International Journal of Mathematics And its Applications
Volume 3, Issue 4–F (2015), 85–91.
ISSN: 2347-1557
Available Online: http://ijmaa.in/

International Journal of Mathematics And its Applications

Efficient Estimator of Population Variance Using Coefficient of Kurtosis and Population Mean of Auxiliary Variable

Research Article

S.K.Yadav[1], Ravendra Kumar[2*], Shakti Kumar[3], Shailendra Verma[4] and Shailendra Kumar[5]

1 Department of Mathematics and Statistics (A Centre of Excellence), Dr.R M L Avadh University, Faizabad, U.P., India.

2 Department of Mathematics, Dr Ram Manohar Lohiya Government Degree College, Aonla, Bareilly, U.P., India.

3 CESP School of Social Sciences, Jawahar Lal Nehru University, New Delhi, India.

4 Department of Business Administration and Interprenureship, Dr.R M L Avadh University, Faizabad, U.P., India.

5 Department of Microbiology, Dr.R M L Avadh University, Faizabad, U.P., India.

Abstract: In the present paper an efficient estimator of population variance of study variable has been proposed using knowledge of coefficient of kurtosis and the population mean of the auxiliary variable. The bias and the mean squared error of the proposed variance have been obtained up to the first order of approximation. The optimum value of the characterizing scalar, which minimizes the mean squared error, has been obtained. The minimum value of the mean squared error has been obtained for this optimum value of the characterizing scalar. A comparison has been done with the mentioned existing estimators of population variance. An empirical study is also carried out to judge the performance of the proposed estimator along with the other estimators of population variance.

Keywords: Main variable, auxiliary variable, bias, mean squared error, efficiency.

1. Introduction

Population variance plays a very important role in day today life such as one is interested in knowing the estimate of variance of a particular crop, blood pressure, temperature etc. It is one of the most important measures of dispersion. Auxiliary information plays a crucial role in drawing conclusion about the parameters of population for the characteristic under study. It is used for improved estimation of population parameters of the main variable under study. At the stage of designing as well as estimation, auxiliary information is used for improved estimation. In the present manuscript, we have used it at the estimation stage only. The study variable (Y) and the auxiliary variable (X) are closely related with each other. When the study variable and the auxiliary variables are positively correlated to each other and the line of regression of the study variable Y on the auxiliary variable X passes through origin, then the ratio type estimators are used for improved estimation of the parameters of the population under consideration. On the other hand the product type estimators are used for improved estimation of parameters when the auxiliary variable X and the study variable Y have negative correlation between them. The regression type estimators are used for the improvement in estimation of population parameters, when the line of regression does not pass through the origin.

* *E-mail: rkpmbly@gmail.com*

Let the population under investigation is finite and it consists of N distinct and identifiable units. Let $(x_i,\, y_i),\; i = 1,\, 2, \ldots, n$ be a random sample of size n from above bivariate population (X, Y) of size N using a SRSWOR scheme. Let $\bar{X}$ and $\bar{Y}$ respectively are the population means of the auxiliary and the study variables, and let $\bar{x}$ and $\bar{y}$ are the corresponding sample means which are unbiased estimators of population means $\bar{X}$ and $\bar{Y}$ respectively.

2. Review of Literature of Variance Estimators

The most appropriate estimator of population variance is the sample variance defined by,

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

This is an unbiased estimator of population variance and it has the variance up to the first degree of approximation as:

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

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

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

where

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2, \qquad s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2, \qquad S_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2,$$

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i, \qquad \bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i, \qquad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \qquad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i.$$

The expressions for the Bias and Mean Square Error (MSE) respectively up to the first order of approximations, 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

$$\lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{r/2}\mu_{02}^{s/2}}, \quad \mu_{rs} = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^r (X_i - \bar{X})^s, \quad \gamma = \frac{1-f}{n} \text{ and } f = \frac{n}{N}.$$

Singh et al. (2011) proposed the following exponential ratio type estimator for the population variance, based on Bahl and Tuteja (1991) exponential ratio type estimator for the population mean, as

$$t_1 = s_y^2 \exp\left[\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2} \right] \tag{6}$$

