

# **Ratio Estimators In Simple Random Sampling When Study Variable Is An Attribute**

Rajesh Singh and Mukesh Kumar

Department of Statistics, Banaras Hindu University, Varanasi-221005, INDIA

Florentin Smarandache

Department of Mathematics, University of New Mexico, Gallup, USA

[rsinghstat@yahoo.com](mailto:rsinghstat@yahoo.com), [mukesh.stat@gmail.com](mailto:mukesh.stat@gmail.com), [smarand@unm.edu](mailto:smarand@unm.edu)

## **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.

## **1. Introduction**

The use of auxiliary information can increase the precision of an estimator when study variable  $y$  is highly correlated with auxiliary variable  $x$ . 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], [3], [4], [5] and [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  $i^{\text{th}}$  unit ( $i=1,2,3,\dots,N$ ).  $\phi_i = 1$ , if  $i^{\text{th}}$  unit of population possesses attribute  $\phi$  and  $\phi_i = 0$ , 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,

$$e_{\phi} = \frac{(p - P)}{P}, \quad e_x = \frac{(\bar{x} - \bar{X})}{\bar{X}}$$

Such that,

$$E(e_i) = 0, (i = \phi, x)$$

and

$$E(e_\phi^2) = fC_p^2, \quad E(e_x^2) = fC_x^2, \quad E(e_x e_\phi) = f\rho_{pb}C_\phi C_x.$$

where,

$$f = \left( \frac{1}{n} - \frac{1}{N} \right), \quad C_p^2 = \frac{S_p^2}{P^2}, \quad C_x^2 = \frac{S_x^2}{\bar{X}^2},$$

and  $\rho_{pb} = \frac{S_{\phi x}}{S_\phi S_x}$  is the point biserial correlation coefficient.

Here,

$$S_\phi^2 = \frac{1}{N-1} \sum_{i=1}^N (\phi_i - P)^2, \quad S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2 \quad \text{and} \quad S_{\phi x} = \frac{1}{N-1} \left( \sum_{i=1}^N \phi_i x_i - NP\bar{X} \right).$$

## 2. The proposed estimator

We first propose the following ratio-type estimator

$$t_1 = \left( \frac{P}{\bar{X}} \right) \bar{X} \tag{2.1}$$

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

given by

$$B(t_1) = f \left( \frac{C_x^2}{2} - \rho_{pb} C_\phi C_x \right) \tag{2.2}$$

$$MSE(t_1) = f \left( C_\phi^2 + C_x^2 - 2\rho_{pb} C_\phi C_x \right) \tag{2.3}$$

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

$$t_2 = H(p, u) \quad (2.4)$$

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

$$H(p, 1) = P, \forall P \quad (2.5)$$

and satisfying following regulations:

- (i) 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  $(p, 1)$ .
- (ii) The function  $H(p, u)$  is a continuous and bounded in  $R_2$ .
- (iii) 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

$$t_2 = H(p, u) = p + (u - 1)H_1 + (u - 1)^2 H_2 + (p - P)(u - 1)H_3 + (p - P)^2 H_4 + \dots \quad (2.6)$$

where,

$$H_1 = \left. \frac{\partial H}{\partial u} \right|_{p=P, u=1}, \quad H_2 = \left. \frac{1}{2} \frac{\partial^2 H}{\partial u^2} \right|_{p=P, u=1},$$

$$H_3 = \left. \frac{1}{2} \frac{\partial^2 H}{\partial p \partial u} \right|_{p=P, u=1}, \quad \text{and} \quad H_4 = \left. \frac{1}{2} \frac{\partial^2 H}{\partial p^2} \right|_{p=P, u=1}.$$

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

$$B(t_2) = f\left(P\rho_{pb}C_p C_x H_3 + C_x^2 H_2 + P^2 C_y^2 H_4\right) \quad (2.7)$$

$$\text{MSE}(t_2) = f(P^2 C_p^2 + H_1^2 C_x^2 + 2H_1 P \rho_{pb} C_p C_x) \quad (2.8)$$

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

$$H_1 = -\rho_{pb} P \frac{C_p}{C_x} \quad (2.9)$$

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

$$\min \text{MSE}(t_2) = f P^2 C_p^2 (1 - \rho_{pb}^2) \quad (2.10)$$

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

$$t_3 = [q_1 P + q_2 (\bar{X} - \bar{x})] \left[ \frac{a\bar{X} + b}{a\bar{x} + b} \right]^\alpha \exp \left[ \frac{(a\bar{X} + b) - (a\bar{x} + b)}{(a\bar{X} + b) + (a\bar{x} + b)} \right]^\beta \quad (2.11)$$

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]).

$$\begin{aligned} t_3 &= [q_1 P(1 + e_0) - q_2 \bar{X}] \left[ 1 - \alpha \theta e_1 + \frac{\alpha(\alpha + 1)}{2} \theta^2 e_1^2 \right] \left[ 1 - \frac{\beta \theta e_1}{2} + \frac{\beta \theta^2 e_1^2}{8} (\beta + 2) \right] \\ &= q_1 P \left\{ 1 + e_0 - B(e_1 + e_0 e_1) + A e_1^2 (1 + e_0) \right\} - q_2 \bar{X} \left\{ e_1 - B e_1^2 \dots \right\} \end{aligned} \quad (2.12)$$

where,  $\theta = \frac{a\bar{X}}{a\bar{X} + b}$ ,  $B = \left( \alpha + \frac{\beta}{2} \right) \theta$  and  $A = \frac{\theta^2}{8} [4\alpha(\alpha + 1) + \beta(\beta + 2) + 4\alpha\beta]$ .

