Ratio Estimators in Simple Random Sampling When Study Variable Is an Attribute

Rajesh Singh, Mukesh Kumar and Florentin Smarandache

Department of Statistics, Banaras Hindu University, Varanasi-221005, India
Department of Mathematics, Chair of University of New Mexico, Gallup, USA

Abstract: In this paper we have suggested a family of estimators for the population mean when study variable itself is qualitative in nature. Expressions for the bias and mean square error (MSE) of the suggested family have been obtained. An empirical study has been carried out to show the superiority of the constructed estimator over others.

Key words: Attribute • Point bi-serial • Mean square error • Simple random sampling

INTRODUCTION

The use of auxiliary information can increase the precision of an estimator when study variable y is highly correlated with auxiliary variable x. In many situations study variable is generally ignored not only by ratio scale variables that are essentially qualitative, or nominal scale, in nature, such as sex, race, colour, religion, nationality, geographical region, political upheavals (see [1]). Taking into consideration the point biserial correlation coefficient between auxiliary attribute and study variable, several authors including [2-6] defined ratio estimators of population mean when the priori information of population proportion of units, possessing some attribute is available. All the others have implicitly assumed that the study variable Y is quantitative whereas the auxiliary variable is qualitative.

In this paper we consider some estimators in which study variable itself is qualitative in nature. For example suppose we want to study the labour force participation (LFP) decision of adult males. Since an adult is either in the labour force or not, LFP is a yes or no decision. Hence, the study variable can take two values, say 1, if the person is in the labour force and 0 if he is not. Labour economics research suggests that the LFP decision is a function of the unemployment rate, average wage rate, education, family income, etc (See [1]).

Consider a sample of size n drawn by simple random sampling without replacement (SRSWOR) from a population size N. Let \( \phi_i \) and \( x_i \) denote the observations on variable \( \phi \) and \( x \) respectively for \( i = 1, 2, 3 \ldots N \). \( \phi_i \) if \( i \)th unit of population possesses attribute \( \phi \) and \( x_i \), 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 possessing attribute \( \phi \) respectively, \( p = \frac{A}{N} \) and \( p = \frac{a}{n} \) denote the proportion of units in the population and sample, respectively, possessing attribute \( \phi \).

Define,

\[
e_t = \frac{(p-P)}{P}, \quad e_x = \frac{(\bar{x} - \bar{X})}{\bar{X}}
\]

Such that,

\[
E(e_t) = 0, \quad (1 = \phi, x)
\]

and

\[
E(e_t^2) = \alpha C_p^2, \quad E(e_x^2) = \alpha C_x^2, \quad E(e_t e_x) = \alpha \rho_{px} C_p C_x
\]

Where,

\[
f = \left( \frac{1}{n}, \frac{1}{N} \right), \quad C_p = \frac{S_p^2}{p^2}, \quad C_x = \frac{S_x^2}{\bar{x}^2},
\]

and \( \rho_{px} = \frac{S_{px}}{S_p S_x} \) is the point biserial correlation coefficient.

Here,

\[
S_p^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2, \quad S_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x} - \bar{x})^2 \quad \text{and} \quad S_{px} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(p - \bar{p})
\]

Corresponding Author: Rajesh Singh, Department of Statistics, Banaras Hindu University, Varanasi-221005, India.
E-mail: rainghstat@yahoo.com, mukesh.stat@gmail.com, smarand@unm.edu.
The Proposed Estimator: We first propose the following ratio-type estimator

\[ t_3 = \left( \frac{p}{x} \right) \bar{x} \]  \hspace{1cm} (2.1)

The bias and MSE of the estimator \( t_3 \), to the first order of approximation is respectively, given by

\[ B(t_3) = -\frac{C_x^2}{2} \rho \rho_b C_f C_x \] \hspace{1cm} (2.2)

\[ \text{MSE}(t_3) = f \left( C_x^2 + C_x^2 \right) \] \hspace{1cm} (2.3)

Following [7], we propose a general family of estimators for \( P \) as

\[ t_3 = H(p,u) \] \hspace{1cm} (2.4)

