An Estimator of the Mean Estimation of Study Variable Using Median of Auxiliary Variable

Sunil Kumar*
Alliance University, Bangalore, India
*sunilbhougal06@gmail.com

Received: 07, October 2014/ Revised: 26, June 2015 / Accepted: 24, August 2015
©IAppStat-SL2015

ABSTRACT

In the present study, I propose a modified estimator for estimating the population mean of the study variable auxiliary information when the population mean and the population median of the auxiliary variable is known. The expression of bias and mean squared error (MSE) of the proposed estimator is derived. Some existing estimators are also discussed. Comparisons of the proposed estimator with the other estimators are carried out. The results obtained are illustrated numerically by using three natural populations considered in the literature.

Keywords: Mean squared error, bias, median, simple random sampling

1. Introduction

1.1 Introduction to the Research Problem

Use of auxiliary information in the estimation of population parameters such as population mean, ratio of two population means, product of two population means, coefficient of variation etc. has been in practice. Ratio, product and regression type estimators are good examples in this context. Cochran (1940) initiated the use of auxiliary information at estimation stage and proposed ratio estimator for population mean. It is well established fact that ratio type estimators provide better efficiency in comparison to simple mean estimator if the study variable and auxiliary variable are positively correlated. If the correlation between the study variable and auxiliary variables negative, product estimator given by Robson (1957) is more efficient than simple mean estimator.

Further improvements are also achieved on the classical ratio estimator by introducing a large number of modified ratio estimators with the use of known parameters like, coefficient of variation, coefficient of kurtosis, coefficient of...

Further, Subramani and Kumarapandiyan (2013) had taken initiative by proposed modified ratio estimator for estimating the population mean of the study variable by using the population median of the auxiliary variable.

The objective of the paper is to proposed modified estimator for estimating the population mean by using the population median of the auxiliary variable.

2. Notations Used

The following are the notations used in the paper:

\[ N: \text{Size of the finite population}; n: \text{Sample size taken from the population of size } N; \ y: \text{Study variable}; x: \text{Auxiliary variable}; y_i: \text{Value of the } i^{th} \text{unit of the sample for the study variable } y; x_i: \text{Value of the } i^{th} \text{unit of the sample for the auxiliary variable } x; \bar{y} = \frac{\sum_{i=1}^{N} y_i}{N}: \text{Population mean for study variable } y \text{ for the entire population}; \]

\[ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}: \text{Sample mean for the auxiliary variable } x \text{ for the sample of size } n; \]

\[ \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}: \text{Sample mean for the study variable } y \text{ for the sample of size } n; S_y^2 = \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{(N - 1)}: \text{Population mean square of the study variable } y \text{ for the entire population}; \]

\[ S_x^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{X})^2}{(N - 1)}: \text{Population mean square of the auxiliary variable } x \text{ for the entire population}; \]

\[ S_{xy} = \frac{\sum_{i=1}^{N} (y_i - \bar{y})(x_i - \bar{X})}{(N - 1)}: \text{Population covariance between } y \text{ and } x \text{ for the entire population}; \]

\[ \rho_{yx} = \frac{S_{xy}}{S_x S_y}: \text{Correlation coefficient between } y \text{ and } x \text{ in the entire population}; C_y = S_y/\bar{y}: \text{Coefficient of variation (CV) of the study variable } y \text{ in the entire population}; C_x = S_x/\bar{X}: \text{Coefficient of variation (CV) of the auxiliary variable } x \text{ in the entire population}; \]

\[ \beta_1 = \frac{N \sum_{i=1}^{N} (x_i - \bar{X})^3}{(N-1)(N-2)S_x^3} \text{ is the coefficient of Skewness of the auxiliary variable}; \]

\[ \beta_2 = \frac{N(N+1) \sum_{i=1}^{N} (x_i - \bar{X})^4}{(N-1)(N-2)(N-3)S_x^4} - \frac{3(N-1)^2}{(N-2)(N-3)} \text{ is the coefficient of Kurtosis of the auxiliary variable}, \]

\[ M_d \text{ is the population median of the auxiliary variable}. \]

3. Procedure and Definitions

Consider a finite population \( U = (U_1, U_2, ..., U_N) \) of size \( N \). Let \( y \) and \( x \) denote the study variable and the auxiliary variable taking values \( y_i \) and \( x_i \) respectively on the \( i^{th} \) unit (\( i = 1, 2, ..., N \)). For estimating the population mean \( \bar{y} \) of \( y \) a simple random
sample of size \( n \) is drawn without replacement from the population \( U \). Then the classical ratio estimator is defined as
\[
t_R = \frac{\bar{y}}{\bar{x}} \bar{X}; \text{ if } \bar{x} \neq 0
\]
and the product estimator is given by
\[
t_p = \frac{\bar{y}}{\bar{x}} \bar{x};
\]
where \( \bar{X} \), the population mean of the auxiliary variable \( x \) is known.

