Volume 26 Issue 02, 2023 ISSN: 1005-3026 https://dbdxxb.cn/ Original Research Paper

ON SOME IMPROVED ESTIMATOR FOR ESTIMATION OF POPULATION VARIANCE USING AUXILIARY INFORMATION

Peeyush Misra and Manish Kumar*
Department of Statistics, D.A.V. (P.G.) College, Dehradun – 248001, Uttarakhand (India)
dr.pmisra.dav@gmail.com and kumarmann03@gmailcom

Abstract: The purpose of this paper is to use auxiliary data to estimate the population variance of the study variable. A estimator for estimating population variance has been proposed to the first order of approximation and bias and mean squared error are produced. The efficiency of the suggested estimator is proven to be higher than that of the usual unbiased estimator. The present study is supported by an empirical study also.

Keywords: Auxiliary Information, Coefficient of Variation, Bias, Mean Squared Error.

1. Introduction

It is commonly known fact that using auxiliary data in sample surveys significantly improves the accuracy of estimators of population parameters like mean, variance and coefficient of variation of the study variable. When auxiliary information is available, researchers want to use it in the estimation process to provide improved estimator of the parameter under study. Das et al (1978), Isaki (1983), Kadilar C. and Cingi H. (2007),Yadav and Kadilar (2013) and Singh , H. P. et al.(2014) Misra and Singh (2014), Misra (2015) and many others have examined the issue of estimating the population variance.

Let,

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i \quad , \quad \bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^r\left(Y_i - \bar{Y}\right)^s$$

Also

$$\sigma_Y^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2 , \; C_Y = \frac{\sigma_Y}{\bar{Y}} \; , C_X = \frac{\sigma_X}{\bar{X}} \text{ and}$$

$$\sigma_{XY} = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) \text{Also } \rho = \frac{\rho_{XY}}{\sigma_X \sigma_Y}, B = \frac{\sigma_{xy}}{\sigma_x^2}$$

Also, let

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \; , \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Submitted: 22/04/2023 **Accepted**: 10/05/2023

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 , \quad s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right), \quad b = \frac{S_{xy}}{S_x^2}$$

where $y_1, y_2, \ldots y_n$ are the observations on y and $x_1, x_2, \ldots x_n$ are the observations on auxiliary variable x for a simple random sample of size n.

For estimating population variance, An improved estimator is proposed as

$$s_Y^2 = \hat{\sigma}_Y^2 = s_Y^2 \exp\left\{k_1\left(\frac{s_x^2}{C_X^2} - \bar{x}^2\right)\right\}\exp\left\{k_2\left(\frac{s_y^2}{C_Y^2} - \bar{y}^2\right)\right\} \tag{1.1}$$

2. Bias and Mean Square Error

Let, $\bar{y} = \bar{Y}\left(1 + e_0\right)$

$\bar{x} = \bar{X}\left(1 + e_1\right)$

$s_y^2 = \sigma_y^2\left(1 + e_2\right)$

$s_x^2 = \sigma_X^2\left(1 + e_3\right)$

$s_{xy} = \sigma_{XY}\left(1 + e_4\right)$

So that $E(e_0) = E(e_1) = E(e_2) = E(e_3) = 0$

$$E\left(e_0^2\right) = \frac{C_Y^2}{n}, E\left(e_1^2\right) = \frac{C_X^2}{n} \tag{2.2}$$

$$E\left(e_2^2\right) = \frac{\beta_2(y) - 1}{n} = \frac{A_Y}{n}, where\ A_Y = \beta_2(y) - 1 \ (2.3) \tag{2.3}$$

$$E(e_0 e_1) = \frac{\rho C_X C_Y}{n} \tag{2.4}$$

$$E\left(e_0 e_2\right) = \frac{\mu_{03}}{n\sigma_Y^2\bar{Y}} \text{and } E\left(e_1 e_2\right) = \frac{\mu_{12}}{n\sigma_Y^2\bar{X}} \tag{2.5}$$

$$E\left(e_1 e_3\right) = \frac{\mu_{30}}{n\sigma_X^2\bar{X}} \tag{2.6}$$

$$E\left(e_3^2\right) = \frac{\beta_2(X) - 1}{n} = \frac{A_X}{n}, where\ A_X = \beta_2(x) - 1 \tag{2.7}$$

$$E\left(e_0 e_3\right) = \frac{\mu_{21}}{n\sigma_X^2\bar{Y}} \tag{2.8}$$

$$E\left(e_2 e_3\right) = \frac{1}{n\sigma_X^2\sigma_Y^2}\left(\mu_{22} - \mu_{20}\mu_{02}\right) = \frac{\delta - 1}{n}, \text{ Where } \delta = \frac{\mu_{22}}{\mu_{20}\mu_{02}} \tag{2.9}$$

