Estimation of Variance Using Known Coefficient of Variation and Median of an Auxiliary Variable

J. Subramani  
*Pondicherry University*

G. Kumarapandiyan  
*Pondicherry University*

Follow this and additional works at: [http://digitalcommons.wayne.edu/jmasm](http://digitalcommons.wayne.edu/jmasm)

Part of the [Applied Statistics Commons](http://digitalcommons.wayne.edu/jmasm), [Social and Behavioral Sciences Commons](http://digitalcommons.wayne.edu/jmasm), and the [Statistical Theory Commons](http://digitalcommons.wayne.edu/jmasm)

Recommended Citation

DOI: 10.22237/jmasm/1367381400  
Available at: [http://digitalcommons.wayne.edu/jmasm/vol12/iss1/11](http://digitalcommons.wayne.edu/jmasm/vol12/iss1/11)
Estimation of Variance Using Known Coefficient of Variation and Median of an Auxiliary Variable

Cover Page Footnote
The authors wish to express their gratitude and thanks for financial assistance received from UGC-MRP and the Department of Science and Technology (DST), New Delhi, India.
Estimation of Variance Using Known Coefficient of Variation and Median of an Auxiliary Variable

J. Subramani  G. Kumarapandiyan
Pondicherry University
R V Nagar, Kalapet, Puducherry

A modified ratio type variance estimator for estimating population variance of a study variable when the population median and coefficient of variation of an auxiliary variable are known is proposed. The bias and mean squared error of the proposed estimator are derived and conditions under which the proposed estimator performs better than the traditional ratio type variance estimators and modified ratio type variance estimators are obtained. Using a numerical study results show that the proposed estimator performs better than the traditional ratio type variance estimator and existing modified ratio type variance estimators.

Key words: Bias, mean squared error, natural populations, ratio type estimators, simple random sampling.

Introduction
Consider a finite population U = \{U_1, U_2, ..., U_N\} of N distinct and identifiable units. Let Y be a real variable with value Y_i measured on U_i, i = 1,2,3, ..., N giving a vector Y = \{Y_1, Y_2, ..., Y_N\}. The goal is to estimate the population mean, \(\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i\) or its variance, \(S_Y^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} (Y_i - \bar{Y})^2\) on the basis of a random sample selected from a population, U. When no additional information is available on an auxiliary variable, the simplest estimator of population mean is a simple random sample mean without replacement. However if an auxiliary variable X is closely related to a study variable Y then ratio and regression estimators can be used to improve the performance of the estimator of the study variable. This article considers the problem of estimating a population variance and uses auxiliary information to improve the efficiency of the population variance estimator

\[ S_Y^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} (Y_i - \bar{Y})^2. \]

Population variance estimation is considered by Isaki (1983) who proposed ratio and regression estimators. Prasad & Singh (1990) considered a ratio type estimator for estimating population variance by improving Isaki’s estimator (1983) with respect to bias and precision. Arcos, et al. (2005) introduced another ratio type estimator that also improved the Isaki’s estimator (1983) and is less biased and more precise than other estimators. Notations used throughout this article are:

- \(N\): population size
- \(n\): sample size
- \(\gamma\): \(1/n\)
- \(Y\): study variable
- \(X\): auxiliary variable
- \(\bar{X}, \bar{Y}\): population means

J. Subramani is an Associate Professor and Head in the Department of Statistics in the Ramanujan School of Mathematical Sciences. Email him at: drjsubramani@yahoo.co.in.
G. Kumarapandiyan is a Research Scholar in the Department of Statistics in the Ramanujan School of Mathematical Sciences. Email him at: kumarstat88@gmail.com.
\( \bar{x}, \bar{y} \): sample means
\( S_y^2, S_x^2 \): population variances
\( s_y^2, s_x^2 \): sample variances
\( C_x, C_y \): coefficients of variation
\( \rho \): correlation coefficient
\( B(.) \): bias of the estimator
\( MSE(.) \): mean squared error of the estimator
\( \hat{S}_R^2 \): traditional ratio type variance estimator of \( S_y^2 \)
\( \hat{S}_{RC}^2 \): estimator of \( S_y^2 \)
\( \hat{S}_K^2 \): existing modified ratio type variance estimator of \( S_y^2 \)
\( \hat{S}_{SG}^2 \): proposed modified ratio type variance estimator of \( S_y^2 \)

