Vol. 13. Issue.3. 2025 (July-Sept.)

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

http://www.bomsr.com
Email:editorbomsr@gmail.com

INTERNATIONAL STANDARD SERIAL NUMBER 2348-0580

RESEARCH ARTICLE

Enhanced Ratio-Type Population Variance Estimator Including Excess Kurtosis and Known Median of an Auxiliary Variable

Chandni Kumari
Assistant Professor, Department of Statistics, University of Lucknow, Lucknow, Uttar Pradesh, India
Email: chandnithakur897@gmail.com

DOI:10.33329/bomsr.13.3.19

Chandni Kumari

Article Info
Article Received: 29/06/2025
Article Accepted: 26/07/2025
Published online: 01/08/2025

Abstract

We propose a ratio-type estimator for population variance that leverages known auxiliary-variable median (M_x) and excess kurtosis (K_x). Derivations of bias and Mean Squared Error (MSE) via first-order Taylor series are presented. Analytical efficiency conditions show it outperforms traditional unbiased estimators and earlier ratio-type methods. A numerical analysis indicates that the suggested estimator performs well in terms of a reduced mean squared error.

Keywords: Median, Kurtosis, Auxiliary Variable, Bias and Mean squared error.

1. Introduction

A key goal of statistical analysis is the accurate assessment of population variance, especially in domains like economics, quality control, agriculture and the social sciences. In order to evaluate variability, create confidence intervals and carry out hypothesis testing, the population variance is necessary. However, gathering complete population data is frequently expensive or impossible in real-world applications, therefore effective estimation from a sample is essential.

Statisticians commonly use auxiliary information—features of a correlated variable that are either known or easier to measure than the research variable—to improve the accuracy of such calculations. The ability of an auxiliary variable's median and kurtosis to enhance population variance estimators has drawn attention among other auxiliary statistics.

The median serves as a reliable indicator of central tendency, especially useful when the accompanying variable shows skewness or includes outliers. In contrast to the mean, the median is less influenced by extreme values, rendering it particularly insightful in real-world datasets that do not follow a normal distribution—for instance, data on household income, biological measurements or response times in psychological studies.

Conversely, kurtosis offers insights into the sharpness of a distribution. In fields like finance, hydrology and environmental monitoring, where extreme occurrences (such as stock market crashes, floods or pollution surges) are critical, understanding excess kurtosis assists in more accurately describing the shape and spread of the population distribution. When this information is available for an auxiliary variable, it can be utilized to improve estimators of population variance by taking into account the anticipated departure from normality. Searls (1964), Sisodia (1961), Kadilar (2006), Khan (2013), Yadav (2016), Kumari & Thakur (2018, 2019, 2020), and Kumar et al. (2025) employed auxiliary information to enhance the work they proposed.

An auxiliary variable's median and kurtosis can be combined to model the population structure more thoroughly and robustly. This method works especially well for survey sampling, where supplementary data from earlier research or census data is easily accessible. According to Kumari and Thakur (2020a, 2020b) and Subramani (2012), using both parameters in the estimation process can significantly lower the Mean squared error (MSE) when compared to traditional unbiased estimators or those that only use one auxiliary statistic. We present a new population variance estimator in this study that takes advantage of an auxiliary variable's excess kurtosis and known median. Using real-world datasets, we show its practical utility, derive its theoretical properties and compare its performance analytically and empirically.

2. Notation & Setup

- A simple random sample of size N with no replacement is taken from a finite population.
- Examine variables y and x, with excess kurtosis K(x) and population median M(x) known.
- Estimate population variance S_y^2 of y.
- $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$, $S_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$
- $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$, $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
- $\mu_{rs} = \frac{1}{N-1}\sum_{i=1}^{n}(y_i - \bar{y})^r (x_i - \bar{x})^s$, $\lambda_{22} = \frac{\mu_{22}}{\mu_{02}\mu_{20}}$
- $K_x = \lambda_{04} = \frac{\mu_{04}}{\mu_{02}^2}$ and $K_y = \lambda_{40} = \frac{\mu_{04}}{\mu_{20}^2}$ are coefficient of kurtosis of the study and auxiliary variable.
- M_x be the population median of x.