Where \( u = \frac{x}{X} \) and \( H(p,u) \) is a parametric equation of \( p \) and \( u \) such that

\[ H(p,1) = P, \forall P \] \hspace{1cm} (2.5)

and satisfying following regulations:

- Whatever be the sample chosen, the point \( (p,u) \) assume values in a bounded closed convex subset \( \mathcal{R}_c \) of the two-dimensional real space containing the point \( (p,1) \).
- The function \( H(p,u) \) is a continuous and bounded in \( \mathcal{R}_c \).
- The first and second order partial derivatives of \( H(p,u) \) exist and are continuous as well as bounded in \( \mathcal{R}_c \).

Expanding \( H(p,u) \) about the point \( (P,1) \) in a second order Taylor series we have

\[ t_3 = H(p,u) \]
\[ - p + (u-1)H_i + (u-1)H_1 + (p-1)H_2 + (p-1)^2H_3 + ... \] \hspace{1cm} (2.6)

Where,

\[ H_1 = \frac{\partial^2 H}{\partial u \partial p} |_{p=P,u=1} \]
\[ H_2 = \frac{\partial^2 H}{\partial u^2} |_{p=P,u=1} \]

The bias and MSE of the estimator \( t_3 \) are respectively given by -

\[ B(t_3) = f \left( P \rho \rho_b C_f C_x \right) \] \hspace{1cm} (2.7)

\[ \text{MSE}(t_3) = f \left( P^2 C_f^2 + H_1^2 \right) \] \hspace{1cm} (2.8)

On differentiating (2.8) with respect to \( H_i \) and equating to zero we obtain

\[ H_1 = \rho \rho_b \frac{C_f}{C_x} \] \hspace{1cm} (2.9)

On substituting (2.9) in (2.8), we obtain the minimum MSE of the estimator \( t_3 \) as

\[ \text{minMSE}(t_2) = P^2 C_f^2 \left( 1-\rho_{2,1}^2 \right) \] \hspace{1cm} (2.10)

We suggest another family of estimators for estimating \( P \) as

\[ t_3 = q_1 + q_2 \left( \bar{X} - \bar{X} \right) \left[ \frac{aX + b}{\bar{X}} \right] \exp \left[ \frac{[aX + b] - (aX + b)}{[aX + b] + (aX + b)} \right] \] \hspace{1cm} (2.11)

Where \( a, b, q_1, q_2, \alpha, \beta, \) and \( \alpha \) are real constants and \( a \) and \( b \) are known as characterising positive scalars. Many ratio-product estimators can be generated from \( t_3 \) by putting suitable values of \( q_1, q_2, \alpha, \beta \), \( a \) and \( b \) (for choice of the parameters refer to [8] and [5]).

\[ t_3 = \left[ q_2 P (1+e_0) - q_2 \bar{x} \right] \left[ 1 - a \theta_0 + \frac{\alpha (\alpha + 1)}{2} \theta^2 \right] \]
\[ \left[ 1 - \frac{B \theta_k}{2} - \frac{B \theta_k^2}{8} \right] \] \hspace{1cm} (2.12)

Where, \( \theta = \frac{\bar{X}}{aX + b} \), \( B = \left( a + \frac{\beta}{2} \right) \theta \) and

\[ A = \frac{B^2}{8} \left[ 4a (a + 1) + B (\beta + 2) + 4a \beta \right] \]

The bias and MSE of the estimator \( t_3 \) to the first order of approximation, are given as

\[ \text{Bias}(t_3) = P(\bar{x} - \bar{x}) \left[ q_2 X B + q_2 P A C_f^2 \right] \] \hspace{1cm} (2.13)
MSE(t_3) = E(t_3 - P)^2
= (q_1 - 1)^2 P^2 + q_1^2 (M_1 + 2M_3) + q_2^2 M_2
+ 2q_1q_2 (M_4 - M_5) - 2q_1M_3 + 2q_2M_5

Where,

M_1 = P^2 f \left( C_p^2 + B_p^2 \right),
M_2 = P^2 f \left( C_x^2 \right),
M_3 = P^2 f \left( A_x^2 + 2B_p C_p C_x \right),
M_4 = P^2 f \left( -B C_x^2 + r C_p C_x \right),
M_5 = P^2 f \left( -B C_x^2 \right).

