Advances in Fuzzy Mathematics.
ISSN 0973-533X Volume 12, Number 1 (2017), pp. 147-155

http://www.ripublication.com

On an Unbiased Jack-Knife Regression Type Estimator

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

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

Abstract

For estimation of finite population mean of study variable, regression type estimator using known mean and variance of an auxiliary variable is proposed, its exact bias and mean square error are found. Further using jack-knife technique on the lines of Gray and Schucany and Sukhatme et al, the exact bias of the proposed regression type estimator is eliminated and thus getting the unbiased jack-knifed regression type estimator with the same mean square error as that of the usual biased linear regression estimator. A comparative study of the proposed unbiased jack-knifed regression type estimator with the other estimators shows that the proposed jack-knifed estimator is better in the sense of unbiasedness and mean square error.

Keywords: Auxiliary information, regression type Estimator, Bias, Mean Square Error.

1. INTRODUCTION

The utilization of additional in the form of population mean and variance (standard deviation) of auxiliary variable for increasing the efficiency of the sampling strategy has been discussed recently by Singh (2003), Bhushan (2007), and Bhushan and Katara (2010) among others. One of the major objections against the use of various ratio and regression type estimators is that they are biased and most of times bias found is only upto the first order of approximation. Since the bias of such estimator is approximate therefore the use of jack – knife technique for bias correction only yields an almost unbiased estimator without any loss in efficiency. In this paper an attempt to utilize additional supplementary information in the form of population mean and variance of auxiliary variable so that their proposed estimator's exact bias could be computed, which upon jack-knifing resulted in an unbiased estimator without any loss in efficiency.

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

$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2 = \frac{N}{N-1}\sigma_x^2, S_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y}) = \rho S_x S_y, B = \frac{S_{xy}}{S_x^{\ 2}} = \rho\frac{S_y}{S_x}$, ρ is the population correlation coefficient between 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, $s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ 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.

If the prior information about population mean $\bar{X}$ of auxiliary variable x is known then the difference estimator (Ref: Cochran (1977)) is given by

$$\bar{y}_D = \bar{y} + \beta(\bar{x} - \bar{X}) \tag{1.1}$$

where β is the characterizing scalar to be determined suitably. Note that for $\beta = 0$ the proposed estimator reduces to the usual mean per unit estimator. The difference estimator is an unbiased estimator of population mean of the study variable for all values of the constant β. The mean square error is given by

$$MSE(\bar{y}_D) = \gamma_n \left(S_y^{\ 2} + k^2 S_x^{\ 2} + 2k\rho S_x S_y\right) \tag{1.2}$$

and the optimizing value of the scalar minimizing mean square error is given by

$$\beta = \frac{S_{xy}}{S_x^2} = \rho\frac{S_y}{S_x} = B \tag{1.3}$$

The minimum mean square error under the optimizing value is given by

$$MSE\left(\bar{y}_D\right)_{\min} = \gamma_n\left(1-\rho^2\right)S_y^{\ 2} \tag{1.4}$$

Since the optimizing value of the characterizing scalar depends on the population parameters, it is replaced by the unbiased estimators of the parameters involved and thus getting the celebrated linear regression estimator

$$\bar{y}_{lr} = \bar{y} + b(\bar{X} - \bar{x}) \quad \text{where } b = \frac{s_{xy}}{s_x^2}. \tag{1.5}$$

The bias and mean square error of $\bar{y}_{lr}$ is given by

$$Bias(\bar{y}_{lr}) = E(\bar{y}_{lr}) - \bar{Y} = \gamma_n B\left(\frac{\mu_{30}}{S_x^2} - \frac{\mu_{21}}{S_{xy}}\right) \quad \text{and} \tag{1.6}$$

$$MSE(\bar{y}_{lr}) = E\left(\bar{y}_{lr} - \bar{Y}\right)^2 = \gamma_n\left(1-\rho^2\right)S_y^2 \quad \text{respectively.} \tag{1.7}$$

Recently, a product estimator, using prior information about population mean and population standard deviation (variance) of auxiliary variable, based on Singh (2003) is $\bar{y}_s = \bar{y}\dfrac{\bar{x}+\sigma_x}{\bar{X}+\sigma_x}$ having MSE given by

$$MSE(\bar{y}_s) = \gamma_n\left(S_y^{\ 2} + R^2\delta^2 S_x^{\ 2} + 2R\delta S_{xy}\right) \quad \text{where } \delta = \frac{\bar{X}}{\bar{X}+\sigma_x}. \tag{1.8}$$

