ISSN 1684-8403
Journal of Statistics
Volume 23, 2016. pp. 1-11

Efficient Estimators of Finite Population Mean using Predictive Estimation in Simple Random Sampling

Rajesh Singh[1], Hemant K. Verma[2] and Prayas Sharma[3]

Abstract

This paper suggests the prediction approach for estimating the population mean under Simple Random Sampling Without Replacement. We propose an Estimator using a predictor and study its properties. The Bias and Mean Square Error (MSE) expression of Proposed Estimator has been derived up to the first order approximation. It has been shown that the Proposed Estimator is more efficient than usual Unbiased Estimator, Exponential Ratio and Exponential Product Estimators and Estimator recently proposed by Singh et al. (2014). In support of the theoretical, numerical illustration has also been carried out.

Keywords

Auxiliary information, Bias, Mean Square Error, Predictive approach, Simple random sampling

1. Introduction

It is well known that in sample surveys supplementary population information is often used at estimation stage to increase the precision which is highly correlated with study variable y. It is customary to make use of auxiliary information on variable x in the estimation of finite population parameters. Several authors including Sharma and Singh (2013, 2014), Singh and Kumar (2011), and Upadhayay and Singh (1999) have suggested Estimators using auxiliary information under different situations.

[1] Department of Statistics, Banaras Hindu University, Uttar Pardesh (U.P.), India
[2] Department of Statistics, Banaras Hindu University, Uttar Pardesh (U.P.), India
[3] Department of Decision Sciences, College of Management and Economics Studies, University of Petroleum and Energy Studies, Dehradun, India

The literature of sampling surveys consists of a great variety of techniques to construct more efficient Estimators using design and model based approach.

In a predictive approach, the non-sampled values are predicted by using the population values of a specified model. Prediction theory or model based theory for sample survey can be considered as a general framework for Statistical Inferences on the character of finite population. Existing Estimators of population parameters encountered in the classical theory, (such as, Expansion, Ratio, Regression Estimators) can be predictors in the general prediction theory under some special model. In Survey Sampling, several authors have used predictive approach either to study or modify the existing Estimators or to produce new one. When we use usual UnBiased, Ratio and Regression Estimators as a predictor for the mean of the unobserved units of the population, it results in the corresponding customary (usual) Estimators of the mean of whole population. Srivastava (1983) used the predictive approach for the usual Product Estimator, and it is observed that the resulting Estimator of the mean of the whole population is not same as the customary (usual) Product Estimator. Some other authors including Agrawal and Roy (1999), Biradar and Singh (1998) and Nayak and Sahoo (2012) suggested some Predictive Estimators for finite population variance. For Two Stage Sampling, Sahoo and Panda (1999) provided a Regression Type Estimator. Further, Sahoo et al. (2009) and Sahoo and Sahoo (2001) introduced a class of Estimators for the finite population mean availing information on two auxiliary variables in Two Stage Sampling. Hossain and Ahmed (2001) and Saini (2013) have also important contribution in estimation using predictive approach in Two Phase Sampling using auxiliary information. In this paper, we consider prediction approach to examine an Estimator as predictor of the mean of the unobserved units of the population using the information of observed units in the sample.

2. Notations and Estimators in Literature

Let $U = (U_1, U_2, U_3, \ldots, U_N)$ denote a finite population of N distinct and identifiable units. For the estimation of population mean $\overline{Y}$ of a study variable *y*, a sample of size $\eta(s)$ be drawn from this population using Simple Random Sampling Without Replacement (SRSWOR) scheme. Further, consider *x* be the auxiliary variable that is correlated with study variable *y*, taking the corresponding values of the units. Any ordered subset of *U* is called a sample from *U*. Let *S* denote the collection of all possible samples from *U*. Let $\eta(s)$ denote effective sample size (the number of distinct units in *s*) for any given $s \in S$, and $\bar{s}$ denote the set of all those units of *U* which are not in *s*. Define,

$$\bar{Y}_s = \frac{1}{\eta(s)} \sum_{i \in s} y_i \text{ and } \bar{Y}_{\bar{s}} = \frac{1}{(N - \eta(s))} \sum_{i \in \bar{s}} y_i$$

Thus for any given $s \in S$, we can write,

$$\bar{Y} = \left[\frac{\eta(s)}{N} \bar{Y}_s + \frac{(N - \eta(s))}{N} \bar{Y}_{\bar{s}} \right]. \tag{2.1}$$