The MSE of the estimator in (6), up to the first order of approximation, is

$$MSE(t_1) = \gamma\, S_y^4 \left[(\lambda_{40} - 1) + \frac{(\lambda_{04} - 1)}{4} - (\lambda_{22} - 1) \right] \tag{7}$$

The exponential product type estimator for the population variance, based on Bahl and Tuteja [2] product type estimator for the population mean, can be defined as

$$t_2 = s_y^2 \exp\left[\frac{s_x^2 - S_x^2}{s_x^2 + S_x^2}\right] \tag{8}$$

The MSE of the estimator in (8), up to the first order of approximation, is

$$MSE(t_2) = \gamma\, S_y^4 \left[(\lambda_{40} - 1) + \frac{(\lambda_{04} - 1)}{4} + (\lambda_{22} - 1)\right]. \tag{9}$$

Upadhyaya and Singh [8] proposed a modified ratio estimator of population variance using the population mean of the auxiliary variable as,

$$t_3 = s_y^2 \left[\frac{\bar{X}}{\bar{x}}\right] \tag{10}$$

The bias and the mean squared error of above estimator up to the first order of approximation respectively are,

$$B(t_3) = \gamma\, S_y^2 [C_x^2 - \lambda_{21} C_x] \tag{11}$$

$$MSE(t_3) = \gamma\, S_y^4 [(\lambda_{40} - 1) + C_x^2 - 2\lambda_{21} C_x] \text{ where } \lambda_{21} = \frac{\mu_{21}}{\mu_{20}\sqrt{\mu_{02}}} \tag{12}$$

Asgar et al. [1] proposed an improved ratio and product type estimators of population variance using population mean of auxiliary variable respectively as,

$$t_4 = s_y^2 \exp\left[\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right] \tag{13}$$

$$t_5 = s_y^2 \exp\left[\frac{\bar{x} - \bar{X}}{\bar{x} + \bar{X}}\right] \tag{14}$$

The bias and mean squared error of above estimators up to the first order of approximation respectively are,

$$B(t_4) = \gamma\, S_y^2 [\frac{C_x^2}{8} - \frac{\lambda_{21}}{2} C_x] \tag{15}$$

$$MSE(t_4) = \gamma\, S_y^4 [(\lambda_{40} - 1) + \frac{C_x^2}{4} - \lambda_{21} C_x] \tag{16}$$

$$B(t_5) = \gamma\, S_y^2 [\frac{C_x^2}{8} + \frac{\lambda_{21}}{2} C_x] \tag{17}$$

$$MSE(t_5) = \gamma\, S_y^4 [(\lambda_{40} - 1) + \frac{C_x^2}{4} + \lambda_{21} C_x] \tag{18}$$

Several authors proposed different estimators by utilizing auxiliary information in different forms. They used it in the form of different parameters of auxiliary variable for estimating the population variance of the main variable under study.

3. Proposed Estimator

Motivated by Khan and Arunachalam [5] and Upadhyaya and Singh [8], we propose an improved ratio estimator of population variance as,

$$\hat{S}_{YK}^2 = s_y^2 \left[\varphi \left\{2 - \left(\frac{\bar{X} + \beta_{2x}}{\bar{x} + \beta_{2x}}\right)\right\} + (1 - \varphi) \left\{2 - \left(\frac{\bar{x} + \beta_{2x}}{\bar{X} + \beta_{2x}}\right)\right\}\right], \tag{19}$$

where φ is a suitably chosen characterizing constant and is obtained by minimizing the MSE of the proposed estimator $\hat{S}_{YK}^2$. In order to study the large sample properties of the proposed estimator $\hat{S}_{YK}^2$, we define $s_y^2 = S_y^2 (1 + e_0)$ and $\bar{x} = \bar{X}(1 + e_1)$ such that $E(e_i) = 0$ for $(i = 0, 1)$ and $E\left(e_0^2\right) = \gamma(\lambda_{40} - 1)$, $E\left(e_1^2\right) = \gamma\, C_x^2$, $E(e_0 e_1) = \gamma\, \lambda_{21} C_x$.

