IMPROVEMENT FOR ESTIMATING POPULATION MEAN IN SIMPLE RANDOM SAMPLING SCHEME

Gagan Kumar1, Priyanka Bharti2, S.K. Yadav3 and Surendra Kumar*4
1Department of Economics, Govt. Degree College, Pihani, Hardoi, India
2Department of Economics, DDU Govt. PG College, Sitapur, India
3Department of Statistics, Babasaheb Bhimrao Ambedkar University, Lucknow, India
*4Department of Mathematics, Govt. Degree College Pihani, Hardoi, India
E Mail: *surendra.kumar776@gmail.com

Received October 09, 2018
Modified March 09, 2019
Accepted April 07, 2019

Abstract
This article, considers the improved population mean estimation without increasing the cost of the survey. An improved estimator for mean of population characteristic has been suggested which makes use of population median of main (study) variable. We study the sampling properties of the proposed estimator up to the approximation of order one. The least value of the mean squared error (MSE) for the optimum value of the constant of the suggested estimator has been obtained. The suggested estimator is compared with the competing estimators. The theoretical findings are justified with a numerical example. Through the numerical example we show the improvement of suggested estimator over other competing estimators.

Key Words: Main Variable, Secondary Variable, Exponential Ratio Estimators, Bias, MSE, Efficiency.

1. Introduction
It is essential to use sampling techniques when the population is very large to save the time and cost. Generally the corresponding statistics is used for estimating the parameter under consideration with the desirable properties of unbiasedness, consistency, efficiency and sufficiency etc. The mean per unit estimator of study variable is an appropriate estimator to be considered good for population mean and has unbiasedness property but it has a large amount of variation which is not desirable. Now we search for the estimator of population mean even biased but should have least MSE. The very purpose is achieved through using auxiliary variable and study variable as well which are in high correlation to each other. But this auxiliary information is collected on additional cost of the survey. Thus we further search for the improved estimators not using auxiliary information but makes use of known parameters of the main variable such as population median which is easily available with not influencing the survey cost.

In the present paper we use known population median of main variable for efficiently estimating the population mean of main variable and study its sampling properties up to approximation of degree one.
2. Review of Literature

Sample mean is the most natural estimator and is unbiased for population mean, given by,

\[ t_o = \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \]  
(1)

Its variance is given by,

\[ V(t_o) = \frac{1-f}{n} S_y^2 = \frac{1-f}{n} \bar{y}^2 C_y^2 \]  
(2)

\[ C_y = \frac{S_y}{\bar{y}}, \quad S_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{y})^2 = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \bar{y})^2, \quad f = \frac{n}{N}. \]

Watson [14] given the conventional regression estimator as,

\[ t_1 = \bar{y} + \beta (\bar{X} - \bar{x}) \]  
(3)

Its variance is given by,

\[ V(t_1) = \frac{1-f}{n} \bar{y}^2 C_y^2 (1 - \rho_{xy}^2) \]  
(4)

\[ \rho_{xy} = \frac{\text{Cov}(x,y)}{S_x S_y}, \quad \text{Cov}(x,y) = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{y})(X_i - \bar{X}), \]

\[ C_{yx} = \rho_{xy} C_x C_y, \quad C_x = \frac{S_x}{\bar{X}}, \]

\[ S_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2 = \frac{1}{N} \sum_{i=1}^{N} (\bar{X}_i - \bar{X})^2. \]

Cochran [3] suggested traditional ratio estimator as,

\[ t_2 = \frac{\bar{y}}{\bar{x}} \]  
(5)

The MSE of \( t_2 \) is expressed as,

\[ \text{MSE}(t_2) = \frac{1-f}{n} \bar{y}^2 [C_y^2 + C_x^2 - 2 C_{yx}] \]  
(6)

Bahl and Tuteja [2] proposed the usual exponential ratio estimator as,

\[ t_3 = \bar{y} \exp \left( \frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}} \right) \]  
(7)

The MSE of the above estimator is as follows,

\[ \text{MSE}(t_3) = \frac{1-f}{n} \bar{y}^2 \left[ C_y^2 + \frac{C_x^2}{4} - C_{yx} \right] \]  
(8)

Kadilar and Cingi [4] presented the following ratio type estimator as,

\[ t_4 = \bar{y} \left( \frac{\bar{X}}{\bar{x}} \right)^2 \]  
(9)
The MSE of $t_4$ in (9) is given by,

$$MSE(t_4) = \frac{1-f}{n} \bar{Y}^2[C_y^2 + 4C_x^2 - 4C_{yx}]$$  \hspace{1cm} (10)

Srivastava [8] proposed a generalized ratio type estimator as,

$$t_5 = \bar{Y} \left( \frac{\bar{x}}{\bar{X}} \right)^\alpha$$  \hspace{1cm} (11)

Where $\alpha$ is the arbitrary constant to be obtained that the MSE of $t_5$ is least.

