remaining existing estimators [see Yadav and Mishra (2015)] so that the proposed estimator is most precise estimator from all the mentioned existing estimators.

Also, a direct comparison of the proposed estimator needs to be verified on a data set, which is done in the next section.

VI. AN EMPIRICAL STUDY

We have taken following populations from Sarjinder Singh and parameters are mentioned in Table 1.

1) *Population I.* All operating banks: Amount (in \$000) of agricultural loans outstanding in different states in 1997. In this, Y_i and X_i are amounts of real and non-real estate farm loans in different states during 1997 respectively.

2) *Population II.* Hypothetical situation of a small village having only30 old persons (age more than 50 years): Approximate duration of sleep (in minutes) and Age (in years) of the persons. Here, Y_i and X_i are duration of sleep and age of old persons respectively.

Table 1: Parameters of the populations

Population	N	n	$\bar{Y}$	C_x	C_y	ρ
I	50	11	555.43	1.2351	1.0529	0.8038
II	30	10	384.20	0.0434	0.1559	-0.8552

Table 2: Percentage Relative Efficiency with respect to the usual unbiased estimator $\bar{y}_s$

Estimators	PRE based on Population I	PRE based on Population II
$\bar{y}_s$	100.00	100.00
$\tilde{y}_p$	*	164.32
$\bar{y}_{Re}$	249.31	*
$\bar{y}_{Pe}$	*	127.26
τ_{Re}	141.98	*
τ_{Pe}	*	124.45
τ	282.56	372.26
t_1	290.42	372.42

where, $PRE(.,\bar{y}_s) = \dfrac{MSE(\bar{y}_s)}{MSE(.)} * 100$

VII. CONCLUSION

In the paper, the proposed regression type estimator as a predictor for unobserved units of the population to estimate the finite population mean comes out to be most efficient as it has maximum PRE than other estimators. The expression of mean square is function of the mean square error of the estimator proposed by Yadav and Mishra (2015). From table 2, it is evident that the estimator t_1 has efficient results than $\bar{y}_s$, $\tilde{y}_p$, $\bar{y}_{Re}$, $\bar{y}_{Pe}$, τ_{Re}, τ_{Pe} and τ. Therefore, the proposed estimator t_1 should be preferred to estimate the population mean under predictive estimation approach.

REFERENCES

[1] Basu, D. : An Essay on the Logical Foundation of Survey Sampling, Part I. Foundations of Statistical Inference, eds. Godambe, V. P. and Sprott, D. A. Toronto: Holt, Rinehart and Winston, 203–242 (1971).
[2] Cassel, C. M., Sarndal, C. E. and Wretman, J. H. : Foundation of Inference in Survey Sampling. New York, USA: Wiley (1977).
[3] Searls, D. T. : The utilization of a known coefficient of variation in the estimation procedure. Journal of the American Statistical Association, 59(308), 1225-1226 (1964).
[4] Srivastava, S. K. : Predictive estimation of finite population mean using product estimator. Metrika, 30(1), 93–99 (1983).
[5] Singh, S.: Advanced sampling theory with applications. Kluwer Academic Publishers, 41, 27-37 (2003).
[6] Singh, H. P., Solanki, R. S. and Singh, A. K. : Predictive estimation of finite population mean using exponential estimators. Statistika: Statistics and Economy Journal, 94(1), 41-53 (2014).
[7] Yadav, S.K., Mishra, S.S. and Kumar, S.: Optimal search for efficient estimator of finite population mean using auxiliary information. American Journal of Operational Research, 4(2), 28-34 (2014).
[8] Yadav, S.K., Mishra, S.S. (2015): Developing Improved Predictive estimator for finite population mean using Auxiliary Information, Statistika Analyses, 95, 76-85.