The Aligarh Journal of Statistics ISSN 0971-0388
Vol. 45 (2025), pp. 53-60

Use of Median for Constructing a Competent Estimator of the Finite Population Variance

Rohini Yadav
Department of Statistics, University of Lucknow, Lucknow, U.P., India
Corresponding author: Rohini Yadav, *rohiniyadav.ism@gmail.com*

AMS Subject Classification: 62D05
Received: 02/08/2023 Accepted: 10/01/2025

Abstract

In the current work, a novel estimator for estimating population variance has been proposed, employing the available information of the median of the supplementary variate specifically regarding the simple random sampling (SRS) technique. The characteristics of a recently introduced estimator have been observed while utilizing a large sample approximation. It is evaluated both theoretically and empirically by comparison with other well-known estimators of the population variance.

Keywords: Main (or study) variable, auxiliary (or supplementary) variable, median (second quartile), mean squared error and percent relative efficiency.

1 Introduction

The principal objective in sampling theory is to use additional information to boost the precision of the estimator of the target variable. The estimator's precision may be increased at any stage during the selection process of the sampling unit or at the time of estimating the parameter(s) of interest. The selection of suitable supplementary (auxiliary) information plays a key role while proposing an efficient estimator of any parameter under different sampling techniques. Similar to estimating the population mean, estimating the population variance also captured the interest of a number of researchers. In agriculture, business, industry, medical sciences, and many other areas where the population is not likely to be symmetric (or skewed), the estimation of the population variance while dealing with a problem has played a prime role. On throwing light on the works of several researchers, one is able to notice the importance of the estimation of variance of the population of the target variate using a priori information under SRS.

Variance plays a very important role in our daily life. It is a well-known fact that two individuals or data sets or populations are not likely the same, which gives the cause of variation. A product manufacturer industry needs to have clear knowledge about the variations in customers' liking or disliking of the quality as well as the price of the product so that one may improve the quality of the product or reduce the price. An economist/investor should possess the adequate understanding of variations in stock/market rate to be able to plan when, where, and how much to invest the money during the financial year so that the profit (or loss) is maximum (or minimum). A physician must have clear insights into variations in the degree of human blood pressure, blood sugar, and body temperature, as well as pulse rate, for adequate prescription. There are some more examples available in our practical life where the importance of estimation of variance of the population of the target variate can be observed.

The estimation of the variance technique was first introduced and its features were investigated by Isaki (1983) with the help of auxiliary information. Prasad and Singh (1990) have given an improved ratio-type estimator for the variance of the finite population of the target variable. Utilizing the details of the coefficient of kurtosis based on supplementary information, Upadhyaya and Singh (1999) provided a new estimator of the variance of the finite population. Under the both sampling techniques SRS and stratified, some new ratio type estimators for the population variance were given by Kadilar and Cingi (2006). On the same line, another elevated estimator for population variance was searched out by Tailor and Sharma (2012) in the existence of supplementary information.

With all these modifications in consideration, Subramani and Kumarapandiyan (2012a, 2012b) introduced the variance estimation technique using quartiles, median as well as some functions of quartiles of an ancillary variable which improves the efficiency of the same over other considered estimators taken in the study. Again, Subramani and Kumarapandiyan (2013) dealt with the variance estimation problem using known value of median or second quartile and other parameters for boosting the efficiency of the same. Further, Singh and Solanki (2013) and Khan and Shabbir (2013) enhanced the work on improved estimators for the variance estimation employing quartile values of additional variable. Apart from these researchers, Yadav et al. (2013a, 2013b) have focused while proposing a generalized estimator of variance of the population based on target variate. They utilized the particulars available on coefficient of variation of supplementary variable for developing the estimators. Yadav et al. (2014) provided a family of estimators of the population variance that contained other known estimators as a subclass of itself and more competent than other estimators present in the literature. After that, Yadav (2021) propounded the variance estimation technique making use of quartiles of supplementary variate which increased accuracy of the estimator.

Some known estimators of variance based on the population like standard unbiased estimator s_y^2 , Isaki (1983) estimator, Kadilar and Cingi (2006) estimator, Subramani and Kumarapandiyan (2012a, 2013) estimators are shown in the following Table 1 accompanying with their biases and MSEs.

