Hemant Verma, Rajesh Singh

Department of Statistics, Banaras Hindu University

Varanasi-221005, India

Florentin Smarandache

University of New Mexico, Gallup, USA

Some Improved Estimators of Population Mean Using Information on Two Auxiliary Attributes

Published in:
Rajesh Singh, Florentin Smarandache (Editors)
ON IMPROVEMENT IN ESTIMATING POPULATION PARAMETER(S) USING AUXILIARY INFORMATION
Educational Publishing (Columbus) & Journal of Matter Regularity (Beijing), USA - China, 2013
ISBN: 978-1-59973-230-5
pp. 17 - 24

Abstract

In this paper, we have studied the problem of estimating the finite population mean when information on two auxiliary attributes are available. Some improved estimators in simple random sampling without replacement have been suggested and their properties are studied. The expressions of mean squared error's (MSE's) up to the first order of approximation are derived. An empirical study is carried out to judge the best estimator out of the suggested estimators.

Key words: Simple random sampling, auxiliary attribute, point bi-serial correlation, phi correlation, efficiency.

Introduction

The role of auxiliary information in survey sampling is to increase the precision of estimators when study variable is highly correlated with auxiliary variable. But when we talk about qualitative phenomena of any object then we use auxiliary attributes instead of auxiliary variable. For example, if we talk about height of a person then sex will be a good auxiliary attribute and similarly if we talk about particular breed of cow then in this case milk produced by them will be good auxiliary variable.

Most of the times, we see that instead of one auxiliary variable we have information on two auxiliary variables e.g.; to estimate the hourly wages we can use the information on marital status and region of residence (see Gujrati and Sangeetha (2007), page-311).

In this paper, we assume that both auxiliary attributes have significant point bi-serial correlation with the study variable and there is significant phi-correlation (see Yule (1912)) between the auxiliary attributes.

Consider a sample of size n drawn by simple random sampling without replacement (SRSWOR) from a population of size N. let y_j, ϕ_{ij} (i=1,2) denote the observations on variable y and ϕ_i (i=1,2) respectively for the j^{th} unit (i=1,2,3,……N) . We note that ϕ_{ij}=1, if j^{th} unit possesses attribute ϕ_{ij}=0 otherwise . Let $A_i = \sum_{j=1}^{N} \phi_{ij}$, $a_i = \sum_{j=1}^{n} \phi_{ij}$; i=1,2 denotes the total number of units in the population and sample respectively, possessing attribute ϕ. Similarly, let $P_i = \frac{A_i}{N}$ and $p_i = \frac{a_i}{n}$;(i=1,2) denotes the proportion of units in the population and sample respectively possessing attribute ϕ_i (i=1,2).

In order to have an estimate of the study variable y, assuming the knowledge of the population proportion P, Naik and Gupta (1996) and Singh et al. (2007) respectively proposed following estimators:

$$t_1 = \bar{y}\left(\frac{P_1}{p_1}\right) \tag{1.1}$$

$$t_2 = \bar{y}\left(\frac{p_2}{P_2}\right) \tag{1.2}$$

$$t_3 = \bar{y}\exp\left(\frac{P_1 - p_1}{P_1 + p_1}\right) \tag{1.3}$$

$$t_4 = \bar{y}\exp\left(\frac{p_2 - P_2}{p_2 + P_2}\right) \tag{1.4}$$

The bias and MSE expression's of the estimator's t_i (i=1, 2, 3, 4) up to the first order of approximation are, respectively, given by

$$B(t_1) = \bar{Y} f_1 C_{p_1}^2 \left[1 - K_{pb_1}\right] \tag{1.5}$$

$$B(t_2) = \bar{Y} f_1 K_{pb_2} C_{p_2}^2 \tag{1.6}$$

$$B(t_3)=\bar{Y}f_1\frac{C_{p_1}^2}{2}\left[\frac{1}{4}-K_{pb_1}\right] \tag{1.7}$$

$$B(t_4)=\bar{Y}f_1\frac{C_{p_2}^2}{2}\left[\frac{1}{4}+K_{pb_2}\right] \tag{1.8}$$

$$MSE(t_1)=\bar{Y}^2 f_1\left[C_y^2+C_{p_1}^2\left(1-2K_{pb_1}\right)\right] \tag{1.9}$$

