Variance Estimation Using Linear Combination of Tri-mean and Quartile Average

Showkat Maqbool[1], Shakeel Javaid[2]

[1]Division of Agricultural Statistics and Economics, FOA, Wadura, Skuast, Kashmir, India

[2]Department of Statistics & O.R., A.M.U., Aligarh, India

Email address:
showkatmaq@gmail.com (S. Maqbool), shakeel.jd@operamail.com (S. Javaid)

Abstract: In this paper, we have proposed a class of modified ratio type variance estimator for estimation of population variance of the study variable, when Tri Mean and Quartile average of the auxiliary variable are known. The bias and mean square error (MSE) of the proposed estimator are obtained. From the numerical study it is observed that the proposed estimator performs better than the existing estimators in the literature.

Keywords: Simple Random Sampling, Bias, Mean Square Error, Tri-mean, Quartile Average, Auxiliary Variable

1. Introduction

The use of an auxiliary variable x in the estimation of the finite population total or mean of a characteristic y is a common occurrence in practice. The theory of basic sample survey as available in standard text books on sampling deals with the case which comprises linear estimators such as mean, total, proportion and simple sample design using simple random sampling with replacement. It continues to supply new and improved procedures for estimation of variances assuming independence of observations. Ratio, difference and regression estimators utilize an auxiliary variable for more efficient estimation of the parameter in question. Such estimators take advantage of the correlation between the x variable and the characteristic y. In a similar manner, then, it seems reasonable that under suitable conditions efficient estimation of the variance of the estimator of the finite population total or mean of the characteristic y is also possible using such estimation techniques. 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 (1970)). A basic requirement of a good survey is that a measure of precision is provided for each estimate computed from survey data collected on the basis of the survey design. An important question is how to choose an appropriate variance estimator. The choice in general is very difficult. Factors like accuracy of the variance estimator, timeliness, cost, simplicity and other administrative conveniences must be considered. Here, we consider a finite population $U = \{U_1, U_2, ..., U_N\}$ of N distinct and identifiable units. Let Y be a real variable with value Y_i measured on U_i, $i=1,2,3,…,N$ giving a vector $Y = \{y_1, y_2, ..., y_N\}$. The goal is to estimate the population

means $\bar{Y} = \dfrac{1}{N} \sum_{i=1}^{N} y_i$ or its variance $S_Y^2 = \dfrac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{y})^2$

on the basis of random sample selected from the population U. Sometimes in sample surveys along with the study variable Y, information on auxiliary variable X, which is positively correlated with Y, is also available. The information on auxiliary variable X, may be utilized to obtain a more efficient estimator of the population. In this paper, our aim is to estimate the population variance on the basis of a random sample of size n selected from the population U. Ogus and Clark (1971) proposed the use of ratio or difference estimators of the variance under a passion sampling design (a design in which each sampling unit is given an independent chance of being selected into the sample without replacement) for the purpose of reducing the effect of the random sample size on the variance estimator. Under a sample design in which one sample unit is selected in each stratum with probability proportional to size (PPS). Shapiro and Bateman (1978) considered reducing the bias of the

estimator of the variance in a one-per stratum design by using a variance estimator as the Yates Grundy variance estimator for a two sample units per stratum design with joint inclusion probabilities. The problem of constructing efficient estimators for the population variance has been widely discussed by various authors such as Isaki (1983), who proposed ratio and regression estimators. Prasad and Singh (1990) considered a ratio type estimator for estimating population variance by improving Isaki's (1983) estimator. Gupta and Shabir (2008) proposed a new hybrid class of estimators and showed that in some cases their efficiency is better than the traditional ratio estimators, where as Kadilar & Cingi (2006,2007) proposed the modified estimators using coefficient of variation (C.V.) and their linear combinations. Subramani & Kumarapandiyan (2015) improved the already existing estimators by introducing modified estimators with the use of known parameters like C.V., Kurtosis, Median, Quartiles and Deciles. Recently, Maqbool *et al.* (2016) proposed a modified ratio estimator using non-conventional location parameters because these parameters take care of outliers in the data. Outliers can be generated by from a simple operational mistake to including small sample from a different population, and they make serious effects of statistical inference. Even one outlying observation can destroy least squares estimation, resulting in parameter estimates that do not provide useful information for the majority of the data.

