İSTATİSTİK: JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION

İSTATİSTİK

Vol. 6, No. 3, December 2013, pp. 116–121

ISSN 1300-4077 | 13 | 3 | 116 | 121

IMPROVED ESTIMATION OF FINITE POPULATION VARIANCE USING QUARTILES

Housila P. Singh, Surya Kant Pal and Ramkrishna S. Solanki*

School of Studies in Statistics, Vikram University, Ujjain-456010, India

Abstract : We have addressed the problem of estimation of finite population variance using known values of quartiles of an auxiliary variable. Some ratio type estimators have been proposed with their properties in simple random sampling. The suggested estimators have been compared with the usual unbiased and ratio estimators. In addition, an empirical study is also provided in support of theoretical findings.

Key words: Study variable, Auxiliary variable, Bias, Mean squared error, Quartiles, Simple random sampling.

History: Submitted: 6 September 2013; Revised: 14 December 2013; Accepted: 17 December 2013

1. Introduction

Variation is present everywhere in our day to day life. It is law of nature that no two things or individuals are exactly alike. For instance, a physician needs a full understanding of variation in the degree of human blood pressure, body temperature and pulse rate for adequate prescription. A manufacture needs constant knowledge of the level of variation in peoples reaction to his product to be able to known whether to reduce or increase his price, or improve the quality of his product. An agriculturist needs an adequate understanding of variations in climate factors especially from place to place (or time to time) to be able to plan on when, how and where to plant his crop. Many more situations can be encountered in practice where the estimation of population variance of the study variable y assumes importance. In survey sampling, known auxiliary information is often used at the estimation stage to increase the precision of the estimators of population variance. For these reasons various authors such as Singh and Solanki (2009-2010), Tailor and Sharma (2012), Solanki and Singh (2013), Singh and Solanki (2013a, b), Subramani and Kumarapandiyan (2013a, b) and Yadav and Kadilar (2013a, b) have paid their attention towards the improved estimator of population variance of the study variable y using information on the known parameters of the auxiliary variable x such as mean, variance, coefficient of skewness, coefficient of kurtosis, correlation coefficient between the study variable y and the auxiliary variable x etc. Recently Subramani and Kumarapandiyan (2012a, b) have considered the problem of estimating the population variance of study variable y using information on variance, quartiles, inter-quartile range, semi-quartile range and semi-quartile average of the auxiliary variable x. In this paper our quest is to estimate the unknown population variance of study variable y by improving the estimators suggested by Subramani and Kumarapandiyan (2012a, b) using same information on an auxiliary variable x. Let $U=(U_1,U_2,...,U_N)$ be finite population of size N and (y,x) are (study, auxiliary) variables taking values (y_i,x_i) respectively for the i-th unit U_i of the finite population U. Let a simple random sample (SRS) of size n be drawn without replacement (WOR) from the finite population U. The usual unbiased estimator s_y^2 and the estimators of the population variance due to Isaki (1983) and Subramani and Kumarapandiyan (2012a, b) are given in the Table 1 along with their biases and mean squared errors (MSEs).

* E-mail:ramkssolanki@gmail.com

TABLE 1. The existing estimators of population variance S_y^2.