Further, a generalized Singh (2003) estimator is given by

$$\bar{y}_{s\alpha} = \bar{y}\left(\frac{\bar{x}+\sigma_x}{\bar{X}+\sigma_x}\right)^{\alpha}, \text{ where } \alpha \text{ is a suitably chosen scalar.} \tag{1.9}$$

The MSE of $\bar{y}_{s\alpha}$ is given by $MSE(\bar{y}_{s\alpha}) = \gamma_n\left(S_y^{\ 2} + \alpha^2 R^2\delta^2 S_x^{\ 2} + 2\alpha R\delta S_{xy}\right)$ where

$$\delta = \frac{\bar{X}}{\bar{X}+\sigma_x} \tag{1.10}$$

and the minimum mean square error is given by $MSE(\bar{y}_{s\alpha})_{\min} = \gamma_n\left(1-\rho^2\right)S_y^{\ 2}$ (1.11)

The Singh (2003) estimators are biased and minimum mean square error is equal to that of linear regression estimator.

2. PROPOSED ESTIMATOR

A proposed estimator of population mean $\bar{Y}$ is $\bar{Y} = \bar{y} + \frac{s_{xy}}{S_x^2}(\bar{X} - \bar{x})$ (2.1)

Let $\bar{y} = \bar{Y} + e_0$, $\bar{x} = \bar{X} + e_1, s_{xy} = S_{xy} + e_2$ with $E(e_0) = E(e_1) = E(e_2) = 0$ (2.2)

Then by using (1), we have $\hat{\bar{Y}} = \bar{Y} + e_0 + B\left[-e_1 - \frac{e_1 e_2}{S_{xy}}\right]$ (2.3)

Using the results given in Sukhatme and Sukhatme (1997) and Bhushan (2007), we have

$$E(e_1 e_2) = \gamma_n \mu_{21},\ E(e_0{}^2) = \gamma_n S_y{}^2, E(e_1{}^2) = \gamma_n S_x{}^2,\ E(e_0 e_1) = \gamma_n S_{xy} \quad (2.4)$$

where $\gamma_n = (N-n)/Nn$. Using (4) and (5) it can be seen that $Bias(\bar{Y}) = E(\bar{Y}) - \bar{Y} = -\gamma_n B \frac{\mu_{21}}{S_{xy}}$ (2.5)

Showing that $\bar{Y}$ is a biased estimator of population mean and its exact bias is given by (2.5). Using (2.3) and neglecting terms of e_i's having powers greater than two, we get the MSE given by

$$MSE\left(\bar{Y}\right) = \gamma_n (1-\rho^2) S_y{}^2 \quad (2.6)$$

which is equal to the $MSE(\bar{y}_D)_{\min} = MSE(\bar{y}_{lr}) = \gamma_n \left(1-\rho^2\right) S_y{}^2$.

3. PROPOSED JACK-KNIFE ESTIMATOR

Let a simple random sample of size $m = 2n$ is drawn without replacement from the population of size N. This sample of size $m = 2n$ is then split up at random into two sub-samples each of size n. Let us define:

$$\hat{\bar{Y}}^{(1)} = \bar{y}_n^{(1)} + \frac{s_{xy}^{(1)}}{S_x^2}(\bar{X} - \bar{x}^{(1)}),\ \hat{\bar{Y}}^{(2)} = \bar{y}_n^{(2)} + \frac{s_{xy}^{(2)}}{S_x^2}(\bar{X} - \bar{x}^{(2)})\ \hat{\bar{Y}}^{(3)} = \bar{y}_{2n} + \frac{s_{xy}}{S_x^2}(\bar{X} - \bar{x}_{2n}) \quad (3.1)$$

where $\bar{y}_n^{(1)}$, $\bar{y}_n^{(2)}$, $\bar{y}_{2n}$, be the respective sample means based on two sub-samples of size n and the entire sample of size $m = 2n$ for the characteristic y under study; $\bar{x}_n^{(1)}$, $\bar{x}_n^{(2)}$, $\bar{x}_{2n}$ be the corresponding means for the auxiliary variable x.

