RATIO ESTMATORS IN SIMPLE RANDOM SAMPLING WHEN STUDY VARIABLE IS AN ATTRIBUTE

RAJESH SINGH, MUKESH KUMAR, FLORENTIN SMARANDACHE

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 ϕ_i and x_i denote the observations on variable ϕ and x respectively for i^{th} unit (i=1,2,3…N). ϕ_i, if i^{th} unit of population possesses attribute ϕ and ϕ_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 ϕ respectively, $P=\frac{A}{N}$ and $p=\frac{a}{n}$ denote the proportion of units in the population and sample, respectively, possessing attribute ϕ.

Define,

$$e_f=\frac{(p-P)}{P},\qquad e_x=\frac{(\bar{x}-\bar{X})}{\bar{X}}$$

Such that,

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

and

$$E\left(e_f^2\right)=fC_p^2,\qquad E\left(e_x^2\right)=fC_x^2,\qquad E\left(e_x e_f\right)=f\rho_{pb}C_f C_x.$$

Where,

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

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

Here,

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

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

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

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

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

$$t_2 = H(p,u) \tag{2.4}$$

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

$$H(p,1) = P, \forall P \tag{2.5}$$

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

$$t_2 = H(p,u) = p + (u-1)H_1 + (u-1)^2H_2 + (p-P)(u-1)H_3 + (p-P)^2H_4 + \ldots \tag{2.6}$$

Where,

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

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

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

$$B(t_2)=f\left(P\rho_{pb}C_pC_xH_3+C_x^2H_2+P^2C_y^2H_4\right) \tag{2.7}$$

$$MSE(t_2)=f\left(P^2C_p^2+H_1^2C_x^2+2H_1P\rho_{pb}C_pC_x\right) \tag{2.8}$$

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

$$H_1=-\rho_{pb}P\frac{C_p}{C_x} \tag{2.9}$$

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

$$\min MSE(t_2)=fP^2C_p^2\left(1-\rho_{pb}^2\right) \tag{2.10}$$

We suggest another family of estimators for estimating P as

$$t_3=\left[q_1P+q_2\left(\bar{X}-\bar{x}\right)\right]\left[\frac{a\bar{X}+b}{a\bar{x}+b}\right]^{a}\exp\left[\frac{\left(a\bar{X}+b\right)-\left(a\bar{x}+b\right)}{\left(a\bar{X}+b\right)+\left(a\bar{x}+b\right)}\right]^{\beta} \tag{2.11}$$

Where α,β,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, α, β, a and b (for choice of the parameters refer to [8] and [5].

$$\begin{aligned} t_3 &=\left[q_1P(1+e_0)-q_2\bar{X}\right]\left[1-\alpha\theta e_1+\frac{\alpha(\alpha+1)}{2}\theta^2e_1^2\right] \\ &\quad \left[1-\frac{\beta\theta e_1}{2}+\frac{\beta\theta^2e_1^2}{8}(\beta+2)\right] \\ &= q_1P\left\{1+e_0-B(e_1+e_0e_1)+Ae_1^2(1+e_0)\right\}-q_2\bar{X}\left\{e_1-Be_1^2\ldots\right\} \end{aligned} \tag{2.12}$$

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

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

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

$$Bias(t_3)=P(q-1)+f\left[\left(q_2\bar{X}B+q_1PA\right)C_x^2-q_1PB\rho C_pC_x\right] \tag{2.13}$$

$MSE(t_3) = E(t_3 - P)^2$

$$= (q_1 - 1)^2 P^2 + q_1^2 (M_1 + 2M_3) + q_2^2 M_2$$
$$+ 2q_1 q_2 (-M_4 - M_5) - 2q_1 M_3 + 2q_2 M_5$$

Where,

$$M_1 = P^2 f\left(C_p^2 + B^2 C_x^2 - 2B\rho C_p C_x\right), \quad M_2 = \bar{X}^2 f\left(C_x^2\right),$$

$$M_3 = P^2 f\left(AC_x^2 - 2B\rho C_p C_x\right), \quad M_4 = P\bar{X} f\left(-BC_x^2 + \rho C_p C_x\right),$$

$$M_5 = \bar{X} P f\left(-BC_x^2\right).$$

On minimising the MSE of t3 with respect to q1 and q2, respectively, we get

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

Where,

$\Delta_1 = (P^2 + M_1 + 2M_3)$, $\Delta_2 = (- M_4 - M_5)$,

$\Delta_3 = (M_2)$, $\Delta_4 = (P^2 + M_3)$

$\Delta_5 = (- M_5)$,

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

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

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

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

If,

$$\left[P^2 - \frac{\Delta_1 \Delta_5^2 + \Delta_3 \Delta_4^2 - 2\Delta_2 \Delta_4 \Delta_5}{\Delta_1 \Delta_3 - \Delta_2^2}\right] \le P^2 f_1 C_p^2$$

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$(t\alpha)_{min} \le$ MSE(reg)

If,

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

Empirical Study:

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

Where

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

n	N	P	$\bar{x}$	ρpb	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

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

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. and N. Sangeetha, 2007. Basic economtrics. Tata McGraw - Hill.
2. Naik,V.D. and P.C. Gupta, 1996. A note on estimation of mean with known population proportion of an auxiliary character. J. Ind. Soc. Agr. Stat., 48(2): 151-158.
3. Jhajj, H.S., M.K. Sharma and L.K. Grover, 2006. A family of estimators of population mean using information on auxiliary attribute. Pak. J. Statist., 22(1): 43-50.

4. Shabbir, J. and S. Gupta, 2007. On estimating the finite population mean with known population proportion of an auxiliary variable, Pakistan J. Statistics, 23(1): 1-9.
5. Singh, R., P. Chauhan, N. Sawan and F. Smarandache, 2008. Ratio estimators in simple random sampling using information on auxiliary attribute. Pak. J. Stat. Oper. Res., 4(1): 47-53.
6. Abd-Elfattah, A.M., E.A. El-Sherpieny, S.M. Mohamed and 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.
7. Srivastava, S.K., 1980. A class of estimators using auxiliary information in sample surveys. Canad. J. Statist., 8: 253-254.
8. Khoshnevisan, M., R. Singh, P. Chauhan, N. Sawan and F. Smarandache, 2007. A general family of estimators for estimating population mean using known value of some population parameter(s). Far east J. Theoretical Statistics, 22(2): 181-191.

Published in "Pakistan Journal of Statistics & Operational Research", Vol. IV, No. 1, pp. 47-53, 2008.