Isaki (1983) suggested a ratio type variance estimator for population variance \( S_y^2 \) when the population variance \( S_x^2 \) of an auxiliary variable \( X \) is known together with its bias and mean squared error as:

\[
\hat{S}_R^2 = \frac{y}{s_x^2} \frac{s_y^2}{s_x^2}
\]

\( B(\hat{S}_R^2) = \gamma S_y^2 \left[ (\beta_{2(s)} - I) - (\lambda_{22} - I) \right] \)

\( MSE(\hat{S}_R^2) = \gamma S_y^2 \left[ (\beta_{2(y)} - I) + (\beta_{2(s)} - I) \right] - 2(\lambda_{22} - I) \)

where
\[
\beta_{2(y)} = \frac{\mu_{20}}{\mu_{22}}, \quad \beta_{2(s)} = \frac{\mu_{40}}{\mu_{22}}, \quad \lambda_{22} = \frac{\mu_{22}}{\mu_{20} \mu_{02}}
\]

and
\[
\mu_n = \frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{Y})^T (x_i - \bar{X})^t.
\]

The ratio type variance estimator in (1) is used to improve the precision of a population variance estimate compared to simple random sampling when a positive correlation exists between \( X \) and \( Y \). Further improvements are achieved on the classical ratio type variance estimator by introducing a number of modified ratio type variance estimators using known parameters such as the coefficients of variation and kurtosis. The problem of constructing efficient estimators for a population variance has been discussed by many authors including: Agarwal & Sithapit (1995), Ahmed, et al. (2000), Al-Jararha & Al-Haj Ebrahim (2012), Arcos, et al. (2005), Bhushan (2012), Das & Tripathi (1978), Garcia & Cebrain (1997), Gupta & Shabbir (2008), Isaki (1983), Kadilar & Cingi (2006a, 2006b), Prasad & Singh (1990), Reddy (1974), Singh & Chaudhary (1986), Singh, et al. (1988, 2003), Subramani & Kumarapandiyan (2012b, 2012c), Upadhyaya & Singh (2006) and Wolter (1985).

Based on works by Sisodia & Dwivedi (1981), Upadhyaya & Singh (1999) and Singh, et al. (2004), Kadilar & Cingi (2006a) suggested four ratio type variance estimators using known values of coefficient of variation \( C_X \) and coefficient of kurtosis \( \beta_{2(s)} \) of an auxiliary variable \( X \) together with their biases and mean squared errors (see Table 1). The modified ratio type variance estimators discussed are biased but have minimal mean squared errors compared to traditional ratio type variance estimators. The list of estimators in Table 1 uses known values of the parameters such as \( S_x^2, C_x, \beta_2 \) and their linear combinations. Subramani & Kumarapandiyan (2012a) used the linear combination of a known value of the population median \( M_d \) and Coefficient of variation \( C_x \) of an auxiliary variable to improve the ratio estimators in estimating a population mean. Based on Subramani & Kumarapandiyan (2012a), a modified ratio type variance estimator using a linear combination of known value of a population median and coefficient of variation of an auxiliary variable is proposed. Results show that the proposed estimator performs better than the traditional ratio type and existing modified ratio type variance estimators.
VARIANCE ESTIMATION USING COEFFICIENT OF VARIATION AND MEDIAN

The performance of the estimator of the study variable can be improved by using known population parameters of an auxiliary variable, which are positively correlated with a study variable. A modified ratio type variance estimator using the linear combination of known value of the population median \( M_d \) and coefficient of variation \( C_x \) of an auxiliary variable is proposed; this modified ratio type variance estimator for population variance \( S^2 \) is defined as

\[
\hat{S}_{yG}^2 = s_y^2 \left[ \frac{C_x s_x^2 + M_d}{C_x s_x^2 + M_d} \right].
\]  

Table 1: Existing Modified Ratio Type Variance Estimators with their Biases and Mean Squared Errors
(Source: Kadilar & Cingi, 2006a)