$$MSE(t_2)=\bar{Y}^2 f_1\left[C_y^2+C_{p_1}^2\left(1+2K_{pb_2}\right)\right] \tag{1.10}$$

$$MSE(t_3)=\bar{Y}^2 f_1\left[C_y^2+C_{p_1}^2\left(\frac{1}{4}-K_{pb_1}\right)\right] \tag{1.11}$$

$$MSE(t_4)=\bar{Y}^2 f_1\left[C_y^2+C_{p_2}^2\left(\frac{1}{4}+K_{pb_2}\right)\right] \tag{1.12}$$

where, $f_1=\frac{1}{n}-\frac{1}{N}$, $S_{\phi_j}^2=\frac{1}{N-1}\sum_{i=1}^{N}\left(\phi_{ji}-P_j\right)^2$, $S_{y\phi_j}=\frac{1}{N-1}\sum_{i=1}^{N}\left(y_i-\bar{Y}\right)\left(\phi_{ji}-P_j\right)$,

$\rho_{pb_j}=\frac{S_{y\phi_j}}{S_y S_{\phi_j}}$, $C_y=\frac{S_y}{\bar{Y}}$, $C_{pj}=\frac{S_{\phi_j}}{P_j}$; $(j=1,2)$, $K_{pb_1}=\rho_{pb_1}\frac{C_y}{C_{p_1}}$, $K_{pb_2}=\rho_{pb_2}\frac{C_y}{C_{p_2}}$.

$s_{\phi_1\phi_2}=\frac{1}{n-1}\sum_{i=1}^{n}\left(\phi_{1i}-p_1\right)\left(\phi_{2i}-p_2\right)$ and $\rho_\phi=\frac{s_{\phi_1\phi_2}}{s_{\phi_1}s_{\phi_2}}$ be the sample phi-covariance and phi-correlation between ϕ_1 and ϕ_2 respectively, corresponding to the population phi-covariance and phi-correlation $S_{\phi_1\phi_2}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\phi_{1i}-P_1\right)\left(\phi_{2i}-P_2\right)$ and $\rho_\phi=\frac{S_{\phi_1\phi_2}}{S_{\phi_1}S_{\phi_2}}$.

In this paper we have proposed some improved estimators of population mean using information on two auxiliary attributes in simple random sampling without replacement. A comparative study is also carried out to compare the optimum estimators with respect to usual mean estimator with the help of numerical data.

2. Proposed Estimators

Following Olkin (1958), we propose an estimator t_1 as

$$t_5 = \bar{y}\left[w_1 \frac{P_1}{p_1} + w_2 \frac{P_2}{p_2}\right] \tag{2.1}$$

where w_1 and w_2 are constants, such that $w_1 + w_2 = 1$.

Consider another estimator t_6 as

$$t_6 = \left[K_{61}\bar{y} + K_{62}(P_1 - p_1)\right]\exp\left[\frac{P_2 - p_2}{P_2 + p_2}\right] \tag{2.2}$$

where K_{61} and K_{62} are constants.

Following Shaoo et al. (1993), we propose another estimator t_7 as

$$t_7 = \bar{y} + K_{71}(P_1 - p_1) + K_{72}(P_2 - p_2) \tag{2.3}$$

where K_{71} and K_{72} are constants.

Bias and MSE of estimators t_5, t_6 and t_7:

To obtain the bias and MSE expressions of the estimators $t_i (i = 5,6,7)$ to the first degree of approximation, we define

$$e_0 = \frac{\bar{y} - \bar{Y}}{\bar{Y}}, \quad e_1 = \frac{p_1 - P_1}{P_1}, \quad e_2 = \frac{p_2 - P_2}{P_2}$$

such that, $E(e_i) = 0$; $i = 0, 1, 2.$

Also,

$$E(e_0^2) = f_1 C_y^2, \ E(e_1^2) = f_1 C_{p_1}^2, \ E(e_2^2) = f_1 C_{p_2}^2,$$

$$E(e_0 e_1) = f_1 K_{pb_1} C_{p_1}^2, \ E(e_0 e_2) = f_1 K_{pb_2} C_{p_2}^2, \quad E(e_1 e_2) = f_1 K_\phi C_{p_2}^2,$$

$$K_{pb_1} = \rho_{pb_1}\frac{C_y}{C_{p_1}}, \ K_{pb_2} = \rho_{pb_2}\frac{C_y}{C_{p_2}}, \ K_\phi = \rho_\phi \frac{C_{p_1}}{C_{p_2}}$$