Now Expression (1) in terms of e_i's, we have

$$\hat{\sigma}_Y^2 = \sigma_Y^2(1+e_2)\exp\left[k_1\left\{\frac{\sigma_X^2(1+e_3)}{\left(\frac{\sigma_X^2}{\overline{X}^2}\right)} - \overline{X}^2(1+e_1)^2\right\}\right]\exp\left[k_2\left\{\frac{\sigma_Y^2(1+e_2)}{\left(\frac{\sigma_Y^2}{\overline{Y}^2}\right)} - \overline{Y}^2(1+e_0)^2\right\}\right].$$

(2.10)

Or

$$\hat{\sigma}_Y^2 = \left(\sigma_Y^2 + \sigma_Y^2 e_2\right)\exp\left\{k_1\overline{X}^2\left(1+e_3-1-e_1^2-2e_1\right)\right\}\exp\left\{k_2\overline{Y}^2\left(1+e_2-1-e_0^2-2e_0\right)\right\}$$

$$\hat{\sigma}_Y^2 = \left(\sigma_Y^2 + \sigma_Y^2 e_2\right)\exp\left\{k_1\overline{X}^2\left(e_3-2e_1-e_1^2\right)\right\}\exp\left\{k_2\overline{Y}^2\left(e_2-2e_0-e_0^2\right)\right\} \qquad . \qquad (2.11)$$

Or

$$\hat{\sigma}_Y^2 = \left(\sigma_Y^2 + \sigma_Y^2 e_2\right)\left\{1 + k_1\overline{X}^2\left(e_3-2e_1-e_1^2\right) + \frac{1}{2}k_1^2\overline{X}^4\left(e_3^2+4e_1^2-4e_1e_3\right)\right\}$$
$$\left\{1 + k_2\overline{Y}^2\left(e_2-2e_0-e_0^2\right) + \frac{1}{2}k_2^2\overline{Y}^4\left(e_2^2+4e_0^2-4e_0e_2\right)\right\} \qquad (2.12)$$

Or

$$\hat{\sigma}_Y^2 = \left(\sigma_Y^2 + \sigma_Y^2 e_2\right)\left\{1 + k_2\overline{Y}^2\left(e_2-2e_0-e_0^2\right) + \frac{1}{2}k_2^2\overline{Y}^4\left(e_2^2+4e_0^2-4e_0e_2\right)\right.$$
$$+ k_1\overline{X}^2\left(e_3-2e_1-e_1^2\right) + k_1k_2\overline{X}^2\overline{Y}^2\left(e_2e_3-2e_0e_3-2e_1e_2+4e_0e_1\right)$$
$$\left. + \frac{1}{2}k_1^2\overline{X}^4\left(e_3^2+4e_1^2-4e_1e_3\right)\right\}$$

$$\hat{\sigma}_Y^2 = \left(\sigma_Y^2 + \sigma_Y^2 e_2\right)\left\{1 + k_1\overline{X}^2\left(e_3-2e_1-e_1^2\right) + k_2\overline{Y}^2\left(e_2-2e_0-e_0^2\right) + \frac{1}{2}k_1^2\overline{X}^4\left(e_3^2+4e_1^2-4e_1e_3\right)\right.$$
$$\left. + \frac{1}{2}k_2^2\overline{Y}^4\left(e_2^2+4e_0^2-4e_0e_2\right) + k_1k_2\overline{X}^2\overline{Y}^2\left(e_2e_3-2e_0e_3-2e_1e_2+4e_0e_1\right)\right\}$$

$$\hat{\sigma}_Y^2 = \sigma_Y^2 + \sigma_Y^2\left\{k_1\overline{X}^2\left(e_3-2e_1-e_1^2\right) + k_2\overline{Y}^2\left(e_2-2e_0-e_0^2\right) + \frac{1}{2}k_1^2\overline{X}^4\left(e_3^2+4e_1^2-4e_1e_3\right)\right.$$
$$\left. + \frac{1}{2}k_2^2\overline{Y}^4\left(e_2^2+4e_0^2-4e_0e_2\right) + k_1k_2\overline{X}^2\overline{Y}^2\left(e_2e_3-2e_0e_3-2e_1e_2+4e_0e_1\right)\right\}$$

$$+\sigma_Y^2 e_2 + \sigma_Y^2\left\{k_1\overline{X}^2\left(e_2e_3 - 2e_1e_2\right) + k_2\overline{Y}^2\left(e_2^2 - 2e_0e_2\right)\right\}$$