The least MSE of above estimator for optimum $\alpha_{opt} = -C_{yx}/C_x^2$ is given by,

$$MSE_{\min}(t_5) = \frac{1-f}{n} \bar{Y}^2C_x^2(1 - \rho_{yx}^2)$$  \hspace{1cm} (12)

Reddy [7] proposed a generalized ratio type estimator as,

$$t_6 = \bar{Y} \left( \frac{X}{\bar{X} + \alpha(x - \bar{X})} \right)$$  \hspace{1cm} (13)

The minimum MSE of above estimator for optimum $\alpha_{opt} = C_{yx}/C_x^2$ is given by,

$$MSE_{\min}(t_6) = \frac{1-f}{n} \bar{Y}^2C_y^2(1 - \rho_{yx}^2)$$  \hspace{1cm} (14)

Subramani [10] suggested ratio type estimator as,

$$t_7 = \frac{\bar{Y} M}{m}$$  \hspace{1cm} (15)

Where $M$ is population median and $m$ is sample median of the main variable.

The MSE of $t_7$ is given by,

$$MSE(t_7) = \frac{1-f}{n} \bar{Y}^2[C_y^2 + R_{12}^2C_m^2 - 2R_{12}C_{ym}]$$  \hspace{1cm} (16)

Where,

$$R_{12} = \frac{\bar{Y}}{M}, \quad C_m = \frac{S_m}{M}, \quad S_{ym} = \frac{1}{C_n} \sum_{i=1}^{N} (\bar{Y}_i - \bar{Y})(m_i - M), \quad C_{ym} = \frac{S_{ym}}{YM} \quad \text{and} \quad S_m = \frac{1}{C_n} \sum_{i=1}^{N} (m_i - M)^2.$$

In the literature various authors including Kumar et al. [6], Subramani [9], Subramani and Kumarapandiyan [11-12], Tailor and Sharma [13], Yadav and Kadilar [16-17], Yadav and Mishra [18], Yadav et al. [19-21], Yan and Tian [22], Abid et al. [1] and Kumar et al. [7] used the auxiliary information and known population parameters of study variable as well form improved estimation of population mean.
3. Proposed Estimator

Getting motivated by Subramani [10], we proposed the new estimator as,

$$ t = \bar{Y} \left[ \alpha \left( 2 - \frac{M}{m} \right) + (1 - \alpha) \left( 2 - \frac{M}{M} \right) \right] $$  \hspace{1cm} (17)

Where $\alpha$ is a characterizing scale and is obtained such that the MSE of $t$ is minimum.

We use the standard approximations given below for studying the properties of the proposed estimator as,

$$ \bar{Y} = \bar{Y}(1 + e_o) \quad \text{and} \quad m = M(1 + e_i) \quad \text{such that} \quad E(e_o) = 0, \quad E(e_i^2) = \frac{1 - f}{n} C^2, \quad E(e_i) = \frac{1 - f}{n} C^2, \quad E(e_o e_i) = \frac{1 - f}{n} C^2, \quad \text{where,} \quad \overline{M} = \frac{1}{n} \sum_{i=1}^{n} m_i $$

The suggested estimator $t$ can be represented in the form of $e_i$'s ($i = 1, 2$) as,

$$ t = \bar{Y}(1 + e_o) \left[ \alpha \left( 2 - \frac{M}{M(1 + e_i)} \right) + (1 - \alpha) \left( 2 - \frac{M}{M} \right) \right] $$

$$ = \bar{Y}(1 + e_o) \left[ \alpha \left( 2 - (1 + e_i)^{-1} \right) + (1 - \alpha)(1 - e_i) \right] $$

$$ = \bar{Y}(1 + e_o) \left[ \alpha \left( 2 - (1 - e_i + e_i^2 + ...) \right) + (1 - \alpha)(1 - e_i) \right] $$

$$ = \bar{Y}(1 + e_o) \left[ \alpha(1 + e_i - e_i^2) + (1 - \alpha)(1 - e_i) \right], \quad \text{up to approximation of order one} $$

$$ = \bar{Y}(1 + e_o)(1 - e_i + 2ae_i - ae_i^2) $$

$$ = \bar{Y}(1 + e_o - e_i e_i + 2ae_i - 2ae_o e_i - ae_i^2) $$

$$ t - \bar{Y} = \bar{Y}[e_o - e_i e_i + 2ae_i - 2ae_o e_i - ae_i^2] $$

$$ t - \bar{Y} = \bar{Y}[e_o + (2\alpha - 1)e_i + (2\alpha - 1)e_o e_i - \alpha e_i^2] \hspace{1cm} (18) $$