2. Notations

Let N = population size, n = sample size, $\gamma = \dfrac{1}{n}$, Y= study variable, X= auxiliary variable. $\overline{X}$, $\overline{Y}$ = population means, $\overline{x}$, $\overline{y}$ = sample means. S_Y^2, S_X^2 = population variances, s_y^2, s_x^2 = sample variances. C_x, C_y = coefficient of variation, ρ = correlation coefficient, $\beta_{1(x)}$ = skewness of the auxiliary variable, $\beta_{2(x)}$ = kurtosis of the auxiliary variable, $\beta_{2(y)}$ = kurtosis of the study variable, M_d= median of the auxiliary variable, $B(.)$=bias of the estimator, MSE(.)= Mean square error, $\hat{S}_R^2$ = ratio type variance estimator, $\hat{S}_{Kc1}^2$, $\hat{S}_{jG}^2$ = existing modified ratio estimators, Q_a= $(Q_3+Q_1)/2$ is population semi-quartile average of the auxiliary variable x, TM=$(Q_1+2Q_2+Q_3)/4$is a Tri-Mean, S_j= proposed estimator by Showkat and Javaid.

3. Ratio Type Variance Estimator Proposed by Isaki (1983)

Isaki (1983) suggested a ratio type variance estimator for the population variance S_Y^2 when the population variance S_x^2 of the auxiliary variable X is known. Its bias and mean square error are given by

$$\hat{S}_R^2 = s_y^2 \frac{S_x^2}{s_x^2} \tag{1}$$

$$\text{Bias}(\hat{S}_R^2) = \gamma S_y^2 \left[\left(\beta_{2(x)} - 1 \right) - \left(\lambda_{22} - 1 \right) \right]$$

$$\text{MSE}(\hat{S}_R^2) = \gamma S_y^4 \left[\left(\beta_{2(y)} - 1 \right) + \left(\beta_{2(x)} - 1 \right) - 2 \left(\lambda_{22} - 1 \right) \right]$$

Here γ is the sampling fraction and λ_{22} is covariance between the study and auxiliary variable.

4. Kadilar and Cingi (2006) Estimators

Kadilar and Cingi (2006) suggested four ratio type variance estimators using known values of C.V. and coefficient of kurtosis of an auxiliary variable

$$\hat{S}_{kc1}^2 = s_y^2 \left[\frac{S_x^2 + C_x}{s_x^2 + C_x} \right] \tag{2}$$

$$\text{Bias}(\hat{S}_{kc1}^2) = \gamma S_y^2 A_1 \left[A_1 \left(\beta_{2(x)} - 1 \right) - \left(\lambda_{22} - 1 \right) \right]$$

$$\text{MSE}(\hat{S}_{kc1}^2) =$$
$$\gamma S_y^4 \left[\left(\beta_{2(y)} - 1 \right) + A_1^2 \left(\beta_{2(x)} - 1 \right) - 2A_1 \left(\lambda_{22} - 1 \right) \right]$$

$$\hat{S}_{kc2}^2 = s_y^2 \left[\frac{S_x^2 + \beta_{2(x)}}{s_x^2 + \beta_{2(x)}} \right] \tag{3}$$

$$\text{Bias}(\hat{S}_{kc2}^2) = \gamma S_y^2 A_2 \left[A_2 \left(\beta_{2(x)} - 1 \right) - \left(\lambda_{22} - 1 \right) \right]$$

$$\text{MSE}(\hat{S}_{kc2}^2) =$$
$$\gamma S_y^4 \left[\left(\beta_{2(y)} - 1 \right) + A_2^2 \left(\beta_{2(x)} - 1 \right) - 2A_2 \left(\lambda_{22} - 1 \right) \right]$$

$$\hat{S}_{kc3}^2 = s_y^2 \left[\frac{S_x^2 \beta_{2(x)} + C_x}{s_x^2 \beta_{2(x)} + C_x} \right] \tag{4}$$

$$\text{Bias}(\hat{S}_{kc3}^2) = \gamma S_y^2 A_3 \left[A_3 \left(\beta_{2(x)} - 1 \right) - \left(\lambda_{22} - 1 \right) \right]$$