Estimator (.)	Bias B(.)	Mean squared error MSE(.)
$s_y^2=(n-1)^{-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$	$--$	$\gamma(\lambda_{40}-1)$
$t_R=s_y^2\,(S_x^2/s_x^2)$ [Isaki (1983)]	$\Phi(1-c)$	$\gamma\left[(\lambda_{40}-1)+(\lambda_{04}-1)(1-2c)\right]$
$t_1=s_y^2\,(S_x^2+Q_1/s_x^2+Q_1)$	$\Phi\theta_1(\theta_1-c)$	$\gamma\left[(\lambda_{40}-1)+\theta_1(\lambda_{04}-1)(\theta_1-2c)\right]$
$t_2=s_y^2\,(S_x^2+Q_2/s_x^2+Q_2)$	$\Phi\theta_2(\theta_2-c)$	$\gamma\left[(\lambda_{40}-1)+\theta_2(\lambda_{04}-1)(\theta_2-2c)\right]$
$t_3=s_y^2\,(S_x^2+Q_3/s_x^2+Q_3)$	$\Phi\theta_3(\theta_3-c)$	$\gamma\left[(\lambda_{40}-1)+\theta_3(\lambda_{04}-1)(\theta_3-2c)\right]$
$t_4=s_y^2\,(S_x^2+Q_r/s_x^2+Q_r)$	$\Phi\theta_4(\theta_4-c)$	$\gamma\left[(\lambda_{40}-1)+\theta_4(\lambda_{04}-1)(\theta_4-2c)\right]$
$t_5=s_y^2\,(S_x^2+Q_d/s_x^2+Q_d)$	$\Phi\theta_5(\theta_5-c)$	$\gamma\left[(\lambda_{40}-1)+\theta_5(\lambda_{04}-1)(\theta_5-2c)\right]$
$t_6=s_y^2\,(S_x^2+Q_a/s_x^2+Q_a)$	$\Phi\theta_6(\theta_6-c)$	$\gamma\left[(\lambda_{40}-1)+\theta_6(\lambda_{04}-1)(\theta_6-2c)\right]$

*t_i, $(i=1,2,\ldots,6)$are due to Subramani and Kumarapandiyan (2012a, b).

where $\bar{Y}=N^{-1}\sum_{i=1}^{N}y_i$ (population mean of y), $\bar{X}=N^{-1}\sum_{i=1}^{N}x_i$ (population mean of x), $S_y^2=(N-1)^{-1}\sum_{i=1}^{N}(y_i-\bar{Y})^2$ (population variance of y), $S_x^2=(N-1)^{-1}\sum_{i=1}^{N}(x_i-\bar{X})^2$ (population variance of x), $s_x^2=(n-1)^{-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ (sample variance of x), $\theta_1=S_x^2(S_x^2+Q_1)^{-1}$, $\theta_2=S_x^2(S_x^2+Q_2)^{-1}$, $\theta_3=S_x^2(S_x^2+Q_3)^{-1}$, $\theta_4=S_x^2(S_x^2+Q_r)^{-1}$, $\theta_5=S_x^2(S_x^2+Q_d)^{-1}$, $\theta_6=S_x^2(S_x^2+Q_a)^{-1}$, $Q_i(i=1,2,3)$ indicates the quartile, $Q_r=(Q_3-Q_1)$ (inter-quartile range), $Q_d=[(Q_3-Q_1)/2]$ (semi-quartile range), $Q_a=[(Q_3+Q_1)/2]$ (semi-quartile average), $\gamma=n^{-1}S_y^4$, $\Phi=n^{-1}S_y^2(\lambda_{04}-1)$, $c=(\lambda_{04}-1)^{-1}(\lambda_{22}-1)$, $\lambda_{rs}=\mu_{rs}\mu_{02}^{-s/2}\mu_{20}^{-r/2}$, $\mu_{rs}=N^{-1}\sum_{i=1}^{N}(y_i-\bar{Y})^r(x_i-\bar{X})^s$ (r,s being non negative integer).

It is observed that the estimators $(t_1,t_2,...,t_6)$ due to Subramani and Kumarapandiyan (2012a, b) have used the quartiles $(Q_i, i=1,2,3)$ and their functions such as inter-quartile range Q_r, semi-quartile range Q_d and semi-quartile average Q_a and in additive form to sample and population variances s_x^2 and S_x^2 respectively of the auxiliary variable x. It is to be noted that the unit of the quartiles and their function as given above is of original variable x, while the unit of S_x^2 and s_x^2 are in the square of the unit of the original variable x. This suggests authors to develop some alternative estimators for the population variance and study their properties.

The remaining part of the paper is organized as follows: In Sec. 2, the improved estimators of population variance have been suggested and the expressions of their asymptotic biases and the mean squared errors are obtained. Sec. 3 addresses the problem of efficiency comparisons of proposed estimators with the usual unbiased estimator and the estimator due to Isaki (1983), while Sec. 4 has focused on empirical study of proposed estimators for the real data set. Sec. 5 finished off the paper with final remarks.

