Estimation of Finite Population Mean of Median Based Using Power Transformation

J. O. MUILI* and A. AUDU²

¹Department of Mathematics, Kebbi State University of Science and Technology Aliero, Nigeria.
²Department of Mathematics, Usmanu Danfodiyo University, Sokoto, Nigeria.

Abstract
This paper deals with the assessment of finite population mean. An estimator is suggested for estimation of finite population mean of study variable. The purpose of this study is to evolve a ratio-type estimator to enhance the proficiency of the existing estimators considered in the study in sample random sampling without replacement using information of auxiliary variable. Expressions of the bias and mean square error (MSE) of the proposed estimator was derived by Taylor series method. The efficiency conditions under which the proposed ratio-type estimator is better than sample mean, ratio estimator, and other estimators considered in this study have been established. Theoretical and empirical findings are incentive and brace the robustness of the proposed estimator for mean estimation. The empirical results shown that the suggested estimator is more efficient than the sample mean, ratio estimator and other estimators.

Introduction
Median is one of the auxiliary variables aid in improving the precision of estimates of the finite population mean. Auxiliary variables associated with the study variables have been identified to be helpful in improving the efficiency of ratio, product and regression estimators both at planning and estimation stages. Cochran¹ invented the use of auxiliary information and developed a ratio estimator for population mean. Ratio type estimator provides effective estimate (minimum value of MSE) in comparison to simple mean estimator provided the variable of interest and auxiliary variable are positively associated. If the association between the study and auxiliary variables is positive, then ratio type estimator is applicable. Product estimator is useful where the association between the study variable and auxiliary variable is negative, and more efficient than sample mean. This concept has been utilized by several researchers in order to increase the precisions of ratio and product type estimators in estimating population mean of study variable.

Article History
Received: 6 October 2021
Accepted: 3 January 2022

Keywords
Auxiliary Variable; Efficiency; Median; Ratio Estimator; Study Variable.
variable using auxiliary information for assessments to maximize precisions. Bahl and Tuteja\textsuperscript{2} initiated exponential estimators with the used of exponential function in simple random sampling. Singh \textit{et al.}\textsuperscript{3} developed exponential ratio estimator with the used of known values of coefficient of variation, correlation coefficient and coefficient of kurtosis. Sanaullah \textit{et al.}\textsuperscript{4}, Riaz \textit{et al.}\textsuperscript{5}, Yadav and Adeware\textsuperscript{6}, Rashid \textit{et al.}\textsuperscript{7}, and Kadilar\textsuperscript{8} have developed different exponential estimators purposely to solve the problem of estimation of finite population mean. Few researchers have made use of median as the only auxiliary information in their works such as Subramani\textsuperscript{9} and Kumar \textit{et al.}\textsuperscript{10} Other researchers have also used linear combinations of median and other auxiliary parameters for the estimation of population mean such as Subramani and Kumarapandiyan\textsuperscript{11,12,13} Subramani and Kumarapandiyan\textsuperscript{14} Yadav \textit{et al.}\textsuperscript{15}, Subzar \textit{et al.}\textsuperscript{16}, Muili \textit{et al.}\textsuperscript{17}, and Muili \textit{et al.}\textsuperscript{18}

The purpose of this research is to evolve a ratio-type estimator to improve the precision of estimation of finite population mean in sample random sampling without replacement with the use of available known information of auxiliary variable.

Let a finite population $\Psi = \{\Psi_1, \Psi_2, \ldots, \Psi_N\}$ having $N$ units where each $\Psi = (X_i, Y_i), i=1,2,3,4,\ldots,N$ has a pair of values. $X$ is the auxiliary variable which $Y$ is the study variable and is correlated with $X$, where $y = \{y_1, y_2, \ldots, y_n\}$ and $x = \{x_1, x_2, \ldots, x_n\}$ are the $n$ sample values. $\bar{y}$ is the sample mean of the study variable and $\bar{x}$ is the sample mean auxiliary variable. Let $s_Y^2$ and $s_X^2$ be the population mean squares of $Y$ and $X$ respectively and $s_y^2$ be sample mean square of study variable and $s_x^2$ be sample mean squares based on the random sample of size $n$ drawn without replacement. $N$: Population size, $Y$: Study variable, $\bar{y}$: Sample mean of study variable and $\bar{x}$ sample mean of auxiliary variables, $f$: Sampling fraction, $p_y$: Coefficient of correlation between $X$ and $Y$, $C_y, C_x$: Coefficient of variations of study and auxiliary variables, $N$: Sample size, $M$: Median of the study variable, $\binom{n}{N}$: Number of possible samples of size $n$ from the population size $N$, $M$: Sample Median of the study variable, $\bar{X}$: auxiliary variable, $\bar{M}$: Average of Sample Median of the study variable,$\bar{Y}$, $\bar{X}$: Population means of study and auxiliary variables.

