Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 13, Number 4 (2017), pp. 1209-1216

http://www.ripublication.com

On Efficient Class of Estimators for Population Mean

Shashi Bhushan[1] and Praveen Kumar Misra[1]

[1]*Department of Mathematics and Statistics, Dr. Shakuntala Misra National Rehabilitation University, India*

Abstract

For estimating finite population mean using coefficient of variation of study variable and the usual linear regression estimator, a generalized class of estimators of which the usual linear regression estimator is a special case, is proposed and its mean square error along with its bias are found. A sub – class of optimum estimators depending on optimum value minimizing the mean square error is found. Further, to enhance the practical utility, a sub – class of estimators based upon estimated optimum value attaining the same minimum mean square error as that of the sub-class of estimators depending on optimum value is obtained. A comparative study of the proposed generalized class of estimators is made with some estimators including the usual linear regression estimator.

Keywords: Linear regression estimator, Bias, Mean Square Error.

1. INTRODUCTION

The use of population coefficient of variation of study variable for increasing the efficiency of the sampling strategy has been discussed by Searle (1964). Also, the use of prior information on population coefficient of variation of auxiliary variable for increasing the efficiency of the sampling strategy the auxiliary variable has been discussed by various authors including Sisodia and Dwivedi (1981) and Pandey and Dubey (1988) by using a modified ratio and product estimator respectively. One of the major objections against the use of various ratio and product type estimators is that they are biased and most of times minimum mean square error never less than that of linear regression estimator. In this paper, we have made an attempt to propose a class of estimators which has got lesser minimum mean square error than that of linear regression estimator and possesses some important classes of estimators as its sub-classes.

Let y be the characteristic under study and x be the auxiliary variable. Thus for a finite population of size N, we denote by

Y_i: the observation on the ith unit of the population for the characteristic y under study $(i = 1, 2, ..., N)$, X_i: the observation on the ith unit of the population for the auxiliary characteristic x under study $(i = 1, 2, ..., N)$,

$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$, $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$, $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2 = \frac{N}{N-1}\sigma_x^2$,

$S_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y}) = \rho S_x S_y$, $\beta = \frac{S_{xy}}{S_x^{\ 2}} = \rho\frac{S_y}{S_x}$, ρ is the population correlation coefficient between x and y, C_x and C_y are the coefficients of variation of x and y, and $\mu_{pq} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^p(Y_i - \bar{Y})^q$: the (p, q)th product moment about mean between x and y.

Also, based on a sample of size n, let $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ be the sample mean of auxiliary characteristic x and characteristic y under study respectively, $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ be the sample variance of characteristic x, $s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ be the sample covariance between x and y and $b = \frac{s_{xy}}{s_x^2}$ be the sample regression coefficient of y on x.

2. PROPOSED ESTIMATOR

In this paper, we propose the following estimator of population mean $\bar{Y}$, using the prior information about the population mean of auxiliary variable $\bar{X}$ and coefficient of variation of the study variable C_y, given by

$$\bar{Y}_g = \{\bar{y} + b(\bar{X} - \bar{x})\}\, g(w) = \left\{\bar{y} + \frac{s_{xy}}{s_x^2}(\bar{X} - \bar{x})\right\} g(w) \tag{2.1}$$

where $w = C_y^2 \big/ C_y^2$ and $g(w)$ satisfies the following regularity conditions

(i) $g(w = 1) = 1$ (2.2)

(ii) $g(w)$ could be expanded using a Taylor's series expansion such that

$$g(w) = g(1) + (w-1)g_1(1) + \frac{(w-1)^2}{2!} g_2(1) + \frac{(w-1)^3}{3!} g_3(w^*) \qquad (2.3)$$

Where $g_1(1) = \left.\frac{\partial g}{\partial w}\right|_1$, $g_2(1) = \left.\frac{\partial^2 g}{\partial w^2}\right|_1$, $g_3(w^*) = \left.\frac{\partial^3 g}{\partial w^3}\right|_{w^*}$, $w^* = 1 + \varphi(w-1)$ and $0 < \varphi < 1$.

