World Applied Sciences Journal 8 (1): 133-136, 2010
ISSN 1818-4952

Ratio Estimators using Multiple Auxiliary Attributes

[1]Muhammad Hanif, [2]Inam ul Haq and [3]Muhammad Qaiser Shahbaz

[1]Department of Mathematics, LUMS, Lahore, Pakistan
[2]National College of Business Administration and Economics, Lahore, Pakistan
[3]Department of Mathematics, COMSATS Institute of IT, Lahore, Pakistan

Abstract: In this paper some ratio estimators for single phase and two phase sampling has been proposed by using information on multiple auxiliary attributes. The proposed estimators are generalization of the estimator proposed by Naik and Gupta [4]. The shrinkage version of the proposed estimators has also been obtained by using the method given by Shahbaz and Hanif [5]. Efficiency comparison has also been made by using the real time data.

Key words: Ratio estimator . auxiliary attributes . two phase sampling . bi-serial correlation coefficient, shrinkage estimator

INTRODUCTION

Auxiliary information has been widely used to increase the efficiency of estimators for estimation of population parameters. Most of the methods utilize quantitative auxiliary variables to increase the precision of estimates. Use of auxiliary attributes has also been used in survey sampling. Naik and Gupta [4] proposed ratio, product and regression estimator using information on a single auxiliary attribute. Haq [2] proposed some estimators using information on two auxiliary attributes. A family of estimators using auxiliary attributes has been proposed by Jhajj *et al*. [3] along side several examples. Hanif *et al*. [1] generalizes the family of estimators proposed by Jhajj *et al*. [3]. Many authors have proposed the use of shrinkage estimators in survey sampling to reduce the error of estimation. Shahbaz and Hanif [5] have proposed a general shrinkage estimator for estimation of population characteristics. The estimator proposed by Shahbaz and Hanif [5] has the form:

$$\hat{t}_S = \frac{\hat{t}}{1+T^{-2}MSE\left(\hat{t}\right)} \quad (1.1)$$

where $\hat{t}$ is any available estimator of parameter T. The Mean Square Error of (1.1) is:

$$mse\left(\hat{t}_S\right) = \frac{MSE\left(\hat{t}\right)}{1+T^{-2}MSE\left(\hat{t}\right)} \quad (1.2)$$

In this paper, we have generalized the estimators proposed in [4] by using information on multiple auxiliary attributes. The estimators have been proposed for single phase and two phase sampling. In order to develop the estimators we introduce some notations in the following section.

PRELIMINARIES AND NOTATIONS

Let $(y_i, \tau_{i1}, \tau_{i2},\ldots,\tau_{ik})$ be the ith sample point from a population of size N, where y_i is value of variable of interest and τ_{ij} (j=1,2…k) is the value of jth auxiliary attribute. In defining the attributes we assume complete dichotomy so that $\tau_{ij} = 1$ if ith unit of population possesses jth attribute and $\tau_j = 0$ otherwise. Let

$$A_j = \sum_{i=1}^{N}\tau_{ij} \text{ and } a_j = \sum_{i=1}^{n}\tau_{ij}$$

be the total number of units in the population and sample respectively possessing attribute τ_j. Let $P_j = N^{-1}A_j$ and $p_j = n^{-1}a_j$ be the corresponding proportion of units possessing a specific attributes τ_j $\bar{y}$ is the mean of main variable at first phase. We also define $\bar{e}_y = \bar{y} - \bar{Y}$ and $\bar{e}_{\tau_j} = p_j - P_j$ so that $E(\bar{e}_y) = 0 = E(\bar{e}_{\tau_j})$. We also define:

$$E(\bar{e}_y^2) = \theta S_y^2$$

$$E(\bar{e}_{\tau_j}^2) = \theta S_{\tau_j}^2$$

$$E(\bar{e}_y \bar{e}_{\tau_j}) = \theta S_y S_{\tau_j} \rho_{pb_j}$$

Corresponding Author: Dr. Muhammad Hanif, Department of Mathematics, LUMS, Lahore, Pakistan

$$E\left(\bar{e}_{\tau_j}\bar{e}_{\tau_\psi}\right)=\theta S_{\tau_j}S_{\tau_\psi}Q_{j\psi} \tag{2.1}$$