Taking expectations on both sides of (18) and using standard results of expectations, we have the bias of $t$ as,

$$ B(t) = \bar{Y} \left[ (2\alpha - 1) \frac{B(m)}{M} + (2\alpha - 1) \lambda C_{ym} - \alpha \lambda C^2_m \right] $$

$$ = \bar{Y} \left[ \alpha_1 \frac{B(m)}{M} + \alpha_1 \lambda C_{ym} - \alpha_1 \lambda C^2_m \right], \quad \alpha_1 = (2\alpha - 1) \hspace{1cm} (19) $$

From equation (2), up to approximation of degree one, we have,

$$ t - \bar{Y} \approx \bar{Y}[e_o + (2\alpha - 1)e_i] = \bar{Y}[e_o + \alpha_i e_i] $$
Improvement for estimating population mean in ...

Squaring above equation and getting expectations both sides, we have the MSE of \( t \) as,

\[
\text{MSE}(t) = \overline{Y}^2 E[e_0 + \alpha_t e_1]^2
\]

\[
= \overline{Y}^2 E[e_0^2 + \alpha_t^2 e_1^2 + 2\alpha_t e_0 e_1]
\]

\[
= \overline{Y}^2 [E(e_0^2) + \alpha_t^2 E(e_1^2) + 2\alpha_t E(e_0 e_1)]
\]

We get MSE of \( t \) by putting, \( E(e_0^2) \), \( E(e_1^2) \) and \( E(e_0 e_1) \) as,

\[
\text{MSE}(t) = \lambda \overline{Y}^2 [C_y^2 + \alpha_t^2 C_m^2 + 2\alpha_t C_{ym}]
\]  

(20)

The \( \text{MSE}(t) \) is least for,

\[
\frac{\partial \text{MSE}(t)}{\partial \alpha_t} = 0 \text{ gives,}
\]

\[
\alpha_t C_m^2 + C_{ym} = 0 \text{ or,}
\]

\[
\alpha_{t(\text{opt})} = -\frac{C_{ym}}{C_m^2}
\]  

(21)

The minimum \( \text{MSE}(t) \) for \( \alpha_{t(\text{opt})} \) is given by,

\[
\text{MSE}_{\text{min}}(t) = \frac{1 - f}{n} \overline{Y}^2 \left[ C_y^2 - \frac{C_{ym}^2}{C_m^2} \right]
\]  

(22)

4. Efficiency Comparison

From equation (22) and (2), we have,

\[
V(t_0) - \text{MSE}_{\text{min}}(t) > 0, \text{ if}
\]

\[
\frac{C_{ym}^2}{C_m^2} > 0, \text{ or if } C_{ym}^2 > 0
\]

Thus \( t \) are better than \( t_0 \).

From equation (22) and (4), we observe,

\[
\text{MSE}(t_1) - \text{MSE}_{\text{min}}(t) > 0 \text{ if}
\]

\[
\frac{C_{ym}^2}{C_m^2} - C_y^2 \rho_{yx}^2 > 0
\]

Thus \( t \) are better than \( t_1 \) if above condition is met.

From equation (22) and (6), we observe,

\[
\text{MSE}(t_2) - \text{MSE}_{\text{min}}(t) > 0 \text{ if}
\]

\[
C_x^2 - 2C_{yx} + \frac{C_{ym}^2}{C_m^2} > 0, \text{ if}
\]
\[ C_x^2 + \frac{C_{ym}^2}{C_m^2} > 2C_{yx} \]

Thus \( t \) are better than \( t_2 \) if above condition is met.

From equation (22) and (8), we observe,
\[ MSE(t_3) - MSE_{\text{min}}(t) > 0 \] if
\[ \frac{C_x^2}{4} - C_{yx} + \frac{C_{ym}^2}{C_m^2} > 0, \text{ or} \]
\[ \frac{C_x^2}{4} + \frac{C_{ym}^2}{C_m^2} > C_{yx} \]

Thus \( t \) are better than \( t_3 \) if above condition is met.

From equation (22) and (10), we observe,
\[ MSE(t_4) - MSE_{\text{min}}(t) > 0, \] if
\[ \frac{C_{ym}^2}{C_m^2} - C_y^2 \rho_{yx}^2 > 0 \]

Thus \( t \) are better than \( t_4 \) if above condition is met.

Proposed estimator is better than Reddy [7] and Kadilar [5] estimators \( t_5 \) and \( t_6 \) respectively of population mean under the same condition as for Srivastava [8] estimator given in above equation.

From equation (22) and (16), we have,
\[ MSE(t_7) - MSE_{\text{min}}(t) > 0, \] if
\[ R_7^2 C_m^2 - 2R_7 C_{ym} + \frac{C_{ym}^2}{C_m^2} > 0, \text{ or} \]
\[ R_7^2 C_m^2 + \frac{C_{ym}^2}{C_m^2} > 2R_7 C_{ym} \]

Thus \( t \) are better than \( t_7 \) if above condition is met.