2. The proposed estimators

Using known values of quartiles (Q_i, $i=1,2,3$) of the auxiliary variable x, we have suggested the following modified estimators of population variance S_y^2 as

$$t_i^*=s_y^2\left(\frac{S_x^2+\alpha L_i^2}{s_x^2+\alpha L_i^2}\right)\ ,(\ i=1,2,...,6) \tag{2.1}$$

where $(S_x^2+\alpha L_i^2)>0$, $(s_x^2+\alpha L_i^2)>0$, $L_1=Q_1$, $L_2=Q_2$, $L_3=Q_3$, $L_4=Q_r$, $L_5=Q_d$, $L_6=Q_a$ and α being a constant such that $0\le\alpha\le1$.

To obtain the biases and MSEs of the estimators t_i^*, $(i=1,2,...,6)$ we write $s_y^2=S_y^2(1+e_0)$, $s_x^2=S_x^2(1+e_1)$ such that $E(e_0)=E(e_1)=0$ and to the first degree of approximation (ignoring finite population correction (f.p.c.) term), we have

$$E(e_0^2)=n^{-1}(\lambda_{40}-1),$$
$$E(e_1^2)=n^{-1}(\lambda_{04}-1),$$
$$E(e_0e_1)=n^{-1}(\lambda_{22}-1).$$

Now expressing (2.1) in terms of e's, we have

$$\begin{aligned} t_i^* &= S_y^2(1+e_0)\left[S_x^2+\alpha L_i^2/\left(S_x^2(1+e_1)+\alpha L_i^2\right)\right], \\ &= S_y^2(1+e_0)\left[1+\left(S_x^2 e/S_x^2+\alpha L_i^2\right)\right], \\ &= S_y^2(1+e_0)(1+\theta_i^* e_1)^{-1}, \end{aligned} \tag{2.2}$$

where $\theta_i^* = S_x^2(S_x^2+\alpha\, L_i^2)^{-1}$.

We assume that $|\theta_i^* e_1|<1$ so that $(1+\theta_i^* e_1)^{-1}$ is expendable. Expending the right hand side of (2.2) and multiplying out, we have

$$\begin{aligned} t_i^* &= S_y^2(1+e_0)(1-\theta_i^* e_1+\theta_i^{*2}e_1^2-...), \\ &= S_y^2(1+e_0-\theta_i^* e_1-\theta_i^* e_0 e_1+\theta_i^{*2}e_1^2+...). \end{aligned}$$

Neglecting terms of e's having power greater than the two, we have

$$t_i^* \cong S_y^2(1+e_0-\theta_i^* e_1-\theta_i^* e_0 e_1+\theta_i^{*2}e_1^2),$$

or

$$(t_i^*-S_y^2) \cong S_y^2(e_0-\theta_i^* e_1-\theta_i^* e_0 e_1+\theta_i^{*2}e_1^2). \tag{2.3}$$

Taking expectation of both sides of (2.3), we get the biases of t_i^* to the first degree of approximation as

$$B(t_i^*) = \Phi\theta_i^*(\theta_i^*-c)\ ,\ (i=1,2,\ldots,6). \tag{2.4}$$

Squaring both sides of (2.3) and neglecting terms of e's having power greater than two, we have

$$(t_i^*-S_y^2)^2 \cong S_y^4(e_0^2+\theta_i^{*2}e_1^2-2\theta_i^* e_0 e_1). \tag{2.5}$$

Taking expectation of both sides of (2.5), we get the MSEs of t_i^* to the first degree of approximation as

$$MSE(t_i^*) = \gamma\left[(\lambda_{40}-1)+\theta_i^*(\lambda_{04}-1)(\theta_i^*-2c)\right]\ ,\ (i=1,2,\ldots,6). \tag{2.6}$$

For the sake of convenience to the readers the biases and MSEs of the proposed estimators t_i^*, $(i=1,2,\ldots,6)$ are summarized in the Table 2.

TABLE 2. The biases and MSEs of t_i^* $(i=1,2,...,6)$ of S_y^2.

