Ratio Estimators in Simple Random Sampling
Using Information on Auxiliary Attribute

Rajesh Singh
Department of Statistics, Banaras Hindu University
Varanasi (U.P.), India
Email: rsinghstat@yahoo.co.in

Pankaj Chauhan
School of Statistics, DAVV, Indore (M.P.), India

Nirmala Sawan
School of Statistics, DAVV, Indore (M.P.), India

Florentin Smarandache
Department of Mathematics, University of New Mexico, Gallup, USA
Email: smarand@unm.edu

Abstract
Some ratio estimators for estimating the population mean of the variable under study, which
make use of information regarding the population proportion possessing certain attribute, are
proposed. Under simple random sampling without replacement (SRSWOR) scheme, the
expressions of bias and mean-squared error (MSE) up to the first order of approximation are
derived. The results obtained have been illustrated numerically by taking some empirical
population considered in the literature.

Key words: Proportion, bias, MSE, ratio estimator.

1. Introduction
The use of auxiliary information can increase the precision of an estimator when
study variable $y$ is highly correlated with auxiliary variable $x$. There exist
situations when information is available in the form of attribute $\phi$, which is highly
correlated with $y$. For example
a) Sex and height of the persons,
b) Amount of milk produced and a particular breed of the cow,
c) Amount of yield of wheat crop and a particular variety of wheat etc.
(see Jhajj et. al. [1]).

Consider a sample of size $n$ drawn by SRSWOR from a population of size $N$. Let
$y_i$ and $\phi_i$ denote the observations on variable $y$ and $\phi$ respectively for $i^{th}$
unit ($i = 1, 2, \ldots, N$). Suppose there is a complete dichotomy in the population with
respect to the presence or absence of an attribute, say $\phi$, and it is assumed that
attribute $\phi$ takes only the two values 0 and 1 according as
\[
\phi_i = 1, \text{ if } i^{th} \text{ unit of the population possesses attribute } \phi \\
\phi_i = 0, \text{ otherwise.}
\]
Let \( A = \sum_{i=1}^{N} \phi_i \) and \( a = \sum_{i=1}^{n} \phi_i \) denote the total number of units in the population and sample respectively possessing attribute \( \phi \). Let \( P = \frac{A}{N} \) and \( p = \frac{a}{n} \) denote the proportion of units in the population and sample respectively possessing attribute \( \phi \).

Taking into consideration the point biserial correlation between a variable and an attribute, Naik and Gupta [2] defined ratio estimator of population mean when the prior information of population proportion of units, possessing the same attribute is available, as follows:

\[
 t_{NG} = \bar{y} \left( \frac{p}{P} \right)
\]  

(1.1)

Here \( \bar{y} \) is the sample mean of variable of interest. The MSE of \( t_{NG} \) up to the first order of approximation is –

\[
 \text{MSE}(t_{NG}) = \left( 1 - \frac{f}{n} \right) \left[ S_y^2 + R_1 S_\phi^2 - 2R_1 S_{y\phi} \right]
\]  

(1.2)

Where \( f = \frac{n}{N} \), \( R_1 = \frac{\bar{Y}}{P} \), \( S_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2 \), \( S_\phi^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\phi_i - P)^2 \), \( S_{y\phi} = \frac{1}{N-1} \sum_{i=1}^{N} (\phi_i - P)(y_i - \bar{Y}) \).

In the present paper, some ratio estimators for estimating the population mean of the variable under study, which make use of information regarding the population proportion possessing certain attribute, are proposed. The expressions of bias and MSE have been obtained. The numerical illustrations have also been done by taking some empirical populations considered in the literature.

