Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.12785/jsap/040209

Improvement of Estimator for Population Variance using Correlation Coefficient and Quartiles of The Auxiliary Variable

Subhash Kumar Yadav[1], *Sant Saran Mishra*[1,*], *Alok Kumar Shukla*[2] *and Vishwas Tiwari*[2]

[1] Department of Mathematics and Statistics (A Centre of Excellence on Advanced Computing) Dr. RML Avadh University, Faizabad-224001, U.P., India
[2] Department of Statistics, D.A-V. College, Kanpur, U.P., India

Received: 22 Jan. 2015, Revised: 27 May 2015, Accepted: 27 May 2015
Published online: 1 Jul. 2015

Abstract: In the present paper, we propose an improved estimator of population variance for the main variable under study under simple random sampling without replacement ($SRSWOR$) scheme utilizing the correlation coefficient between the study variable and an auxiliary variable along with the inter-quartile range of the auxiliary variable. The expressions for the bias and Mean Square Error (MSE) of proposed estimator have been corrected up to first order of approximation. A comparison has been made with many of the existing estimators of population variance both theoretically and through numerical examples using real secondary data. An improvement of the proposed estimator has been shown over all of the estimators considered in the paper.

Keywords: Ratio estimator, quartiles, bias, mean squared error, efficiency

1 Introduction

Auxiliary information is being utilized by many researchers for improved estimation of population parameters of the main variable under study in the theory of survey sampling. It use has been widely discussed in sampling theory. The auxiliary information is used for improvement at both the stage of designing (for stratified, systematic or probability proportional to size sampling designs) and estimation. It is used in sampling theory to obtain the improved sampling designs and to achieve higher precision in the estimates of the population parameters under consideration such as the mean or the variance of the main variable under study. Here we have utilized it at estimation stage only.The auxiliary variable (X) is closely related with the main variable (Y) under study. When the study variable and the auxiliary variable are positively correlated and the lines of regression of Y on X passes through origin, then the ratio type estimators are used for improved estimation of population parameters. On the other hand the product type estimators are used when X on Y are negatively correlated to each other. Regression estimators are used when the line of regression does not pass through the origin.
The estimation of the population variance is one of the burning issues in survey sampling and a lot of efforts have been made for the improvement in the precision of the estimates of the population variance. In literature of sampling theory, a great variety of techniques using the auxiliary information by means of ratio, product and regression methods have been used for the estimation of population variance.

Let the finite population under consideration consist of N distinct and identifiable units and let $(x_i, y_i), i = 1, 2, \cdots, n$ be a bivariate sample of size n taken from (X,Y) using a $SRSWOR$ scheme. Let $\bar{X}$ and $\bar{Y}$ respectively be the population means of the auxiliary and the study variables, and let $\bar{x}$ and $\bar{y}$ be the corresponding sample means which are unbiased estimators of $\bar{X}$ and $\bar{Y}$ respectively. Let ρ be the correlation coefficient between X and Y and Q_r be the interquartile range of the auxiliary variable X In the present study, we have proposed a ratio type estimator of population mean of study variable utilizing ρ and Q_r. We assume that a reliable estimate of ρ is available in advance.
In the present paper, we have proposed an improved ratio type estimator for the estimation of the population variance.

* Corresponding author e-mail: sant_x2003@yahoo.co.in

The main aim of this paper is to develop a new ratio estimator and to improve the efficiency of ratio type estimators for the population variance. For this we have proposed the generalized estimator for population variance of which there are many estimators of population variance as its particular estimators as special case for different values of the characterizing scalar.

