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 biserial • 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 $i^{th}$ unit ($i=1,2,3...N$). $\phi_i$ is the $i^{th}$ unit of population possesses attribute $\phi$ and $\phi_i$ otherwise. Let $\bar{A} = \sum_{i=1}^{N} \phi_i$ and $\bar{a} = \sum_{i=1}^{n} \phi_i$ 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,

$e_t = \frac{(p - \bar{p})}{p}$, \hspace{1cm} $e_x = \frac{(\bar{x} - \bar{X})}{X}$

Such that,

$E(e_t) = 0$, $(1 = \phi, x)$

and

$E\left(\left(\frac{e_t}{e_x}\right)^2\right) = \frac{C_p^2}{p^2}$, \hspace{1cm} $E\left(\left(\frac{e_t}{e_x}\right)^2\right) = \frac{C_\phi^2}{\bar{x}^2}$, \hspace{1cm} $E(e_x e_t) = \hat{p} \rho _{\phi} C_p C_\phi$

Where,

$f = \left(1 - \frac{1}{n}\right) N$, \hspace{1cm} $C_p^2 = \frac{S_p^2}{p^2}$, \hspace{1cm} $C_\phi^2 = \frac{S_\phi^2}{\bar{x}^2}$, \hspace{1cm} and $\rho _{\phi} = \frac{S_{\phi X}}{S_p S_\phi}$

and $\rho _{\phi}$ is the point biserial correlation coefficient.

Here,

$S_\phi^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\phi_i - \bar{\phi})^2$, \hspace{1cm} $S_p^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$\hspace{1cm} and $S_{\phi X} = \frac{1}{N-1} \sum_{i=1}^{N} (\phi_i x_i - \bar{\phi} \bar{x})$
The Proposed Estimator: We first propose the following ratio-type estimator

$$t_1 = \frac{p}{\bar{X}}$$

The bias and MSE of the estimator $t_1$, to the first order of approximation is respectively, given by

$$B(t_1) = \left( \frac{C^2_r + \rho P C_1 C_X}{2} \right)$$

$$MSE(t_1) = \left( \frac{C^2_r + \rho P C_1 C_X}{2} \right)$$

(2.2)

(2.3)

Following [7], we propose a general family of estimators for $P$ as

$$t_2 = H(p,u)$$

(2.4)

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

$$H(p,1) = P; \forall P$$

(2.5)

and satisfying following regulations:

- Whatever be the sample chosen, the point $(p,u)$ assume values in a bounded closed convex subset $R_2$ of the two-dimensional real space containing the point $(p,1)$.
- The function $H(p,u)$ is a continuous and bounded in $R_2$.
- The first and second order partial derivatives of $H(p,u)$ exist and are continuous as well as bounded in $R_2$.

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

$$t_2 = H(p,u)$$

$$= p + (u-1)H_1 + (u-1)^2H_2 + (p-P)(u-1)H_3 + (p-P)^2H_4 + ...$$

(2.6)

Where,

$$\begin{align*}
H_1 &= \left. \frac{\partial H}{\partial u} \right|_{p=P,u=1}, \\
H_2 &= \left. \frac{\partial^2 H}{\partial u^2} \right|_{p=P,u=1}, \\
H_3 &= \left. \frac{\partial^3 H}{\partial p \partial u} \right|_{p=P,u=1}, \\
H_4 &= \left. \frac{\partial^4 H}{\partial p^2 \partial u^2} \right|_{p=P,u=1}.
\end{align*}$$

The bias and MSE of the estimator $t_2$ are respectively given by

$$B(t_2) = \left( \frac{P \rho P C_1 C_X H_3 + C^2_r H_1 + \rho P^2 C^2_r H_4}{2} \right)$$

$$MSE(t_2) = \left( \frac{P^2 C^2_r + H_1 C^2_1 + 2H_2 \rho P \rho P C_1 C_X}{2} \right)$$

(2.7)

(2.8)

On differentiating (2.8) with respect to $H_1$ and equating to zero we obtain

$$H_1 = \rho P \rho P C_1 C_X$$

(2.9)

On substituting (2.9) in (2.8), we obtain the minimum MSE of the estimator $t_2$ as

$$\text{minMSE}(t_2) = \left( \frac{P^2 C^2_r + \rho P C_1 C_X}{2} \right)$$

(2.10)