It can be easily seen that

$$B\left(\hat{\bar{Y}}^{(1)}\right) = -\gamma_n B \frac{\mu_{21}}{S_{xy}},\ B\left(\hat{\bar{Y}}^{(2)}\right) = -\gamma_n B \frac{\mu_{21}}{S_{xy}},\ B\left(\hat{\bar{Y}}^{(3)}\right) = -\gamma_{2n} B \frac{\mu_{21}}{S_{xy}} = B_1(say) \quad (3.2)$$

Let us define:

$$\hat{\hat{Y}}' = \frac{\hat{\hat{Y}}^{(1)} + \hat{\hat{Y}}^{(2)}}{2}$$ be an alternative estimator of the population mean $\bar{Y}$. (3.3)

The bias of $\hat{\hat{Y}}'$ is given by: $B\left(\hat{\hat{Y}}'\right) = -\gamma_n B \dfrac{\mu_{21}}{S_{xy}} = B_2 (say)$

In this paper we propose to use the following jack-knife estimator for estimating the population mean $\bar{Y}$ as:

$$\hat{\hat{Y}}_J = \frac{\hat{\hat{Y}}^{(3)} - R\hat{\hat{Y}}'}{1-R} \qquad \text{with } R = B_1/B_2 \tag{3.4}$$

Taking expectation of (2.4), we have $E\left(\hat{\hat{Y}}_J\right) = \dfrac{E\left(\hat{\hat{Y}}^{(3)}\right) - RE\left(\hat{\hat{Y}}'\right)}{1-R} = \bar{Y}$ (3.5)

Showing that $\hat{\hat{Y}}_J$ is an unbiased estimate of population mean $\bar{Y}$.

$$MSE\left(\hat{\hat{Y}}_J\right) = E\left(\frac{\hat{\hat{Y}}^{(3)} - R\hat{\hat{Y}}'}{1-R}\right)^2 = \frac{1}{(1-R)^2}\left\{E\left(\hat{\hat{Y}}^{(3)} - \bar{Y}\right)^2 + R^2 E\left(\hat{\hat{Y}}' - \bar{Y}\right)^2 - 2RE\left(\hat{\hat{Y}}^{(3)} - \bar{Y}\right)\left(\hat{\hat{Y}}' - \bar{Y}\right)\right\} \tag{3.6}$$

Now consider

$$E\left(\hat{\hat{Y}}^{(3)} - \bar{Y}\right)^2 = MSE\left(\hat{\hat{Y}}^{(3)}\right) = \gamma_{2n}\left(1-\rho^2\right)S_y^2 \tag{3.7}$$

Also

$$E\left(\hat{\hat{Y}}' - \bar{Y}\right)^2 = E\left(\frac{\hat{\hat{Y}}^{(1)} + \hat{\hat{Y}}^{(2)}}{2} - \bar{Y}\right)^2 = \frac{1}{4}E\left\{\left(\hat{\hat{Y}}^{(1)} - \bar{Y}\right) + \left(\hat{\hat{Y}}^{(2)} - \bar{Y}\right)\right\}^2$$

$$= \frac{1}{4}\left\{E\left(\hat{\hat{Y}}^{(1)} - \bar{Y}\right)^2 + E\left(\hat{\hat{Y}}^{(2)} - \bar{Y}\right)^2 + 2E\left(\hat{\hat{Y}}^{(1)} - \bar{Y}\right)\left(\hat{\hat{Y}}^{(2)} - \bar{Y}\right)\right\} \tag{3.8}$$

Since

$$MSE\left(\hat{\hat{Y}}^{(i)}\right) = \gamma_n\left(1-\rho^2\right)S_y^2 ; i = 1, 2 \tag{3.9}$$

To evaluate it, consider

$\bar{y}_n^{(i)} = \bar{Y} + e_0^{(i)}$, $\bar{x}_n^{(i)} = \bar{X} + e_1^{(i)}$, $s_{xy}^{(i)} = S_{xy} + e_2^{(i)}$, $s_x^{(i)2} = S_x^2 + e_3^{(i)}$; $i = 1, 2$

With $E\left(e_0^{(i)}\right) = E\left(e_1^{(i)}\right) = E\left(e_2^{(i)}\right) = E\left(e_3^{(i)}\right) = 0;\ i = 1, 2$