Table 1: Some known estimators of variance of the population

Estimators(.)	**Bias(.)**	**Mean Squared Error(.)**
$t_0 = s_y^2$	-	$\gamma S_y^4(\lambda_{40} - 1)$
$t_R = s_y^2 \left(\frac{S_x^2}{s_x^2}\right)$ Isaki (1983)	$\gamma S_y^2(\lambda_{04} - 1)(1 - C)$	$\gamma S_y^4[(\lambda_{40} - 1) + (\lambda_{04} - 1)(1 - 2C)]$
$t_1 = s_y^2 \left[\frac{S_x^2 - C_x}{s_x^2 - C_x}\right]$ Kadilar and Cingi(2006)	$\gamma S_y^2(\lambda_{04} - 1)\theta_1(\theta_1 - C)$	$\gamma S_y^4[(\lambda_{40} - 1) + \theta_1(\lambda_{04} - 1)(\theta_1 - 2C)]$
$t_2 = s_y^2 \left[\frac{S_x^2 - \beta_2(x)}{s_x^2 - \beta_2(x)}\right]$ Kadilar and Cingi(2006)	$\gamma S_y^2(\lambda_{04} - 1)\theta_2(\theta_2 - C)$	$\gamma S_y^4[(\lambda_{40} - 1) + \theta_2(\lambda_{04} - 1)(\theta_2 - 2C)]$
$t_3 = s_y^2 \left[\frac{S_x^2\beta_2(x) - C_x}{s_x^2\beta_2(x) - C_x}\right]$ Kadilar and Cingi(2006)	$\gamma S_y^2(\lambda_{04} - 1)\theta_3(\theta_3 - C)$	$\gamma S_y^4[(\lambda_{40} - 1) + \theta_3(\lambda_{04} - 1)(\theta_3 - 2C)]$
$t_4 = s_y^2 \left[\frac{S_x^2 C_x - \beta_2(x)}{s_x^2 C_x - \beta_2(x)}\right]$ Kadilar and Cingi(2006)	$\gamma S_y^2(\lambda_{04} - 1)\theta_4(\theta_4 - C)$	$\gamma S_y^4[(\lambda_{40} - 1) + \theta_4(\lambda_{04} - 1)(\theta_4 - 2C)]$
$t_5 = s_y^2 \left[\frac{S_x^2 + Q_2}{s_x^2 + Q_2}\right]$ Subramani and Kumarapandiyan (2012a)	$\gamma S_y^2(\lambda_{04} - 1)\theta_5(\theta_5 - C)$	$\gamma S_y^4[(\lambda_{40} - 1) + \theta_5(\lambda_{04} - 1)(\theta_5 - 2C)]$
$t_6 = s_y^2 \left[\frac{S_x^2 C_x + Q_2}{s_x^2 C_x + Q_2}\right]$ Subramani and Kumarapandiyan (2013)	$\gamma S_y^2(\lambda_{04} - 1)\theta_6(\theta_6 - C)$	$\gamma S_y^4[(\lambda_{40} - 1) + \theta_6(\lambda_{04} - 1)(\theta_6 - 2C)]$

where the population means of Y and X are respectively expressed as :

$$\bar{Y} = N^{-1} \sum_{i=1}^{N} Y_i$$

and

$$\bar{X} = N^{-1} \sum_{i=1}^{N} X_i$$

; the population variances of Y and X are respectively defined as:

$$S_y^2 = (N-1)^{-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$$

and

$$S_x^2 = (N-1)^{-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

the covariance between X and Y:

$$S_{xy} = (N-1)^{-1}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})$$

the coeff. of correlation between Y and X:

$$\rho = S_{xy}/(S_x S_y)$$

CV(y):

$$C_y = S_y/\bar{Y}$$

and CV(x):

$$C_x = S_x/\bar{X}$$

the coeff. of kurtosis of x:

$$\beta_2(x) = \mu_{04}/\mu_{02}^2$$

the sample variances of y and x are respectively defined as:

$$s_y^2 = (n-1)^{-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

and

$$s_x^2 = (n-1)^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

the second quartile (or median), Q_2;