$$\text{MSE}(\hat{S}_{kc3}^2) =$$
$$\gamma S_y^4 \left[\left(\beta_{2(y)} - 1 \right) + A_3^2 \left(\beta_{2(x)} - 1 \right) - 2A_3 \left(\lambda_{22} - 1 \right) \right]$$

$$\hat{S}_{kc4}^2 = s_y^2 \left[\frac{S_x^2 C_x + \beta_{2(x)}}{s_x^2 C_x + \beta_{2(x)}} \right] \tag{5}$$

$$\text{Bias}(\hat{S}_{kc4}^2) = \gamma S_y^2 A_4 \left[A_4 \left(\beta_{2(x)} - 1 \right) - \left(\lambda_{22} - 1 \right) \right]$$

$$\text{MSE}(\hat{S}_{kc4}^2) =$$
$$\gamma S_y^4 \left[\left(\beta_{2(y)} - 1 \right) + A_4^2 \left(\beta_{2(x)} - 1 \right) - 2A_4 \left(\lambda_{22} - 1 \right) \right]$$

Here

$$A_1 = \frac{S_x^2}{s_x^2 + c_x}, \quad A_2 = \frac{S_x^2}{s_x^2 + \beta_{2(x)}}, \quad A_3 = \frac{S_x^2 \beta_{2(x)}}{s_x^2 \beta_{2(x)} + c_x}, \quad A_4 = \frac{S_x^2 C_x}{s_x^2 c_x + \beta_{2(x)}}$$

5. Recent Developments

Subramani and Kumarapandiyan (2015) proposed a generalized modified ratio type estimator for estimating population variance using the known parameters of the auxiliary variable and their estimator is given as

$$\hat{S}_{jG}^2 = s_y^2 \left[\frac{S_x^2 + \alpha w_i}{s_x^2 + \alpha w_i} \right] \qquad (6)$$

The bias and MSE of the estimator in (6) as suggested by the Subramani and kumarapandiyan (2015) is given as

$$\text{Bias}(\hat{S}_{jG}^2) = \gamma S_y^2 A_{jG} \left[A_{jG}(\beta_{2(x)} - 1) - (\lambda_{22} - 1) \right]$$

$$\text{MSE}(\hat{S}_{jG}^2) =$$
$$\gamma S_y^4 \left[(\beta_{2(y)} - 1) + A_{jG}^2 (\beta_{2(x)} - 1) - 2A_{jG}(\lambda_{22} - 1) \right]$$

$$A_{jG} = \frac{S_x^2}{s_x^2 + \alpha \omega_i}$$

When the study variable Y and the auxiliary variable X are negatively correlated and the population parameters of the auxiliary variable are known, Subramani & Kumarapandiyan (2015) proposed the following generalized modified product type variance estimator as

$$\hat{S}_{jG1}^2 = s_y^2 \left[\frac{S_x^2 + \tau w_i}{s_x^2 + \tau w_i} \right], i = 1, 2, ---.$$

The notations in the estimator $\hat{S}_{jG1}^2$ are explained in detail in Subramani & Kumarapandiyan (2015).

When $\alpha = 0$ in Equation (6), $\hat{S}_{jG1}^2$ reduces to Isaki (1983) estimator.

When $\alpha = 1$ in Equation (6), $\hat{S}_{jG1}^2$ reduces to Kadilar and Cingi (2006) estimator.

6. Proposed Estimator

The accuracy of the variance estimator is an important criterion for the choice of a variance estimator and it can be assessed by mean square error (MSE) of the variance estimator that has the best statistical properties for the proposed analysis and interpretation of data. Compromises may have to be made because different analysis of the same data may suggest different variance estimator. In short, it is, therefore seen that the process of evaluating alternative variance estimator and selecting a specific estimator for use in particular complex survey in a difficult and knotty problem, involving both subjective and objective elements. Keen judgment in making intelligent compromises between the consideration of accuracy, cost, timeliness and simplicity is required. The performance of the estimator of the study variable can be improved by using known population parameters of an auxiliary variable, which are positively correlated with a study variable. We have proposed a new modified ratio type variance estimator of the auxiliary variable by using linear combination of Tri-mean and population semi inter quartile 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)).

$$\hat{S}_{Sj}^2 = s_y^2 \left[\frac{S_x^2 + (TM + Q_a)}{s_x^2 + (TM + Q_a)} \right] \qquad (7)$$

We have derived here the bias and mean square error of the proposed estimator $\hat{S}_{SJ}^2$ to first order of approximation as given below. The Taylor series linearization approach derives a linear approximation to the survey estimator and then obtains variance estimate for this linear combination taking account of the sampling design.