5. Empirical Study

For the verification of the theoretical developments, we used two populations of Subramani [10]. Table-1 represents various parameters of these populations and Table-2 represents the MSEs of competing and suggested estimators respectively. Table-2 also depicts the percentage relative efficiency (PRE) of the suggested estimator over competing estimators given by,
\[ PRE(t) = \frac{MSE(t_i)}{MSE(t)} \times 100, \ i = 1, 2, ..., 7 \]
Table 1: Constants and Parameters of two populations

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Pop-1</th>
<th>Pop-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>34</td>
<td>20</td>
</tr>
<tr>
<td>$n$</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>$N C_n$</td>
<td>278256</td>
<td>15504</td>
</tr>
<tr>
<td>$\bar{y}$</td>
<td>856.4118</td>
<td>41.5</td>
</tr>
<tr>
<td>$\bar{M}$</td>
<td>736.9811</td>
<td>40.0552</td>
</tr>
<tr>
<td>$M$</td>
<td>767.5</td>
<td>40.5</td>
</tr>
<tr>
<td>$\bar{X}$</td>
<td>208.8824</td>
<td>441.95</td>
</tr>
<tr>
<td>$R_7$</td>
<td>1.1158</td>
<td>1.0247</td>
</tr>
<tr>
<td>$C_x^2$</td>
<td>0.125014</td>
<td>0.008338</td>
</tr>
<tr>
<td>$C_x^2$</td>
<td>0.088563</td>
<td>0.007845</td>
</tr>
<tr>
<td>$C_{sm}$</td>
<td>0.100833</td>
<td>0.006606</td>
</tr>
<tr>
<td>$C_{ym}$</td>
<td>0.07314</td>
<td>0.005394</td>
</tr>
<tr>
<td>$C_{yx}$</td>
<td>0.047257</td>
<td>0.005275</td>
</tr>
<tr>
<td>$\rho_{yx}$</td>
<td>0.4491</td>
<td>0.6522</td>
</tr>
</tbody>
</table>

Table 2: Mean squared error of various estimators and PRE of Proposed estimator over others

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Pop-1</th>
<th>PRE</th>
<th>Popln-2</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_0$</td>
<td>15640.97</td>
<td>173.76</td>
<td>2.15</td>
<td>228.72</td>
</tr>
<tr>
<td>$t_1$</td>
<td>12486.75</td>
<td>138.72</td>
<td>1.24</td>
<td>131.91</td>
</tr>
<tr>
<td>$t_2$</td>
<td>14895.27</td>
<td>165.48</td>
<td>1.48</td>
<td>157.45</td>
</tr>
<tr>
<td>$t_3$</td>
<td>12498.01</td>
<td>138.85</td>
<td>1.30</td>
<td>138.30</td>
</tr>
<tr>
<td>$t_4$</td>
<td>12486.75</td>
<td>138.72</td>
<td>1.24</td>
<td>131.91</td>
</tr>
<tr>
<td>$t_5$</td>
<td>12486.75</td>
<td>138.72</td>
<td>1.24</td>
<td>131.91</td>
</tr>
<tr>
<td>$t_6$</td>
<td>12486.75</td>
<td>138.72</td>
<td>1.24</td>
<td>131.91</td>
</tr>
<tr>
<td>$t_7$</td>
<td>10926.53</td>
<td>121.39</td>
<td>1.09</td>
<td>115.96</td>
</tr>
<tr>
<td>$t$</td>
<td>9001.36</td>
<td>100.00</td>
<td>0.94</td>
<td>100.00</td>
</tr>
</tbody>
</table>
6. Results and Discussion

In Table 2, it is observed that for both the populations, the MSEs of the competing estimators are in between $[10926.53, 15640.97]$ and $[1.09, 2.15]$ while that of the suggested estimator is 9001.36 and 0.94 respectively which are least MSE among all other estimators considered in completion in this manuscript, which was the aim of the present research that search for such estimator even biased but should have least MSE without enhancing the survey cost. The PRE of the suggested estimators for both the populations over the competing estimators are in between $[121.39, 173.76]$ and $[115.96, 228.72]$ respectively.

7. Conclusion

In the present study, we suggested a new ratio type estimator for enhancing the estimate of population mean using known value population median of main variable. We tried to develop the improved estimator without using auxiliary information, which is collected on some addition cost of the survey. Thus our aim was to find the improved estimator without increasing the cost of the survey. It is evident from Table 2 that the proposed estimator is best among other competing estimators of population mean. Almost all other estimators made use of supplementary information which increases the cost of the survey. Thus our aim of searching more improved estimator without increasing survey cost is achieved. Therefore it is recommended to use this estimator for efficient estimation of population mean of main variable without increasing the survey cost.

References