Then

$$\hat{\bar{Y}}^{(i)} = \bar{Y} + e_0^{(i)} + B\left(-e_1^{(i)} - \frac{e_1^{(i)}e_2^{(i)}}{S_{xy}}\right); i = 1, 2 \tag{3.10}$$

Hence

$$E\left(\hat{\bar{Y}}^{(1)} - \bar{Y}\right)\left(\hat{\bar{Y}}^{(2)} - \bar{Y}\right) = E\left\{\left(e_0^{(1)} - Be_1^{(1)}\right)\left(e_0^{(2)} - Be_1^{(2)}\right)\right\} \quad \text{(to the first order of approximation)}$$

$$= E\left(e_0^{(1)}e_0^{(2)}\right) - BE\left(e_0^{(1)}e_1^{(2)} + e_0^{(2)}e_1^{(1)}\right) + B^2E\left(e_1^{(1)}e_1^{(2)}\right) \tag{3.11}$$

Substituting the following results given in Sukhatme and Sukhatme (1997), we have

$$E\left(e_0^{(1)}e_0^{(2)}\right) = -\frac{1}{N}S_y^2, \; E\left(e_1^{(1)}e_1^{(2)}\right) = -\frac{1}{N}S_x^2, \text{ and } \; E\left(e_0^{(1)}e_1^{(2)}\right) = E\left(e_0^{(2)}e_1^{(1)}\right) = -\frac{1}{N}S_{xy}. \tag{3.12}$$

We have $E\left(\hat{\bar{Y}}^{(1)} - \bar{Y}\right)\left(\hat{\bar{Y}}^{(2)} - \bar{Y}\right) = -\frac{1}{N}\left(S_y^2 - 2BS_{xy} + B^2S_x^2\right) = -\frac{1}{N}\left(1-\rho^2\right)S_y^2$ (3.13)

Putting the values from (3.9) and (3.13) in (3.8) we have

$$E\left(\hat{\bar{Y}}' - \bar{Y}\right)^2 = \frac{1}{4}\left[2\gamma_n - \frac{2}{N}\right]\left(1-\rho^2\right)S_y^2 = \gamma_n\left(1-\rho^2\right)S_y^2 \tag{3.14}$$

Now consider

$$E\left(\hat{\bar{Y}}^{(3)} - \bar{Y}\right)\left(\hat{\bar{Y}}' - \bar{Y}\right) = E\left\{\left(\hat{\bar{Y}}^{(3)} - \bar{Y}\right)\left(\frac{\left(\hat{\bar{Y}}^{(1)} + \hat{\bar{Y}}^{(2)}\right)}{2} - \bar{Y}\right)\right\}$$

$$= \frac{1}{2}\left[E\left(\hat{\bar{Y}}^{(3)} - \bar{Y}\right)\left(\hat{\bar{Y}}^{(1)} - \bar{Y}\right) + E\left(\hat{\bar{Y}}^{(3)} - \bar{Y}\right)\left(\hat{\bar{Y}}^{(2)} - \bar{Y}\right)\right] \tag{3.15}$$

$$E\left(\hat{\bar{Y}}^{(3)} - \bar{Y}\right)\left(\hat{\bar{Y}}^{(i)} - \bar{Y}\right) = E\left(e_0 - Be_1\right)\left(e_0^{(i)} - Be_1^{(i)}\right)$$

$$= E\left(e_0e_0^{(i)}\right) - BE\left(e_0e_1^{(i)} + e_0^{(i)}e_1\right) + B^2E\left(e_1e_1^{(i)}\right) \quad ; \; i = 1, 2$$

Substituting the following results given in Sukhatme and Sukhatme (1997) and Bhushan (2007)

$$E\left(e_0e_0^{(i)}\right) = \gamma_{2n}S_y^2, \; E\left(e_1e_1^{(i)}\right) = \gamma_{2n}S_x^2, \text{ and } \; E\left(e_0e_1^{(i)}\right) = E\left(e_0^{(i)}e_1\right) = \gamma_{2n}S_{xy} \; ; \; i = 1, 2 \tag{3.16}$$

$$E\left(\hat{\bar{Y}}^{(3)} - \bar{Y}\right)\left(\hat{\bar{Y}}^{(i)} - \bar{Y}\right) = \gamma_{2n}\left(1-\rho^2\right)S_y^2; i = 1, 2 \tag{3.17}$$