3. Proposed Estimator

Define the ratio-type estimator:

$$t_c = \alpha\, s_y^2 + (1-\alpha)\, s_y^2 \log\frac{[K_x s_x^2 + M_x^2]}{[K_x S_x^2 + M_x^2]} \qquad (1)$$

Where α is a constant.

To obtain the bias of the proposed estimator, we have

$$t_c = \alpha\, S_y^2 (1+\varepsilon_0) + (1-\alpha)\, S_y^2 (1+\varepsilon_0) \log\frac{[K_x S_x^2 (1+\varepsilon_1) + M_x^2]}{[K_x S_x^2 + M_x^2]}$$

$$= \alpha\, S_y^2 (1+\varepsilon_0) + (1-\alpha)\, S_y^2 (1+\varepsilon_0) \log(1+\eta\,\varepsilon_1), \text{ where } \boldsymbol{\eta} = \frac{[K_x S_x^2]}{[K_x S_x^2 + M_x^2]}$$

$$= \alpha\, S_y^2 (1+\varepsilon_0) + (1-\alpha)\, S_y^2 (1+\varepsilon_0) \left(\eta\,\varepsilon_1 - \frac{(\eta\,\varepsilon_1)^2}{2}\right)$$

$$= \alpha\, S_y^2 (1+\varepsilon_0) + (1-\alpha)\, S_y^2 \left(\eta\,\varepsilon_1\varepsilon_0 + \eta\,\varepsilon_1 - \frac{(\eta\,\varepsilon_1)^2}{2}\right) \qquad (2)$$

Substracting S_y^2 on both the sides of above equation, we get

$$t_c - S_y^2 = \alpha\, S_y^2 (1+\varepsilon_0) + (1-\alpha)\, S_y^2 \left(\eta\,\varepsilon_1\varepsilon_0 + \eta\,\varepsilon_1 - \frac{(\eta\,\varepsilon_1)^2}{2}\right) - S_y^2$$

$$t_c - S_y^2 = (\alpha-1)\, S_y^2 + \left(\alpha\, S_y^2\, \varepsilon_0\right) + (1-\alpha)\, S_y^2 \left(\eta\,\varepsilon_1\varepsilon_0 + \eta\,\varepsilon_1 - \frac{(\eta\,\varepsilon_1)^2}{2}\right) \qquad (3)$$

Taking Expectation on both the sides, we have

$$\mathrm{E}\left(t_c - S_y^2\right) = E\left\{(\alpha-1)S_y^2\right\} + E\left(\alpha\, S_y^2\, \varepsilon_0\right) + (1-\alpha)\, S_y^2\ \mathrm{E}\left(\eta\,\varepsilon_1\varepsilon_0 + \eta\,\varepsilon_1 - \frac{(\eta\,\varepsilon_1)^2}{2}\right)$$

$$= E\left\{(\alpha-1)S_y^2\right\} + E\left(\alpha\, S_y^2\, \varepsilon_0\right) + S_y^2\, (1-\alpha)\left\{\eta\, E(\varepsilon_1\varepsilon_0) - E\,\frac{(\eta\,\varepsilon_1)^2}{2}\right\} \qquad (4)$$

Using the formulae, in above expression we obtained the expression of bias.