2. The suggested estimator

Following Ray and Singh [3], we propose the following estimator –

\[
 t_1 = \frac{\bar{y} + b_\phi (P - p)}{p} = R^* P
\]  

(2.1)

Where \( b_\phi = \frac{s_{y\phi}}{s_\phi} \), \( R^* = \frac{\bar{Y} + b_\phi (P - p)}{P} \), \( s_\phi^2 = \left( \frac{1}{n-1} \right) \sum_{i=1}^{n} (\phi_i - P)^2 \) and \( s_{y\phi} = \left( \frac{1}{n-1} \right) \sum_{i=1}^{n} (\phi_i - P)(y_i - \bar{Y}) \).

**Remark 1**: When we put \( b_\phi = 0 \) in (2.1) the proposed estimator turns to the Naik and Gupta [2] ratio estimator \( t_{NG} \) given in (1.1).
MSE of this estimator can be found by using Taylor series expansion given by –

\[
f(p, \bar{y}) \cong f(P, \bar{y}) + \frac{\partial f(c, d)}{\partial c} \bigg|_{P, \bar{y}} (p - P) + \frac{\partial f(c, d)}{\partial \bar{y}} \bigg|_{P, \bar{y}} (\bar{y} - \bar{Y})
\]  

(2.2)

Where \( f(p, \bar{y}) = R^* \) and \( f(P, \bar{Y}) = R_1 \).

Expression (2.2) can be applied to the proposed estimator in order to obtain MSE equation as follows:

\[
R^* - R_1 \cong \frac{\partial}{\partial p} \left( \frac{(\bar{y} + b_\psi (P - p))}{p^3} \right) \bigg|_{P, \bar{y}} (p - P) + \frac{1}{p} \bigg|_{P, \bar{y}} (\bar{y} - \bar{Y})
\]

\[
E(R^* - R_1)^2 \cong \frac{(Y + B_\psi P)^2}{p^4} V(p) - \frac{2(Y + B_\psi P)}{p^3} \text{Cov}(p, \bar{y}) + \frac{1}{p^2} V(\bar{y})
\]

(2.3)

Where \( B_\psi = \frac{S_{\psi y}}{S_\psi^2} = \frac{\rho_{pb} S_{\gamma}}{S_\psi^2} \).

\[\rho_{pb} = \frac{S_{\psi \gamma}}{S_\psi S_\gamma}, \text{ is the point biserial correlation coefficient.}\]

Now,

\[
\text{MSE}(t_1) = p^2 E(R_1 - R_\psi)^2
\]

\[
\cong \frac{(Y + B_\psi P)^2}{p^4} V(p) - \frac{2(Y + B_\psi P)}{p^3} \text{Cov}(p, \bar{y}) + V(\bar{y})
\]

(2.4)

Simplifying (2.4), we get MSE of \( t_1 \) as

\[
\text{MSE}(t_1) \cong \left(1 - \frac{f}{n}\right) \left[R_1^2 S_\psi^2 + S_\gamma^2 (1 - \rho_{pb}^2)\right]
\]

(2.5)

Several authors have used prior value of certain population parameters \( s \) to find more precise estimates. Searls (1964) used Coefficient of Variation (CV) of study character at estimation stage. In practice this CV is seldom known. Motivated by Searls (1964) work, Sen (1978), Sisodiya and Dwivedi (1981), and Upadhyaya and Singh (1984) used the known CV of the auxiliary character for estimating population mean of a study character in ratio method of estimation. The use of prior value of Coefficient of Kurtosis in estimating the population variance of study character \( y \) was first made by Singh et. al. (1973). Later, used by and Searls and Intarapanich (1990), Upadhyaya and Singh (1999), Singh (2003) and Singh et. al. (2004) in the estimation of population mean of study character. Recently Singh and Tailor (2003) proposed a modified ratio estimator by using the known value of correlation coefficient.
In next section, we propose some ratio estimators for estimating the population mean of the variable under study using known parameters of the attribute \( \phi \) such as coefficient of variation \( C_p \), Kurtosis \( (\beta_2(\phi)) \) and point biserial correlation coefficient \( \rho_{pb} \).