2 Review of Literature of Variance Estimatorss

The basic estimator of population variance is the sample variance given by:

$$t_0 = s_y^2 \tag{1}$$

It is unbiased, and its variance up to the first degree of approximation is:

$$V(t_0) = \gamma s_y^4 (\lambda_{40} - 1) \tag{2}$$

Isaki (1983) used the auxiliary information and proposed the following ratio estimator of population variance as:

$$t_R = s_y^2 \left(\frac{S_x^2}{s_x^2} \right) \tag{3}$$

where
$s_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2, \quad s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2, \quad S_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2$
$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i, \quad \bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i, \quad \bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i$
The expressions for the Bias and Mean Square Error (MSE) respectively for the estimator in (3), up to the first order of approximation, are given by

$$B(t_R) = \gamma S_y^2 [(\lambda_{04} - 1) - (\lambda_{22} - 1)] \tag{4}$$

$$MSE(t_R) = \gamma S_y^4 [(\lambda_{40} - 1) + (\lambda_{04} - 1) - 2(\lambda_{22} - 1)] \tag{5}$$

where $\lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{r/2} \mu_{02}^{s/2}}, \quad \mu_{rs} = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^r (X_i - \bar{X})^s, \quad \gamma = \frac{1-f}{n}$ and $f = \frac{n}{N}$
Many authors used auxiliary information in the form of population parameters of the auxiliary information and proposed different estimators of population variance of the study variable. Table-1, early given by Subramani and Kumarpandiyan (2012), represents different estimators of population variance with their Bias, MSE and corresponding constants.

Table 1: Bias, MSE and Corresponding Constants for Various Estimators of population variance

Estimator	Bias	MSE	Constant
$\hat{S}_1^2 = s_y^2 \left[\frac{S_x^2 + C_x}{s_x^2 + C_x} \right]$ Kadilar and Cingi (2006)	$\gamma S_y^2 R_1 [R_1(\lambda_{04} - 1) - (\lambda_{22} - 1)]$	$\gamma S_y^4 [(\lambda_{40} - 1) + R_1^2(\lambda_{04} - 1) - 2R_1(\lambda_{22} - 1)]$	$R_1 = \left[\frac{S_x^2}{S_x^2 + C_x} \right]$
$\hat{S}_2^2 = s_y^2 \left[\frac{S_x^2 + \beta_{2(x)}}{s_x^2 + \beta_{2(x)}} \right]$ Upadhyaya and Singh (1999)	$\gamma S_y^2 R_2 [R_2(\lambda_{04} - 1) - (\lambda_{22} - 1)]$	$\gamma S_y^4 [(\lambda_{40} - 1) + R_2^2(\lambda_{04} - 1) - 2R_2(\lambda_{22} - 1)]$	$R_2 = \left[\frac{S_x^2}{S_x^2 + \beta_{2(x)}} \right]$
$\hat{S}_3^2 = s_y^2 \left[\frac{S_x^2 \beta_{2(x)} + C_x}{s_x^2 \beta_{2(x)} + C_x} \right]$ Kadilar and Cingi (2006)	$\gamma S_y^2 R_3 [R_3(\lambda_{04} - 1) - (\lambda_{22} - 1)]$	$\gamma S_y^4 [(\lambda_{40} - 1) + R_3^2(\lambda_{04} - 1) - 2R_3(\lambda_{22} - 1)]$	$R_3 = \left[\frac{S_x^2 \beta_{2(x)}}{S_x^2 \beta_{2(x)} + C_x} \right]$
$\hat{S}_4^2 = s_y^2 \left[\frac{S_x^2 C_x + \beta_{2(x)}}{s_x^2 C_x + \beta_{2(x)}} \right]$ Kadilar and Cingi (2006)	$\gamma S_y^2 R_4 [R_4(\lambda_{04} - 1) - (\lambda_{22} - 1)]$	$\gamma S_y^4 [(\lambda_{40} - 1) + R_4^2(\lambda_{04} - 1) - 2R_4(\lambda_{22} - 1)]$	$R_4 = \left[\frac{S_x^2 C_x}{S_x^2 C_x + \beta_{2(x)}} \right]$

