An Improvement in Variance Estimation Using Information on Median and Coefficient of Kurtosis of an Auxiliary Variable

G. N. Singh, A. K. Pandey*, C. Singh and M. Usman

Department of Applied Mathematics, Indian Institute of Technology (Indian School of Mines), Dhanbad-826004, India

Received: 7 Feb. 2018, Revised: 8 Mar. 2018, Accepted: 29 Mar. 2018
Published online: 1 Jul. 2018

Abstract: The objective of the present work is to develop an efficient estimation procedure to enhance the precision of estimate of a finite population variance under simple random sampling without replacement (SRSWOR) scheme. Utilizing information on median and coefficient of kurtosis of an auxiliary variable, an improved ratio-type estimator of population variance has been suggested. The properties of the proposed estimation procedure have been examined and empirical studies are performed to show the dominance over some contemporary estimators. Suitable recommendations are made to the survey practitioners.

Keywords: Study variable, auxiliary variable, variance estimation, median, coefficient of kurtosis, bias, mean square error.

1 Introduction

The estimation of finite population variance attracts the attention of survey practitioners for many practical applications. For example, a physician requires sufficient knowledge about different pathological parameters of human body such as variations in degree of blood pressure, pulse rate, blood sugar level to provide adequate prescriptions to the patients. Similarly farmers need adequate information regarding the patterns of variations in various weather parameters for cultivation of different crops. The use of auxiliary information has been a popular technique for enhancing the precision of estimates for long time and its application in estimation of finite population variance goes back to [3] and followed by [15], [4], [13], [1], and [2] among others.

Several authors have introduced the modified versions of ratio, product and linear regression methods of estimation for finite population variance by utilizing the information on auxiliary variable. Singh et al. [12] utilized known coefficient of kurtosis of auxiliary variable to estimate the population variance of the study variable. In follow up Upadhyaya and Singh [19,20], Kadilar and Cingi [5,6], Khan and Shabbir [7] and Subramani and Kumarapandeyan [16,17,18] proposed the modified estimation procedures of population variance which made the use of various known parameters of auxiliary variable. Motivated with the above works, the aim of the present work is to propose an improved estimation procedure for estimation of finite population variance by using information on known population median and coefficient of kurtosis of an auxiliary variable under SRSWOR scheme. Properties of the suggested estimation procedure are deeply examined and supplemented with empirical studies.

2 Description of notations and some existing estimators of population variance

Let $y_i$ and $x_i$ be the values of study variable $y$ and auxiliary variable $x$ respectively for the $i^{th}$ unit of a finite population of size $N$. To estimate the population variance $\sigma_y^2$ of study variable $y$, a random sample of size $n$ under without replacement scheme is drawn from the population and surveyed for the study variable $y$ under the assumption that the information on auxiliary variable $x$ readily available for all the units of the population. The following notations have been adopted for the further use:

* Corresponding author e-mail: avipbhu@gmail.com
\(X(\bar{Y})\) : Population means of auxiliary (study) variable \(x(y)\).
\(\bar{x}(\bar{y})\) : Sample means of auxiliary (study) variable \(x(y)\).
\(s_x^2(s_y^2)\) : Sample variance of the variable \(x(y)\).
\(C_y, C_x\) : Population coefficients of variation for the variables shown in subscripts.
\(\rho_{yx}\) : Population correlation coefficient between the variables \(y\) and \(x\).
\(B(\cdot)\) : Bias of the estimator.
\(M(\cdot)\) : Mean square error (MSE) of the estimator.
\(Q_i\) : Quartile defined as below in Table 1. The usual unbiased estimator of population variance and the expression of their variance is presented as basis for proposition of an estimator. The estimators and their expression of the bias and mean square errors are shown.