3. Suggested Estimators

We suggest following estimator –

\[
 t = \frac{\bar{y} + b_\phi(P - p)}{(m_1 P + m_2)} \tag{3.1}
\]

Where \( m_1 (\neq 0), m_2 \) are either real number or the functions of the known parameters of the attribute such as \( C_p, (\beta_2(\phi)) \) and \( \rho_{pb} \).

The following scheme presents some of the important estimators of the population mean, which can be obtained by suitable choice of constants \( m_1 \) and \( m_2 \):

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Values of ( m_1 )</th>
<th>Values of ( m_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_1 = \frac{\bar{y} + b_\phi(P - p)}{p} )</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( t_2 = \frac{\bar{y} + b_\phi(P - p)}{(p + \beta_2(\phi))}\left[P + \beta_2(\phi)\right] )</td>
<td>1</td>
<td>( \beta_2(\phi) )</td>
</tr>
<tr>
<td>( t_3 = \frac{\bar{y} + b_\phi(P - p)}{(p + C_p)}\left[P + C_p\right] )</td>
<td>1</td>
<td>( C_p )</td>
</tr>
<tr>
<td>( t_4 = \frac{\bar{y} + b_\phi(P - p)}{(p + \rho_{pb})}\left[P + \rho_{pb}\right] )</td>
<td>1</td>
<td>( \rho_{pb} )</td>
</tr>
<tr>
<td>( t_5 = \frac{\bar{y} + b_\phi(P - p)}{(p\beta_2(\phi) + C_p)}\left[P\beta_2(\phi) + C_p\right] )</td>
<td>( \beta_2(\phi) )</td>
<td>( C_p )</td>
</tr>
<tr>
<td>( t_6 = \frac{\bar{y} + b_\phi(P - p)}{(pC_p + \beta_2(\phi))}\left[pC_p + \beta_2(\phi)\right] )</td>
<td>( C_p )</td>
<td>( \beta_2(\phi) )</td>
</tr>
<tr>
<td>( t_7 = \frac{\bar{y} + b_\phi(P - p)}{(pC_p + \rho_{pb})}\left[pC_p + \rho_{pb}\right] )</td>
<td>( C_p )</td>
<td>( \rho_{pb} )</td>
</tr>
<tr>
<td>( t_8 = \frac{\bar{y} + b_\phi(P - p)}{(p \rho_{pb} + C_p)}\left[p \rho_{pb} + C_p\right] )</td>
<td>( \rho_{pb} )</td>
<td>( C_p )</td>
</tr>
<tr>
<td>( t_9 = \frac{\bar{y} + b_\phi(P - p)}{(p \beta_2(\phi) + \rho_{pb})}\left[p \beta_2(\phi) + \rho_{pb}\right] )</td>
<td>( \beta_2(\phi) )</td>
<td>( \rho_{pb} )</td>
</tr>
<tr>
<td>( t_{10} = \frac{\bar{y} + b_\phi(P - p)}{(p \rho_{pb} + \beta_2(\phi))}\left[p \rho_{pb} + \beta_2(\phi)\right] )</td>
<td>( \rho_{pb} )</td>
<td>( \beta_2(\phi) )</td>
</tr>
</tbody>
</table>
Following the approach of section 2, we obtain the MSE expression for these proposed estimators as –

\[
\text{MSE}(t_i) \geq \left( \frac{1-f}{n} \right) \left[ R_i S_y^2 + S_y^2 (1 - \rho_{pb}^2) \right], \quad (i = 1, 2, 3, \ldots, 10) \tag{3.2}
\]

Where

\[
R_1 = \frac{\overline{Y}}{P}, \quad R_2 = \frac{\overline{Y}}{P + \beta_2(\phi)}, \quad R_3 = \frac{\overline{Y}}{P + C_p}, \quad R_4 = \frac{\overline{Y}}{P + \rho_{pb}},
\]

\[
R_5 = \frac{\overline{Y} \beta_2(\phi)}{P \beta_2(\phi) + C_p}, \quad R_6 = \frac{\overline{Y} C_p}{P C_p + \beta_2(\phi)}, \quad R_7 = \frac{\overline{Y} C_p}{P C_p + \rho_{pb}}, \quad R_8 = \frac{\overline{Y} \rho_{pb}}{P \rho_{pb} + C_p},
\]

\[
R_9 = \frac{\overline{Y} \beta_2(\phi)}{P \beta_2(\phi) + \rho_{pb}}, \quad R_{10} = \frac{\overline{Y} \rho_{pb}}{P \rho_{pb} + \beta_2(\phi)}.
\]