\begin{align*}
\hat{y} = & \frac{1}{n} \sum_{i=1}^{n} y_i \quad \text{(1)} \\
\bar{y} = & \frac{1}{N} \sum_{i=1}^{N} y_i \\
\bar{x} = & \frac{1}{n} \sum_{i=1}^{n} x_i \\
\bar{X} = & \frac{1}{N} \sum_{i=1}^{N} X_i \\
\bar{M} = & \frac{1}{N} \sum_{i=1}^{N} M_i \\
\end{align*}

\begin{align*}
\hat{X} = & \frac{1}{N} \sum_{i=1}^{N} X_i \\
\hat{Y} = & \frac{1}{n} \sum_{i=1}^{n} Y_i \\
\hat{M} = & \frac{1}{n} \sum_{i=1}^{n} M_i \\
\end{align*}

\begin{align*}
S^2_Y = & \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2 \\
S^2_X = & \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2 \\
S^2_y = & \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \\
S^2_x = & \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \\
\end{align*}

\begin{align*}
S^2_{YX} = & \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})(X_i - \bar{X}) \\
S^2_{xy} = & \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) \\
\end{align*}

\begin{align*}
\rho_Y = & \frac{S^2_{YX}}{\sqrt{S^2_Y \cdot S^2_X}} \\
\rho_y = & \frac{S^2_{xy}}{\sqrt{S^2_y \cdot S^2_x}} \\
\end{align*}

\begin{align*}
\text{Bias and mean square error of}\text{ (5)}\text{ are:} \\
\text{Bias}\left(\hat{Y}\right) = & \gamma \bar{Y}^2 \left(C_y^2 - C_x^2\right) \quad \text{(6)} \\
\text{MSE}\left(\hat{Y}\right) = & \gamma \bar{Y}^2 \left(C_y^2 + C_x^2 - 2C_y C_x\right) \quad \text{(7)} \\
\end{align*}

Bahl and Tuteja\textsuperscript{2} pioneer exponential estimator of population for estimation of finite population mean given as:

\begin{align*}
\hat{Y}_{\text{BR}} = & \bar{Y} \exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right) \quad \text{(8)} \\
\text{Bias and mean square error (MSE) of Bahl and Tuteja}\textsuperscript{2} \text{are given in (9) and (10) respectively} \\
\text{Bias}\left(\hat{Y}_{\text{BR}}\right) = & \gamma \bar{Y}^2 \left(\frac{3}{8} C_y^2 - \frac{1}{2} C_x^2\right) \quad \text{(9)} \\
\text{MSE}\left(\hat{Y}_{\text{BR}}\right) = & \gamma \bar{Y}^2 \left(\frac{3}{4} C_y^2 - C_x^2\right) \quad \text{(10)}
\end{align*}
Subramani\(^9\) developed a ratio-type estimator for estimation of finite population mean given by:
\[
\hat{\bar{y}}_r = \bar{y} \left( \frac{M}{m} \right)^n \quad \ldots(11)
\]

**Bias** of \(\hat{\bar{y}}_r\) is given by:
\[
Bias(\hat{\bar{y}}_r) = yF^2 \left( \frac{C_m^2 - RC_m}{M} \right) \quad \ldots(12)
\]

**MSE** of \(\hat{\bar{y}}_r\) is given by:
\[
MSE(\hat{\bar{y}}_r) = yF^2 \left( C_m^2 + RC_m - 2RC_m \right) \quad \ldots(13)
\]

where \(R = \frac{\bar{y}}{\overline{M}}\).