$$\hat{\sigma}_Y^2 - \sigma_Y^2 = \sigma_Y^2\left(e_2 + \overline{X}^2k_1e_3 - 2\overline{X}^2k_1e_1 + \overline{Y}^2k_2e_2 - 2\overline{Y}^2k_2e_0\right)$$
$$+\frac{1}{2}\sigma_Y^2\left\{k_1^2\overline{X}^4\left(e_3^2 + 4e_1^2 - 4e_1e_3\right) + k_2^2\overline{Y}^4\left(e_2^2 + 4e_0^2 - 4e_0e_2\right) - 2k_1\overline{X}^2\left(e_1^2 - e_2e_3 + 2e_1e_2\right)\right.$$
$$\left. - 2k_2\overline{Y}^2\left(e_0^2 + 2e_0e_2 - e_2^2\right) + 2k_1k_2\overline{X}^2\overline{Y}^2\left(e_2e_3 - 2e_0e_3 - 2e_1e_2 + 4e_0e_1\right)\right\} \quad (2.13)$$

In the first order of approximation, the bias $\hat{\sigma}_Y^2$ is given by considering the expectation on both sides of (2.1).

$$Bias\left(\hat{\sigma}_Y^2\right) = E\left(\hat{\sigma}_Y^2 - \sigma_Y^2\right)$$

$$Bias\left(\hat{\sigma}_Y^2\right) = \frac{\sigma_Y^2}{2}\left[k_1^2\overline{X}^4\left\{\frac{A_X}{n} + 4\frac{C_X^2}{n} - 4\frac{\mu_{30}}{n\sigma_X^2\overline{X}}\right\} + k_2^2\overline{Y}^4\left\{\frac{A_Y}{n} + 4\frac{C_Y^2}{n} - 4\frac{\mu_{03}}{n\sigma_Y^2\overline{Y}}\right\}\right)$$
$$-2k_1\overline{X}^2\left\{\frac{C_X^2}{n} - \frac{(\delta-1)}{n} + 2\frac{\mu_{12}}{n\sigma_Y^2\overline{X}}\right\} - 2k_2\overline{Y}^2\left\{\frac{C_Y^2}{n} + 2\frac{\mu_{03}}{n\sigma_Y^2\overline{Y}} - \frac{A_Y}{n}\right\}$$
$$+2k_1k_2\overline{X}^2\overline{Y}^2\left\{\frac{(\delta-1)}{n} - 2\frac{\mu_{21}}{n\sigma_X^2\overline{Y}} - 2\frac{\mu_{12}}{n\sigma_Y^2\overline{X}} + 4\frac{\rho C_X C_Y}{n}\right\}\right|$$

$$Bias\left(\hat{\sigma}_Y^2\right) = \left|\frac{\sigma_Y^2}{2n}k_1^2\overline{X}^4\left(A_X + 4C_X^2 - 4\frac{\mu_{30}}{\overline{X}\sigma_X^2}\right) + k_2^2\overline{Y}^4\left(A_Y + 4C_Y^2 - 4\frac{\mu_{03}}{\overline{Y}\sigma_Y^2}\right)\right.$$
$$-2k_1\overline{X}^2\left\{C_X^2 + 2\frac{\mu_{12}}{\overline{X}\sigma_Y^2} - (\delta-1)\right\} - 2k_2\overline{Y}^2\left\{C_Y^2 + 2\frac{\mu_{03}}{\overline{Y}\sigma_Y^2} - A_Y\right\}$$
$$\left.+2k_1k_2\overline{X}^2\overline{Y}^2\left\{(\delta-1) - 2\frac{\mu_{21}}{\overline{Y}\sigma_X^2} - 2\frac{\mu_{12}}{\overline{X}\sigma_Y^2} + 4\rho C_X C_Y\right\}\right| \quad (2.14)$$

squaring (2.13) on both sides and then taking expectation, we have m.s.e to the first order of approximation is :

$$MSE\left(\hat{\sigma}_Y^2\right) = E\left(\hat{\sigma}_Y^2 - \sigma_Y^2\right)^2$$

$$= E\left\{\sigma_Y^2\left(e_2 + \overline{X}^2k_1e_3 - 2\overline{X}^2k_1e_1 + \overline{Y}^2k_2e_2 - 2\overline{Y}^2k_2e_0\right)\right\}^2$$