Expressing (2.1) in terms of e's we have,

$$t_5 = \bar{Y}(1 + e_0)\left[w_1 \frac{P_1}{P_1(1 + e_1)} + w_2 \frac{P_2}{P_2(1 + e_2)}\right]$$

$$t_5 = \overline{Y}(1+e_0)\left[w_1(1+e_1)^{-1} + w_2(1+e_2)^{-1}\right] \tag{3.1}$$

Expanding the right hand side of (3.1) and retaining terms up to second degrees of e's, we have,

$$t_5 = \overline{Y}\left[1 + e_0 - w_1e_1 - w_2e_2 + w_1e_1^2 + w_2e_2^2 - w_1e_0e_1 - w_2e_0e_2\right] \tag{3.2}$$

Taking expectations of both sides of (3.1) and then subtracting $\overline{Y}$ from both sides, we get the bias of estimator t_5 upto the first order of approximation as

$$\mathrm{Bias}(t_5) = \overline{Y}f_1\left[w_1C_{p_1}^2\left(1-K_{pb_1}\right) + w_2C_{p_2}^2\left(1-K_{pb_2}\right)\right] \tag{3.3}$$

From (3.2), we have,

$$\left(t_5 - \overline{Y}\right) \cong \overline{Y}\left[e_0 - w_1e_1 - w_2e_2\right] \tag{3.4}$$

Squaring both sides of (3.4) and then taking expectations, we get the MSE of t_5 up to the first order of approximation as

$$\mathrm{MSE}(t_5) = \overline{Y}^2 f_1\left[C_y^2 + w_1^2C_{p_1}^2 + w_2^2C_{p_2}^2 - 2w_1K_{pb_1}C_{p_1}^2 - 2w_2K_{pb_2}C_{pb_2}^2 + 2w_1w_2K_\phi C_{p_2}^2\right] \tag{3.5}$$

Minimization of (3.5) with respect to w_1 and w_2, we get the optimum values of w_1 and w_2, as

$$w_{1(opt)} = \frac{K_{pb_1}C_{p_1}^2 - K_\phi C_{p_2}^2}{C_{p_1}^2 - K_\phi C_{p_2}^2} = w_1^*(\mathrm{say})$$

$$\begin{aligned} w_{2(opt)} &= 1 - w_{1(opt)} \\ &= 1 - \frac{K_{pb_1}C_{p_1}^2 - K_\phi C_{p_2}^2}{C_{p_1}^2 - K_\phi C_{p_2}^2} \\ &= \frac{C_{p_1}^2\left[1-K_{pb_1}\right]}{C_{p_1}^2 - K_\phi C_{p_2}^2} = w_2^*(\mathrm{say}) \end{aligned}$$

Similarly, we get the bias and MSE expressions of estimator t_6 and t_7 respectively, as

$$\mathrm{Bias}(t_6) = K_{61}\overline{Y}\left[1 + f_1C_{p_2}^2\left(\frac{3}{8} - \frac{1}{2}K_{pb_2}\right)\right] + \frac{1}{2}K_{22}P_1f_1K_\phi C_{p_2}^2 \tag{3.6}$$

$$\mathrm{Bias}(t_7) = 0 \tag{3.7}$$

And

$$MSE(t_6) = K_{61}^2 \bar{Y}^2 A_1 + K_{62}^2 P_1^2 A_2 - 2K_{61}K_{62}P_1\bar{Y}A_3 + (1 - 2K_{61})\bar{Y}^2 \tag{3.8}$$

where $A_1 = 1 + f_1\left(C_y^2 + C_{p_2}^2\left(\frac{1}{4} - K_{pb_2}\right)\right)$