$\hat{S}_5^2 = s_y^2\left[\frac{S_x^2+Q_1}{s_x^2+Q_1}\right]$ Subramani and Kumarpandiyan (2012)	$\gamma S_y^2 R_5[R_5(\lambda_{04}-1)-(\lambda_{22}-1)]$	$\gamma S_y^4[(\lambda_{40}-1)+R_5^2(\lambda_{04}-1)-2R_5(\lambda_{22}-1)]$	$R_5=\left[\frac{S_x^2}{S_x^2+Q_1}\right]$
$\hat{S}_6^2 = s_y^2\left[\frac{S_x^2+Q_3}{s_x^2+Q_3}\right]$ Subramani and Kumarpandiyan (2012)	$\gamma S_y^2 R_6[R_6(\lambda_{04}-1)-(\lambda_{22}-1)]$	$\gamma S_y^4[(\lambda_{40}-1)+R_6^2(\lambda_{04}-1)-2R_6(\lambda_{22}-1)]$	$R_6=\left[\frac{S_x^2}{S_x^2+Q_3}\right]$
$\hat{S}_7^2 = s_y^2\left[\frac{S_x^2+Q_r}{s_x^2+Q_r}\right]$ Subramani and Kumarpandiyan (2012)	$\gamma S_y^2 R_7[R_7(\lambda_{04}-1)-(\lambda_{22}-1)]$	$\gamma S_y^4[(\lambda_{40}-1)+R_7^2(\lambda_{04}-1)-2R_7(\lambda_{22}-1)]$	$R_7=\left[\frac{S_x^2}{S_x^2+Q_r}\right]$
$\hat{S}_8^2 = s_y^2\left[\frac{S_x^2+Q_d}{s_x^2+Q_d}\right]$ Subramani and Kumarpandiyan (2012)	$\gamma S_y^2 R_8[R_8(\lambda_{04}-1)-(\lambda_{22}-1)]$	$\gamma S_y^4[(\lambda_{40}-1)+R_8^2(\lambda_{04}-1)-2R_8(\lambda_{22}-1)]$	$R_8=\left[\frac{S_x^2}{S_x^2+Q_d}\right]$
$\hat{S}_9^2 = s_y^2\left[\frac{S_x^2+Q_a}{s_x^2+Q_a}\right]$ Subramani and Kumarpandiyan (2012)	$\gamma S_y^2 R_9[R_9(\lambda_{04}-1)-(\lambda_{22}-1)]$	$\gamma S_y^4[(\lambda_{40}-1)+R_9^2(\lambda_{04}-1)-2R_9(\lambda_{22}-1)]$	$R_9=\left[\frac{S_x^2}{S_x^2+Q_a}\right]$
$\hat{S}_{10}^2 = s_y^2\left[\frac{S_x^2\rho+Q_3}{s_x^2\rho+Q_3}\right]$ Khan and Shabbir (2013)	$\gamma S_y^2 R_{10}[R_{10}(\lambda_{04}-1)-(\lambda_{22}-1)]$	$\gamma S_y^4[(\lambda_{40}-1)+R_{10}^2(\lambda_{04}-1)-2R_{10}(\lambda_{22}-1)]$	$R_{10}=\left[\frac{S_x^2\rho}{S_x^2\rho+Q_3}\right]$
$\hat{S}_{11}^2 = s_y^2\left[\frac{S_x^2\rho+Q_r}{s_x^2\rho+Q_r}\right]$ Yadav et al. (2014)	$\gamma S_y^2 R_{11}[R_{11}(\lambda_{04}-1)-(\lambda_{22}-1)]$	$\gamma S_y^4[(\lambda_{40}-1)+R_{11}^2(\lambda_{04}-1)-2R_{11}(\lambda_{22}-1)]$	$R_{11}=\left[\frac{S_x^2\rho}{S_x^2\rho+Q_r}\right]$