The mean squared error expressions of the ratio and product estimators are
\[
MSE(t_R) = \left( \frac{1-f}{n} \right) \bar{y}^2 \left( C_y^2 + C_x^2 \left( 1 - 2 \rho_{yx} \frac{C_y}{C_x} \right) \right)
\]
\[
MSE(t_p) = \left( \frac{1-f}{n} \right) \bar{y}^2 \left( C_y^2 + C_x^2 \left( 1 + 2 \rho_{yx} \frac{C_y}{C_x} \right) \right)
\]

Further, a list of modified ratio estimators is given in Table 1 is used for assessing the performance of the proposed estimator along with their bias and mean squared error expressions.

Table 1: Existing modified ratio type estimators with their biases and mean squared errors

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Bias, ( B(.) )</th>
<th>Mean squared error, MSE(.)</th>
<th>Constants or ( R_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_1 = \bar{y} \left( \frac{\bar{X} + C_x}{\bar{x} + C_x} \right) ) Sisodia and Dwivedi (1981)</td>
<td>( \frac{1-f}{n} \bar{y} \left( \theta_1^2 C_x^2 \right) - 2 \theta_1 C_x C_y \rho_{yx} )</td>
<td>( \left( \frac{1-f}{n} \right) \bar{y}^2 \left( C_y^2 \right) )</td>
<td>( \theta_1 = \frac{\bar{X}}{\bar{X} + C_x} )</td>
</tr>
<tr>
<td>( t_2 = \bar{y} \left( \frac{\bar{x} + \beta_2}{\bar{x} + \beta_2} \right) ) Singh et al (2004)</td>
<td>( \frac{1-f}{n} \bar{y} \left( \theta_2^2 C_x^2 \right) - 2 \theta_2 C_x C_y \rho_{yx} )</td>
<td>( \left( \frac{1-f}{n} \right) \bar{y}^2 \left( C_y^2 \right) )</td>
<td>( \theta_2 = \frac{\bar{X}}{\bar{X} + \beta_2} )</td>
</tr>
<tr>
<td>( t_3 = \bar{y} \left( \frac{\bar{x} + \beta_1}{\bar{x} + \beta_1} \right) ) Yan and Tian (2010)</td>
<td>( \frac{1-f}{n} \bar{y} \left( \theta_3^2 C_x^2 \right) - 2 \theta_3 C_x C_y \rho_{yx} )</td>
<td>( \left( \frac{1-f}{n} \right) \bar{y}^2 \left( C_y^2 \right) )</td>
<td>( \theta_3 = \frac{\bar{X}}{\bar{X} + \beta_1} )</td>
</tr>
<tr>
<td>( t_4 = \bar{y} \left( \frac{\bar{x} + \rho_{yx}}{\bar{x} + \rho_{yx}} \right) ) Singh and Tailor (2003)</td>
<td>( \frac{1-f}{n} \bar{y} \left( \theta_4^2 C_x^2 \right) - 2 \theta_4 C_x C_y \rho_{yx} )</td>
<td>( \left( \frac{1-f}{n} \right) \bar{y}^2 \left( C_y^2 \right) )</td>
<td>( \theta_4 = \frac{\bar{X}}{\bar{X} + \rho_{yx}} )</td>
</tr>
<tr>
<td>( t_5 = \bar{y} \left( \frac{\bar{X} C_x + \beta_2}{\bar{x} C_x + \beta_2} \right) ) Upadhyaya and Singh (1999)</td>
<td>( \frac{1-f}{n} \bar{y} \left( \theta_5^2 C_x^2 \right) - 2 \theta_5 C_x C_y \rho_{yx} )</td>
<td>( \left( \frac{1-f}{n} \right) \bar{y}^2 \left( C_y^2 \right) )</td>
<td>( \theta_5 = \frac{\bar{X} C_x}{\bar{X} C_x + \rho_{yx}} )</td>
</tr>
<tr>
<td>( t_6 = \bar{y} + b(\bar{x} - \bar{x}) ) Kadilar and Cingi (2004)</td>
<td>( \frac{1-f}{n} S_x^2 \sqrt{R_6} )</td>
<td>( \left( \frac{1-f}{n} \right) \left( R_6 \bar{X} S_x^2 \right) + S_x^2 (1 - \rho_{yx}^2) )</td>
<td>( R_6 = \bar{y} )</td>
</tr>
</tbody>
</table>
Table 1 (continued):