Note that for $g(w) = 1$ the proposed estimator reduces to the usual linear regression estimator

$$\bar{y}_{lr} = \bar{y} + \frac{s_{xy}}{s_x^2}(\bar{X} - \bar{x}) \ . \qquad (2.4)$$

Let $\bar{y} = \bar{Y}(1+e_0)$, $\bar{x} = \bar{X}(1+e_1)$, $s_{xy} = S_{xy}(1+e_2)$, $s_x^2 = S_x^2(1+e_3)$, $s_y^2 = S_y^2(1+e_4)$

with $E(e_0) = E(e_1) = E(e_2) = E(e_3) = E(e_4) = 0$ (2.5)

Then, using Taylor's series expansion in (2.1), (2.2) and (2.3), we have

$$\bar{Y}_g = \left\{\bar{y} + \frac{s_{xy}}{s_x^2}(\bar{X} - \bar{x})\right\} g(w)$$

$$= \left\{\bar{y} + \frac{s_{xy}}{s_x^2}(\bar{X} - \bar{x})\right\}\left\{1 + (w-1)g_1(1) + \frac{(w-1)^2}{2!} g_2(1) + \frac{(w-1)^3}{3!} g_3(w^*)\right\}$$

$$= \bar{Y} + \bar{Y}e_0 + \beta\bar{X}\left(-e_1 - e_1e_2 + e_1e_3\right) + \bar{Y}\left(e_4 - 2e_0 - e_0e_4 + e_0^2\right)g_1(1)$$

$$+\beta\bar{X}\left(e_1e_4 - 2e_0e_1\right)g_1(1) + \frac{\bar{Y}}{2}\left(e_4^2 - 4e_0e_4 + 4e_0^2\right)g_2(1) \qquad (2.6)$$

Assuming N to be so large that finite population correction could be ignored and using the results given in Sukhatme and Sukhatme (1997) and Bhushan (2007), we have

$$E(e_0{}^2) = n^{-1}C_y{}^2,\ E(e_1{}^2) = n^{-1}C_x{}^2,\ E(e_4{}^2) = n^{-1}(\beta_{02} - 1),\ E(e_0e_1) = n^{-1}\rho C_x C_y,$$
$$E(e_1e_3) = n^{-1}\lambda_{30},\ E(e_1e_4) = n^{-1}\lambda_{12},\ E(e_1e_2) = n^{-1}\lambda_{21},\ E(e_0e_4) = n^{-1}\lambda_{03} \qquad (2.7)$$

where $\gamma_n = (N-n)/Nn$, $\beta_{02} = \mu_{04} / \mu_{02}{}^2$, $\lambda_{03} = \mu_{03}\big/\bar{Y}S_y^2$, $\lambda_{12} = \mu_{12}\big/\bar{X}S_y^2$, $\lambda_{30} = \mu_{30}\big/\bar{X}S_x^2$ and $\lambda_{21} = \mu_{21}\big/\bar{X}S_{xy}$.

Using (2.6) and (2.7) it can be easily seen that

$$Bias(\bar{Y}_g) = E(\bar{Y}_g) - \bar{Y}$$

$$= n^{-1}\left[B\bar{X}\left(\lambda_{30}-\lambda_{21}\right)+\left\{\bar{Y}\left(C_y^2-\lambda_{03}\right)+B\bar{X}\left(2\rho C_x C_y-\lambda_{12}\right)\right\}g_1(1)\right.$$
$$\left.+\frac{\bar{Y}}{2}\left\{\beta_{02}-1+4\left(C_y^2-\lambda_{03}\right)\right\}g_2(1)\right] \qquad (2.8)$$