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Bias ( B(.) )</th>
<th>Mean Squared Error ( \text{MSE} (.) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{S}_{KC1}^2 ) = ( s_y^2 \left[ \frac{S_x^2 + C_x}{s_x^2 + C_x} \right] )</td>
<td>( \gamma S_y^2 A_1 \left[ A_1 \left( \beta_{2(x)} - I \right) - (\lambda_{22} - I) \right] )</td>
<td>( \gamma S_y^2 \left[ \left( \beta_{2(y)} - I \right) + A_1^2 \left( \beta_{2(x)} - I \right) - 2A_1 \left( \lambda_{22} - I \right) \right] )</td>
</tr>
<tr>
<td>( \hat{S}<em>{KC2}^2 ) = ( s_y^2 \left[ \frac{S_x^2 + \beta</em>{2(x)}}{s_x^2 + \beta_{2(x)}} \right] )</td>
<td>( \gamma S_y^2 A_2 \left[ A_2 \left( \beta_{2(x)} - I \right) - (\lambda_{22} - I) \right] )</td>
<td>( \gamma S_y^2 \left[ \left( \beta_{2(y)} - I \right) + A_2^2 \left( \beta_{2(x)} - I \right) - 2A_2 \left( \lambda_{22} - I \right) \right] )</td>
</tr>
<tr>
<td>( \hat{S}<em>{KC3}^2 ) = ( s_y^2 \left[ \frac{S_x^2 \beta</em>{2(x)} + C_x}{s_x^2 \beta_{2(x)} + C_x} \right] )</td>
<td>( \gamma S_y^2 A_3 \left[ A_3 \left( \beta_{2(x)} - I \right) - (\lambda_{22} - I) \right] )</td>
<td>( \gamma S_y^2 \left[ \left( \beta_{2(y)} - I \right) + A_3^2 \left( \beta_{2(x)} - I \right) - 2A_3 \left( \lambda_{22} - I \right) \right] )</td>
</tr>
<tr>
<td>( \hat{S}<em>{KC4}^2 ) = ( s_y^2 \left[ \frac{S_x^2 C_x + \beta</em>{2(x)}}{s_x^2 C_x + \beta_{2(x)}} \right] )</td>
<td>( \gamma S_y^2 A_4 \left[ A_4 \left( \beta_{2(x)} - I \right) - (\lambda_{22} - I) \right] )</td>
<td>( \gamma S_y^2 \left[ \left( \beta_{2(y)} - I \right) + A_4^2 \left( \beta_{2(x)} - I \right) - 2A_4 \left( \lambda_{22} - I \right) \right] )</td>
</tr>
</tbody>
</table>

Notes: \( A_1 = \frac{S_x^2}{S_x^2 + C_x}, A_2 = \frac{S_x^2}{S_x^2 + \beta_{2(x)}}, A_3 = \frac{S_x^2 \beta_{2(x)}}{s_x^2 \beta_{2(x)} + C_x}, A_4 = \frac{S_x^2 C_x}{s_x^2 C_x + \beta_{2(x)}} \)

Proposed Estimator

The performance of the estimator of the study variable can be improved by using known population parameters of an auxiliary variable, which are positively correlated with a study variable. A modified ratio type variance estimator using the linear combination of known value of the population median \( M_d \) and coefficient of variation \( C_x \) of an auxiliary variable is proposed; this modified ratio type variance estimator for population variance \( S^2 \) is defined as

\[
\hat{S}_{yG}^2 = s_y^2 \left[ \frac{C_x s_x^2 + M_d}{C_x s_x^2 + M_d} \right].
\]

The bias and mean squared error of \( \hat{S}_{yG}^2 \) to the first degree of approximation are:

\[
B(\hat{S}_{yG}^2) = \gamma S_y^2 A_{yG} \left[ A_{yG} \left( \beta_{2(x)} - I \right) - (\lambda_{22} - I) \right]
\]

(3)

\[
\text{MSE}(\hat{S}_{yG}^2) = \gamma S_y^2 \left[ \left( \beta_{2(y)} - I \right) + A_{yG}^2 \left( \beta_{2(x)} - I \right) - 2A_{yG} \left( \lambda_{22} - I \right) \right]
\]