$$\theta_1 = \left[\frac{S_x^2}{S_x^2 - C_x}\right]$$

$$\theta_2 = \left[\frac{S_x^2}{S_x^2 - \beta_2(x)}\right]$$

$$\theta_3 = \left[\frac{\beta_2(x) S_x^2}{\beta_2(x) S_x^2 - C_x}\right]$$

$$\theta_4 = \left[\frac{C_x S_x^2}{C_x S_x^2 - \beta_2(x)}\right]$$

$$\theta_5 = \left[\frac{S_x^2}{S_x^2 + Q_2}\right]$$

$$\theta_6 = \left[\frac{C_x S_x^2}{C_x S_x^2 + Q_2}\right]$$

$$\gamma = 1/n, \qquad C = (\lambda_{04} - 1)^{-1}(\lambda_{22} - 1);$$

$\lambda_{rs} = \mu_{rs}(\mu_{02}^{(s/2)}\mu_{20}^{(r/2)})^{-1}$, $\mu_{rs} = N^{-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^r(X_i - \bar{X})^s$; (where r, s being positive integers).

It has been discovered that the estimators due to Subramani and Kumarapandiyan (2012a, 2013) have used information available on second quartile Q_2 (or median) along with the CV of x, the sample variance s_x^2 and the population variance S_x^2 . After monitoring the form of these mentioned estimators, it is found that the unit of Q_2 (or median), S_x^2 and s_x^2 are not same, i.e., the units of S_x^2 and s_x^2 are in the square of the unit of the original variable x while Q_2 is of original variable x. This motivates the author to construct new estimator for population variance and throw some light on its properties whose precision is higher.

2 A New Estimator

Getting ideas and motivation from Subramani and Kumarapandiyan (2012a, 2013), a new estimator has been constructed for S_y^2 utilizing the knowledge of square of second quartile (or median M_x^2) and coefficient of kurtosis $\beta_2(x)$ which is expressed as follows:

$$T_M = s_y^2\left\{\frac{\beta_2(x)S_x^2 + M_x^2}{\beta_2(x)s_x^2 + M_x^2}\right\} \tag{2.1}$$

The bias and MSE of the estimator T_M are derived upto the first order approximation under the large sample assumptions by using the following transformations: $s_y^2 = S_y^2(1+\xi_0)$, $s_x^2 = S_x^2(1+\xi_1)$ such that $E(\xi_0) = E(\xi_1) = 0$

Under the above transformations, we have

$$E(\xi_0^2) = (\lambda_{40} - 1)/n$$

$$E(\xi_1^2) = (\lambda_{04} - 1)/n$$

$$E(\xi_0\xi_1) = (\lambda_{22} - 1)/n$$

Re-writing (2.1) in terms of $\xi' s$, the expression is converted as follows:

$$T_M = S_y^2(1+\xi_0)\left\{\frac{\beta_2(x)S_x^2 + M_x^2}{\beta_2(x)S_x^2(1+\xi_1) + M_x^2}\right\}$$

$$T_M = S_y^2(1+\xi_0)(1+\theta^*\xi_1)^{-1} \tag{2.2}$$

where $\theta^* = \left(\frac{\beta_2(x)S_x^2}{\beta_2(x)S_x^2 + M_x^2}\right)$

We presuppose that $|\theta^*\xi_1| < 1$, so that$(1+\theta^*\xi_1)^{-1}$ is expandable in terms of power series. Now, its form reduces to

$$T_M = S_y^2(1+\xi_0)(1 - \theta^*\xi_1 + \theta^{*2}\xi_1^2 - .....)$$

Overlooking $\xi' s$ term containing power higher than 2 (two), the expression becomes as follows:

$$(T_M - S_y^2) \approx S_y^2[(\xi_0 - \theta^*\xi_1 + \theta^{*2}\xi_1^2 - \theta^*\xi_0\xi_1)] \tag{2.3}$$

Employing the properties of the expectation on the equation (2.3), Bias(T_M) has been obtained as expressed in the following equation:

$$Bias(T_M) = \frac{S_y^2}{n}[\theta\beta_2^*(x)(\theta - C)] \tag{2.4}$$

After getting bias of the suggested estimator, again consider the equation (2.3), squaring it and excluding those terms of $\xi' s$ whose power is more than 2, it provides

$$(T_M - S_y^2)^2 \approx S_y^4[\xi_0 - \theta^*\xi_1]^2 \tag{2.5}$$

Now applying expectation on above equation to achieve the MSE of the estimator T_M as expressed in the equation (2.6):

$$MSE(T_M) = \frac{S_y^4}{n}[\beta_2^*(y) + \theta^*\beta_2^*(x)\{\theta^* - 2C\}] \tag{2.6}$$

