Variance Estimation Using Median of the Auxiliary Variable

J. Subramani*, G. Kumarapandiyan

Department of Statistics, Ramanujan School of Mathematical Sciences, Pondicherry University, R V Nagar, Kalapet, 605014, Puducherry
kumarstat88@gmail.com

Abstract The present paper deals with a modified ratio type variance estimator for estimation of population variance of the study variable, when the population median of the auxiliary variable is known. The bias and the mean squared error of the proposed estimator are obtained and also derived the conditions for which the proposed estimator performs better than the traditional ratio type variance estimator suggested by Isaki[10] and the modified ratio type variance estimators suggested by Kadilar and Cingi[11]. Further we have compared the efficiencies of the proposed estimator with that of traditional ratio type variance estimator and existing modified ratio type variance estimators for certain known populations. From the numerical study it is observed that the proposed estimator performs better than the traditional ratio type variance estimator and existing modified ratio type variance estimators.

Keywords Bias, Mean Squared Error, Natural Populations, Simple Random Sampling

1. 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 problem is to estimate the population mean \( \bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i \) on the basis of a random sample selected from the population \( U \) and / or its variance \( S_Y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2 \). When there is no additional information on the auxiliary variable available, the simplest estimator of population mean is the simple random sample mean without replacement. However if an auxiliary variable \( X \) closely related to the study variable \( Y \) is available then one can use Ratio or Regression estimators to improve the performance of the estimator of the study variable. In this paper, we consider the problem of estimation of the population variance and use the auxiliary information to improve the efficiency of the estimator of population variance \( S_Y^2 \). Estimation of population variance is considered by Isaki[10] where ratio and regression estimators are proposed. Prasad and Singh[14] have considered a ratio type estimator for estimation of population variance by improving Isaki’s estimator[10] with respect to bias and precision. Arcos et al. [14] have introduced another ratio type estimator, which has also improved the Isaki’s estimator[10], which is almost unbiased and more precise than the other estimators.

Before discussing further about the traditional ratio type variance estimator, modified ratio type variance estimators and the proposed modified ratio type variance estimator, the notations to be used in this paper are described below:

- \( N \) – Population size
- \( n \) – Sample size
- \( \gamma = 1/n \)
- \( Y \) – Study variable
- \( X \) – Auxiliary variable
- \( \bar{X}, \bar{Y} \) – Population means
- \( \bar{x}, \bar{y} \) – Sample means
- \( S_X^2, S_Y^2 \) – Population variances
- \( s_X^2, s_Y^2 \) – Sample variances
- \( C_X, C_Y \) – Coefficient of variations
- \( \rho \) – Coefficient of correlation
- \( B(.) \) – Bias of the estimator
- \( \text{MSE}(.) \) – Mean squared error of the estimator
- \( S_R^2 \) – Traditional Ratio type variance estimator of \( S_Y^2 \)
- \( S_{RC}^2 \) – Existing modified ratio type variance estimator of \( S_Y^2 \)
- \( S_{SK}^2 \) – Proposed modified ratio type variance estimator of \( S_Y^2 \)

Isaki[10] suggested a ratio type variance estimator for the population variance \( S_Y^2 \) when the population variance \( S_X^2 \) of the auxiliary variable \( X \) is known together with its bias and mean squared error as given below:

\[
S_R^2 = \frac{s_Y^2}{s_X^2}
\]

(1)
The modified ratio type variance estimators discussed above are biased but have minimum mean squared errors compared to the traditional ratio type variance estimator. The list of estimators given in Table 1 uses the known values of the parameters like $S^2_x$, $C_x$, $\beta_2$ and their linear combinations. Subramani and Kumarapandian[21] used the known value of the population median $M_d$ of the auxiliary variable to improve the ratio estimators in estimation of population mean. Further we know that the value of median is unaffected and robustness by the extreme values or the presence of outliers in the population data. The above discussed points have motivated us to introduce a modified ratio type variance estimator using the known value of the population median of the auxiliary variable. As a result, it is observed that the proposed estimator performs better than the traditional ratio type variance estimator as well as the existing modified ratio type variance estimators listed in Table 1. The materials of the present study are arranged as given below. The proposed estimator with known population median is presented in section 2 where as the conditions in which the proposed estimator performs better than the existing estimators are derived in section 3. The performances of the proposed and the existing estimators are assessed for certain natural populations in section 4 and the conclusion is presented in section 5.