Estimator (.)	Bias B(.)	Mean squared error MSE(.)
$t_1^* = s_y^2(S_x^2+\alpha Q_1^2/s_x^2+\alpha Q_1^2)$	$\Phi\theta_1^*(\theta_1^*-c)$	$\gamma\left[(\lambda_{40}-1)+\theta_1^*(\lambda_{04}-1)(\theta_1^*-2c)\right]$
$t_2^* = s_y^2(S_x^2+\alpha Q_2^2/s_x^2+\alpha Q_2^2)$	$\Phi\theta_2^*(\theta_2^*-c)$	$\gamma\left[(\lambda_{40}-1)+\theta_2^*(\lambda_{04}-1)(\theta_2^*-2c)\right]$
$t_3^* = s_y^2(S_x^2+\alpha Q_3^2/s_x^2+\alpha Q_3^2)$	$\Phi\theta_3^*(\theta_3^*-c)$	$\gamma\left[(\lambda_{40}-1)+\theta_3^*(\lambda_{04}-1)(\theta_3^*-2c)\right]$
$t_4^* = s_y^2(S_x^2+\alpha Q_r^2/s_x^2+\alpha Q_r^2)$	$\Phi\theta_4^*(\theta_4^*-c)$	$\gamma\left[(\lambda_{40}-1)+\theta_4^*(\lambda_{04}-1)(\theta_4^*-2c)\right]$
$t_5^* = s_y^2(S_x^2+\alpha Q_d^2/s_x^2+\alpha Q_d^2)$	$\Phi\theta_5^*(\theta_5^*-c)$	$\gamma\left[(\lambda_{40}-1)+\theta_5^*(\lambda_{04}-1)(\theta_5^*-2c)\right]$
$t_6^* = s_y^2(S_x^2+\alpha Q_a^2/s_x^2+\alpha Q_a^2)$	$\Phi\theta_6^*(\theta_6^*-c)$	$\gamma\left[(\lambda_{40}-1)+\theta_6^*(\lambda_{04}-1)(\theta_6^*-2c)\right]$

The absolute biases of the proposed estimations t_i^* $(i=1,2,...,6)$ are less than that of Isaki (1983) ratio type estimator t_R, i.e

$$|B(t_i^*)| < |B(t_R)|\ \text{ if }\ |\theta_i^*(\theta_i^*-c)| < |1-c|\,. \tag{2.7}$$

3. Efficiency comparisons

In this section we have derived the conditions under which the proposed estimators t_i^*, $(i = 1, 2, ..., 6)$ are more efficient than the usual unbiased estimator s_y^2 and estimator t_R envisaged by Isaki (1983). It is observed from Table 1 and (2.6) that the

$$\text{i. } MSE(t_i^*) < MSE(s_y^2)\text{ , if } 0 < c < (\theta_i^*/2) \tag{3.1}$$
$$\text{ii. } MSE(t_i^*) < MSE(s_y^2)\text{ , if } \min\{1,(2c-1)\} \le \theta_i^* \le \max\{1,(2c-1)\} \tag{3.2}$$

4. Empirical study

The performance of the proposed ratio-type estimators t_i^*, $(i = 1, 2, ..., 6)$ of population variance S_y^2 are assessed with that of usual unbiased estimator (s_y^2) and traditional ratio estimator t_R for the population data set [Singh and Chaudhary (1986, p. 108)] summarized in Table 3.

TABLE 3. The population data set.

N	70	C_y	0.6254	Q_1	80.1500
n	25	S_x	140.8572	Q_2	160.3000
$\overline{Y}$	96.7000	C_x	0.8037	Q_3	225.0250
$\overline{X}$	175.2671	λ_{04}	7.0952	Q_r	144.8750
ρ	0.7293	λ_{40}	4.7596	Q_d	72.4375
S_y	60.7140	λ_{22}	4.6038	Q_a	152.5875