(4)

where \( A_{yG} = \frac{S_x^2}{S_x^2 + M_d} \).
Efficiency Comparison of the Proposed Estimator

The bias and mean squared error of the traditional ratio type variance estimator are:

\[
B\left(\hat{S}_R^2\right) = \gamma S_y^2 \left[ (\beta_{2(x)} - I) - (\lambda_{22} - I) \right]
\]

(5)

\[
MSE\left(\hat{S}_R^2\right) = \gamma S_y^2 \left[ (\beta_{2(y)} - I) + (\beta_{2(x)} - I) \right] - 2(\lambda_{22} - I)
\]

(6)

For convenience and for the ease of comparisons, the biases mean squared errors and constants of the modified ratio type variance estimators shown in Table 1 are represented in a single class as:

\[
B\left(\hat{S}_{KCI}^2\right) = \gamma S_y^2 A_i \left[ A_i (\beta_{2(x)} - I) - (\lambda_{22} - I) \right],
\]

\[ \text{for } i = 1, 2, 3, 4 \]

(7)

\[
MSE\left(\hat{S}_{KCI}^2\right) = \gamma S_y^2 \left[ (\beta_{2(y)} - I) + A_i^2 (\beta_{2(x)} - I) \right] - 2A_i(\lambda_{22} - I)
\]

\[ \text{for } i = 1, 2, 3, 4 \]

(8)

where

\[
A_i = \frac{S_i^2}{S_x^2 + C_x},
\]

A_2 = \frac{S_i^2}{S_x^2 + \beta_{2(x)}},

A_3 = \frac{S_i^2 \beta_{2(x)}}{S_x^2 \beta_{2(x)} + C_x},

A_4 = \frac{S_i^2 C_x}{S_x^2 C_x + \beta_{2(x)}}

The bias and mean squared error of the proposed modified ratio type variance estimator are:

\[
B\left(\hat{S}_{JG}^2\right) = \gamma S_y^2 A_p \left[ A_{JG} (\beta_{2(x)} - I) - (\lambda_{22} - I) \right]
\]

(9)

\[
MSE\left(\hat{S}_{JG}^2\right) = \gamma S_y^2 \left[ (\beta_{2(y)} - I) + A_{JG} (\beta_{2(x)} - I) \right] - 2A_{JG}(\lambda_{22} - I)
\]

(10)

where \( A_{JG} = \frac{S_x^2 C_x}{S_x^2 C_x + M_d} \)

The condition for which the proposed estimator \( \hat{S}_{JG}^2 \) is more efficient than the traditional ratio type variance estimator derived from (6) and (10) is:

\[
MSE\left(\hat{S}_{JG}^2\right) < MSE\left(\hat{S}_R^2\right)
\]

\[
\text{if } \lambda > I + \frac{(A_{JG} + I)(\beta_{2(x)} - I)}{2}
\]

(11)

The conditions for which the proposed estimator \( \hat{S}_{JG}^2 \) is more efficient than the existing modified ratio type variance estimators in Table 1, \( \hat{S}_{KCI}^2; i = 1, 2, 3 \) and \( 4 \) derived from (8) and (10) are:

\[
MSE\left(\hat{S}_{JG}^2\right) < MSE\left(\hat{S}_{KCI}^2\right)
\]

\[
\text{if } \lambda > I + \frac{(A_{JG} + A_i)(\beta_{2(x)} - I)}{2}
\]

\[ \text{for } i = 1, 2, 3 \text{ and } 4 \]

(12)

Numerical Study

The performance of the proposed modified ratio type variance estimator was assessed and compared with a traditional ratio type and existing modified ratio type variance estimators (see Table 1) for natural populations. Populations 1 and 2 are a real data set taken from the Italian Bureau for Environment Protection (APAT) 2004 Report on Waste 2004. (Data and reports are available at: http://www.osservatorionazionalerifiuti.it.) For
VARIANCE ESTIMATION USING COEFFICIENT OF VARIATION AND MEDIAN