$$= E\left|\sigma_Y^4\left\{e_2^2 + k_1^2\overline{X}^4\left(e_3^2 + 4e_1^2 - 4e_1e_3\right) + k_2^2\overline{Y}^4\left(e_2^2 + 4e_0^2 - 4e_1e_2\right) + 2\overline{X}^2k_1\left(e_2e_3 - 2e_1e_2\right)\right.\right.$$

$$\left.\left. + 2\overline{Y}^2k_2\left(e_2^2 - 2e_0e_2\right) + 2k_1k_2\overline{X}^2\overline{Y}^2\left(e_2e_3 - 2e_0e_3 - 2e_1e_2 + 4e_0e_1\right)\right\}\right|$$

$$= \sigma_Y^4E\left(e_2^2\right) + \sigma_Y^4\left[\overline{X}^4k_1^2E\left(e_3^2 + 4e_1^2 - 4e_1e_3\right) + \overline{Y}^4k_2^2E\left(e_2^2 + 4e_0^2 - 4e_1e_2\right)\right.$$

$$+ 2\overline{X}^2kE\left(e_2e_3 - 2e_1e_2\right) + 2\overline{Y}^2k_2E\left(e_2^2 - 2e_0e_2\right)$$

$$+ 2k_1k_2\overline{X}^2\overline{Y}^2E\left(e_2e_3 - 2e_0e_3 - 2e_1e_2 + 4e_0e_1\right)$$

or,

$$MSE\left(\hat{\sigma}_Y^2\right) = \sigma_Y^4\frac{A_Y}{n} + \sigma_Y^4\left[\overline{X}^4k_1^2\left(\frac{A_X}{n} + 4\frac{C_X^2}{n} - 4\frac{\mu_{30}}{n\sigma_X^2\overline{X}}\right) + \overline{Y}^4k_2^2\left(\frac{A_Y}{n} + 4\frac{C_Y^2}{n} - 4\frac{\mu_{03}}{n\sigma_Y^2\overline{Y}}\right)\right.$$

$$- 2\overline{X}^2k_1\left(\frac{(\delta-1)}{n} + 2\frac{\mu_{12}}{n\sigma_Y^2\overline{X}}\right) + 2\overline{Y}^2k_2\left(\frac{A_Y}{n} - 2\frac{\mu_{03}}{n\sigma_Y^2\overline{Y}}\right)$$

$$\left. 2k_1k_2\overline{X}^2\overline{Y}^2\left(\frac{(\delta-1)}{n} - 2\frac{\mu_{21}}{n\sigma_X^2\overline{Y}} - 2\frac{\mu_{12}}{n\sigma_Y^2\overline{X}} + 4\frac{\rho C_XC_Y}{n}\right)\right] \quad (2.15)$$

or

$$MSE\left(\hat{\sigma}_Y^2\right) = \sigma_Y^4\frac{A_Y}{n} + \sigma_Y^4\left[k_1^2\delta_{11} + k_2^2\delta_{22} + 2k_1\delta_{10} + 2k_2\delta_{02} - 2k_1k_2\delta_{12}\right] \quad (2.16)$$

Where

$$\delta_{11} = \overline{X}^4\left(\frac{A_X}{n} + 4\frac{C_X^2}{n} - 4\frac{\mu_{30}}{n\sigma_X^2\overline{X}}\right),$$

$$\delta_{22} = \overline{Y}^4\left(\frac{A_Y}{n} + 4\frac{C_Y^2}{n} - 4\frac{\mu_{03}}{n\sigma_Y^2\overline{Y}}\right)$$

$$\delta_{10} = \overline{X}^2\left(\frac{(\delta-1)}{n} + 2\frac{\mu_{12}}{n\sigma_Y^2\overline{X}}\right),$$

$$\delta_{02} = \overline{Y}^2\left(\frac{A_Y}{n} - 2\frac{\mu_{03}}{n\sigma_Y^2\overline{Y}}\right)$$

$$\delta_{12} = \overline{X}^2\overline{Y}^2\left(\frac{(\delta-1)}{n} - 2\frac{\mu_{21}}{n\sigma_X^2\overline{Y}} - 2\frac{\mu_{12}}{n\sigma_Y^2\overline{X}} + 4\frac{\rho C_XC_Y}{n}\right)\Bigg]$$

The two normal equations for k_1 &k_2 is as follows

$$\delta_{11}k_1 + \delta_{12}k_2 + \delta_{10} = 0$$

(2.17)

$$\delta_{12}k_1 + \delta_{22}k_2 + \delta_{02} = 0 \quad (2.18)$$

on solving the two normal equations, we have

$$k_1 = \frac{\delta_{22}\delta_{10} - \delta_{02}\delta_{12}}{\delta_{12}^2 - \delta_{11}\delta_{22}} \quad (2.19)$$

and

$$k_2 = \frac{\delta_{11}\delta_{02} - \delta_{12}\delta_{10}}{\delta_{12}^2 - \delta_{11}\delta_{22}} \quad 2.20)$$