**Proposed Estimator**

Motivated by Subramani\(^9\) of finite population mean based on the information on auxiliary variable for estimation of population mean of study variable is proposed by:
\[
\hat{\bar{y}}_p = \bar{y} \log \left( \frac{M}{m} \right)^n \quad \ldots(14)
\]

**Derivation of Bias and Mean Square Error of the Proposed Estimator** \(\hat{\bar{y}}_p\)

Note that: \(e_i = \bar{y} - \overline{M}\) and \(s_i = \frac{M - M}{M}\) such that
\[
\hat{e}_i = \bar{y} \left( 1 + e_i \right), \quad m = M \left( 1 + e_i \right),
\]

\[
E(e_i) = 0, \quad E(s_i) = \frac{M - M}{M} = \frac{Bias(m)}{M},
\]

\[
E(e_i^2) = yC_m, \quad E(s_i^2) = yC_m, \quad E(e_i) = yC_m
\]

Expressing (14) in terms of \(\hat{e}_i\) and \(C_m\), we have
\[
\hat{\bar{y}}_p = \bar{y} \left( 1 + e_i \right) \left( 1 + e_i \right)^n \quad \ldots(16)
\]
\[
\hat{\bar{y}}_p = \bar{y} \left( 1 + e_i \right) \left[ 1 - \frac{\alpha(\alpha + 1)}{2} e_i^2 \right] \quad \ldots(17)
\]

Simplifying (17) gives (18) as:
\[
\hat{\bar{y}}_p = \bar{y} \left( 1 + e_i - \alpha e_i - \frac{\alpha(\alpha + 1)}{2} e_i^2 \right) \quad \ldots(18)
\]

Subtracting \(\hat{y}\) from and taking expectation of both sides
\[
E(\hat{\bar{y}}_p - \bar{y}) = \bar{y}E(e_i - \alpha e_i - \frac{\alpha(\alpha + 1)}{2} e_i^2) \quad \ldots(19)
\]

Applying the results of (15) in (19), gives the bias as:
\[
Bias(\hat{\bar{y}}_p) = \bar{y} \left( \frac{\alpha(\alpha + 1)}{2} + \frac{\alpha(\alpha + 1)}{2} e_i^2 \right) \quad \ldots(20)
\]

Squaring and taking expectation of (19) as
\[
MSE(\hat{\bar{y}}_p) = \bar{y}^2 E(e_i - \alpha e_i)^2 \quad \ldots(21)
\]

Expanding (21)
\[
MSE(\hat{\bar{y}}_p) = \bar{y}^2 E(e_i^2 + \alpha^2 e_i^2 - 2\alpha e_i e_i) \quad \ldots(22)
\]

Applying the results of (15) to (22), gives
\[
MSE(\hat{\bar{y}}_p) = \bar{y}F^2 \left( C_m^2 + \alpha^2 C_m^2 - 2\alpha C_m \right) \quad \ldots(23)
\]

Differentiation of (23) w.r.t \(\alpha\) as:
\[
\frac{\partial \left( MSE(\hat{\bar{y}}_p) \right)}{\partial \alpha} = \bar{y}F^2 \left( C_m^2 + \alpha^2 C_m^2 - 2\alpha C_m \right) = 0 \quad \ldots(24)
\]

Making \(\alpha\) the subject of the formula, gives
\[
\alpha = \frac{C_m}{C_m^2} \quad \ldots(25)
\]

Substituting (25) in (23), gives
\[
MSE(\hat{\bar{y}}_p) = \bar{y}F^2 \left( C_m^2 - \frac{C_m^2}{C_m^2} \right) \quad \ldots(26)
\]

**EFFICIENCY COMPARISON**

Efficiency of the suggested estimator is compared with efficiencies of the existing estimators in the study

Proposed estimator \(\hat{\bar{y}}_p\) of the finite population mean is better than \(\hat{\bar{y}}\) if,
\[
MSE(\hat{\bar{y}}_p) < MSE(\hat{\bar{y}}) \quad \ldots(27)
\]

\[
\bar{y}F^2 \left( C_m^2 - \frac{C_m^2}{C_m^2} \right) < \bar{y}F^2 C_m^2 \quad \ldots(28)
\]