$\mathrm{E}(\varepsilon_0) = \mathrm{E}(\varepsilon_1) = 0$, $\mathrm{E}(\varepsilon_0)^2 = \delta\,(\lambda_{40}-1)$, $\mathrm{E}(\varepsilon_1)^2 = \delta\,(\lambda_{04}-1)$, $\mathrm{E}(\varepsilon_0\varepsilon_1) = \delta\,(\lambda_{22}-1)$

$$\mathrm{Bias}(t_c) = (\alpha-1)\, S_y^2 + (1-\alpha)\, S_y^2\, \delta\left\{\eta\,(\lambda_{22}-1) - \frac{\eta^2}{2}(\lambda_{04}-1)\right\} \qquad (5)$$

To obtain the mean squared error of the proposed estimator, let us take the square of equation (3) on both the sides,

$$\left(t_c - S_y^2\right)^2 = (\alpha-1)^2 S_y^4 + \left(\alpha^2 S_y^4\, \varepsilon_0^2\right) + (\alpha-1)^2 S_y^4 \left(\eta\,\varepsilon_1\varepsilon_0 + \eta\,\varepsilon_1 - \frac{(\eta\,\varepsilon_1)^2}{2}\right)^2$$
$$+ 2\, S_y^4\, \alpha(\alpha-1)(\varepsilon_0) - 2S_y^4 (\alpha-1)^2 \left(\eta\,\varepsilon_1\varepsilon_0 + \eta\,\varepsilon_1 - \frac{(\eta\,\varepsilon_1)^2}{2}\right)$$
$$- 2\, S_y^4\, \alpha(\alpha-1)(\eta\,\varepsilon_1\varepsilon_0) \qquad (6)$$

Taking expectation on both the sides of equation (6), we get

$$E\left(t_c - S_y^2\right)^2 = (\alpha-1)^2 S_y^4 + E\left(\alpha^2 S_y^4\, \varepsilon_0^2\right) + (\alpha-1)^2 S_y^4\ \mathrm{E}\left(\eta\,\varepsilon_1\varepsilon_0 + \eta\,\varepsilon_1 - \frac{(\eta\,\varepsilon_1)^2}{2}\right)^2$$
$$+ 2\, S_y^4\, \alpha(\alpha-1)E(\varepsilon_0) - 2S_y^4 (\alpha-1)^2\, E\left(\eta\,\varepsilon_1\varepsilon_0 + \eta\,\varepsilon_1 - \frac{(\eta\,\varepsilon_1)^2}{2}\right)$$
$$- 2\, S_y^4\, \alpha(\alpha-1)\, \eta E(\varepsilon_1\varepsilon_0) \qquad (7)$$

On putting the values of expectations, we may get the desired expression for mean squared error of our proposed estimator.

$$\text{MSE}(t_c) = (\alpha-1)^2 S_y^4 + \alpha^2 S_y^4\,\delta\,(\lambda_{40}-1) + (\alpha-1)^2\, S_y^4\,\delta(\eta^2\,(\lambda_{04}-1))$$
$$-2S_y^4\,(\alpha-1)^2\,\delta\left\{\eta\,(\lambda_{22}-1) - \frac{\eta^2}{2}\,(\lambda_{04}-1)\right\} - 2\,S_y^4\,\alpha(\alpha-1)\,\eta\,\delta(\lambda_{22}-1) \quad (8)$$

The expression of minimum MSE is obtained by partially differentiating equation (8) with respect to α and equating it equals to 0, we obtained the minimum MSE.

$$\left(\frac{\partial}{\partial\alpha}\right)\text{MSE}(t_c) = \left(\frac{\partial}{\partial\alpha}\right)(\alpha-1)^2 S_y^4 + \left(\frac{\partial}{\partial\alpha}\right)\alpha^2\, S_y^4\,\delta(\lambda_{40}-1) + \left(\frac{\partial}{\partial\alpha}\right)(\alpha-1)^2\, S_y^4\,\delta(\eta^2\,(\lambda_{04}-1))$$
$$-2\left(\frac{\partial}{\partial\alpha}\right)S_y^4\,(\alpha-1)^2\,\delta\left\{\eta\,(\lambda_{22}-1) - \frac{\eta^2}{2}\,(\lambda_{04}-1)\right\} - 2\left(\frac{\partial}{\partial\alpha}\right)S_y^4\,\alpha(\alpha-1)\,\eta\,\delta(\lambda_{22}-1) = 0$$

$$\alpha_{opt} = \frac{1+\delta\,\{2\,\eta^2\,(\lambda_{04}-1) - \eta\,(\lambda_{22}-1)\}}{1+\ \delta\{(\lambda_{40}-1)\ +2\,\eta^2\,(\lambda_{04}-1) - 4\eta\,(\lambda_{22}-1)\}}$$