Where $Q_i (i=1,2,3)$ are the quartiles, the three points dividing the whole distribution into four equal parts. The used functions of quartiles are the inter quartile range, $Q_r = Q_3 - Q_1$, the semi-quartile range $Q_d = \frac{Q_3-Q_1}{2}$ and the quartile average $Q_a = \frac{Q_3+Q_1}{2}$.
Thus the MSE for the estimators given in Table-1 may be written as,

$$MSE(\hat{S}_i^2) = \gamma S_y^4[(\lambda_{40}-1)+R_i^2(\lambda_{04}-1)-2R_i(\lambda_{22}-1)](i=1,2,\cdots,11) \tag{6}$$

3 Proposed Estimator

Motivated by Yadav et al. (2014), we propose an improved ratio estimator of population variance as,

$$t = s_y^2\left[\alpha+(1-\alpha)\frac{S_x^2\rho+Q_r}{s_x^2\rho+Q_r}\right] \tag{7}$$

where α is a suitably chosen constant and is obtained by minimizing the MSE of the proposed estimator t.
In order to study the large sample properties of the proposed estimator t.
We define $s_y^2 = S_y^2(1+e_0)$ and $s_x^2 = S_x^2(1+e_1)$ such that $E(e_i)=0$ for $(i=0,1)$ and $E(e_0^2) = \frac{1-f}{n}(\lambda_{40}-1)$, $E(e_1^2) = \frac{1-f}{n}(\lambda_{04}-1)$, $E(e_0e_1) = \frac{1-f}{n}(\lambda_{22}-1)$.
Expressing t in terms of e_i's $(i=0,1)$, we have

$$\begin{aligned} t &= s_y^2(1+e_0)\left[\alpha+(1-\alpha)(1+Re_1)^{-1}\right] \ where \ R=\left[\frac{S_x^2\rho}{S_x^2\rho+Q_r}\right] \\ &= s_y^2(1+e_0)\left[\alpha+\alpha_1(1+Re_1)^{-1}\right] \ where \ \alpha_1=(1-\alpha) \end{aligned}$$

After simplifying and retaining terms up to the first order of approximation, we have:

$$t = S_y^2\left(1+e_0-R\alpha_1e_1-R\alpha_1e_0e_1+R^2\alpha_1^2e_1^2\right) \tag{8}$$

Subtracting S_y^2 on both the sides, we obtain

$$t-S_y^2 = S_y^2\left(e_0-R\alpha_1e_1-R\alpha_1e_0e_1+R^2\alpha_1^2e_1^2\right) \tag{9}$$

Taking expectation on both sides of (9), we have the Bias of proposed estimator t as:

$$B(t) = \lambda S_y^4[R^2\alpha_1^2(\lambda_{04}-1) - R_i\alpha_1(\lambda_{22}-1)] \tag{10}$$

From (9), we have up to the first order of approximation as,

$$t - S_y^2 = S_y^2(e_0 - R\alpha_1 e_1) \tag{11}$$

Squaring on both the sides, simplifying and taking expectations on both sides we have MSE of t as,

$$MSE(t) = \gamma S_y^4[(\lambda_{40}-1) + R^2\alpha_1^2(\lambda_{04}-1) - 2R\alpha_1(\lambda_{22}-1)] \tag{12}$$

which is minimum for,

$$\alpha_1 = \frac{(\lambda_{22}-1)}{R(\lambda_{04}-1)} \tag{13}$$

And the minimum MSE is,

$$MSE_{min}(t) = \gamma S_y^4\left[(\lambda_{40}-1) - \frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\right] \tag{14}$$

4 Efficiency Comparison

From (14) and (2), we have that the proposed estimator t is more efficient than the estimator t_0, if

$$V(t_0) - MSE(t) > 0,\ if\ (\lambda_{22}-1) > 0 \tag{15}$$

From (14) and (5), we infer that the proposed estimator is better than the estimator t_R as it has

$$V(t_R) - MSE(t) > 0,\ if\ (\lambda_{22}-1) > (\lambda_{04}-1) \tag{16}$$

From (14) and (6), we have that the proposed estimator t is more efficient than the estimators $\hat{S}_i^2 (i = 1,2,\cdots,11)$ in Table-1 under the condition, if

$$V(\hat{S}_i^2) - MSE(t) > 0, if\ (\lambda_{22}-1) > (\lambda_{04}-1), i = 1,2,\cdots,11 \tag{17}$$