The MSE of the estimator T_M given by above equation (2.6) is optimized for $\theta^* = C$. Thus, the optimum (or minimum) MSE of T_M is obtained as

$$MSE(T_M)_{opt.} = \frac{S_y^4}{n}[\beta_2^*(y) - C^2\beta_2^*(x)] \tag{2.7}$$

3 Theoretical Comparison for Efficiency

After obtaining the expression for MSE of the recommended estimator T_M , a comparison has been performed between the MSEs of the estimator T_M and the other stated estimators (as given in Table 1) and found the circumstances in which the suggested estimator T_M executed excelling to the standard unbiased estimator s_y^2 , Isaki (1983) estimator, Kadilar and Cingi (2006) estimator, Subramani and Kumarapandiyan (2012a, 2013) t_k (k=1,2...6) estimators. The expressions of inequalities are as follows:

I. $MSE(T_M) < MSE(s_y^2)$, iff

$$C > \frac{\theta^*}{2} \tag{3.1}$$

II. $MSE(T_M) < MSE(t_R)$, iff

$$min.\{1, (2C - 1)\} < \theta^* < max.\{1, (2C - 1)\} \tag{3.2}$$

III. $MSE(T_M) < MSE(t_k)$; k=1,2...6 iff

$$min.[\theta_k, (2C - \theta_k)] < \theta^* < max.[\theta_k, (2C - \theta_k)]; k = 1, 2...6 \tag{3.3}$$

From the above equations (3.1), (3.2) and (3.3), it is observed that the proposed estimator T_M outperformed the considered estimators in present study.

4 Exploratory Comparison for Efficiency

Any study is incomplete without judging its precision or efficiency on different population data sets. To support the results came from theoretical comparison done in earlier section, now the numerical comparison is performed over the considered estimators (as mentioned in Table 1) using two population data sets, first is given by Murthy (1967) and second by Singh and Chaudhary (1986).

	Population data set 1	Population data set 2
N	80	70
n	20	25
$\bar{Y}$	51.8264	96.7000
$\bar{X}$	11.2646	175.2671
ρ	0.9413	0.7293
S_y	18.3549	60.7140
C_y	0.3542	0.6254
S_x	8.4563	140.85
C_x	0.7507	0.8037
λ_{04}	2.8664	7.0952
λ_{40}	2.2667	4.7596
λ_{22}	2.2209	4.6038
Q_2	10.300	160.30

The PREs of the estimators t_k (k=1,2....6) and T_M have been computed with respect to the standard unbiased estimator by implementing the formula provided below so that the outcomes of the efficiency comparison of the considered estimators would be achieved.

$$PRE(t_k, s_y^2) = \frac{MSE(s_y^2)}{MSE(t_k)} * 100\,; (k = 1, 2, ...6)$$

The values are represented in the Table -2 as given below:

Table 2

Estimators(.)	Population 1	Population 2
$t_R = s_y^2 \left(\frac{S_x^2}{s_x^2}\right)$	183.2345	142.0218
$t_1 = s_y^2 \left[\frac{S_x^2 - C_x}{s_x^2 - C_x}\right]$	179.6210813	142.01093
$t_2 = s_y^2 \left[\frac{S_x^2 - \beta_2(x)}{s_x^2 - \beta_2(x)}\right]$	169.2398	141.9261
$t_3 = s_y^2 \left[\frac{S_x^2 \beta_2(x) - C_x}{s_x^2 \beta_2(x) - C_x}\right]$	181.9786	142.0202
$t_4 = s_y^2 \left[\frac{S_x^2 C_x - \beta_2(x)}{s_x^2 C_x - \beta_2(x)}\right]$	164.4934	141.9028
$t_5 = s_y^2 \left[\frac{S_x^2 + Q_2}{s_x^2 + Q_2}\right]$	226.8671	144.1754
$t_6 = s_y^2 \left[\frac{S_x^2 C_x + Q_2}{s_x^2 C_x + Q_2}\right]$	238.1734	144.6995
$T_M = s_y^2 \left\{\frac{\beta_2(x) S_x^2 + Q_2^2}{\beta_2(x) s_x^2 + Q_2^2}\right\}$	270.6075	185.8162

The findings of the above table showed that the attainment of the suggested estimator is preferable than the standard unbiased estimator s_y^2 , Isaki (1983) estimator and t_k (k=1,2....6) estimators.