<table>
<thead>
<tr>
<th>Expression</th>
<th>Formula</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_7 = \frac{\bar{Y} + b(\bar{X} \cdot \bar{x}) - \bar{X} \cdot \bar{C}_x}{\bar{X} + \bar{C}_x}$</td>
<td>$t_7 = \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_7 + \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_7 + S_y^2 (1 - \rho_{yx})$</td>
<td></td>
</tr>
<tr>
<td>Kadilar and Cingi (2004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_8 = \frac{\bar{Y} + b(\bar{X} \cdot \bar{x})}{\bar{X} + \bar{\beta}_x}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kadilar and Cingi (2004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_9 = \frac{\bar{Y} + b(\bar{X} \cdot \bar{x})}{\bar{X} \cdot \bar{C}_x + \bar{\beta}_x}$</td>
<td>$t_9 = \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_0 + \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_0 + S_y^2 (1 - \rho_{yx})$</td>
<td></td>
</tr>
<tr>
<td>Kadilar and Cingi (2004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{10} = \frac{\bar{Y} + b(\bar{X} \cdot \bar{x})}{\bar{X} + \bar{\beta}_1}$</td>
<td>$t_{10} = \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_{10} + \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_{10} + S_y^2 (1 - \rho_{yx})$</td>
<td></td>
</tr>
<tr>
<td>Yan and Tian (2010)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{11} = \frac{\bar{Y} + b(\bar{X} \cdot \bar{x})}{\bar{X} \cdot \bar{\beta}_1 + \bar{\beta}_2}$</td>
<td>$t_{11} = \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_{11} + \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_{11} + S_y^2 (1 - \rho_{yx})$</td>
<td></td>
</tr>
<tr>
<td>Yan and Tian (2010)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{12} = \frac{\bar{Y} + b(\bar{X} \cdot \bar{x})}{\bar{X} + \bar{\rho}_{yx} + \bar{C}_x}$</td>
<td>$t_{12} = \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_{12} + \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_{12} + S_y^2 (1 - \rho_{yx})$</td>
<td></td>
</tr>
<tr>
<td>Kadilar and Cingi (2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{13} = \frac{\bar{Y} + b(\bar{X} \cdot \bar{x})}{\bar{X} \cdot \bar{C}<em>x + \bar{\rho}</em>{yx} + \bar{C}_x}$</td>
<td>$t_{13} = \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_{13} + \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_{13} + S_y^2 (1 - \rho_{yx})$</td>
<td></td>
</tr>
<tr>
<td>Kadilar and Cingi (2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{14} = \frac{\bar{Y} + b(\bar{X} \cdot \bar{x})}{\bar{X} \cdot \bar{\rho}_{yx} + \bar{C}_x}$</td>
<td>$t_{14} = \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_{14} + \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_{14} + S_y^2 (1 - \rho_{yx})$</td>
<td></td>
</tr>
<tr>
<td>Kadilar and Cingi (2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{15} = \frac{\bar{Y} + b(\bar{X} \cdot \bar{x})}{\bar{X} \cdot \bar{\rho}_{yx} + \bar{\beta}_x}$</td>
<td>$t_{15} = \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_{15} + \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_{15} + S_y^2 (1 - \rho_{yx})$</td>
<td></td>
</tr>
<tr>
<td>Kadilar and Cingi (2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{16} = \frac{\bar{Y}(\bar{X} + M_d)}{\bar{X} + M_d}$</td>
<td>$t_{16} = \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_{16} + \frac{1 - f}{n} \frac{S^2}{\bar{Y}} R_{16} + S_y^2 (1 - \rho_{yx})$</td>
<td></td>
</tr>
<tr>
<td>Subramani and Kumarapandian (2013)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. Proposed Estimator
Following Subramani and Kumarapandiyan (2013), I have proposed an estimator for estimating the population mean when the population mean and population median of the auxiliary variable is known
\[
t_s = \bar{y} \left\{ \alpha \left( \frac{\bar{X} + M_d}{\bar{x} + M_d} \right) + (1 - \alpha) \left( \frac{\bar{x} + M_d}{\bar{X} + M_d} \right) \right\}
\]
where \( M_d \) is the population median of the auxiliary variable \( X \).
To the first degree of approximation, I have obtained the expression of bias and mean squared error (MSE) of the proposed estimator as
\[
B(t_s) = \left( \frac{1 - f}{n} \right) \bar{y} \left\{ (1 - 2\alpha)\phi \rho_{yx} C_y C_x + \alpha \phi^2 C_x^2 \right\}
\]
and
\[
MSE(t_s) = \left( \frac{1 - f}{n} \right) \bar{y} \left\{ C_y^2 + (1 - 2\alpha)^2 \phi^2 C_x^2 + 2(1 - 2\alpha)\phi \rho_{yx} C_y C_x \right\}
\]
where \( \phi = \frac{\bar{x}}{\bar{X} + M_d} \).
The \( MSE(t_s) \) will be minimum when
\[
\alpha = \frac{1}{2} \left( 1 - \frac{\rho_{yx} C_y}{\phi C_x} \right) = \alpha_0 \text{(say)}
\]
By Substituting the minimum value of \( \alpha \) in the proposed estimator, one can get the asymptotically optimum estimator (AOE) as
\[
t_s(\text{opt}) = \bar{y} \left\{ \alpha_0 \left( \frac{\bar{X} + M_d}{\bar{x} + M_d} \right) + (1 - \alpha_0) \left( \frac{\bar{x} + M_d}{\bar{X} + M_d} \right) \right\}
\]
Thus, the optimum MSE of \( t_s \) is
\[
\text{Min.} \; MSE(t_s) = \left( \frac{1 - f}{n} \right) \bar{y} \left( 1 - \rho_{yx}^2 \right) C_y^2.
\]
5. Efficiency Comparison
For comparison of proposed estimator with the existing estimators, I have derived the conditions for which the proposed estimator is more efficient than the existing modified ratio estimators as
\[
\text{Min.} \; MSE(t_s) \leq MSE(t_i; i = 1, 2, 3, 4, 5) \text{ if } \rho_{yx} \leq \frac{q_i C_x}{C_y}.
\]
\[
\text{Min.} \; MSE(t_s) \leq MSE(t_i; i = 6 \text{ to } 15) \text{ if } R_i^2 S_x^2 \geq 0.
\]
\[
\text{Min.} \; MSE(t_s) \leq MSE(t_{16}) \text{ if } \rho_{yx} \leq \frac{R_{16} C_x}{C_y}.
\]
\[
\text{Min.} \; MSE(t_s) \leq MSE(t_R) \text{ if } \rho_{yx} \leq \frac{C_x}{C_y}.
\]
From the above conditions, it is noted that the proposed estimator is more efficient among other discussed estimators if the above conditions holds true.
6. Empirical Study
To demonstrate the performance of the suggested estimator empirically in comparison to other estimators. I have used three natural population data sets. The descriptions of the populations are given below.