$$A_2 = f_1 C_{p_1}^2$$

$$A_3 = f_1\left(k_{pb_1} C_{p_1}^2 - \frac{1}{2} K_\phi C_p^2\right)$$

And the optimum values of K_{61} and K_{62} are respectively, given as

$$K_{61(opt)} = \frac{A_2}{A_1 A_2 - A_3^2} = K_{61}^*(\text{say})$$

$$K_{62(opt)} = \frac{\bar{Y}A_3}{P_1\left(A_1 A_2 - A_3^2\right)} = K_{62}^*(\text{say})$$

$$MSE(t_7) = \bar{Y}^2 f_1 C_y^2 + K_{71}^2 P_1^2 f_1 C_{p_1}^2 + K_{72}^2 P_2^2 f_1 C_{p_2}^2 - 2K_{71}P_1\bar{Y}f_1 K_{pb_1} C_{p_1}^2 - 2K_{72}P_2\bar{Y}f_1 K_{pb_2} C_{p_2}^2 + 2K_{71}K_{72}P_1 P_2 f_1 K_\phi C_{p_2}^2 \tag{3.9}$$

And the optimum values of K_{71} and K_{72} are respectively, given as

$$K_{71(opt)} = \frac{\bar{Y}}{P_1}\left(\frac{K_{pb_1}C_{p_1}^2 - K_{pb_2}K_\phi C_{p_2}^2}{C_{p_1}^2 - K_\phi^2 C_{p_2}^2}\right) = K_{71}^*(\text{say})$$

$$K_{72(opt)} = \frac{\bar{Y}}{P_2}\left(\frac{K_{pb_2}C_{p_1}^2 - K_{pb_1}K_\phi C_{p_1}^2}{C_{p_1}^2 - K_\phi^2 C_{p_2}^2}\right) = K_{72}^*(\text{say})$$

4. Empirical Study

Data: (Source: Government of Pakistan (2004))

The population consists rice cultivation areas in 73 districts of Pakistan. The variables are defined as:

Y= rice production (in 000' tonnes, with one tonne = 0.984 ton) during 2003,

P_1 = production of farms where rice production is more than 20 tonnes during the year 2002, and

P_2 = proportion of farms with rice cultivation area more than 20 hectares during the year 2003.

For this data, we have

N=73, $\bar{Y}$ =61.3, P_1=0.4247, P_2=0.3425, S_y^2=12371.4, $S_{\phi_1}^2$ =0.225490, $S_{\phi_2}^2$ =0.228311,

ρ_{pb_1} =0.621, ρ_{pb_2} =0.673, ρ_ϕ=0.889.

The percent relative efficiency (PRE's) of the estimators t_i (i=1,2,…7) with respect to unusual unbiasedestimator $\bar{y}$ have been computed and given in Table 4.1.

Table 4.1 : PRE of the estimators with respect to $\bar{y}$

Estimator	**PRE**
$\bar{y}$	100.00
t_1	162.7652
t_2	48.7874
t_3	131.5899
t_4	60.2812
t_5	165.8780
t_6	197.7008
t_7	183.2372

Conclusion

In this paper we have proposed some improved estimators of population mean using information on two auxiliary attributes in simple random sampling without replacement. From the Table 4.1 we observe that the estimator t_6 is the best followed by the estimator t_7 .

References

Government of Pakistan, 2004, Crops Area Production by Districts (Ministry of Food, Agriculture and Livestock Division, Economic Wing, Pakistan).

Gujarati, D. N. and Sangeetha (2007): Basic econometrics. Tata McGraw – Hill.

Malik, S. and Singh, R. (2013): A family of estimators of population mean using information on point bi-serial and phi correlation coefficient. IJSE

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. Agri. Stat., 48(2), 151-158.

Olkin, I. (1958): Multivariate ratio estimation for finite populations, Biometrika, 45, 154–165.

Singh, R., Chauhan, P., Sawan, N. and Smarandache, F.(2007): Auxiliary information and a priori values in construction of improved estimators. Renaissance High press.

Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. (2008): Ratio Estimators in Simple Random Sampling Using Information on Auxiliary Attribute. Pak. Jour. Stat. Oper. Res. Vol. IV, No.1, pp. 47-53.

Singh, R., Kumar, M. and Smarandache, F. (2010): Ratio estimators in simple random sampling when study variable is an attribute. WASJ 11(5): 586-589.

Yule, G. U. (1912): On the methods of measuring association between two attributes. Jour. of the Royal Soc. 75, 579-642.