2. Proposed Estimator

As we stated earlier one can always improve the performance of the estimator of the study variable by using the known population parameters of the auxiliary variable, which are positively correlated with that of study variable. In this section we have suggested a modified ratio type variance estimator given in (1) is used to improve the precision of the estimate of the population variance compared to simple random sampling when there exists a positive correlation between $X$ and $Y$. Further improvements are also achieved on the classical ratio estimator by introducing a number of modified ratio estimators with the use of known parameters like, Co-efficient of Variation and Co-efficient of Kurtosis. The problem of constructing efficient estimators for the population variance has been widely discussed by various authors such as Das and Tripathi[7], Isaki[10], Singh et al.[17,19], Agarwal and Sihapat[1], Garcia and Cebrian[8], Arcos et al.[4], Ahmed et al.[2], Al-Jarabah and Al-Haj Ebrahim[3], Bhushan[5], Prasad and Singh[14], Reddy[15], Singh and Chaudhary[16], Upadhyaya and Singh[23], Wolter[24], Kadilar and Cingi[11,12] and Gupta and Shabbir[9].

Motivated by Sisodia and Dwivedi[20], Singh et al.[18] and Upadhyaya and Singh[22], Kadilar and Cingi[11] suggested four ratio type variance estimators using known values of Co-efficient of variation $C_x$ and Co-efficient of Kurtosis $\beta_2(x)$ of an auxiliary variable $X$ together with their biases and mean squared errors as given in the Table 1:

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Bias - $B(.)$</th>
<th>Mean squared error $MSE(.)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S^2_{KC1} = \frac{s^2_x(C_x + s^2_x)}{s^2_x + C_x}$</td>
<td>$\gamma S^2_y A_1 \left[ (\beta_2(y) - 1) + \lambda \right]$</td>
<td>$\gamma S^2_y \left[ (\beta_2(y) - 1) + A^1 \left( \beta_2(x) - 1 \right) - 2A_1(\lambda - 1) \right]$</td>
</tr>
<tr>
<td>$S^2_{KC2} = \frac{s^2_x + \beta_2(x)}{s^2_x + C_x}$</td>
<td>$\gamma S^2_y A_2 \left[ (\beta_2(x) - 1) + \lambda \right]$</td>
<td>$\gamma S^2_y \left[ (\beta_2(y) - 1) + A^2 \left( \beta_2(x) - 1 \right) - 2A_2(\lambda - 1) \right]$</td>
</tr>
<tr>
<td>$S^2_{KC3} = \frac{s^2_x(\beta_2(x) + C_x)}{s^2_x + C_x}$</td>
<td>$\gamma S^2_y A_3 \left[ (\beta_2(x) - 1) + \lambda \right]$</td>
<td>$\gamma S^2_y \left[ (\beta_2(y) - 1) + A^3 \left( \beta_2(x) - 1 \right) - 2A_3(\lambda - 1) \right]$</td>
</tr>
<tr>
<td>$S^2_{KC4} = \frac{s^2_x C_x + \beta_2(x)}{s^2_x C_x + C_x}$</td>
<td>$\gamma S^2_y A_4 \left[ (\beta_2(x) - 1) + \lambda \right]$</td>
<td>$\gamma S^2_y \left[ (\beta_2(y) - 1) + A^4 \left( \beta_2(x) - 1 \right) - 2A_4(\lambda - 1) \right]$</td>
</tr>
</tbody>
</table>

where $\beta_2(y) = \frac{\mu_4}{\mu_2^2}$, $\beta_2(x) = \frac{\mu_4}{\mu_2^2}$, $\lambda = \frac{\mu_2}{\mu_2 \mu_4}$ and

$$B(\frac{s^2_x}{s^2_x}) = \gamma S^2_y \left[ (\beta_2(y) - 1) - (\lambda - 1) \right]$$

$$A_4 = \frac{S^2_x C_x}{s^2_x C_x + \beta_2(x)}$$

The modified ratio type variance estimators discussed above are biased but have minimum mean squared errors compared to the traditional ratio type variance estimator. The list of estimators given in Table 1 uses the known values of the parameters like $S^2_x$, $C_x$, $\beta_2$ and their linear combinations. Subramani and Kumarapandian[21] used the known value of the population median $M_d$ of the auxiliary variable to improve the ratio estimators in estimation of population mean. Further we know that the value of median is unaffected and robustness by the extreme values or the presence of outliers in the population data. The above discussed points have motivated us to introduce a modified ratio type variance estimator using the known value of the population median of the auxiliary variable. As a result, it is observed that the proposed estimator performs better than the traditional ratio type variance estimator as well as the existing modified ratio type variance estimators listed in Table 1. The materials of the present study are arranged as given below. The proposed estimator with known population median is presented in section 2 where as the conditions in which the proposed estimator performs better than the existing estimators are derived in section 3. The performances of the proposed and the existing estimators are assessed for certain natural populations in section 4 and the conclusion is presented in section 5.