4. Efficiency comparisons

In sample survey, the most important consideration for any one of the estimator is their efficiency related to existing estimators. If the proposed estimator has minimum mean squared error compared to other estimators which are available in literature, then our proposed estimator will have considered the best over conventional estimators. Let us compare the efficiency of our proposed formula to the other estimators as below.

1. Unbiased estimator for population variance

   $t_0 = s_y^2$

   $\text{MSE}(t_0) = S_y^4\,\delta\,(\lambda_{40}-1)$

   $\text{MSE}(t_0) > \text{MSE}(t_c)$

2. Isaki (1983) ratio estimator for population variance

   $t_1 = s_y^2\left({S_x^2}/{s_x^2}\right)$

   $\text{MSE}(t_1) = S_y^4\,\delta\,[(\lambda_{40}-1) + (\lambda_{04}-1) - 2\,(\lambda_{22}-1)]$

   $\text{MSE}(t_1) > \text{MSE}(t_c)$

3. Linear regression estimator

   $t_2 = s_y^2 + b_0(S_x^2 - s_x^2)$

   $\text{MSE}(t_2) = S_y^4\,\delta\left[(\lambda_{40}-1) - \left({(\lambda_{22}-1)^2}/{(\lambda_{04}-1)}\right)\right]$

   $\text{MSE}(t_2) > \text{MSE}(t_c)$

4. $t_3 = s_y^2\exp\left({(S_x^2 - s_x^2)}/{(S_x^2 + s_x^2)}\right)$

   $\text{MSE}(t_3) = S_y^4\,\delta\left[(\lambda_{40}-1) + \frac{(\lambda_{04}-1)}{4} - (\lambda_{22}-1)\right]$

   $\text{MSE}(t_3) > \text{MSE}(t_c)$

5. Kumar et al. (2011) $t_4 = s_y^2 \exp\left({(s_x^2 - S_x^2)}/{(s_x^2 + S_x^2)}\right)$

   MSE(t_4) =$S_y^4\ \delta\left[(\lambda_{40}-1) + \frac{(\lambda_{04}-1)}{4} - (\lambda_{22}-1)\right]$

   MSE(t_4) > MSE(t_c)

6. Yadav & Kadilar (2013) $t_5 = s_y^2 \exp\left({(S_x^2 - s_x^2)}/{(S_x^2 + (-1)s_x^2)}\right)$

   MSE($t_{5\,min}$) =$S_y^4\ \delta\left[(\lambda_{40}-1) - \left({(\lambda_{22}-1)^2}/{(\lambda_{04}-1)}\right)\right]$

   MSE(t_5) > MSE(t_c)

7. Misra et al. (2024) $t_6 = s_y^2 + \alpha \log\frac{[s_x^2]}{[S_x^2]}$

   MSE($t_{6\,min}$) =$S_y^4\ \delta\left[(\lambda_{40}-1) - \left({(\lambda_{22}-1)^2}/{(\lambda_{04}-1)}\right)\right]$

   MSE(t_6) > MSE(t_c)

5. Empirical Studies

Perform on real datasets.