Thus, showing that $\bar{Y}_g$ is a biased estimator of population mean and its exact bias is given by (2.8). Using (2.6) and neglecting terms of e_i's having powers greater than two, we get the MSE given by

$$MSE\left(\bar{Y}_g\right)=\gamma_n\bar{Y}^2\left[\left(1-\rho^2\right)C_y^2+\left\{\beta_{02}-1+4\left(C_y^2-\lambda_{03}\right)\right\}g_1^2(1)\right.$$
$$\left.-2\left\{\lambda_{03}-BR^{-1}\lambda_{12}-2C_y^2+2BR^{-1}\rho C_x C_y\right\}g_1(1)\right] \qquad (2.9)$$

which is minimum for the optimum value of $g_1(1)$ given by

$$g_1(1)_{opt}=\frac{\lambda_{03}-BR^{-1}\lambda_{12}-2C_y^2+2BR^{-1}\rho C_x C_y}{\beta_{02}-1+4\left(C_y^2-\lambda_{03}\right)}. \qquad (2.10)$$

Therefore, the minimum mean square error of $\bar{Y}_g$ is given by

$$MSE\left(\bar{Y}_g\right)_{\min}=\gamma_n\bar{Y}^2\left\{(1-\rho^2)C_y{}^2-\frac{\left\{\lambda_{03}-BR^{-1}\lambda_{12}-2C_y^2+2BR^{-1}\rho C_x C_y\right\}^2}{\left\{\beta_{02}-1+4\left(C_y^2-\lambda_{03}\right)\right\}}\right\} \qquad (2.11)$$

3. ESTIMATION OF POPULATION MEAN UNDER ESTIMATED OPTIMUM VALUE

The mean square error of $\hat{\bar{Y}}_g$ is minimized for the choice of $g_1(1)=g_1(1)_{opt}$ where

$$g_1(1)_{opt}=\frac{\lambda_{03}-BR^{-1}\lambda_{12}-2C_y^2+2BR^{-1}\rho C_x C_y}{\beta_{02}-1+4\left(C_y^2-\lambda_{03}\right)}$$

$$=\frac{\dfrac{\mu_{03}}{\bar{Y}S_y^2}-\dfrac{S_{xy}}{S_x^2}\left(\dfrac{\bar{Y}}{\bar{X}}\right)^{-1}\dfrac{\mu_{12}}{\bar{X}S_y^2}-2\dfrac{S_y^2}{\bar{Y}^2}+2\dfrac{S_{xy}}{S_x^2}\left(\dfrac{\bar{Y}}{\bar{X}}\right)^{-1}\dfrac{S_{xy}}{\bar{X}\bar{Y}}}{\dfrac{\mu_{04}}{\mu_{02}^2}-1+4\left(\dfrac{S_y^2}{\bar{Y}^2}-\dfrac{\mu_{03}}{\bar{Y}S_y^2}\right)}$$

and the minimum mean square error is given by (2.11).

The parameters involved in $g_1(1)_{opt}$ may not be known in practice; hence it is better to replace them by their estimators based on the sample observations. Let us denote by $g_1(1)$ as the estimator of $g_1(1)_{opt}$ such that

$$g_1(1) = \frac{\frac{\mu_{03}}{\bar{y}s_y^2} - \frac{s_{xy}}{s_x^2}\left(\frac{\bar{y}}{\bar{x}}\right)^{-1}\frac{\mu_{12}}{\bar{x}s_y^2} - 2\frac{s_y^2}{\bar{y}^2} + 2\frac{s_{xy}}{s_x^2}\left(\frac{\bar{y}}{\bar{x}}\right)^{-1}\frac{s_{xy}}{\bar{x}\,\bar{y}}}{\frac{\mu_{04}}{s_y^4} - 1 + 4\left(\frac{s_y^2}{\bar{y}^2} - \frac{\mu_{03}}{\bar{y}s_y^2}\right)} \tag{3.1}$$

where, defining $m_{pq} = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})^p(y_i - \bar{y})^q$, $\mu_{12} = \frac{n.m_{12}}{n-2}$, $\mu_{03} = \frac{n.m_{03}}{n-2}$ and $\mu_{04} = \frac{n^2.m_{04} - 3(2n-1)s_y^4}{n^2 - 3n + 3}$ are the estimators with their expected values μ_{12}, μ_{03} and μ_{04} respectively (for large population size N and also in case of simple random sampling with replacement).