Population I: [Source: Murthy (1967)]
Y: Output for 80 factories in a region
X: Fixed Capital
\[ N = 80; n = 20; \bar{Y} = 51.8264; \bar{X} = 11.2646; \rho_{yx} = 0.9413; S_y = 18.3569; C_y = 0.3542; S_x = 8.4563; C_x = 0.7507; \beta_1 = 1.05; \beta_2 = -0.06339; \quad M_d = 7.575 \]

Population II: [Source: Murthy (1967)]
Y: Output for 80 factories in a region
X: Data on number of workers
\[ N = 80; n = 20; \bar{Y} = 51.8264; \bar{X} = 2.8513; \rho_{yx} = 0.9150; S_y = 18.3569; C_y = 0.3542; S_x = 2.7042; C_x = 0.9484; \beta_1 = 0.6978; \beta_2 = 1.3005; \quad M_d = 1.48 \]

Population III: [Source: Cochran (1977)]
Y: Number of persons
X: Number of rooms
\[ N = 10; n = 4; \bar{Y} = 101.00; \bar{X} = 58.80; \rho_{yx} = 0.6515; S_y = 14.6523; C_y = 0.1449; S_x = 7.5339; C_x = 0.1281; \beta_1 = 0.5764; \beta_2 = -0.3814; \quad M_d = 58 \]

Here, I have computed mean squared error (MSE) and the constants of the estimators.