Population I

X = Fixed capital, Y = output of 80 factories, N = 80, n = 20, $\bar{X}$ = 11.265 , $\bar{Y}$ = 51.826, S_x^2 = 71.504 , S_y^2 = 336.979, S_{xy} = 146.068, $\lambda_{04} = K_x$ = 2.866, $\lambda_{40} = K_y$ = 2.267 , λ_{22} = 2.221, ρ_{xy}= 0.941, C_y= 0.354, C_x= 0.751, M_x= 10.300

Population II

X = acreage under wheat crop in 1973, Y = acreage under wheat crop in 1974, N = 70, n = 25, $\bar{X}$ = 175.2671 , $\bar{Y}$ = 96.700, S_x^2 = 19840.7508 , S_y^2 = 3686.1898, $\lambda_{04} = K_x$ = 7.0952, $\lambda_{40} = K_y$ = 4.7596 , λ_{22} = 4.6038, ρ_{xy}= 0.7293, C_y= 0.6254, C_x= 0.8037, M_x= 72.4375

Table 1: Mean Square error (MSE) of estimators

Estimator	MSE of Population I	MSE of Population II
t_1	5395	1313625.00
t_2	2942	924946.50
t_3	1993	1385199
t_4	2182	586861
t_5	1993	1385199
t_6	1993	1385199
t_c	1802	1384204

6. Conclusions

Knowing the auxiliary median and kurtosis helps the estimator reduce variance estimation error. It is useful in survey contexts with such auxiliary data because of its proven theoretical superiority and empirical advantages.

References

[1]. Kumari, C., & Thakur, R. K. (2020). An improved estimation of population variance using the coefficient of kurtosis and median of an auxiliary variable. *International Journal of Engineering Sciences & Research Technology, 9*(4), 168–175.

[2]. Kumari, C., & Thakur, R. K. (2020). Estimation of population variance using known coefficient of kurtosis and median of an auxiliary variable. *International Journal of Advanced Research, 8*(4), 803–810.

[3]. Kumari, C., & Thakur, R. K. (2020). The utilization of auxiliary information in the estimation procedure of population variance. *International Journal of Multidisciplinary Educational Research, 9*(5), 36–45.

[4]. Kumari, C. (2019). Optimal two-parameter logarithmic estimators for estimating the population variance. *Global Journal of Pure and Applied Mathematics, 15*(5), 527–537.

[5]. Kumari, C. (2018). A new log-type estimators for estimating the population variance. *International Journal of Computational and Applied Mathematics, 13*(1), 43–54.

[6]. Kadilar, C., & Cingi, H. (2006). An improvement in estimating the population mean by using the correlation coefficient. *Hacettepe Journal of Mathematics and Statistics, 35*(1), 103–109.

[7]. Murthy, M. N. (1967). *Sampling theory and methods*. Statistical Publishing Society.

[8]. Subramani, J., & Kumarapandiyan, G. (2012). Variance estimation using median of the auxiliary variable. *International Journal of Probability and Statistics, 1*, 6–40. https://doi.org/10.5923/j.ijps.20120103.02

[9]. Subramani, J., & Kumarapandiyan, G. (2012). Variance estimation using quartiles and their functions of an auxiliary variable. *International Journal of Statistics and Applications, 2*, 67–72. https://doi.org/10.5923/j.statistics.20120205.04

[10]. Subramani, J., & Kumarapandiyan, G. (2013). Estimation of variance using known coefficient of variation and median of an auxiliary variable. *Journal of Modern Applied Statistical Methods, 12*(1), Article 11. https://doi.org/10.22237/jmasm/1367381400

[11]. Subramani, J. (2015). Generalized modified ratio type estimator for estimation of population variance. *Sri Lankan Journal of Applied Statistics, 16*(1), 69–79.

[12]. Sukhatme, P. V. (1944). Moments and product moments of moment-statistics for samples of the finite and infinite populations. *Sankhyā: The Indian Journal of Statistics, 6*, 363–382.

[13]. Sukhatme, P. V., & Sukhatme, B. V. (1970). *Sampling theory of surveys with applications*. Asia Publishing House.