For sample observations

\[
\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2, \quad s_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2, \quad s_{yx} = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})
\]

For population observations

\[
\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i, \quad \bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i, \quad s_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{X})^2, \quad C_x = (s_x / \bar{X})
\]
\[
S_y^2 = (N-1)^{-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2, \quad S_{xy} = (N-1)^{-1} \sum_{i=1}^{N} (y_i - \bar{Y})(x_i - \bar{X}), \quad C_y = (S_y / \bar{Y})
\]

### 2.1 Some existing estimators of population variance

In this section, we have revisited to some existing estimators of the population variance \(S_y^2\), which will provide the strong basis for proposition of an estimator. The estimators and their expression of the bias and mean square errors are shown below in Table 1. The usual unbiased estimator of population variance and the expression of their variance is presented as

\[
d_0 = s_y^2
\]

and

\[
V(d_0) = \zeta S_y^4 (\kappa_y - 1)
\]
Following the previously discussed estimation procedures, we propose a general class of estimators of population variance $s_y^2$ which utilizes the readily available information on median and coefficient of kurtosis of an auxiliary variable $x$ and defined as

$$d_{PS} = s_y^2 \left\{ \lambda + (1 - \lambda) \frac{\left( s_x^2 \kappa_x + \frac{M_x^2}{s_x^2 \kappa_x + M_x^2} \right)}{(s_x^2 \kappa_x + M_x^2)} \right\}$$  \hspace{1cm} (4)
where $\lambda$ is a scalar quantity to be determined under certain criterions so as to minimise the MSE of the estimator $d_{PS}$. To obtain the bias and mean square error of the proposed estimator $d_{PS}$ up to the first order of approximations, we use the following transformations:

$$s_{y}^{2} = S_{y}^{2}(1 + e_{1}), \quad s_{x}^{2} = S_{x}^{2}(1 + e_{2})\text{ and } E(e_{i}) = 0 \text{ for } (i = 1, 2).$$

We use the following expected values for further derivations:

$$E(e_{1}^{2}) = \zeta(\kappa_{(y)} - 1), \quad E(e_{2}^{2}) = \zeta(\kappa_{(x)} - 1)\text{ and } E(e_{1}e_{2}) = \zeta(\delta_{22} - 1)$$

To derive the expression of bias and mean square error of the proposed class of estimators $d_{PS}$, the structure of estimator is expressed in terms of $e_{i}$’s as

$$d_{PS} = S_{y}^{2}(1 + e_{1}) \left\{ \lambda + (1 - \lambda) \dfrac{\kappa_{(y)}S_{x}^{2} + M_{d}^{2}}{\kappa_{(y)}S_{x}^{2}(1 + e_{2}) + M_{d}^{2}} \right\}$$

$$= S_{y}^{2}(1 + e_{1}) \left[ \lambda + (1 - \lambda)(1 + L^{*}e_{2})^{-1} \right]$$

where $L^{*} = \dfrac{S_{y}^{2}\kappa_{(x)}}{S_{y}^{2}\kappa_{(x)} + M_{d}^{2}}$ and $|L^{*}e_{1}| < 1$ so that the term $(1 + L^{*}e_{2})^{-1}$ is convergent.

Now expanding the expression in equation (5) binomially and simplifying we have

$$d_{PS} = S_{y}^{2} \left( 1 + e_{1} - L^{*}(1 - \lambda)e_{2} - L^{*}(1 - \lambda)e_{1}e_{2} + L^{*2}(1 - \lambda)e_{2}^{2} + L^{*}(1 - \lambda)e_{1}e_{2} - \ldots \right)$$

(6)

Taking expectation and retaining the terms up to the order $n^{-1}$, we have the expression of bias and mean square error of the estimator $d_{PS}$ as

$$B(d_{PS}) = E(d_{PS} - S_{y}^{2})$$

$$= (1 - \lambda) \zeta S_{y}^{2}(\kappa_{(y)} - 1)L^{*}[L^{*} - \psi]$$

(7)

$$M(d_{PS}) = E[d_{PS} - S_{y}^{2}]^{2}$$

$$= S_{y}^{4} \zeta \left[ (\kappa_{(y)} - 1) + (1 - \lambda)^{2}L^{*2}(\kappa_{(x)} - 1) - 2(1 - \lambda)L^{*}(\delta_{22} - 1) \right]$$

(8)

The expression of mean square error of the estimator $d_{PS}$ in equation (8) consist the unknown scalar $\lambda$, hence to obtain the optimum value of $\lambda$, the expression of MSE ($d_{PS}$) is minimized with respect to $\lambda$ and subsequently the optimum value of $\lambda$ say $\lambda_{opt}$ is obtained as

$$\lambda_{opt} = 1 - \dfrac{(\delta_{22} - 1)}{L^{*}(\kappa_{(x)} - 1)}$$

(9)