Sample mean $\bar{Y}_s$ is known because it is based on the units of the sample *s* whose *y* values have been observed. Therefore, our main focus is to make a prediction of mean $\bar{Y}_{\bar{s}}$ of unobserved units of the population on the basis of observed units in sample *s*.
In case of Simple Random Sampling with sample size n (i.e. $\eta(s) = n$) eq. (2.1) can be expressed as

$$\bar{Y} = \left[\frac{n}{N} \bar{y} + \frac{(N - n)}{N} \bar{Y}_{\bar{s}} \right]. \tag{2.2}$$

From eq. (2.2), an Estimator of population mean $\bar{Y}$ can be written as

$$t = \left[\frac{n}{N} \bar{y} + \frac{(N - n)}{N} T \right] \tag{2.3}$$

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

Srivastava (1983) has described a prediction approach for predicting the mean $\bar{Y}_{\bar{s}}$ of the unobserved units of the population.
If we have no additional information available on population then we use mean as Estimator. Thus, an appropriate choice of *T* is $\bar{y}$ giving,

$$t = \left[\frac{n}{N} \bar{y} + \frac{(N - n)}{N} \bar{y} \right] = \bar{y}, \text{ the customary mean per unit Estimator.} \tag{2.4}$$

If we have information on auxiliary variable *x* which is highly positively correlated with *y*, then we use Ratio Estimator. Thus, an appropriate choice of *T* is $\frac{\bar{y}}{\bar{x}} \bar{X}_{\bar{s}}$.
Then, we get,

$$t=\left[\frac{n}{N}\bar{y}+\frac{(N-n)}{N}\frac{\bar{y}}{\bar{x}}\bar{X}_{\bar{s}}\right]=\frac{\bar{y}}{\bar{x}}\bar{X}=\bar{y}_R \text{ (say)} \qquad (2.5)$$

where,

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

Eq. (2.5) is nothing but a customary Ratio Estimator of $\bar{Y}$.

The Bias and Mean Square Error expression of customary Ratio Estimator $\bar{y}_R$ is given as (Singh et al., 2014),

$$\text{Bias}(\bar{y}_R)=\theta\bar{Y}^2C_x^2(1-C)$$
$$\text{MSE}(\bar{y}_R)=\theta\bar{Y}^2\left\{C_y^2+C_x^2(1-2C)\right\}$$

where,

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

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

If we have information on auxiliary variable *x* which is correlated with *y* and one intends to use this in the form of Regression Estimator, an appropriate choice of *T* is $\bar{y}+b\left(\bar{X}_{\bar{s}}-\bar{x}\right)$(see Srivastava, 1983),

where,

b is the Regression coefficient estimated from the sample *s*. Therefore,

$$t=\left[\frac{n}{N}\bar{y}+\frac{(N-n)}{N}\left\{\bar{y}+b\left(\bar{X}_{\bar{s}}-\bar{x}\right)\right\}\right]=\bar{y}+b\left(\bar{X}-\bar{x}\right)=\bar{y}_{lr}\text{ (say)} \qquad (2.6)$$

which is a customary Regression Estimator of $\bar{Y}$.

The Bias and Mean Square Error expression of customary Regression Estimator $\bar{y}_{lr}$ is given as (Srivastava, 1983)

$$\text{Bias}(\bar{y}_{lr})=0$$
$$\text{MSE}(\bar{y}_{lr})=\theta\bar{Y}^2\left(C_y^2-C^2C_x^2\right).$$

If auxiliary information is available in such a way that it can be utilized in product form then an appropriate choice of *T* is $\dfrac{\bar{y}\,\bar{x}}{\bar{X}_{\bar{s}}}$ giving,

$$t = \left[\frac{n}{N} \bar{y} + \frac{(N-n)}{N} \frac{\bar{y}\,\bar{x}}{\bar{X}_{\bar{s}}} \right] = \bar{y} \frac{n\bar{X} + (N-2n)\bar{x}}{N\bar{X} - n\bar{x}} = t_1 \text{ (say)} \tag{2.7}$$

which is not a customary Product Estimator $\bar{y}_P = \frac{\bar{y}\,\bar{x}}{\bar{X}}$ of $\bar{Y}$. Motivated by Srivastava (1983), Singh et al. (2014) suggested predictive estimation of population mean using the form of Bhal and Tuteja (1991) Ratio-type and Product-type Exponential Estimators as predictor Estimators given as

$$t_{Re} = \left[\frac{n}{N} \bar{y} + \left(\frac{N-n}{N} \right) \bar{y} \exp\left(\frac{\bar{X}_{\bar{s}} - \bar{x}}{\bar{X}_{\bar{s}} + \bar{x}} \right) \right] \tag{2.8}$$

$$t_{Pe} = \left[\frac{n}{N} \bar{y} + \left(\frac{N-n}{N} \right) \bar{y} \exp\left(\frac{\bar{x} - \bar{X}_{\bar{s}}}{\bar{x} + \bar{X}_{\bar{s}}} \right) \right] \tag{2.9}$$