The results are given in the following table.

Table 2: The mean squared errors and constants of the existing and proposed estimators

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Population I</th>
<th></th>
<th>Population II</th>
<th></th>
<th>Population III</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>Constants</td>
<td>MSE</td>
<td>Constants</td>
<td>MSE</td>
<td>Constants</td>
</tr>
<tr>
<td>( t_1 ): Sisodia and Dwivedi(1981)</td>
<td>15.25812</td>
<td>0.937521</td>
<td>17.18812</td>
<td>0.750401</td>
<td>20.2396</td>
<td>0.997826</td>
</tr>
<tr>
<td>( t_2 ): Singh et al (2004)</td>
<td>19.33825</td>
<td>1.005659</td>
<td>12.84257</td>
<td>0.686762</td>
<td>20.14213</td>
<td>0.990292</td>
</tr>
<tr>
<td>( t_3 ): Yan and Tian (2010)</td>
<td>14.01128</td>
<td>0.914735</td>
<td>21.36603</td>
<td>0.803387</td>
<td>20.35575</td>
<td>1.006529</td>
</tr>
<tr>
<td>( t_4 ): Singh and Tailor (2003)</td>
<td>14.45027</td>
<td>0.922882</td>
<td>17.68491</td>
<td>0.757056</td>
<td>20.12623</td>
<td>0.989041</td>
</tr>
<tr>
<td>( t_5 ): Upadhyaya and Singh (1999)</td>
<td>19.45917</td>
<td>1.007553</td>
<td>12.1351</td>
<td>0.675254</td>
<td>19.45448</td>
<td>0.928916</td>
</tr>
</tbody>
</table>
Table 2 (continued):

| \( t_6 \); Kadilar and Cingi (2004) | 58.20263 | 4.60082 | 92.65628 | 18.17641 | 43.70438 | 1.719388 |
| \( t_7 \); Kadilar and Cingi (2004) | 51.3313 | 4.313367 | 53.07362 | 13.6396 | 43.59507 | 1.71565 |
| \( t_8 \); Kadilar and Cingi (2004) | 58.84691 | 4.626857 | 44.78744 | 12.48287 | 43.21808 | 1.702697 |
| \( t_9 \); Kadilar and Cingi (2004) | 4.130342 | 1.001632 | 2.317192 | 0.974223 | 26.33849 | 0.95739 |
| \( t_{10} \); Yan and Tian (2010) | 48.93561 | 4.208533 | 60.5325 | 14.60269 | 44.0341 | 1.730613 |
| \( t_{11} \); Yan and Tian (2010) | 58.81598 | 4.625611 | 35.18871 | 10.99178 | 45.04985 | 1.764745 |
| \( t_{12} \); Kadilar and Cingi (2006) | 49.78532 | 4.246012 | 53.98248 | 13.76056 | 43.15575 | 1.700546 |
| \( t_{13} \); Kadilar and Cingi (2006) | 47.40103 | 4.139986 | 52.63652 | 13.58105 | 39.85643 | 1.58251 |
| \( t_{14} \); Kadilar and Cingi (2006) | 50.94479 | 4.296627 | 50.78761 | 13.33051 | 43.53689 | 1.713657 |
| \( t_{15} \); Kadilar and Cingi (2006) | 58.88745 | 4.628491 | 42.40512 | 12.12991 | 42.96371 | 1.693901 |
| \( t_{16} \); Subramani and Kumarapandiyan (2013) | 2.782544 | 0.597921 | 11.13659 | 0.658301 | 19.89921 | 0.503425 |
| \( t_5 \); (Proposed Estimator) | 0.068028 | 0.597921 | 0.006226 | 0.658301 | 6.267063 | 0.503425 |

From the above table, it is envisaged that the proposed optimum estimator is more efficient than other existing estimator mentioned in Table 1 in terms of less mean squared error (MSE).

7. Conclusion
In this paper, I have proposed a modified estimator based on simple random sampling without replacement by using auxiliary variable, under the situation when population mean and median of the auxiliary variable is known. It is found that the performance of the proposed estimator in terms of bias and mean squared error is more efficient than all other existing estimator for certain known population parameters of auxiliary variable. The above results are supported theoretically and empirically by three natural populations.
Acknowledgements
Author wishes to thank the learned referees for their critical and constructive comments regarding improvement of the paper.

References