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

$$\text{Bias}(t_3) = P(q - 1) + f \left[ (q_2 \bar{X} B + q_1 P A) C_x^2 - q_1 P B \rho_{pb} C_p C_x \right] \quad (2.13)$$

$$\begin{aligned}
\text{MSE}(t_3) &= E(t_3 - P)^2 \\
&= (q_1 - 1)^2 P^2 + q_1^2 (M_1 + 2M_3) + q_2^2 M_2 \\
&\quad + 2q_1 q_2 (-M_4 - M_5) - 2q_1 M_3 + 2q_2 M_5 \quad (2.14)
\end{aligned}$$

where,

$$\begin{aligned}
M_1 &= P^2 f(C_p^2 + B^2 C_x^2 - 2B\rho C_p C_x), & M_2 &= \bar{X}^2 f(C_x^2), \\
M_3 &= P^2 f(AC_x^2 - 2B\rho C_p C_x), & M_4 &= P\bar{X} f(-BC_x^2 + \rho C_p C_x), \\
M_5 &= \bar{X} P f(-BC_x^2)
\end{aligned}$$

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

$$q_1^* = \frac{\Delta_1 \Delta_4 - \Delta_2 \Delta_5}{\Delta_1 \Delta_3 - \Delta_2^2} \quad \text{and} \quad q_2^* = \frac{\Delta_1 \Delta_5 - \Delta_2 \Delta_4}{\Delta_1 \Delta_3 - \Delta_2^2} \quad (2.15)$$

where,

$$\begin{aligned}
\Delta_1 &= (P^2 + M_1 + 2M_3), & \Delta_2 &= (-M_4 - M_5), \\
\Delta_3 &= (M_2), & \Delta_4 &= (P^2 + M_3), \\
\Delta_5 &= (-M_5),
\end{aligned}$$

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

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

### 3. Efficiency Comparisons

First, we compare the efficiency of proposed estimator  $t_3$  with usual estimator.

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

If,

$$\left[ P^2 - \frac{\Delta_1\Delta_5^2 + \Delta_3\Delta_4^2 - 2\Delta_2\Delta_4\Delta_5}{\Delta_1\Delta_3 - \Delta_2^2} \right] \leq P^2 f_1 C_p^2 \quad (3.1)$$

On solving we observed that above conditions holds always true.

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

$$\text{MSE}(\text{reg})\text{MSE}(t_\alpha)_{\min} \leq \text{MSE}(\text{reg})$$

If,

$$\left[ P^2 - \frac{\Delta_1\Delta_5^2 + \Delta_3\Delta_4^2 - 2\Delta_2\Delta_4\Delta_5}{\Delta_1\Delta_3 - \Delta_2^2} \right] \leq P^2 f_1 C_p^2 (1 - \rho_{pb}^2) \quad (3.2)$$

### 4. Empirical study

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

Where

Y – Home ownership

X – Income (thousands of dollars)

n	N	P	$\bar{X}$	$\rho_{pb}$	Cp	Cx
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**

Estimators	$\bar{y}$	$t_1$	$t_2$	$t_3$		
				$\alpha = 1, \beta = 1$	$\alpha = 1, \beta = 0$	$\alpha = 0, \beta = 1$
PRE	100	189.384	511.794	515.798	517.950	<b>518.052</b>

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

1. Gujarat, D. N., Sangeetha, 2007. Basic econometrics. Tata McGraw – Hill.
2. Naik, V.D. and Gupta, P.C., 1996. A note on estimation of mean with known population proportion of an auxiliary character. Jour. Ind. Soc. Agr. Stat., 48(2), 151-158.
3. Jhaji, H. S., Sharma, M. K. and Grover, L. K., 2006. A family of estimators of population mean using information on auxiliary attribute. Pak. J. Statist., 22 (1), 43-50.
4. Shabbir, J. and Gupta, S., 2007. On estimating the finite population mean with known population proportion of an auxiliary variable, Pakistan Journal of Statistics 23 (1) 1–9.



5. Singh, R., Chauhan, P., Sawan, N. and Smarandache, F., 2008. Ratio estimators in simple random sampling using information on auxiliary attribute. *Pak. J. Stat. Oper. Res.* 4(1), 47-53.
6. A.M. Abd-Elfattah, E.A. El-Sherpieny, S.M. Mohamed, O.F. Abdou, 2010. Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute, *Appl. Math. Comput.*  
[doi:10.1016/j.amc.2009.12.041](https://doi.org/10.1016/j.amc.2009.12.041)
7. Srivastava, S. K., 1980. A class of estimators using auxiliary information in sample surveys. *Canad. J. Statist.* 8:253–254.
8. Khoshnevisan, M. Singh, R., Chauhan, P., Sawan, N. and Smarandache, F., 2007. A general family of estimators for estimating population mean using known value of some population parameter(s). *Far east journal of theoretical statistics*, 22(2), 181-191.