The Bias and Mean Square Error expressions of t_{Re} and t_{Pe} up to the first order of approximation are given as

$$\text{Bias}(t_{Re}) = \frac{1}{8} \theta \bar{Y} C_x^2 [3 - 4(C + f)] \tag{2.10}$$

$$\text{Bias}(t_{Pe}) = \frac{1}{8} \theta \bar{Y} C_x^2 \left[4C - \frac{1}{(1-f)} \right] \tag{2.11}$$

$$\text{MSE}(t_{Re}) = \theta \bar{Y}^2 \left[C_y^2 + \frac{1}{4}(1 - 4C) C_x^2 \right] \tag{2.12}$$

$$\text{MSE}(t_{Pe}) = \theta \bar{Y}^2 \left[C_y^2 + \frac{1}{4}(1 + 4C) C_x^2 \right] \tag{2.13}$$

For proofs see Singh et al. (2014) and notations used here, are defined earlier.

3. Proposed Estimator Under Prediction Approach

Adapting prediction approach to examine an Estimator as predictor of the mean of the unobserved units of the population using the information of observed units in the sample, we propose an obvious choice for *T* as

$$t_{\bar{p}} = \bar{y} \exp\left(\frac{\bar{X}_{\bar{s}} - \bar{x}}{\bar{X}_{\bar{s}} + \bar{x}} \right) + b(\bar{X}_{\bar{s}} - \bar{x}) \tag{3.1}$$

For this choice of *T*, the Estimator is defined as

$$t_p = \frac{n}{N}\bar{y} + \left(\frac{N-n}{n}\right)t_{\bar{p}}$$

$$= \frac{n}{N}\bar{y} + \left(\frac{N-n}{n}\right)\left\{\bar{y}\exp\left(\frac{N\left(\bar{X}-\bar{x}\right)}{N\left(\bar{X}+\bar{x}\right)-2n\bar{x}}\right)+b\left(\frac{N\left(\bar{X}-\bar{x}\right)}{N-n}\right)\right\} \tag{3.2}$$

To obtain the Bias and Mean Square Error expressions of the Estimator t_p to the first degree of approximation, we define,

$e_0 = \frac{\bar{y}-\bar{Y}}{\bar{Y}}$ and $e_1 = \frac{\bar{x}-\bar{X}}{\bar{X}}$

such that, $E(e_0) = E(e_1) = 0$

also, $E(e_0^2) = \theta C_y^2$, $E(e_1^2) = \theta C_x^2$, $E(e_0 e_1) = \theta C C_x^2$

Expressing eq. (3.2) in terms of e's, we have,

$$t_p = \bar{Y}(1+e_0)\left[\frac{n}{N} + \left(\frac{N-n}{N}\right)\exp\left\{\frac{-Ne_1}{2(N-n)+(N-2n)e_1}\right\}\right] - b\bar{X}e_1$$

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

$$= \bar{Y}(1+e_0)\left[f + (1-f)\exp\left\{\frac{-e_1}{2(1-f)}\left(1+\frac{(1-2f)}{2(1-f)}e_1\right)^{-1}\right\}\right] - b\bar{X}e_1 \tag{3.3}$$

Expanding the right hand side of eq. (3.3) and retaining terms up to second degree of e's, we get,

$$t_p = \bar{Y}\left[1 + e_0 - \frac{1}{2}e_1 + \frac{1}{8}\frac{(3-4f)}{(1-f)}e_1^2 - \frac{1}{2}e_0 e_1\right] - b\bar{X}e_1$$

$$\left(t_p - \bar{Y}\right) = \bar{Y}\left[e_0 - \frac{1}{2}e_1 + \frac{1}{8}\frac{(3-4f)}{(1-f)}e_1^2 - \frac{1}{2}e_0 e_1\right] - b\bar{X}e_1 \tag{3.4}$$