On minimising the MSE of t_3 with respect to q_1 and q_2, respectively, we get

\[ q_1^* = \frac{\Delta_1 \Delta_4 - \Delta_3 \Delta_4}{\Delta_1 \Delta_3 - \Delta_2^2} \quad \text{and} \quad q_2^* = \frac{\Delta_1 \Delta_5 - \Delta_2 \Delta_4}{\Delta_1 \Delta_3 - \Delta_2^2} \]

Where,

\[ \Delta_1 = (P^2 + M_1 + 2M_3), \quad \Delta_2 = (-M_4 - M_5), \]
\[ \Delta_3 = (M_4), \quad \Delta_4 = (P^2 + M_3) \]
\[ \Delta_5 = (-M_5) \]

On putting these values of q_1 and q_2 in equation (2.14) we obtain the minimum MSE of t_3, as:

\[ \text{MSE}(t_3)_{\text{min}} = \left[ p^2 \frac{\Delta_1 \Delta_4^2 + \Delta_3 \Delta_4^2 - 2\Delta_2 \Delta_4 \Delta_5}{\Delta_1 \Delta_3 - \Delta_2^2} \right] \quad (2.16) \]

Efficiency Comparisons: First, we compare the efficiency of proposed estimator t_3 with usual estimator.

\[ \text{MSE}(t_3)_{\text{min}} \leq V(\bar{Y}) \]

If,

\[ \left[ p^2 \frac{\Delta_1 \Delta_4^2 + \Delta_3 \Delta_4^2 - 2\Delta_2 \Delta_4 \Delta_5}{\Delta_1 \Delta_3 - \Delta_2^2} \right] \leq p^2 f C_p^2 \]

On solving we observed that above condition holds always true.

Next we compare the efficiency of proposed estimator t_3 with regression estimator.

\[ \text{MSE}(\text{reg}) \leq \text{MSE}(t_3)_{\text{min}} \leq \text{MSE}(\text{reg}) \]

If,

\[ \left[ p^2 \frac{\Delta_1 \Delta_4^2 + \Delta_3 \Delta_4^2 - 2\Delta_2 \Delta_4 \Delta_5}{\Delta_1 \Delta_3 - \Delta_2^2} \right] \leq p^2 f C_p^2 \left( 1 - \rho_{p_{\bar{y}}}^2 \right) \]

Empirical Study:

Data Statistics: We have taken the data from [1].

Where

Y – Home ownership
X – Income (thousands of dollars)

<table>
<thead>
<tr>
<th>n</th>
<th>N</th>
<th>P</th>
<th>$x$</th>
<th>$\rho_{pb}$</th>
<th>$C_p$</th>
<th>$C_x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>40</td>
<td>0.525</td>
<td>14.4</td>
<td>0.897</td>
<td>0.963</td>
<td>0.3085</td>
</tr>
</tbody>
</table>

The following Table shows PRE of different estimator's with respect to usual estimator.

<table>
<thead>
<tr>
<th>Estimators</th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{Y}$</td>
<td>$t_1$</td>
<td>$t_2$</td>
<td>$\alpha = 1, \beta = 1$</td>
</tr>
<tr>
<td>$t_1$</td>
<td>$t_2$</td>
<td>$\alpha = 1, \beta = 0$</td>
<td></td>
</tr>
<tr>
<td>$t_3$</td>
<td>$\alpha = 0, \beta = 1$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When we examine Table 1, we observe that the proposed estimators $t_1$, $t_2$ and $t_3$ all performs better than the usual estimator $\bar{Y}$. Also, the proposed estimator $t_3$ is the best among the estimators considered in the paper for the choice $\alpha = 0, \beta = 1$.

REFERENCES