5 Conclusion

The current study revealed a novel constructive estimator to estimate the variance of the finite population under SRS through employing second quartile or median M_x information. Its features have been examined under the approximation of large sample. But to study the characteristics of the estimator is not enough, its efficiency comparison, theoretical as well as numerical, must required to achieve any fruitful conclusion. Some expressions have been established for this purpose, showing that the recommended estimator functioned far superior to the standard unbiased estimator, Isaki (1983) estimator, Kadilar and Cingi (2006) estimator, Subramani and Kumarapandiyan (2012a, 2013) t_k (k=1,2...6) estimators. It has been further assured with the help of the population data sets that the new proposed estimator really executed and performed better over considered estimators under achieved conditions.

6 Funding

There is no funding for the research

7 Conflicts of Interest

The author declares no conflict of interest.

References

[1] Prasad, B. and Singh, H.P. (1990). Some improved ratio-type estimators of finite population variance in sample surveys. *Comm. in Stat. Theo. and Meth*, 19: 1127–1139.

[2] Isaki, C. T. (1983). Variance estimation using auxiliary information .*Journal of the American Statistical Association*, 78(381):117–123.

[3] Kadilar, C. and Cingi, H. (2006). Improvement in variance estimation using auxiliary information. *Hacettepe Journal of Mathematics and Statistics*, 35(1):111–115.

[4] Khan, M. and Shabbir, J. (2013). A RatioType Estimator for the Estimation of Population Variance using Quartiles of an Auxiliary Variable. *Journal of Statistics Applications and Probability*, 2(3), 319-325.

[5] Singh, D. and Chaudhary, F. S. (1986). Theory and analysis of sample survey designs. New-Age International Publisher.

[6] Singh, H. P. and Solanki, R. S. (2013). A new procedure for variance estimation in simple random sampling using auxiliary information. *Statistical papers*, 54:479–497.

[7] Subramani, J. and Kumarapandiyan, G. (2012a). Estimation of population mean using coefficient of variation and median of an auxiliary variable. *International Journal of Probability and Statistics*, 1(4), 111-118.

[8] Subramani, J. and Kumarapandiyan, G. (2012b). Variance estimation using quartiles and their functions of an auxiliary variable. *International Journal of Statistics and Applications1*, 2(5), 67-72.

[9] Subramani, J. and Kumarapandiyan, G. (2013). Estimation of variance using known coefficient of variation and median of an auxiliary variable. *Journal of Modern Applied Statistical Methods*, 12, 58-64.

[10] Upadhyaya, L. N. and Singh, H. P. (1999). An estimator for population variance that utilizes the kurtosis of an auxiliary variable in sample surveys. *Vikram Mathematical Journal*, 19(1), 14-17.

[11] Murthy, M. N. (1967). Sampling Theory and Methods, Statistical Publishing Society Calcutta, India.

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

[13] Yadav, R. (2021). A Generalized Ratio-Type Estimator of Finite Population Variance using Quartiles and their Functions. *Journal of Statistics Applications and Probability*, 10(2), 363-368.

[14] Yadav, R., Upadhyaya, L. N., Singh, H. P. and Chatterjee, S. (2013a). A generalized family of transformed ratio-product estimators for variance in sample surveys. *Communications in Statistics-Theory and Methods*, 42(10), 1839-1850.

[15] Yadav, R., Upadhyaya, L. N., Singh, H. P. and Chatterjee, S. (2013b). A general procedure of estimating the population variance when coefficient of variation of an auxiliary variable is known in sample surveys. *Quality and Quantity*, 47(4), 2331-2339.

[16] Yadav, R., Upadhyaya, L. N., Singh, H. P. and Chatterjee, S. (2014). A class of estimators of population variance using auxiliary information in sample surveys. *Communications in Statistics-Theory and Methods*, 43(6), 1248-1260.