\(C_m^2 > 0\)

Proposed estimator \(\hat{\bar{y}}_p\) is better than Linear Regression\(^9\) estimator \(\hat{\bar{y}}\) if,
\[
MSE(\hat{\bar{y}}_p) < MSE(\hat{\bar{y}}) \quad \ldots(29)
\]

\[
\bar{y}F^2 \left( C_m^2 - \frac{C_m^2}{C_m^2} \right) - \bar{y}F^2 C_m^2 < 0 \quad \ldots(30)
\]

Proposed estimator \(\hat{\bar{y}}_p\) is better than Cochran\(^1\) Ratio estimator \(\hat{\bar{y}}\) if,
\[
MSE(\hat{\bar{y}}_p) < MSE(\hat{\bar{y}}) \quad \ldots(31)
\]

\[
\bar{y}F^2 \left( C_m^2 - \frac{C_m^2}{C_m^2} \right) < \bar{y}F^2 C_m^2 + C_m^2 - 2C_m \quad \ldots(32)
\]
Proposed estimator \( \hat{y}_{p} \) is better than Bahl and Tuteja\(^2\) estimator \( \hat{y}_{pt} \) if,
\[
\text{MSE}(\hat{y}_{p}) < \text{MSE}(\hat{y}_{pt})
\]
\[
\frac{C_{y}^2}{C_{n}^2} + \frac{1}{4} < \frac{C_{y}^2}{C_{n}^2} + \frac{1}{4} \quad \text{... (33)}
\]
\[
C_{n}^2 > C_{y}^2 \left( C_{n}^2 - \frac{1}{4} C_{y}^2 \right) \quad \text{... (34)}
\]
Proposed estimator \( \hat{y}_{p} \) is better than Subramani\(^9\) estimator \( \text{MSE}(\hat{y}_{p}) < \text{MSE}(\hat{y}_{q}) \)
\[
\frac{C_{y}^2}{C_{n}^2} \frac{1}{4} < \frac{C_{y}^2}{C_{n}^2} \frac{1}{4} \quad \text{... (35)}
\]

When conditions (28), (30), (32), (34), and (36) are satisfied, conclude that the proposed estimator is better (efficient) than sample mean, linear regression,\(^{19}\) Cochran,\(^1\) Bahl and Tuteja,\(^2\) and Subramani\(^9\) estimators considered in the study.

**VERIFIABLE STUDY**

To assess the performance of the suggested estimator, a natural population is used as: Source: Subramani\(^9\)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Population 1</th>
<th>Population 2</th>
<th>Population 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>34</td>
<td>34</td>
<td>20</td>
</tr>
<tr>
<td>( n )</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>( nC_{n} )</td>
<td>278256.0</td>
<td>278256.0</td>
<td>15504.0</td>
</tr>
<tr>
<td>( \hat{Y} )</td>
<td>856.412</td>
<td>856.412</td>
<td>41.50</td>
</tr>
<tr>
<td>( \bar{M} )</td>
<td>736.981</td>
<td>736.981</td>
<td>40.055</td>
</tr>
<tr>
<td>( M )</td>
<td>767.50</td>
<td>767.50</td>
<td>40.50</td>
</tr>
<tr>
<td>( \bar{X} )</td>
<td>208.882</td>
<td>199.441</td>
<td>441.950</td>
</tr>
<tr>
<td>( R )</td>
<td>1.1158</td>
<td>1.1158</td>
<td>1.0247</td>
</tr>
<tr>
<td>( C_{y} )</td>
<td>0.12501</td>
<td>0.12501</td>
<td>0.00834</td>
</tr>
<tr>
<td>( C_{x} )</td>
<td>0.08856</td>
<td>0.09677</td>
<td>0.00785</td>
</tr>
<tr>
<td>( C_{y} )</td>
<td>0.10083</td>
<td>0.10083</td>
<td>0.00661</td>
</tr>
<tr>
<td>( C_{w} )</td>
<td>0.073140</td>
<td>0.073140</td>
<td>0.0053940</td>
</tr>
<tr>
<td>( P_{xy} )</td>
<td>0.04726</td>
<td>0.04898</td>
<td>0.00528</td>
</tr>
<tr>
<td>( P_{xy} )</td>
<td>0.449</td>
<td>0.445</td>
<td>0.652</td>
</tr>
</tbody>
</table>