Let $e_0 = \dfrac{s_y^2 - S_y^2}{S_y^2}$ and $e_1 = \dfrac{s_x^2 - S_x^2}{S_x^2}$. Further we can write $s_y^2 = S_y^2(1 + e_0)$ and $s_x^2 = S_x^2(1 + e_0)$ and from the definition of e_0 and e_1 we obtain:

$$E[e_0] = E[e_1] = 0, \ E[e_0^2] = \frac{1-f}{n}(\beta_{2(y)} - 1),$$

$$E[e_1^2] = \frac{1-f}{n}(\beta_{2(x)} - 1), \ E[e_0 e_1] = \frac{1-f}{n}(\lambda_{22} - 1).$$

The proposed estimator $\hat{S}_{SJ}^2$ is given below:

$$\hat{S}_{Sj}^2 = s_y^2 \left[\frac{S_x^2 + (TM + Q_a)}{s_x^2 + (TM + Q_a)} \right]$$

$$\Rightarrow \hat{S}_{SJ}^2 = s_y^2(1 + e_0) \left[\frac{S_x^2 + H}{s_x^2 + e_1 S_x^2 + H} \right], \text{ where } H = TM + Q_a$$

$$\Rightarrow \hat{S}_{SJ}^2 = \frac{S_y^2(1 + e_0)}{(1 + A_{SJ} e_1)}, \text{ where } A_{SJ} = \frac{S_X^2}{S_X^2 + (TM + Q_a)}$$

$$\Rightarrow \hat{S}_{SJ}^2 = S_y^2(1 + e_0)(1 + A_{SJ} e_1)^{-1}$$

$$\Rightarrow \hat{S}_{SJ}^2 = S_y^2(1 + e_0)(1 - A_{SJ} e_1 + A_{SJ}^2 e_1^2 - A_{SJ}^3 e_1^3 + \ldots)$$

Expanding and neglecting the terms more than the 3rd order, we get

$$\hat{S}_{SJ}^2 = S_y^2 + S_y^2 e_0 - S_y^2 A_{SJ} e_1 - S_y^2 A_{SJ} e_0 e_1 + S_y^2 A_{SJ}^2 e_1^2$$

$$\Rightarrow \hat{S}_{SJ}^2 - S_y^2 = S_y^2 e_0 - S_y^2 A_{SJ} e_1 - S_y^2 A_{SJ} e_0 e_1 + S_y^2 A_{SJ}^2 e_1^2 \quad (8)$$

By taking expectation on both sides of (8), we get

$$E(\hat{S}_{SJ}^2 - S_y^2) = S_y^2 E(e_0) - S_y^2 A_{SJ} E(e_1) - S_y^2 A_{SJ} E(e_0 e_1) + S_y^2 A_{SJ}^2 E(e_1^2)$$

$$Bias(\hat{S}_{SJ}^2) = S_y^2 A_{SJ}^2 E(e_1^2) - S_y^2 A_{SJ} E(e_0 e_1)$$

$$Bias(\hat{S}_{SJ}^2) = \gamma S_y^2 A_{SJ} [A_{SJ}(\beta_{2(x)} - 1) - (\lambda_{22} - 1)] \quad (9)$$

Squaring both sides of (8) and (9), neglecting the terms more than 2nd order and taking expectation, we get

$$E(\hat{S}_{SJ}^2 - S_y^2)^2 = S_y^4 E(e_0^2) + S_y^4 A_{SJ}^2 E(e_1^2) - 2 S_y^4 A_{SJ} E(e_0 e_1)$$

The bias and mean square error of the above estimator in (7) after simplification is given as

$$\mathrm{Bias}(\hat{S}_{SJ}^2) = \gamma S_y^2 A_{SJ} \left[A_{SJ}(\beta_{2(x)} - 1) - (\lambda_{22} - 1) \right]$$

$$\mathrm{MSE}(\hat{S}_{SJ}^2) = \gamma S_y^4 \left[(\beta_{2(y)} - 1) + A_{SJ}^2 (\beta_{2(x)} - 1) - 2 A_{SJ}(\lambda_{22} - 1) \right]$$

where $A_{SJ} = \dfrac{S_X^2}{S_X^2 + (TM + Q_a)}$.