Expressing $\hat{S}^2_{YK}$ in terms of e_is $(i = 0,\, 1)$, we have

$$\begin{aligned}
\hat{S}^2_{YK} &= S_y^2(1+e_0)\left[\varphi\left\{2-\left(\frac{\bar{X}+\beta_{2x}}{\bar{X}(1+e_1)+\beta_{2x}}\right)\right\}+(1-\varphi)\left\{2-\left(\frac{\bar{X}(1+e_1)+\beta_{2x}}{\bar{X}+\beta_{2x}}\right)\right\}\right] \\
&= S_y^2(1+e_0)\left[\varphi\left\{2-\left(\frac{\bar{X}+\beta_{2x}}{\bar{X}e_1+\bar{X}+\beta_{2x}}\right)\right\}+(1-\varphi)\left\{2-\left(\frac{\bar{X}e_1+\bar{X}+\beta_{2x}}{\bar{X}+\beta_{2x}}\right)\right\}\right] \\
&= S_y^2(1+e_0)\left[\varphi\left\{2-\left(\frac{1}{\frac{\bar{X}e_1}{\bar{X}+\beta_{2x}}+1}\right)\right\}+(1-\varphi)\left\{2-\left(\frac{\bar{X}e_1}{\bar{X}+\beta_{2x}}+1\right)\right\}\right] \\
&= S_y^2(1+e_0)\left[\varphi\left\{2-\left(\frac{1}{1+\theta e_1}\right)\right\}+(1-\varphi)\left\{2-(1+\theta e_1)\right\}\right] \text{ where } \theta = \frac{\bar{X}}{\bar{X}+\beta_{2x}} \\
&= S_y^2(1+e_0)[\varphi\{2-(1+\theta e_1)^{-1}\}+(1-\varphi)\,(1-\theta e_1)] \\
&= S_y^2(1+e_0)[\varphi\{2-(1-\theta e_1+\theta^2 e_1^2-...)\}+(1-\varphi)\,(1-\theta e_1)] \\
&= S_y^2(1+e_0)[\varphi(1+\theta e_1-\theta^2 e_1^2+...)+(1-\varphi)\,(1-\theta e_1)]
\end{aligned}$$

Multiplying the terms in above equation, simplifying and retaining the terms in e_i's up to the first order of approximation, we have,

$$\hat{S}^2_{SK} = S_y^2[1+e_0-\theta(1-2\varphi)e_1-\varphi\theta^2 e_1^2-\theta(1-2\varphi)e_0e_1] \tag{20}$$

Subtracting S_y^2 on both sides of (20), we get

$$\hat{S}^2_{SK}-S_y^2 = S_y^2[e_0-\theta(1-2\varphi)e_1-\varphi\theta^2 e_1^2-\theta(1-2\varphi)e_0e_1] \tag{21}$$

Taking expectations on both sides of equation (21), we have,

$$E(\hat{S}^2_{SK}-S_y^2) = S_y^2[E(e_0)-\theta(1-2\varphi)E(e_1)-\varphi\theta^2 E(e_1^2)-\theta(1-2\varphi)E(e_0e_1)]$$

Putting the value of different expectations, we get the bias of $\hat{S}^2_{SK}$ up to the first order of approximation as,

$$B(\hat{S}^2_{SK}) = \gamma\, S_y^2[\theta(2\varphi-1)\lambda_{21}C_x-\varphi\theta^2 C_x^2] \tag{22}$$

From equation (21) for mean squared error of $\hat{S}^2_{SK}$, up to the first order of equation we have,

$$\hat{S}^2_{SK}-S_y^2 = S_y^2[e_0-\theta(1-2\varphi)e_1]$$

Squaring above equation, we get

$$\begin{aligned}
(\hat{S}^2_{SK}-S_y^2)^2 &= S_y^4[e_0-\theta(1-2\varphi)e_1]^2 \\
(\hat{S}^2_{SK}-S_y^2)^2 &= S_y^4[e_0^2+\theta^2(1-2\varphi)^2e_1^2-2\theta(1-2\varphi)e_0e_1]
\end{aligned}$$