Table 1 shows the values of the populations’ parameters

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Population 1</th>
<th>Population 2</th>
<th>Population 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Mean ((\bar{y}))</td>
<td>15641.31</td>
<td>15641.31</td>
<td>2.154018</td>
</tr>
<tr>
<td>Linear Regression(^{19}) [( \hat{y} )]</td>
<td>12486.6</td>
<td>12539.76</td>
<td>1.237775</td>
</tr>
<tr>
<td>Ratio Estimator(^{19}) [( \hat{y} )]</td>
<td>14896.74</td>
<td>15492.29</td>
<td>1.455215</td>
</tr>
<tr>
<td>Bahl and Tuteja(^2) [( \hat{y} )]</td>
<td>12498.85</td>
<td>12539.89</td>
<td>1.297952</td>
</tr>
<tr>
<td>Subramani(^9) [( \hat{y} )]</td>
<td>10926.77</td>
<td>10926.77</td>
<td>1.090159</td>
</tr>
<tr>
<td>Proposed Estimator ([\hat{y}_p])</td>
<td>9003.55</td>
<td>9003.55</td>
<td>1.016205</td>
</tr>
</tbody>
</table>

Table 2 shows MSE of the estimators using the three set of populations. The result revealed that the suggested estimator has minimum mean square error compared to the conventional estimators.
This implies that the proposed estimator is better and can produce better estimates of population mean than the conventional estimators, Bahl and Tuteja$^2$ and Subramani.$^9$

Table 3: Percentage Relative Efficiency of the Estimators

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Population 1</th>
<th>Population 2</th>
<th>Population 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Mean ($\bar{x}$)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Linear Regression$^{19}$ ($\hat{x}$)</td>
<td>125.2648</td>
<td>124.7337</td>
<td>174.0234</td>
</tr>
<tr>
<td>Ratio Estimator$^{2}$ ($\bar{x}_r$)</td>
<td>104.9982</td>
<td>100.9619</td>
<td>148.0206</td>
</tr>
<tr>
<td>Bahl and Tuteja$^2$ ($\hat{x}_r$)</td>
<td>125.142</td>
<td>124.7324</td>
<td>165.9551</td>
</tr>
<tr>
<td>Subramani$^9$ ($\hat{x}_r$)</td>
<td>143.1467</td>
<td>143.1467</td>
<td>197.5875</td>
</tr>
<tr>
<td>Proposed Estimator ($\hat{x}_r$)</td>
<td>173.7238</td>
<td>173.7238</td>
<td>211.9669</td>
</tr>
</tbody>
</table>

Table 3 shows PRE of the proposed and some existing estimators using the three set of populations. The result revealed that the suggested estimator has the highest value of PRE compared to the conventional estimators, Bahl and Tuteja$^2$ and Subramani.$^9$ This implies that the suggested estimator is more efficient and can produce better estimates of population mean than the conventional estimators, Bahl and Tuteja$^2$ and Subramani.$^9$

Results and Discussion

Proficient ratio estimator of finite population mean is suggested. The properties of the suggested estimator were obtained. Table 2 shows MSE of the suggested and some existing estimators using the three set of populations. The result revealed that the suggested estimator has minimum MSE compared to the conventional estimators, Bahl and Tuteja$^2$ and Subramani.$^9$ Estimators. Table 3 shows PRE of the estimators using the three set of populations. The result revealed that the suggested estimator has highest PRE compared to the conventional estimators. This implies that the suggested estimator is more efficient and can produce better estimates of population mean than the conventional estimators, Bahl and Tuteja$^2$ and Subramani.$^9$

Conclusion

Based on the empirical study conducted on the efficiency comparison of the suggested estimator with related estimators, it is obtained that the suggested estimator is highly efficient and can produce better estimates of finite population mean than the conventional and existing estimators considered in the study. Future scope of this study can be study under different sampling schemes like stratified sampling, successive sampling or cluster sampling.

Funding

The authors received no financial support for the research, authorship and publication of this article.

Conflict of Interest

The authors declare no conflict of interest.

References