We suggest another family of estimators for estimating $P$ as

$$t_3 = \left[ q_1 P + q_2 \left( \frac{\bar{X}}{\bar{X}} \right) \right] \left[ \frac{a\bar{X} + b}{a\bar{X} + b} \right]^{\alpha} \left[ \frac{(a\bar{X} + b) - (a\bar{X} + b)}{(a\bar{X} + b) + (a\bar{X} + b)} \right]$$

(2.11)

Where, $\alpha, \beta, q_1$, and $q_2$ 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].

$$t_3 = \left[ g_1 P \left( 1 + e_0 \right) - g_2 \bar{X} \left[ 1 - \theta e_0 + \frac{\alpha (\alpha + 1)}{2} \theta^2 e_0^2 \right] \right]$$

$$\left[ 1 - \frac{\theta e_0}{2} \right]^2 + \frac{\theta^2 e_0^2}{8}$$

$$= g_1 P \left[ 1 + e_0 - \beta \left( e_0 + e_0 \alpha \right) + A e_0^2 \left( 1 + e_0 \right) \right] - g_2 \bar{X} \left( q_1 - B e_0 \right)$$

(2.12)

Where, $\theta = \frac{a\bar{X} + b}{a + b}$, $B = \frac{\beta}{2} \theta$ and

$$A = \frac{\theta^2}{8} \left[ 4a (a+1) + \beta (b+2) + 4ab \right].$$

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

$$\text{Bias}(t_3) = P (q_1) + \left[ \left( q_2 \bar{X} + q_1 P A \right) C_1 C_X \right]$$

(2.13)
MSE(t_0) = E(t_1 - P)^2
= (q_1 - 1)^2 P^2 + q_1^2 (M_1 + 2M_2) + q_2^2 M_2
+ 2q_1q_2 (-M_4 - M_5) - 2q_1M_3 + 2q_2M_5

Where,
M_1 = P^2 \{ C_p^2 + B^2C_x^2 \}, \quad M_2 = \sum_i f(x_i^2).
M_3 = P^2 \{ AC_x^2 - 2BP(x)C_pC_x \}, \quad M_4 = PXT \{ -BC_x^2 + f(x)C_pC_x \},
M_5 = \sum f(x).

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

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

Where,
\Delta_1 = (P^2 + M_1 + 2M_2), \quad \Delta_2 = (- M_4 - M_5),
\Delta_3 = (M_4), \quad \Delta_4 = (P^2 + M_3)
\Delta_5 = (- M_4).

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

MSE(t_3)_{\text{min}} = \frac{\Delta_1 \Delta_2 \Delta_3 \Delta_4 \Delta_5}{\Delta_1 \Delta_3 - \Delta_2^2}

(2.16)

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

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

If

\left[ \frac{P^2 \Delta_1 \Delta_2 \Delta_3 \Delta_4 \Delta_5}{\Delta_1 \Delta_3 - \Delta_2^2} \right] \leq P^2 f \left( C_p^2 \right)

On solving we observed that above conditions holds always true.

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

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

If

\left[ \frac{P^2 \Delta_1 \Delta_2^2 \Delta_3 \Delta_4 \Delta_5 - 2 \Delta_2 \Delta_4 \Delta_5}{\Delta_1 \Delta_3 - \Delta_2^2} \right] \leq P^2 f \left( C_p^2 \right)

Empirical Study:

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

Where
Y = Home ownership
X = Income (thousands of dollars)

n N = P \times \rho_{pb} \quad C_p \quad C_x

| 11 | 40 | 0.525 | 14.4 | 0.897 | 0.963 | 0.3085 |

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

Table 1: Percent relative efficiency (PRE) of estimators with respect to usual estimator

<table>
<thead>
<tr>
<th>t_0</th>
<th>t_1</th>
<th>t_2</th>
<th>\alpha = 1, \beta = 1</th>
<th>\alpha = 1, \beta = 0</th>
<th>\alpha = 0, \beta = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE</td>
<td>560</td>
<td>499.384</td>
<td>511.794</td>
<td>517.798</td>
<td>517.798</td>
</tr>
</tbody>
</table>

When we examine Table 1, we observe that the proposed estimator t_3 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