Further substituting the value of $\lambda_{opt}$ in equation (8), we get the optimum MSE ($d_{PS}$) as

$$MSE_{min}(d_{PS}) = S_{y}^{4} \zeta \left[ (\kappa_{(y)} - 1) - \dfrac{(\delta_{22} - 1)^{2}}{(\kappa_{(x)} - 1)} \right]$$

(10)

3.1 Practicability of the proposed estimator $d_{PS}$

In order to make the proposed class of estimators $d_{PS}$ practicable, the unknown scalar $\lambda$ will be replaced by $\lambda_{opt}$ and $\lambda_{opt}$ consist of several unknown parameters such as $\delta_{22}$ and $S_{y}^{2}$. Hence for practical application, these unknown parameters may be replaced by their guess values available from the past surveys or pilot surveys. If such guess values are not readily available then they may be estimated by their corresponding sample estimates, see [8],[9],[10],[14], [11].
4 Theoretical Comparisons

The estimator \( d_{ps} \) is more efficient than the existing estimators \( d_j (j = 1, 2, \ldots, 14) \) under the following conditions which are obtained by comparing their respective variance/MSEs. Using equations (1) and (10), we have the following results

\[ \text{MSE}_{\min}(d_{ps}) < V(d_0), \text{ if } (\hat{\delta}_{22} - 1) > 0 \]

From equations (3) and (10), we have the following conditions:

\[ \text{MSE}_{\min}(d_{ps}) < \text{MSE}_{\min}(d_{ps}), \text{ if } (\hat{\delta}_{22} - 1) > (\kappa(x) - 1) \]

5 Empirical studies

To examine the validity of theoretical comparisons of proposed estimator and other available estimators discussed in this work, we borrowed the following numerical values of different population parameters for real population datasets available in [16].

Population I

\[
\begin{align*}
N &= 80, \; n = 40, \; \bar{Y} = 51.8264, \; \bar{X} = 11.2646, \; \rho_{xy} = 0.9413, \; s_y = 18.3569, \; C_y = 0.3542, \\
S_x &= 8.4563, \; C_x = 0.7507, \; \kappa(x) = 2.8664, \; \kappa(y) = 2.2667, \; \delta_{22} = 2.2209, \; M_d = 7.5750, \\
Q_1 &= 5.1500, \; Q_3 = 16.975, \; Q_r = 11.825, \; Q_d = 5.9125, \; Q_a = 11.0625.
\end{align*}
\]

Population II

\[
\begin{align*}
N &= 70, \; n = 20, \; \bar{Y} = 96.70, \; \bar{X} = 175.2671, \; \rho_{xy} = 0.7293, \; S_y = 160.7140, \; C_y = 0.6279, \\
S_x &= 140.8572, \; C_x = 0.8037, \; \kappa(x) = 7.0952, \; \kappa(y) = 4.7596, \; \delta_{22} = 4.6038, \; M_d = 121.50, \\
Q_1 &= 80.1500, \; Q_3 = 225.0250, \; Q_r = 144.8750, \; Q_d = 72.4375, \; Q_a = 152.5875.
\end{align*}
\]

To assess the performance of the proposed estimator, the bias and mean square errors of all the discussed estimators are calculated and presented in Table 2. Further the percent relative efficiencies (PREs) of proposed estimator and other existing estimators are calculated with respect to the natural sample variance estimator using the following formula and shown in Table 3.

\[ \text{PRE}(t, s_y^2) = \frac{V(s_y^2)}{\text{MSE}(t)} \times 100 \quad (11) \]

where \( t \) is the estimator of our interest.

6 Interpretations of results and conclusions

From Table 2 it is observed that the bias and mean square errors of the proposed estimator is less than the bias and mean square error of other existing estimators and subsequently from Table 3 it is visible that the percent relative efficiencies of the proposed estimator are appreciably higher than other discussed estimators for both the population values. From the above interpretations we may conclude that the structure of the proposed estimator is justified and it may be considered for practical applications by the survey practitioners.