For these optimum values of k_1& k_2, the minimum mse from (2.14) is given by

$$MSE\left(s_Y^2\right)\min = \sigma_Y^4 \frac{A_Y}{n} - \sigma_Y^4 \frac{\left(\delta_{11}\delta_{02}^2 + \delta_{22}\delta_{10}^2 - 2\delta_{02}\delta_{10}\delta_{12}\right)}{\left(\delta_{11}\delta_{22} - \delta_{12}^2\right)} \quad (2.21)$$

3. Theoretical Comparison

a) We compare the proposed estimator s_y^2 with respect to usual unbiased estimator of population variance s_y^2 and the condition for which the proposed estimator will be efficient is given by

$$MSE\left(s_y^2\right)_{\min} - MSE\left(s_y^2\right) < 0 \quad (3.1)$$

b) We compare the proposed estimator s_y^2 with respect to Misra et al (2017) of population variance s_{yk}^2 and the condition for which the proposed estimator will be efficient is given by

$$2\gamma_1 y C_y < \gamma_{2y} + 2 \quad (3.2)$$

4. Empirical Study

For empirical study, we consider the data as
Data given in Cochran (1977, pg. 34) related to the weekly food expenditures of 33 low-income families are used to calculate the suitable values.

$n = 33,$

$\overline{X} = 72.54,$

$\overline{Y} = 27.49$

$\sigma_X^2 = 108.69,$

$\sigma_Y^2 = 99.613033,$

$C_y = 0.363036,$

$C_X = 0.143719,$

$\gamma_{1Y} = 1.4651,$

$\gamma_{2y} = 2.7146,$

MSE's of the estimators

a) $MSE\left(s_y^2\right) = 1417.63112$

$MSE\left(s_y^2\right)_{\min} = 150.79192$

b) $MSE\left(s_{yk}^2\right) = 1065.13127$

$MSE\left(s_y^2\right)_{\min} = 150.79192$

5. Conclusion

The performance of the proposed estimator has been established theoretically and empirically both. To evaluate the effectiveness of the estimators, use the mean squared error criterion. Comparing the suggested estimator to study and compare the standard unbiased estimator for population variance has performed better than in terms of MSE.

The percent relative efficiency (PRE) is 940% and 710 % above the usual unbiased estimator for population variance and hence the proposed estimator exhibits its improved efficiency.

Acknowledgement

The authors respectfully thank the editor in chief and the anonymous referees for their insightful suggestions for the betterment and improvement of the paper.

References

[1] Cochran W.G. (1977). Sampling Techniques, 3rd edition, John Wiley and Sons, New York.

[2] Das, A.K. and Tripathi, T. P. (1978): Use of auxiliary information in estimating the finite population variance. Sankhya, 40, C, 139-148.

[3] Isaki, C. T. (1983): Variance estimation using auxiliary information. *Jour. Amer. Statist.*

Assoc., 78, 117-123.

[4] Kadilar C. and Cingi H. (2007). Improvement in variance estimation in simple random Sampling *Communication in Statistics –Theory and Methods*, 36, 2075-2081.

[5] Misra P. and Singh R.K. (2014). Estimation of population variance using a generalized double sampling estimator. *Sri Lankan Journal of Applied Statistics*, 15(3), 211- 220.

[6] Misra P. (2015). On Improved Estimation of Population Variance in Double Sampling using Auxiliary Information. *International Journal of Essential Sciences*, 9, 23-35.

[7] Misra, S., Kumari. D., and Yadav. K.D. (2017) : An Improved Estimator of Population Variance using known Coefficient of Variation. *Journal of Statistics Applications & Probability Letters.* Vol 4(1),11-16. DOl:10. 18576/jsapl/040102.

[8] Singh, H.P., Pal, S.K. and Solanki, R.S. (2014): A new procedure for estimation of finite population variance using auxiliary information. *Jour. Reliab. Statist. Stud.,* 7(2), 149-160

[9] Sukhatme P.V. and Sukhatme B.V. (1970). Sampling theory of surveys with applications, 3rd revised edition, IOWA State University Press, Ames, U.S.A.

[10] 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. *Commun. Statist. Theo. Method*. 43(6), 1248-1260.

[11] Yadav, S.K. and Kadilar, C. (2013): Improved exponential type ratio estimator of population variance. *Rev.Colum.de Estadist*., 36(1), 145-152.