2. Proposed Estimator

As we stated earlier one can always improve the performance of the estimator of the study variable by using the known population parameters of the auxiliary variable, which are positively correlated with that of study variable. In this section we have suggested a modified ratio type
variance estimator using the population median of the auxiliary variable.

The proposed modified ratio type variance estimator for population variance \( S_y^2 \) is defined as

\[
S_{SK}^2 = \frac{s_y^2 (S_x^2 + M_d)}{s_x^2 + M_d}
\]

(2)

where \( M_d \) is the population median of the auxiliary variable \( X \).

The bias and mean squared error of \( S_{SK}^2 \) to the first degree of approximation are derived and given below:

\[
B(S_{SK}^2) = \gamma S_y^2 A_{SK} [A_{SK} (\beta_{2x(2)} - 1) - (\lambda_{22} - 1)]
\]

(3)

\[
\text{MSE}(S_{SK}^2) = \gamma S_y^2 \left[ \frac{1}{A_{SK}} \left( \beta_{2y(2)} - 1 \right) + A_{SK}^2 (\beta_{2x(2)} - 1) - 2 A_{SK} (\lambda_{22} - 1) \right]
\]

where \( A_{SK} = \frac{s_x^2}{s_x^2 + M_d} \)

\[
(4)
\]

### 3. Efficiency Comparison of Proposed Estimator

As we mentioned earlier the bias and mean squared error of the traditional ratio type variance estimator are given below:

\[
B(S^2) = \gamma S_y^2 A [A (\beta_{2y(2)} - 1) - (\lambda_{22} - 1)]
\]

(5)

\[
\text{MSE}(S^2) = \gamma S_y^2 \left[ \frac{1}{A} \left( \beta_{2y(2)} - 1 \right) + A^2 (\beta_{2x(2)} - 1) - 2 A (\lambda_{22} - 1) \right]
\]

(6)

For want of space; for the sake of convenience to the readers and for the ease of comparisons, the bias, the mean squared errors and the constants of the modified ratio type variance estimators given in Table 1 are represented in single class as given below:

\[
B(S_{KCI}^2) = \gamma S_y^2 A_i [A_i (\beta_{2y(2)} - 1) - (\lambda_{22} - 1)];
\]

\( i = 1, 2, 3 \) and 4

\[
\text{MSE}(S_{KCI}^2) = \gamma S_y^2 \left[ \frac{1}{A_i} \left( \beta_{2y(2)} - 1 \right) + A_i^2 \left( \beta_{2x(2)} - 1 \right) - 2 A_i (\lambda_{22} - 1) \right]
\]

\( i = 1, 2, 3 \) and 4

where \( A_1 = \frac{s_x^2}{s_x^2 + C_x}, A_2 = \frac{s_x^2}{s_x^2 + \beta_{2x(2)}}, A_3 = \frac{s_x^2}{s_x^2 + \beta_{2y(2)}} \)

and

\[
A_4 = \frac{s_x^2}{s_x^2 C_x + \beta_{2x(2)}}
\]

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

\[
B(S_{SK}^2) = \gamma S_y^2 A_{SK} [A_{SK} (\beta_{2y(2)} - 1) - (\lambda_{22} - 1)]
\]

(9)

\[
\text{MSE}(S_{SK}^2) = \gamma S_y^2 \left[ \frac{1}{A_{SK}} \left( \beta_{2y(2)} - 1 \right) + A_{SK}^2 (\beta_{2x(2)} - 1) - 2 A_{SK} (\lambda_{22} - 1) \right]
\]

(10)

where

\[
A_{SK} = \frac{s_x^2}{s_x^2 + M_d}
\]

From the expressions given in (6) and (10) we have derived the condition for which the proposed estimator \( S_{SK}^2 \) is more efficient than the traditional ratio type variance estimator and it is given below:

\[
\text{MSE}(S_{SK}^2) < \text{MSE}(S^2) \text{ if } \lambda > 1 + \frac{(\lambda_{SK} + 1)}{2} (\beta_{2y(2)} - 1)
\]

(11)