Acknowledgements

Authors are thankful to the reviewers for their valuable suggestions which helped in improving the quality of this paper. Authors are also thankful to Indian Institute of Technology (Indian School of Mines), Dhanbad for providing financial and infrastructural support to complete this work.
Table 2: Bias $B(\cdot)$ and MSE $M(\cdot)$ of existing and proposed estimators

<table>
<thead>
<tr>
<th>Estimator and Source</th>
<th>Bias $B(\cdot)$</th>
<th>MSE $M(\cdot)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Population I</td>
<td>Population II</td>
</tr>
<tr>
<td>$d_1 = s_5^2(S_5^2/s_5^2)$ [4]</td>
<td>8.1569</td>
<td>236.1542</td>
</tr>
<tr>
<td>$d_2 = s_5^2 (S_5^2 + \kappa(s_5^2)) / (s_5^2 + \kappa(s_5^2))$ [20]</td>
<td>6.9686</td>
<td>235.8633</td>
</tr>
<tr>
<td>$d_3 = s_5^2 ([S_5^2 - C_5] / (s_5^2 - C_5))$ [5]</td>
<td>8.4963</td>
<td>236.1871</td>
</tr>
<tr>
<td>$d_4 = s_5^2 (S_5^2 C_5 / (s_5^2 C_5))$ [5]</td>
<td>9.5235</td>
<td>236.4454</td>
</tr>
<tr>
<td>$d_5 = s_5^2 (S_5^2 \kappa_3 / (s_5^2 \kappa_3 - C_5))$ [5]</td>
<td>8.7239</td>
<td>236.1588</td>
</tr>
<tr>
<td>$d_6 = s_5^2 (S_5^2 \kappa_3 / (s_5^2 \kappa_3 - C_5))$ [5]</td>
<td>10.0225</td>
<td>236.5166</td>
</tr>
<tr>
<td>$d_7 = s_5^2 ([S_5^2 + M_5] / (s_5^2 + M_5))$ [16]</td>
<td>5.3329</td>
<td>233.2011</td>
</tr>
<tr>
<td>$d_8 = s_5^2 (S_5^2 + Q_5) / (s_5^2 + Q_5)$ [17]</td>
<td>6.1309</td>
<td>232.8888</td>
</tr>
<tr>
<td>$d_9 = s_5^2 (S_5^2 + Q_5) / (s_5^2 + Q_5)$ [17]</td>
<td>2.9355</td>
<td>227.0994</td>
</tr>
<tr>
<td>$d_{10} = s_5^2 (S_5^2 + Q_5) / (s_5^2 + Q_5)$ [17]</td>
<td>4.1277</td>
<td>230.2846</td>
</tr>
<tr>
<td>$d_{11} = s_5^2 (S_5^2 + Q_5) / (s_5^2 + Q_5)$ [17]</td>
<td>5.8707</td>
<td>233.2011</td>
</tr>
<tr>
<td>$d_{12} = s_5^2 (S_5^2 + C_5) / (s_5^2 + C_5)$ [17]</td>
<td>4.3276</td>
<td>229.9762</td>
</tr>
<tr>
<td>$d_{13} = s_5^2 (S_5^2 + Q_5) / (s_5^2 + Q_5)$ [17]</td>
<td>2.7208</td>
<td>223.8262</td>
</tr>
<tr>
<td>$d_{14} = s_5^2 (S_5^2 C_5 + M_5) / (s_5^2 C_5 + M_5)$ [18]</td>
<td>4.5924</td>
<td>232.4854</td>
</tr>
<tr>
<td>$d_{ps} = s_5^2 \lambda + (1 - \lambda) \left( \frac{S_5^2 \kappa_3 + M_5^2}{S_5^2 \kappa_3 + M_5^2} \right)$</td>
<td>1.5164</td>
<td>107.2051</td>
</tr>
</tbody>
</table>