We have computed the θ_i^*, $(i = 1, 2, ..., 6)$ and the percent relative efficiencies (PREs) of the estimators t_i^*, $(i = 1, 2, ..., 6)$ with respect to usual unbiased estimator (s_y^2) and traditional ratio estimator t_R in certain range of $\alpha \in (0.0, 1.0)$ by using following formulae respectively as

$$PRE(t_i^*, s_y^2) = \frac{MSE(s_y^2)}{MSE(t_i^*)} \times 100 = \frac{(\lambda_{40}-1)}{[(\lambda_{40}-1)+\theta_i^*(\lambda_{04}-1)(\theta_i^*-2c)]} \times 100, \tag{4.1}$$
$$PRE(t_i^*, t_R) = \frac{MSE(t_R)}{MSE(t_i^*)} \times 100 = \frac{[(\lambda_{40}-1)+(\lambda_{04}-1)(1-2c)]}{[(\lambda_{40}-1)+\theta_i^*(\lambda_{04}-1)(\theta_i^*-2c)]} \times 100, \tag{4.2}$$

and findings are summarized in Tables 4, 5 and 6.

TABLE 4. The values of θ_i^*, $(i = 1, 2, ..., 6)$ for $\alpha \in (0.0, 1.0)$.

α	θ_1^*	θ_2^*	θ_3^*	θ_4^*	θ_5^*	θ_6^*
0.0	1.000	1.000	1.000	1.000	1.000	1.000
0.1	0.969	0.885	0.797	0.904	0.974	0.895
0.2	0.939	0.794	0.662	0.825	0.950	0.810
0.3	0.911	0.720	0.566	0.759	0.926	0.740
0.4	0.885	0.659	0.495	0.703	0.904	0.681
0.5	0.861	0.607	0.439	0.654	0.883	0.630
0.6	0.837	0.563	0.395	0.612	0.863	0.587
0.7	0.815	0.524	0.359	0.575	0.844	0.549
0.8	0.794	0.491	0.329	0.542	0.825	0.516
0.9	0.774	0.462	0.303	0.512	0.808	0.486
1.0	0.755	0.436	0.282	0.486	0.791	0.460

It is observed from Table 4 that the condition (3.1) is satisfied by all the values of θ_i^*, $(i = 1, 2, ..., 6)$ for the given population data set in certain range $\alpha \in (0.0, 1.0)$, i.e. all the proposed estimators t_i^*,

$(i=1,2,...,6)$ are more efficient than the usual unbiased estimator s_y^2 when $\alpha \in (0.0, 1.0)$ for the given population data set. It is also noticed from Table 4 that the condition (3.2) is satisfied by all the values of θ_i^*, $(i=1,2,...,6)$ for the given population data set, i.e. all the proposed estimators t_i^*, $(i=1,2,...,6)$ are more efficient than the traditional ratio estimator t_R when $\alpha \in (0.0,\ 1.0)$ for the given population data set.

TABLE 5. PREs of Estimators t_i^*, $(i=1,2,...,6)$ with respect to s_y^2.

α	$PRE(t_1^*, s_y^2)$	$PRE(t_2^*, s_y^2)$	$PRE(t_3^*, s_y^2)$	$PRE(t_4^*, s_y^2)$	$PRE(t_5^*, s_y^2)$	$PRE(t_6^*, s_y^2)$
0.0	142.02	142.02	142.02	142.02	142.02	142.02
0.1	150.57	174.38	199.34	168.87	149.02	171.58
0.2	158.85	199.97	226.56	191.53	155.85	195.78
0.3	166.81	217.30	**230.28**	208.80	162.48	213.25
0.4	174.38	226.95	223.06	220.57	168.87	224.13
0.5	181.51	**230.60**	212.47	227.46	175.00	229.51
0.6	188.17	230.11	201.75	230.45	180.83	**230.80**
0.7	194.33	227.03	192.01	**230.57**	186.34	229.28
0.8	199.97	222.47	183.50	228.71	191.53	226.00
0.9	205.08	217.19	176.16	225.55	196.37	221.68
1.0	**209.67**	211.65	169.84	221.62	**200.87**	216.85

TABLE 6. PREs of Estimators t_i^*, $(i=1,2,...,6)$ with respect to t_R.