Taking expectations both sides of above equation and putting the values of different expectations, we get the mean squared error of $\hat{S}^2_{SK}$, up to the first order of approximation as,

$$MSE(\hat{S}^2_{SK}) = \gamma\, S_y^4[(\lambda_{40}-1)+\theta^2(1-2\varphi)^2C_x^2-2\theta(1-2\varphi)\lambda_{21}C_x] \tag{23}$$

The optimum value of the characterizing scalar φ, which minimizes the $MSE(\hat{S}^2_{SK})$ is obtained by minimizing the $MSE(\hat{S}^2_{SK})$ using the method of maxima and minima as, $\frac{\partial\,[MSE(\hat{S}^2_{SK})]}{\partial\varphi} = 0$ gives,

$$\varphi_{opt} = \frac{1}{2}\left[1-\frac{\lambda_{21}}{\theta C_x}\right] \tag{24}$$

The minimum mean squared error of $\hat{S}^2_{SK}$ for this optimum value of φ is,

$$MSE_{\min}(\hat{S}^2_{SK}) = \gamma\, S_y^4[(\lambda_{40}-1)-\lambda_{21}^2] \tag{25}$$

4. Efficiency Comparison

From equation (25) and (2), we have

$$V(t_0) - MSE_{\min}(\hat{S}^2_{SK}) = \gamma\, S_y^4 \lambda_{21}^2 > 0 \tag{26}$$

Thus we see from above equation that the proposed estimator $\hat{S}^2_{SK}$ is better that the estimator t_0 as it has lesser mean squared error. From equation (25) and (5), we have $MSE(t_R) - MSE_{\min}(\hat{S}^2_{SK}) = \gamma\, S_y^4[(\lambda_{04}-1) - 2(\lambda_{22}-1) + \lambda_{21}^2] > 0$. If,

$$[(\lambda_{04}-1) + \lambda_{21}^2] > 2(\lambda_{22}-1) \tag{27}$$

Thus under the above condition, the proposed estimator performs better than the estimator t_R. From equation (25) and (7), we have $MSE(t_1) - MSE_{\min}(\hat{S}^2_{SK}) = \gamma\, S_y^4 \left[\frac{(\lambda_{04}-1)}{4} - (\lambda_{22}-1) + \lambda_{21}^2\right] > 0$. If,

$$\frac{(\lambda_{04}-1)}{4} + \lambda_{21}^2 > (\lambda_{22}-1) \tag{28}$$

Under the above condition, the proposed estimator $\hat{S}^2_{SK}$ will have lesser mean squared error as compared to the estimator t_1. Thus it will be preferable over t_1. From equation (25) and (9), we have

$$MSE(t_2) - MSE_{\min}(\hat{S}^2_{SK}) = \gamma\, S_y^4 \left[\frac{(\lambda_{04}-1)}{4} + (\lambda_{22}-1) + \lambda_{21}^2\right] > 0, \tag{29}$$

Thus we see that the proposed estimator $\hat{S}^2_{SK}$ is having lesser mean squared error as compared to estimator t_2. Thus it will be preferable over t_2 for the estimation of population variance. From equation (25) and (12), we have

$$MSE(t_3) - MSE_{\min}(\hat{S}^2_{SK}) = \gamma\, S_y^4[(C_x^2 - 2\lambda_{21}C_x + \lambda_{21}^2] = \gamma\, S_y^4[C_x - \lambda_{21}]^2 > 0, \tag{30}$$

Thus the proposed estimator $\hat{S}^2_{SK}$ is better than the estimator t_3 as it has lesser mean squared error as compared to t_3. From equation (25) and (16), we have

$$MSE(t_4) - MSE_{\min}(\hat{S}^2_{SK}) = \gamma\, S_y^4[\frac{C_x^2}{4} - \lambda_{21}C_x + \lambda_{21}^2] = \gamma\, S_y^4[\frac{C_x}{2} - \lambda_{21}]^2 > 0 \tag{31}$$

Above equation shows that the proposed estimator $\hat{S}^2_{SK}$ is better than the estimator t_4 as it has lesser mean squared error as compared to t_4. From equation (25) and (18), we have

$$MSE(t_5) - MSE_{\min}(\hat{S}^2_{SK}) = \gamma\, S_y^4[\frac{C_x^2}{4} + \lambda_{21}C_x + \lambda_{21}^2] = \gamma\, S_y^4[\frac{C_x}{2} + \lambda_{21}]^2 > 0 \tag{32}$$

Thus we see that the proposed estimator $\hat{S}^2_{SK}$ is better than the estimator t_5 as it has lesser mean squared error as compared to t_5.