Table 3: PREs of various estimators with respect to $d_0 = s_5^2$

<table>
<thead>
<tr>
<th>Estimator and Source</th>
<th>PREs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Population I</td>
</tr>
<tr>
<td>$d_1 = s_5^2(S_5^2/s_5^2)$ [4]</td>
<td>183.2345</td>
</tr>
<tr>
<td>$d_2 = s_5^2 (S_5^2 + \kappa(s_5^2)) / (s_5^2 + \kappa(s_5^2))$ [20]</td>
<td>196.5956</td>
</tr>
<tr>
<td>$d_3 = s_5^2 ([S_5^2 - C_5] / (s_5^2 - C_5))$ [5]</td>
<td>179.6211</td>
</tr>
<tr>
<td>$d_4 = s_5^2 (S_5^2 - \kappa(s_5^2)) / (s_5^2 - \kappa(s_5^2))$ [5]</td>
<td>169.2398</td>
</tr>
<tr>
<td>$d_5 = s_5^2 (S_5^2 \kappa_3 - C_5) / (s_5^2 \kappa_3 - C_5)$ [5]</td>
<td>181.9786</td>
</tr>
<tr>
<td>$d_6 = s_5^2 (S_5^2 \kappa_3 - C_5) / (s_5^2 \kappa_3 - C_5)$ [5]</td>
<td>164.4933</td>
</tr>
<tr>
<td>$d_7 = s_5^2 (S_5^2 + M_5) / (s_5^2 + M_5)$ [16]</td>
<td>216.6165</td>
</tr>
<tr>
<td>$d_8 = s_5^2 (S_5^2 + Q_5) / (s_5^2 + Q_5)$ [17]</td>
<td>206.6417</td>
</tr>
<tr>
<td>$d_9 = s_5^2 (S_5^2 + Q_5) / (s_5^2 + Q_5)$ [17]</td>
<td>247.2471</td>
</tr>
<tr>
<td>$d_{10} = s_5^2 (S_5^2 + Q_5) / (s_5^2 + Q_5)$ [17]</td>
<td>232.1285</td>
</tr>
<tr>
<td>$d_{11} = s_5^2 (S_5^2 + Q_5) / (s_5^2 + Q_5)$ [17]</td>
<td>209.8595</td>
</tr>
<tr>
<td>$d_{12} = s_5^2 (S_5^2 + Q_5) / (s_5^2 + Q_5)$ [17]</td>
<td>229.5416</td>
</tr>
<tr>
<td>$d_{13} = s_5^2 (S_5^2 + Q_5) / (s_5^2 + Q_5)$ [17]</td>
<td>249.8426</td>
</tr>
<tr>
<td>$d_{14} = s_5^2 (S_5^2 C_5 + M_5) / (s_5^2 C_5 + M_5)$ [18]</td>
<td>226.1175</td>
</tr>
<tr>
<td>$d_{ps} = s_5^2 \lambda + (1 - \lambda) \left( \frac{S_5^2 \kappa_3 + M_5^2}{S_5^2 \kappa_3 + M_5^2} \right)$</td>
<td>270.6320</td>
</tr>
</tbody>
</table>

References


G. N. Singh is a Professor of Statistics in the Department of Applied Mathematics, Indian Institute of Technology (ISM) Dhanbad, India. He obtained his Ph.D. degree in 1990 from Banaras Hindu University, Varanasi, India. He has more than 27 years of teaching experience in the field of statistics. He served as faculty in Panjab University, Chandigarh and Kurukshetra University, India. He has more than 30 years of research experience in the various field of Statistics which covers Sample Surveys, Statistical Inference, Data Analysis, Data Mining etc. He has published number of research papers in Indian and Foreign journals of repute. He presented his research problems in international and national conferences and delivered various invited talks in academic forum. He has produced 12 Ph.D, 5 M.phil and 4 research projects.

Awadhesh Kumar Pandey is a research scholar in the Department of Applied Mathematics, Indian Institute of Technology (ISM) Dhanbad, India. He is pursuing Ph.D. in Applied Statistics. His research interest is in the areas of Sample Survey and Statistical Inference.
Chandraketu Singh is a research scholar in the Department of Applied Mathematics, Indian Institute of Technology (ISM) Dhanbad, India. He is pursuing Ph.D. in Applied Statistics. His research interest is in the areas of Sample Survey, Bio-statistics and Statistical Inference.

Mahamood Usman is a research scholar in the Department of Applied Mathematics, Indian Institute of Technology (ISM) Dhanbad, India. He is pursuing Ph.D. in Applied Statistics. His research interest is in the areas of Sample Survey, Missing Data Analysis and Statistical Inference.