5 Numerical Illustration

To judge the performances of different estimators, we have considered the following real populations:

Population I: Italian bureau for the environment protection-APAT Waste 2004

Y: Total amount (tons) of recyclable-waste collection in Italy in 2003

X: Total amount (tons) of recyclable-waste collection in Italy in 2002

$N = 103$, $n = 40$, $\bar{Y} = 626.2123$, $\bar{X} = 557.1909$, $\rho = 0.9936$, $S_y = 913.5498$, $C_y = 1.4588$, $S_x = 818.1117$, $C_x = 1.4683$, $\lambda_{04} = 37.3216$, $\lambda_{40} = 37.1279$, $\lambda_{22} = 37.2055$, $Q_1 = 142.9950$, $Q_3 = 665.6250$, $Q_r = 522.6300$, $Q_d = 261.3150$, $Q_a = 404.3100$

Population II: Italian bureau for the environment protection-APAT Waste 2004

Y: Total amount (tons) of recyclable-waste collection in Italy in 2003

X: Number of inhabitants in 2003

$N = 103$, $n = 40$, $\bar{Y} = 626.2123$, $\bar{X} = 556.5594$, $\rho = 0.7298$, $S_y = 913.5498$, $C_y = 1.4588$, $S_x = 610.1643$, $C_x = 1.0963$, $\lambda_{04} = 17.8738$, $\lambda_{40} = 37.1279$, $\lambda_{22} = 17.2220$, $Q_1 = 259.3830$, $Q_3 = 628.0235$, $Q_r = 368.6405$, $Q_d = 184.3293$, $Q_a = 443.7033$

Population III: Murthy (1967)

Y: Output for 80 factories in a region

X: Fixed capital

$N = 80$, $n = 20$, $\bar{Y} = 51.8264$, $\bar{X} = 11.2646$, $\rho = 0.9413$, $S_y = 18.3549$, $C_y = 0.3542$, $S_x = 8.4563$, $C_x = 0.7507$, $\lambda_{04} = 2.8664$, $\lambda_{40} = 2.2667$, $\lambda_{22} = 2.2209$, $Q_1 = 5.1500$, $Q_3 = 16.975$, $Q_r = 11.825$, $Q_d = 5.9125$, $Q_a = 11.0625$

Population IV: Singh and Cahudhary (1986)

$N = 70$, $n = 25$, $\bar{Y} = 96.7000$, $\bar{X} = 175.2671$, $\rho = 0.7293$, $S_y = 60.7140$, $C_y = 0.6254$, $S_x = 140.8572$, $C_x = 0.8037$, $\lambda_{04} = 7.0952$, $\lambda_{40} = 4.7596$, $\lambda_{22} = 4.6038$, $Q_1 = 80.1500$, $Q_3 = 225.0250$, $Q_r = 144.8750$, $Q_d = 72.4375$, $Q_a = 152.5875$