From the expressions given in (8) and (10) we have derived the conditions for which the proposed estimator \( S_{SK}^2 \) is more efficient than the existing modified ratio type variance estimators given in Table 1, \( S_{KCI}^2 \); \( i = 1, 2, 3 \) and 4 and are given below:

\[
\text{MSE}(S_{SK}^2) < \text{MSE}(S_{KCI}^2) \text{ if } \lambda > 1 + \frac{(\lambda_{SK} + A_i)}{2} (\beta_{2y(2)} - 1); \text{ i = 1, 2, 3 and 4}
\]

(12)

### 4. Numerical Study

The performance of the proposed modified ratio type variance estimator is assessed with that of traditional ratio type estimator and existing modified ratio type variance estimators listed in Table 1 for certain natural populations. The populations 1 and 2 are the real data set taken from the Report on Waste 2004 drew up by the Italian bureau for the environment protection-APAT. Data and reports are available in the following website address http://www.osservatorionazionalerifiuti.it[25]. In the data set, for each of the Italian provinces, 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 (\( X_2 \)) and the number of inhabitants in 2003 (\( X_3 \)). The population 3 is taken from Murthy[13] given in page 228 and population 4 is taken from Cochran[6] given in page 152. The population parameters and the constants computed from the above populations are given below:

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Population 1</th>
<th>Population 2</th>
<th>Population 3</th>
<th>Population 4</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>Y</td>
<td>626.2123</td>
<td>62.6122</td>
<td>51.8264</td>
<td>116.1635</td>
</tr>
<tr>
<td>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>0.7928</td>
<td>0.9413</td>
<td>0.6904</td>
</tr>
<tr>
<td>C_x</td>
<td>1.4588</td>
<td>1.4588</td>
<td>0.3542</td>
<td>0.8508</td>
</tr>
<tr>
<td>C_y</td>
<td>913.5498</td>
<td>91.3549</td>
<td>18.3569</td>
<td>98.8286</td>
</tr>
<tr>
<td>C_{xy}</td>
<td>1.4588</td>
<td>1.4588</td>
<td>0.3542</td>
<td>0.8508</td>
</tr>
<tr>
<td>( S_y )</td>
<td>818.1117</td>
<td>610.1643</td>
<td>84.563</td>
<td>102.9709</td>
</tr>
<tr>
<td>( C_{xy} )</td>
<td>1.4683</td>
<td>1.0963</td>
<td>0.7507</td>
<td>1.0435</td>
</tr>
<tr>
<td>( \beta_{2x(2)} )</td>
<td>37.3216</td>
<td>17.8758</td>
<td>2.6866</td>
<td>9.5878</td>
</tr>
<tr>
<td>( \beta_{2y(2)} )</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>373.8200</td>
<td>7.5750</td>
<td>64.0000</td>
</tr>
<tr>
<td>( A_1 )</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
</tr>
<tr>
<td>( A_3 )</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
</tr>
<tr>
<td>( A_{SK} )</td>
<td>0.99995</td>
<td>0.99898</td>
<td>0.9942</td>
<td>0.9940</td>
</tr>
</tbody>
</table>

The biases and mean squared errors of the existing and proposed modified ratio type variance estimator for the populations given above are given in the following Tables:
Table 3. Biases of the existing and proposed modified ratio type variance estimators

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Bias ( B(\cdot) )</th>
</tr>
</thead>
</table>

Table 4. Mean squared error of the existing and proposed modified ratio type variance estimators

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Mean Squared Error ( MSE(\cdot) )</th>
</tr>
</thead>
</table>

From the values of Table 3, it is observed that the bias of the proposed modified ratio type variance estimator is less than the biases of the traditional and existing modified ratio type variance estimators. Similarly from the values of Table 4, it is observed that the mean squared error of the proposed modified ratio type variance estimator is less than the mean squared errors of the traditional and existing modified ratio type variance estimators.

5. Conclusions

In this paper we have proposed a modified ratio type variance estimator using known value of Median of the auxiliary variable. The bias and mean squared error of the proposed modified ratio type variance estimator are obtained and compared with that of traditional ratio type variance estimator and existing modified ratio type variance estimators. Further we have derived the conditions for which the proposed estimator is more efficient than the traditional and existing estimators. We have also assessed the performances of the proposed estimator for some known populations. It is observed 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 certain known populations. Hence we strongly recommend that the proposed modified ratio type variance estimator may be preferred over the traditional ratio type variance estimator and existing modified ratio type variance estimators for the use of practical applications.

ACKNOWLEDGEMENTS

The authors record their gratitude to the editor and the reviewer for their constructive comments, which have improved the presentation of the paper. The second author wishes to record his gratitude and thanks to the Vice Chancellor, Pondicherry University for the financial assistance to carry out this research work through the University Fellowship.

REFERENCES

Wiley Eastern Limited