4. Efficiency comparisons

It is well known that under simple random sampling without replacement (SRSWOR) the variance of the sample mean is

\[
V(\overline{y}) = \left( \frac{1-f}{n} \right) S_y^2 \tag{4.1}
\]

From (4.1) and (3.2), we have

\[
V(\overline{y}) - \text{MSE}(t_i) \geq 0, \quad (i = 1, 2, \ldots, 10)
\]

\[
\Rightarrow \rho_{pb}^2 > \frac{S_y^2}{R_i^2} \tag{4.2}
\]

When this condition is satisfied, proposed estimators are more efficient than the sample mean.

Now, we compare the MSE of the proposed estimators with the MSE of Naik and Gupta (1996) estimator \( t_{NG} \). From (3.2) and (1.1) we have

\[
\text{MSE}(t_{NG}) - \text{MSE}(t_i) \geq 0, \quad (i = 1, 2, \ldots, 10)
\]

\[
\Rightarrow \rho_{pb}^2 \geq \frac{S_y^2}{S_y^2} \left[ R_i^2 - R_6^2 + 2R_6 K_{yp} \right] \tag{4.3}
\]

where

\[
K_{yp} = \rho_{yp} \frac{C_y}{C_p}.
\]
5. Empirical Study

The data for the empirical study is taken from natural population data set considered by Sukhatme and Sukhatme (1970):

\[ y \] = Number of villages in the circles and
\[ \phi \] = A circle consisting more than five villages

\[ N = 89, \bar{Y} = 3.36, \ p = 0.1236, \ \rho_{pb} = 0.766, \ C_y = 0.604, \ C_p = 2.19, \ \beta_2(\phi) = 6.23181. \]

In table 5.1 percent relative efficiencies (PRE) of various estimators are computed with respect to \( \bar{y} \).

Table 5.1: PRE of different estimators of \( \bar{Y} \) with respect to \( \bar{y} \)

<table>
<thead>
<tr>
<th>Estimator</th>
<th>PRE (. ( \bar{y} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{y} )</td>
<td>100</td>
</tr>
<tr>
<td>( t_{NG} )</td>
<td>11.61</td>
</tr>
<tr>
<td>( t_1 )</td>
<td>7.36</td>
</tr>
<tr>
<td>( t_2 )</td>
<td>236.55</td>
</tr>
<tr>
<td>( t_3 )</td>
<td>227.69</td>
</tr>
<tr>
<td>( t_4 )</td>
<td>208.09</td>
</tr>
<tr>
<td>( t_5 )</td>
<td>185.42</td>
</tr>
<tr>
<td>( t_6 )</td>
<td>230.72</td>
</tr>
<tr>
<td>( t_7 )</td>
<td>185.27</td>
</tr>
<tr>
<td>( t_8 )</td>
<td>230.77</td>
</tr>
<tr>
<td>( t_9 )</td>
<td>152.37</td>
</tr>
<tr>
<td>( t_{10} )</td>
<td>237.81</td>
</tr>
</tbody>
</table>

From table 5.1, we observe that the proposed estimators \( t_i (i = 2, \ldots, 10) \) which uses some known values of population proportion performs better than the usual sample mean \( \bar{y} \) and Naik and Gupta (1996) estimator \( t_{NG} \).

Conclusion

We have suggested some ratio estimators for estimating \( \bar{Y} \) which uses some known value of population proportion. For practical purposes the choice of the estimator depends upon the availability of the population parameters.

References