Table 2: Parameters and Constants of the Populations

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Population1</th>
<th>Population2</th>
<th>Population3</th>
<th>Population4</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>103</td>
<td>103</td>
<td>80</td>
<td>49</td>
</tr>
<tr>
<td>n</td>
<td>40</td>
<td>40</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>( \bar{Y} )</td>
<td>626.2123</td>
<td>62.6212</td>
<td>51.8264</td>
<td>116.1633</td>
</tr>
<tr>
<td>( \bar{X} )</td>
<td>557.1909</td>
<td>556.5541</td>
<td>11.2646</td>
<td>98.6765</td>
</tr>
<tr>
<td>( \rho )</td>
<td>0.9936</td>
<td>0.7298</td>
<td>0.9413</td>
<td>0.6904</td>
</tr>
<tr>
<td>( S_Y )</td>
<td>913.5498</td>
<td>91.3549</td>
<td>18.3569</td>
<td>98.8286</td>
</tr>
<tr>
<td>( C_Y )</td>
<td>1.4588</td>
<td>1.4588</td>
<td>0.3542</td>
<td>0.8508</td>
</tr>
<tr>
<td>( S_X )</td>
<td>818.1117</td>
<td>610.1643</td>
<td>8.4563</td>
<td>102.9709</td>
</tr>
<tr>
<td>( C_X )</td>
<td>1.4683</td>
<td>1.0963</td>
<td>0.7507</td>
<td>1.0435</td>
</tr>
<tr>
<td>( \beta_2(x) )</td>
<td>37.3216</td>
<td>17.8738</td>
<td>2.8664</td>
<td>5.9878</td>
</tr>
<tr>
<td>( \beta_2(y) )</td>
<td>37.1279</td>
<td>37.1279</td>
<td>2.2667</td>
<td>4.9245</td>
</tr>
<tr>
<td>( \lambda_{22} )</td>
<td>37.2055</td>
<td>17.2220</td>
<td>2.2209</td>
<td>4.6977</td>
</tr>
<tr>
<td>( M_d )</td>
<td>308.0500</td>
<td>373.820</td>
<td>7.5750</td>
<td>64.0000</td>
</tr>
<tr>
<td>( A_1 )</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9896</td>
<td>0.9999</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9615</td>
<td>0.9994</td>
</tr>
<tr>
<td>( A_3 )</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9964</td>
<td>1.0000</td>
</tr>
<tr>
<td>( A_4 )</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9493</td>
<td>0.9995</td>
</tr>
<tr>
<td>( A_{IG} )</td>
<td>0.9996</td>
<td>0.9990</td>
<td>0.8763</td>
<td>0.9942</td>
</tr>
</tbody>
</table>

each Italian province, three variables are considered: the total amount (tons) of recyclable-waste collection in Italy in 2003 (\( Y \)), the total amount of recyclable-waste collection in Italy in 2002 (\( X_1 \)) and the number of inhabitants in 2003 (\( X_2 \)). Population 3 is from Murthy (1967, p. 228) and population 4 is from Cochran (1977, p. 152). The population parameters and constants computed from these populations are shown in Table 2. The biases and mean squared errors for the existing and proposed modified ratio type variance estimator for the populations are shown in Tables 3 and 4.

Table 3 shows that the proposed modified ratio type variance estimator has less bias compared to the biases of the traditional and existing modified ratio type variance estimators. It is observed that the mean squared error of the proposed modified ratio type variance estimator is lower than the mean squared errors of the traditional and existing modified ratio type variance estimators.

Conclusion
This article proposed a modified ratio type variance estimator using a known value of median of an auxiliary variable. The bias and mean squared error of the proposed estimator were obtained and compared with the traditional ratio type and existing modified ratio type variance estimators. Further the conditions for which the proposed estimator is more efficient than the traditional and existing estimators were derived. The performance of the proposed estimator was assessed using four known populations. Results show that the bias and mean squared error of the proposed estimator are less than the biases and mean squared errors of the traditional and existing estimators for the known populations considered. Based on results, the proposed modified ratio type variance estimator may be preferred over traditional ratio type and existing modified ratio type variance estimators for the use in practical applications.
Acknowledgements
The authors wish to express their gratitude and thanks for financial assistance received from UGC-MRP and the Department of Science and Technology (DST), New Delhi, India.

References