Taking expectations of both sides in eq. (3.4), we get the Bias of the t_p up to the first order of approximation as below.

$$\text{Bias}(t_p) = \theta\bar{Y}C_x^2\left[\frac{(3-4f)}{8(1-f)} - \frac{C}{2}\right] \tag{3.5}$$

From eq. (3.4), we have,

$$\left(t_p-\overline{Y}\right)\cong\overline{Y}\left(e_0-\frac{1}{2}e_1\right)-b\overline{X}e_1 \tag{3.6}$$

Squaring both sides of eq. (3.6) and then taking expectations, we get the Mean Square Error expression of t_p up to first order of approximation as

$$MSE\left(t_p\right)=\theta\overline{Y}^2\left[C_y^2+\left(\frac{1}{4}-C\right)C_x^2\right]+b^2\theta\overline{X}^2C_x^2-2b\theta\overline{X}\overline{Y}C_x^2\left(C-\frac{1}{2}\right) \tag{3.7}$$

Partially differentiating eq. (3.7) with respect to *b* and equating to zero, we get the optimum value of b as

$$b_{opt}=\frac{\overline{Y}}{\overline{X}}\left(C-\frac{1}{2}\right) \tag{3.8}$$

The Proposed Estimator t_p will perform better except the value of C=1/2 and for this value of C=1/2, it will behave like Exponential Ratio type Estimator.

Replacing *b* in eq. (3.7) with b_{opt}, we get the minimum MSE expression of t_p as

$$MSE_{min}\left(t_p\right)=\theta\overline{Y}^2\left[C_y^2-C^2C_x^2\right] \tag{3.9}$$

Remark : Note that $MSE_{min}(t_p)=MSE(\overline{y}_{lr})$.

4. Efficiency Comparison

Predictive Estimator t_p is more efficient than $\overline{y}$ if

$V(\overline{y})-MSE\left(t_p\right)>0$

Or if, $C_y^2-\left[C_y^2-C^2C_x^2\right]>0$

Or if, $C^2C_x^2>0$ i.e. $C>0$ (4.1)

Since the left hand side $C^2C_x^2$ is always positive unless C = 0 which may occur if the correlation equals zero(but here we are using correlated auxiliary variable). The other case is when $C_x^2=0$ which occurs if $Var(X)=0$, i.e. all values of X are the same.

However, the design of the experiment requires data on auxiliary related variable. So, $\rho=0$ is excluded.

Predictive Estimator t_p is more efficient than t_{Re} if

$MSE\left(t_{Re}\right)-MSE\left(t_p\right)>0$

Or if, $4C^2-4C+1>0$

Or if, $C > \frac{1}{2}$ (4.2)

Predictive Estimator t_p is more efficient than t_{Pe} if

$MSE(t_{Pe}) - MSE(t_p) > 0$

Or if, $4C^2 + 4C + 1 > 0$

Or if, $C > -\frac{1}{2}$ (4.3)

5. Empirical Study

We use three real data sets for empirical study given in Table 1. These are already used by Kadilar and Cigni (2006) (Population I), Khosnevisan et al. (2007) (Population II), and Gupta and Shabbir (2008) (Population III), respectively, to verify the theoretical result.

Table 1 exhibits Bias and percent relative efficiencies of the existing Estimators $\bar{y}$, $\bar{y}_{lr}$, t_{Re}, t_{Pe} and Proposed Estimator t_p.
where,

$$PRE(T_i, \bar{y}) = \frac{MSE(\bar{y})}{MSE(T_i)} \times 100; \qquad T = [T_i] = (\bar{y},\ \bar{y}_{lr},\ t_{Re}, t_{Pe},\ t_P)$$

Analyzing Table 2 we conclude that the proposed Estimators t_p is less Biased as compared to t_{Re}, t_{Pe} and equally efficient to the Regression Estimator $\bar{y}_{lr}$ but more efficient than the usual Unbiased Estimator, Exponential Ratio, Exponential Product Estimator and Estimator due to Singh et al. (2014) in the sense of having the smallest Mean Square Error.

6. Conclusion

The present paper considered the problem of estimating the population mean $\bar{Y}$ of study variable *y* using prediction approach. We have derived the expressions for Biases and Mean Square Errors of the Proposed Estimators up to the first degree of approximation. In theoretical efficiency comparisons, it has been shown that under certain conditions the Proposed Estimator is more efficient than usual Unbiased Estimator, Singh et al. (2014) Ratio-type and Product-type Exponential Estimators and similar to usual Regression Estimator $\bar{y}_{lr}$. The theoretical results have also been supported empirically with the help of three real populations earlier considered by Kadilar and Cigni (2006), Khosnevisan et al. (2007), and Gupta and Shabbir (2008).