7. Numerical Illustration

The performance of the proposed estimator is assessed with that of simple random sampling without replacement (SRSWOR) sample variance and 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:

$N= 80$, $S_x=8.4542$, $A_2=0.9615$, $n=20$, $C_x=0.7507$, $A_3=0.9964$, $\overline{X} = 11.2624$, $\beta_{2(x)} = 2.8664$, $A_4 =0.9493$, $\overline{Y} = 51.8264$, $\beta_{2(y)} = 2.2667$, $A_{jG}=0.8763$, $\rho = 0.9413$, $\beta_{1(x)} = 1.05$, $\lambda_{22} = 2.2209$, $S_y=18.3569$, $Md=7.5750$, $Q_1=9.318$, $C_y=0.3542$, $A_1=0.9896$, $Q_2 =7.5750$, $Q_3 =16.975$, $Q_D =5.9125$, $Q_a=11.0625$, $Q_R=11.82$, $TM=9.318$, Tri-mean (TM)=9.318

The results obtained are shown in Table 1.

Table 1. *Bias and Mean Square Error of the existing and the proposed estimators.*

Estimators	Bias	Mean Square Error
Isaki (1983)	10.8762	3925.1622
Kadilar&Cingi(2006) 1	10.4399	3850.1552
Kadilar&Cingi(2006) 2	9.2918	3658.4051
Kadilar&Cingi(2006) 3	10.7222	3898.5560
Kadilar&Cingi(2006) 4	8.8117	3580.8342
Subramani&Kumarapandiyan(2015)	6.1235	3180.7740
Proposed Estimator (Maqbool & Shakeel)	3.0289	2820.06

8. Conclusion

The paper proposes a ratio type variance estimator using known values of an auxiliary variable. From Table 1 we see that the bias and MSE of the existing estimators ranges from 6.12 to10.87 and 3180 to 3925 respectively, while as the proposed estimator suggested by Maqbool & Shakeel has a bias of 3.02 and MSE 2820.06 which show that the bias and mean square error are less than the already existing estimators in the literature and the percentage relative efficiency comes out be 139.18%. Hence the proposed estimator may be preferred over existing estimators for use in practical applications. In future work, we hope to adopt the estimator presented here to stratified random sampling and hope to develop a variance estimator using two auxiliary variables.

Acknowledgement

References

[1] Cochran, W. G. (194). Sampling Techniques. Third Edition, Wiley Eastern limted.

[2] Gupta, Sat and Shabbir, Javid (2008). Variance estimation in simple random sampling using auxiliary information. Hacettepe journal of Mathematics and Statistics, 37 (1), 57-67.

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

[4] Isaki, C. T. (1983). Variance estimation using auxiliary information. Journal of the American Statistical Association, 78,117-123.

[5] Kadilar, C. & Cingi, H. (2006). Ratio estimators for population variance in simple and stratified sampling. Applied mathematics and Computation, 173,1047-1058.

[6] Kadilar, C. & Cingi, H. (2007). Improvement in Variance estimation in simple random sampling. Communications in Statistics: Theory and methods, 36,2075-2081.

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

[8] Ogus, J. L., and Clark, D. F. (1971). The annual survey of manufactures: A report on methodology. US bureau of the census technical paper 24, US Government printing office, Washington DC.

[9] Prasad, B., and Singh, H. P. (1990). Some improved ratio type estimators of finite population variance in sample surveys. Communication in Statistics: Theory and methods, 19, 1127-1139.

[10] Maqbool, S., Raja, T. A., and Shakeel Javaid (2016). Generalized modified ratio estimator using non-conventional location parameter, Int. J. Agricult. Stat. Sci, 12 (1), 95-97.

[11] Shapiro, G. M., and Bateman, D. V. (1978). A better alternative to the collapsed stratum variance estimate. Proceedings of the social statistics section, American Statistical Association, 451-456.

[12] Sumramani, J. and Kumarapandiyan, G. (2015). Generalized modified ratio type estimator for estimation of population variance. Sri-Lankan journal of applied Statistics,vol16-1,69-90.

[13] Wolter, K. M. (1985). Introduction to variance estimation. Springer- Verlag.