Now, the estimator $\hat{\hat{Y}}_g$ takes the following form $\hat{\hat{Y}}_g$ under the estimated value $g_1(1)$ of $g_1(1)_{opt}$

$$\hat{\hat{Y}}_g = \left\{\bar{y} + \frac{s_{xy}}{s_x^2}(\bar{X} - \bar{x})\right\} g(w)$$

$$= \left\{\bar{y} + \frac{s_{xy}}{s_x^2}(\bar{X} - \bar{x})\right\}\left\{1 + (w-1)g_1(1)_{opt} + \frac{(w-1)^2}{2!}g_2(1) + \frac{(w-1)^3}{3!}g_3(w^*)\right\} \tag{3.2}$$

In order to obtain the mean square error of $\hat{\hat{Y}}_g$, let us denote by

$$\bar{y} = \bar{Y} + e_0,\ \bar{x} = \bar{X} + e_1,\ s_{xy} = S_{xy} + e_2,$$

$$s_x^2 = S_x^2 + e_3,\ s_y^2 = S_y^2 + e_4,\ \mu_{12} = \mu_{12} + e_5,\ \mu_{03} = \mu_{03} + e_6\ \mu_{04} = \mu_{04} + e_7$$

with $E(e_0) = E(e_1) = E(e_2) = E(e_3) = E(e_4) = E(e_5) = E(e_6) = E(e_7) = 0$ (3.3)

putting these values in (4.2), taking square and taking expectation on both sides we have

$$MSE(\hat{\hat{Y}}_g) = E(\hat{\hat{Y}}_g - \bar{Y})^2$$

$$= \bar{Y}^2 E\left\{ e_0 + \left(e_4 - 2e_0\right) \frac{\lambda_{03} - BR^{-1}\lambda_{12} - 2C_y^2 + 2BR^{-1}\rho C_x C_y}{\beta_{02} - 1 + 4\left(C_y^2 - \lambda_{03}\right)} - BR^{-1}e_1 \right\}^2$$

(to the first order of approximation)

Now, substituting the results given in (2.7), we have

$$MSE(\hat{\bar{Y}}_g) = \gamma_n \bar{Y}^2 \left\{ (1-\rho^2)C_y{}^2 - \frac{\left\{\lambda_{03} - BR^{-1}\lambda_{12} - 2C_y^2 + 2BR^{-1}\rho C_x C_y\right\}^2}{\left\{\beta_{02} - 1 + 4\left(C_y^2 - \lambda_{03}\right)\right\}} \right\}$$

$$= MSE\left(\bar{Y}_g\right)_{\min} \tag{3.4}$$

4. CONCLUDING REMARKS

1. Some particular members belonging to the generalized estimator $\bar{Y}_g$ are

(i) $$\bar{Y}_1 = \left\{\bar{y} + b\left(\bar{X} - \bar{x}\right)\right\}\left\{1 + \frac{k_1\left(C_y^2 - C_y^2\right)}{C_y^2}\right\}$$

(ii) $$\bar{Y}_2 = \left\{\bar{y} + b\left(\bar{X} - \bar{x}\right)\right\}\left(\frac{C_y^2}{C_y^2}\right)^{k_2}$$

(iii) $$\bar{Y}_3 = \left\{\bar{y} + b\left(\bar{X} - \bar{x}\right)\right\}\left\{\frac{C_y^2}{C_y^2 + k_3\left(C_y^2 - C_y^2\right)}\right\}$$

(iv) $$\bar{Y}_4 = \left\{\bar{y} + b\left(\bar{X} - \bar{x}\right)\right\}\left\{2 - \left(\frac{C_y^2}{C_y^2}\right)^{k_4}\right\}$$ etc

where k_1, k_2, k_3, k_4 are the characterizing scalars to be chosen suitably. It may be easily checked that the estimators $\bar{Y}_i$ (i=1,2,3,4) satisfy all the conditions from (i) and (ii) in (2.2) and (2.3), and hence belong to the class of estimators represented by the generalized estimator $\bar{Y}_g$.