α	$PRE(t_1^*, t_R)$	$PRE(t_2^*, t_R)$	$PRE(t_3^*, t_R)$	$PRE(t_4^*, t_R)$	$PRE(t_5^*, t_R)$	$PRE(t_6^*, t_R)$
0.0	100.00	100.00	100.00	100.00	100.00	100.00
0.1	106.02	122.78	140.36	118.91	104.93	120.82
0.2	111.85	140.80	159.53	134.86	109.74	137.86
0.3	117.45	153.00	**162.14**	147.02	114.41	150.15
0.4	122.78	159.80	157.06	155.31	118.91	157.81
0.5	127.81	**162.37**	149.60	160.16	123.22	161.60
0.6	132.50	162.03	142.06	162.27	127.32	**162.51**
0.7	136.83	159.85	135.20	**162.35**	131.21	161.44
0.8	140.80	156.64	129.21	161.04	134.86	159.13
0.9	144.40	152.93	124.04	158.81	138.27	156.09
1.0	**147.63**	149.03	119.59	156.04	**141.44**	152.69

It is observed from Tables 5 and 6 that the proposed estimators t_i^*, $(i=1,2,...,6)$ performed better than the usual unbiased estimator s_y^2 and the usual ratio estimator t_R for all the values of $\alpha \in (0.0, 1.0)$. The bold numbers indicate the maximum gain in efficiencies of the proposed estimators t_i^*, (i =1, 2,..., 6) with respect to the usual unbiased estimator s_y^2 and the usual ratio estimator t_R for the corresponding value of α. The proposed estimator t_6^* (based on semi-quartile average Q_a) performed best among all the estimators discussed here for the given population data set.

5. Conclusion

In this paper we have proposed the ratio type class of estimators of population variance. The bias and mean squared error formulae of the proposed ratio type class of estimators of population variance are obtained and compared with that of the usual unbiased estimator and traditional ratio estimator. Further we have derived the conditions for which the proposed estimators are more efficient than the usual unbiased estimator and traditional ratio estimator. We have also assessed the performance of the proposed estimators for known natural population data set and found that the performances of proposed estimators are better than the traditional unbiased and ratio estimators for certain cases.

Acknowledgments: Authors are thankful to the learned referees for their valuable suggestions regarding improvement to the paper.

References

[1] Singh, D. and Chaudhary, F.S. (1986): Theory and Analysis of Sample Survey Designs. New Age International Publisher, New Delhi, India.

[2] Singh, H.P. and Solanki, R.S. (2009-2010): Estimation of finite population variance using auxiliary information in presence of random non-response. Gujarat Statistical Review, 36 & 37 (1 & 2): 46-58.

[3] Singh, H.P. and Solanki, R.S. (2013a): A new procedure for variance estimation in simple random sampling using auxiliary information. Statistical Papers 54(2): 479-497.

[4] Singh, H.P. and Solanki, R.S. (2013b): Improved estimation of finite population variance using auxiliary information. Communications in Statistics-Theory and Methods 42(15): 2718-2730.

[5] Solanki, R.S. and Singh, H.P. (2013): An improved class of estimators for the population variance. Model Assisted Statistics and Applications 8(3): 229-238.

[6] Subramani, J. and Kumarapandiyan, G. (2012a): Variance estimation using median of the auxiliary variable. International Journal of Probability and Statistics 1(3): 62-66.

[7] Subramani, J. and Kumarapandiyan, G. (2012b): Variance estimation using quartiles and their functions of an auxiliary variable. International Journal of Statistics and Applications 2(5): 67-72.

[8] Subramani, J. and Kumarapandiyan, G. (2013a): Estimation of variance using known co-efficient of variation and median of an auxiliary variable. Journal of Modern Applied Statistical Methods 12(1): 58-64.

[9] Subramani, J. and Kumarapandiyan, G. (2013b): Estimation of variance using deciles of an auxiliary variable. Proceedings of International Conference on Frontiers of Statistics and Its Applications, Bonfring Publisher: 143-149.

[10] Tailor, R. and Sharma, B. (2012): Modified estimators of population variance in presence of auxiliary information. Statistics in Transition-new series 13(1): 37-46.

[11] Yadav, S.K. and Kadilar, C. (2013a): A class of ratio-cum-dual to ratio estimator of population variance. Journal of Reliability and Statistical Studies 6(1): 29-34.

[12] Yadav, S.K. and Kadilar, C. (2013b): Improved Exponential type ratio estimator of population variance. Colombian Journal of Statistics 36(1): 145-152.