Table 2: Bias and Mean Square Error of Different Estimators

Estimator	Bias				MSE			
	Pop-I	Pop-II	Pop-III	Pop-IV	Pop-I	Pop-II	Pop-III	Pop-IV
$\hat{S}_1^2$	2420.6810	135.9827	10.4399	364.3702	67038384403	35796605	3850.1552	1415839
$\hat{S}_2^2$	2379.9609	135.8179	9.2918	363.9722	670169790	35796503	3658.4051	1414994
$\hat{S}_2^2$	2379.9609	135.8179	9.2918	363.9722	670169790	35796503	3658.4051	1414994
$\hat{S}_3^2$	2422.3041	135.9929	10.7222	364.4139	670393032	35796611	3898.5560	1415931
$\hat{S}_4^2$	2393.4791	135.8334	8.8117	363.8627	670240637	35796512	3580.8342	1414762
$\hat{S}_5^2$	2259.9938	133.4494	8.1749	359.3822	669558483	35795045	3480.5516	1427990
$\hat{S}_6^2$	1667.7818	129.8456	3.9142	350.4482	667000531	35792955	2908.6518	1408858
$\hat{S}_7^2$	1829.6315	132.3799	5.5038	355.3634	667623576	35794395	3098.4067	1419946
$\hat{S}_8^2$	2125.7591	134.1848	7.8275	359.8641	668911625	35795495	3427.1850	1429077
$\hat{S}_9^2$	1963.6570	131.6458	5.7705	354.8875	668182833	35793951	3133.3256	1418424
$\hat{S}_{10}^2$	1663.3086	127.6040	3.6276	348.1975	666910707	35791562	2878.5603	1398150
$\hat{S}_{11}^2$	1114.184	80.1433	3.9396	228.1034	645476858	21882440	2240.9762	905945
t	**0**	**0**	**0**	**0**	**623478984**	**21451662**	**1992.2024**	**569065**

6 Results and Conclusion

In the present manuscript, we have proposed the generalized ratio type estimator of population mean, we have obtained its bias and mean square error up to the first order of approximation. Also we have found the optimal value of the characterizing scalar and the minimum value of the mean square error. From the results in table-2 and the theoretical discussions in Section-4, we conclude that the proposed estimator performs much better than all of the other mentioned estimators in table-1 with respect to both the Bias and MSE. One thing more which is to be quoted is that the knowledge of the correlation coefficient ρ should be a priory available. Generally it is usually available from prior studies, or through a pilot study. If it is not known, it is replaced by its estimate and the value of the proposed estimator remains unchanged. Hence the proposed estimator should be preferred over the estimators given in table-1 for the estimation of population variance of the main variable under study.

Acknowledgement

The authors are very much thankful to the editor in chief and the unknown learned referee for critically examining the manuscript and giving valuable suggestions to improve it.

References

[1] C. T.Isaki, Variance estimation using auxiliary information, Journal of American Statistical Association, **78**, 117- 123 (1983).
[2] C. Kadilar and H. Cingi, Improvement in variance estimation using auxiliary information, Hacettepe Journal of mathematics and Statistics,**35**, 111-15 (2006).
[3] C. Kadilar and H. Cingi, Ratio estimators for population variance in simple and Stratified sampling, Applied Mathematics and Computation, **173**, 1047-1058 (2006).
[4] M. Khan and J. Shabbir, A Ratio Type Estimator for the Estimation of Population Variance using Quartiles of an Auxiliary Variable, Journal of Statistics Applications and Probability,**2, No.3**,319- 325 (2013).
[5] M. N. Murthy, Sampling Theory and Methods, Statistical Publishing Society Calcutta, India, (1967).
[6] D. Singh, and F. S. Chaudhary, Theory and analysis of sample survey designs, New-Age International Publisher, (1986).
[7] J. Subramani and G. Kumarapandiyan, Variance estimation using quartiles and their functions of an auxiliary variable, International Journal of Statistics and Applications,**2**,67-72 (2012).
[8] L. N. Upadhyaya and H. P. Singh, Use of auxiliary information in the estimation of population variance, mathematical forum**4**, 33-36 (1983).
[9] http://www.osservatorionalerifiuti.it/ElencoDocPub.asp?A_TipoDoc=6.
[10] S.K. Yadav, S.S. Mishra and S. Gupta, Improved Variance Estimation Utilizing Correlation Coefficient and Quartiles of an Auxiliary Variable, Communicated to American Journal of Mathematics and Mathematical Sciences, (2014).