5. Numerical Illustration

To examine the performance of the proposed estimator along with the other estimators, the empirical study has been carried out using two real populations. The Source, description and parameters for two populations are given below;

Population I- Source: Murthy [6]

X: output

Y: Number of workers

$N = 25$, $n = 25$, $\bar{X} = 283.875$, $\bar{Y} = 33.8465$, $C_y = 0.352$, $C_x = 0.746$, $\rho_{yx} = 0.9136$, $\lambda_{40} = 2.2667$, $\lambda_{04} = 3.65$, $\lambda_{21} = 1.0475$, $\lambda_{22} = 2.3377$.

Population II- Source: Gujarati [3]

X: Average (miles per gallon)

Y: Top Speed (miles per hour)

$N = 81$, $n = 21$, $\bar{X} = 112.4568$, $\bar{Y} = 2137.086$, $C_y = 0.1248$, $C_x = 0.48$, $\rho_{yx} = -0.691135$, $\lambda_{40} = 3.59$, $\lambda_{04} = 6.82$, $\lambda_{21} = 1.4137$, $\lambda_{22} = 2.110$.

Table 1. Percent relative efficiency of different estimators with respect to t_0.

Estimator	Population-I	Population-II
t_0	100.00	100.00
t_R	102.04	$*$
t_1	214.15	$*$
t_2	$*$	50.24
t_3	489.40	177.00
t_4	131.54	202.86
t_5	$*$	77.86
$\hat{S}^2_{\gamma K}$	**747.56**	**521.32**

$*$ Data is not applicable.

6. Results and Conclusion

In the present manuscript, we have proposed a generalized type estimator of population variance using information on coefficient of kurtosis and the mean of the auxiliary variable. The bias and the mean squared error of the proposed estimator have been obtained up the first order of approximation. The optimum value of the characterizing scalar is obtained and for this optimum value of characterizing scalar, the minimum value of mean squared error of proposed estimator has been obtained. Form the theoretical discussion in section-4 and the numerical study given in table-1, we see that the proposed estimator is much more efficient in estimating population variance of the study variable as compared to other mentioned estimators of population variance. Therefore it is strongly recommended that the proposed estimator should be preferred over other estimators for the estimation of population variance.

References

[1] Amber Asghar, Aamir Sananullah and Muhammad Hanif, *Generalized Exponential Type Estimators for Population variance in Survey Sampling*, Revista Colombiana deEstadstica, 37(1)(2014), 213-224.

[2] S.Bhal and R.K.Tuteja, *Ratio and product type exponential estimator*, Information and Optimization Sciences, 12(1991), 159-163.

[3] D.Gujarati, *Basic Econometrics*, 4 Edn, The McGraw-Hill Companies, (2004).

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

[5] M.Khan and A.Arunachalam, *Estimation of Population Variance using the Knowledge of Kurtosis of an Auxiliary Variable under Simple Random Sampling*, International Journal of Applied Science and Mathematics, 1(2)(2014), 63-67.

[6] M.Murthy, *Sampling Theory and Methods*, Calcutta Statistical Publishing Society, Kolkatta, India, (1967).

[7] R.S.Singh, P.Chauhan, N.Sawan and F.Smarandache, *Improved exponential estimator for population variance using two auxiliary variables*, Italian Journal of Pure and Applied Mathematics, 28(2011), 101-108.

[8] L.N.Upadhyaya and G.N.Singh, Chain-type estimators using transformed auxiliary variable in two-phase sampling. A.M.S.E., 38(1-2)(2001), 1-9.