From (3.15) and (3.17) we have

$$E\left(\hat{\bar{Y}}^{(3)}-\bar{Y}\right)\left(\hat{\bar{Y}}'-\bar{Y}\right)=\gamma_{2n}\left(1-\rho^2\right)S_y^2 \tag{3.18}$$

Substituting the values from (3.7), (3.14) and (3.18) in (3.6), we have

$$MSE\left(\hat{\bar{Y}}_J\right)=\frac{1}{(1-R)^2}\left(1+R^2-2R\right)\gamma_{2n}\left(1-\rho^2\right)S_y^2=\gamma_{2n}\left(1-\rho^2\right)S_y^2 \tag{3.19}$$

4. CONCLUDING REMARKS

1. The proposed estimator $\bar{Y}$ given in (2.1) is a biased estimator and its exact bias is given by (2.6). The mean square error of the proposed class is given by (2.7). Therefore, the proposed estimator $\bar{Y}$ is preferred to usual ratio estimator, product estimator, mean per unit estimator and generalized Singh (2003) estimator in the sense of lesser mean square error.

2. Since the exact bias of the proposed estimator has been computed thus we use the generalized jack knife technique due to Quenouille (1956) and Gray and Schucany (1976) to obtain a perfectly unbiased estimator $\bar{Y}_J$ given in (3.4) having the same mean square error as that of the $\bar{y}_{lr}$ and the minimum mean square error of $\bar{y}_D$. Therefore, the proposed estimator $\bar{Y}_J$ is preferred to usual linear regression estimator, ratio estimator, product estimator, mean per unit estimator and generalized Singh (2003) estimator in the sense of unbiasedness and lesser mean square error.

3. It may be pointed out here that the jack – knife estimator $\bar{Y}_J$ given in (3.4) is exactly unbiased while the jack – knife counterparts of the linear regression estimator, ratio estimator, product estimator, and generalized Singh (2003) estimator would only be almost unbiased and not exactly unbiased. Thus, the proposed jack – knife estimator $\bar{Y}_J$ is preferred to usual linear regression estimator, ratio estimator, product estimator, mean per unit estimator and generalized Singh (2003) estimator in the sense of unbiasedness and lesser mean square error.

4. The MSE of product estimator based on Singh (2003) is given by

$$MSE(\bar{y}_s)=\gamma_n\left(S_y{}^2+R^2\delta^2S_x{}^2+2R\delta S_{xy}\right) \text{ where } \delta=\frac{\bar{X}}{\bar{X}+\sigma_x} \text{ such that}$$

$$MSE(\bar{y}_s)-MSE(\bar{Y})=\gamma_n(\rho S_y+R\delta S_x)^2\geq 0$$

showing that the proposed estimator is better than the Singh (2003) estimator. Further, the MSE of generalized Singh (2003) estimator is given by

$MSE(\bar{y}_{s\alpha}) = \gamma_n \left(S_y^{\ 2} + \alpha^2 R^2 \delta^2 S_x^{\ 2} + 2\alpha R \delta S_{xy} \right)$ where $\delta = \dfrac{\bar{X}}{\bar{X} + \sigma_x}$

Therefore, we have

$$MSE(\bar{y}_{s\alpha}) - MSE(\bar{Y}) = \gamma_n (\rho S_y + \alpha R \delta S_x)^2 \geq 0$$

and the minimum mean square error is given by

$$MSE(\bar{y}_{s\alpha})_{\min} = \gamma_n \left(1 - \rho^2\right) S_y^{\ 2}$$

such that $MSE(\bar{y}_{s\alpha})_{\min} = MSE(\bar{Y})$.

showing that the proposed estimator is better than the generalized Singh (2003) estimator as the optimizing value may or may not be attained in practice.

REFERENCES

[1] *Abu-Dayyeh,W.A. (2003) :Some estimators of a finite population mean using auxiliary information, Applied Mathematics and Computation,* ***139****, 287-298.*

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

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

[4] *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.*

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

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

[7] *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.*

[8] *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.*

[9] *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.*

[10] *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.*

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

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

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

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

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

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