2. The proposed generalized class of estimators $\bar{Y}_g$ is biased and its exact bias is given by (2.8). The mean square error of the proposed class is given by (2.9). The

minimum mean square error of the proposed generalized class is given by (2.11). Thus, we have

$$MSE\left(\bar{y}_{lr}\right)-MSE\left(\bar{Y}_g\right)_{\min}=\gamma_n\bar{Y}^2\frac{\left\{\lambda_{03}-BR^{-1}\lambda_{12}-2C_y^2+2BR^{-1}\rho C_xC_y\right\}^2}{\left\{\beta_{02}-1+4\left(C_y^2-\lambda_{03}\right)\right\}}\geq 0$$

Therefore, the proposed generalized class of estimators $\bar{Y}_k$ is preferred to usual linear regression estimator, ratio estimator, product estimator and mean per unit estimator in the sense of lesser mean square error.

3. Also, the parameters involved in the $g_1(1)$ may be estimated by the corresponding sample values in order to get a class of estimators depending upon estimated optimum value. Then, the resultant estimator based on the estimated optimum values of the parameter is shown to have the same mean square error up to the first order of approximation.

REFERENCES

[1] *Bhushan, S. (2013).Improved Sampling Strategies in Finite Population.Scholars Press, Germany.*

[2] *Bhushan S. (2012). Some Efficient Sampling Strategies based on Ratio Type Estimator, Electronic Journal of Applied Statistical Analysis, 5(1), 74 – 88.*

[3] *Bhushan S., Gupta R. and Pandey S. K. (2015). Some log-type classes of estimators using auxiliary information, International Journal of Agricultural and Statistical Sciences, 11(2), 487 – 491.*

[4] *Bhushan S. and Katara, S. (2010). On Classes of Unbiased Sampling Strategies, Journal of Reliability and Statistical Studies, 3(2), 93-101.*

[5] *Bhushan, S. and Kumar S. (2016).Recent advances in Applied Statistics and its applications. LAP Publishing.*

[6] *Bhushan S., Pandey A. and Singh R.K. (2009) "Improved Classes of Regression Type Estimators"; International Journal of Agricultural and Statistical Sciences (ISSN: 0973 – 1903), 5(1), 73 – 84.*

[7] *Bhushan S. and Pandey A. (2010). Modified Sampling Strategies using Correlation Coefficient for Estimating Population Mean, Journal of Statistical Research of Iran, 7(2), 121- 131.*

[8] *Bhushan S., Singh, R. K. and Katara, S. (2009). Improved Estimation under Midzuno – Lahiri – Sen-type Sampling Scheme, Journal of Reliability and Statistical Studies, 2(2), 59 – 66.*

[9] *Bhushan S., Masaldan R. N. and Gupta P. K. (2011). Improved Sampling Strategies based on Modified Ratio Estimator, International Journal of Agricultural and Statistical Sciences, 7(1), 63-75.*

[10] *Cochran W.G. (1977): Sampling Techniques; Third edition, John Wiley and Sons, New York.*

[11] *Gray and Schucany: Generalized Jack – Knife Statistic.*

[12] *Jessen, R.S.(1978): Statistical Survey Techniques, John Wiley &sons, New York, 1978.*

[13] *Singh, D., Chaudhary F.S.(1986): Theory and Analysis of sample Survey Design, New Age Publication, New Delhi, India 1986.*

[14] *Singh G. N. (2003): On the improvement of product method of estimation in sample surveys, Jour. Ind. Soc. Agr. Stat., **56(3)**, 267-275.*

[15] *Sukhatme P. V. and Sukhatme B. V. (1997): Sampling Theory of Surveys with Applications, Piyush Publications, Delhi.*