Acknowledgements

Authors are thankful to unknown referees for their valuable comments, which helped in improving the quality of the paper.

Table1: Parameters of populations

	Parameters						
Populations	**N**	**n**	$\bar{Y}$	$\bar{X}$	ρ	C_y	C_x
I	106	20	2212.59	27421.7	0.86	5.22	2.10
II	20	8	19.55	18.8	-0.92	0.355	0.394
III	104	20	6.254	13931.68	0.865	1.866	1.653

Table 2: Bias and PRE of Estimators

	Population I		Population II		Population III	
Estimators	**Bias**	**PRE**	**Bias**	**PRE**	**Bias**	**PRE**
$\bar{y}$	0	100	0	100	0	100
$\bar{y}_{lr}$	0	384.03	0	651.04	0	397.18
t_{Re}	-9497.42	143.99	2.98	42.94	-4.43	232.60
t_{Pe}	11001.89	72.13	-3.15	348.58	7.06	50.96
t_p	-8694.44	384.03	357	651.04	-3.03	397.18

References

1. Agrawal, M. C. and Roy, D. C. (1999). Efficient Estimators of population variance with Regression-type and Ratio-type predictor-input. *Metron*, **57(3)**, 4-13.
2. Bahl, S. and Tuteja, R. K. (1991). Ratio and Product type Exponential Estimators. *Journal of Information and Optimization Sciences*, **12(1)**, 159-164.
3. Biradar, R. S. and Singh, H. P. (1998). Predictive estimation of finite population variance. *Calcutta Statistical Association Bulletin*, **48**, 229-235.
4. Gupta, S. and Shabbir, J. (2008). On improvement in estimating the population mean in Simple Random Sampling. *Journal of Applied Statistics*, **35(5)**, 559-566.
5. Hossian, M. I. and Ahmed, M. S. (2001). A class of Predictive Estimators in Two-Stage Sampling using auxiliary information. *Information and Management Sciences*, **12(1)**, 49-55.
6. Kadilar, C. and Cingi, H. (2006). New Ratio Estimators using correlation coefficient. *Interstat*, **4(March)**, 1-11.
7. Khoshnevisan, M., Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. (2007). A general family of Estimators for estimating population mean using known value of some population parameter(s). *Far East Journal of Theoretical Statistics*, **22(2)**, 181-191.
8. Nayak, L. and Sahoo, L. (2012). Some alternative Predictive Estimators of population variance. *Revista Colombiana de Estadistica*, **35(3)**, 509-521.
9. Sahoo, L. N., Das, B. C. and Sahoo, J. (2009). A class of Predictive Estimator in Two Stage Sampling. *Journal of the Indian Society of Agricultural Statistics*, **63(2)**, 175-180.
10. Sahoo, L. N. and Panda, P. (1999). A Predictive Regression-type Estimator in Two-Stage Sampling. *Journal of the Indian Society of Agricultural Statistics*, **52(3)**, 303-308.
11. Sahoo, L. N. and 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)**, 258-264.
12. Saini, M. (2013). A class of Predictive Estimators in Two-Stage Sampling when auxiliary character is estimated at SSU level. *International Journal of Pure and Applied Mathematics*, **85(2)**, 285-295.
13. Sharma, P. and Singh, R. (2013). Improved Estimators for Simple Random Sampling and Stratified Random Sampling under second order of approximation. *Statistics in Transition*, **14(3)**, 379-390.

14. Sharma, P. and Singh, R. (2014). Improved Ratio Type Estimators using two auxiliary variables under second order approximation. *Mathematical Journal of Interdisciplinary Sciences*, **2(2)**, 179-190.
15. Singh, H. P., Solanki, R. S. and Singh, A. K. (2014). Predictive estimation of finite population mean using Exponential Estimators. *STATISTIKA - Statistics and Economic Journal*, **94(1)**, 41-53.
16. Singh, R. and Kumar, M., (2011). A note on transformations on auxiliary variable in Survey Sampling. *MASA*, **6(1)**, 17-19.
17. Srivastava, S. K. (1983). Predictive estimation of finite population mean using Product Estimator. *Metrika*, **30**, 93-99.
18. Upadhyaya, L. N. and Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal*, **41(5)**, 627–636.