where

$$\theta = n^{-1} - N^{-1}$$

$$S_{y\tau_j}=\left(N-1\right)^{-1}\sum_{j=1}^{N}\left(y_i-\bar{Y}\right)\left(\tau_{ij}-P_j\right)$$

$$\rho_{Pbj}=S_{y\tau_j}\Big/\left(S_yS_{\tau_j}\right)$$

is the bi-serial correlation coefficient and $Q_{j\psi}$ $(-1\le Q_{j\psi}\le +1)$ is coefficient of association. When sample is drawn in phases then we define n_1 and n_2 as the size of first and second phase sample respectively. Further, let $p_{j(1)}$, $p_{j(2)}$ be the proportion of units possessing attribute τ_j in first-phase and second-phase sample respectively. The mean of main variable of interest at second phase will be denoted by $\bar{y}_2$. Also let us define

$$\bar{e}_{y_2}=\bar{y}_2-\bar{Y}$$

$$\bar{e}_{\tau_{j(1)}}=p_{j(1)}-P_j$$

$$\bar{e}_{\tau_{j(2)}}=p_{j(2)}-P_j$$

for (j=1,2…k). We also define following expectations:

$$\left.\begin{aligned}&E\left(\bar{e}_{y_2}\right)=\theta_2S_y^2,\; E\left(\bar{e}_{\tau_{j(1)}}-\bar{e}_{\tau_{j(2)}}\right)^2=\theta_3S_{\tau_j}^2\\&E\left[\bar{e}_{y_2}\left(\bar{e}_{\tau_{j(2)}}-\bar{e}_{\tau_{j(1)}}\right)\right]=\theta_3S_yS_{\tau_j}\rho_{pbj}\\&E\left[\left(\bar{e}_{\tau_{j(2)}}-\bar{e}_{\tau_{j(1)}}\right)\left(\bar{e}_{\tau_{\psi(2)}}-\bar{e}_{\tau_{\psi(1)}}\right)\right]=\theta_3S_{\tau_j}S_{\tau_\psi}Q_{j\psi}\end{aligned}\right\} \tag{2.2}$$

where $\theta_h=n_h^{-1}-N^{-1}$, $\theta_3 = \theta_2 - \theta_1$ and n_h is size of sample drawn at *h*th phase. After defining these notations we propose the new estimators in the following sections.

ESTIMATORS FOR SINGLE PHASE SAMPLING

In this section we have proposed some new estimators for single phase sampling using information on multiple auxiliary attributes. The proposed estimator is:

$$t_{1(1)}=\bar{y}\left(\frac{P_1}{p_1}\right)\left(\frac{P_2}{p_2}\right)\cdots\left(\frac{P_k}{p_k}\right)=\bar{y}\prod_{j=1}^{k}\left(P_j/p_j\right) \tag{3.1}$$

Using notations from section-2 and expanding; the mean square error of (3.1) is:

$$MSE\left(t_{1(1)}\right)\approx E\left(\bar{e}_y-\sum_{j=1}^{k}\frac{\bar{Y}}{P_j}\bar{e}_{\tau_j}\right)^2$$

Now using expectations from (2.1), the mean square error is given as:

$$MSE\left(t_{1(1)}\right)\approx\theta\bar{Y}^2\left[\begin{aligned}&C_y^2+\sum_{j=1}^{k}C_{\tau_j}^2-2\sum_{j=1}^{k}C_yC_{\tau_j}\rho_{Pb_j}\\&+2\sum_{j\neq\psi=1}^{k}C_{\tau_j}C_{\tau_\psi}Q_{j\psi}\end{aligned}\right] \tag{3.2}$$

The estimate of above mean square error is readily obtained by replacing population parameters with sample estimates in (3.2).

PROPOSED ESTIMATORS FOR TWO PHASE SAMPLING

In this section we have proposed some ratio estimators for two phase sampling using various mechanisms. These estimators are proposed below:

Estimator for partial information case: We propose following estimator for two phase sampling when information of a set of auxiliary attributes is available for the complete population. We proposed two different estimators for this situation. The first of these two estimators is:

$$t_{2(2)}=\bar{y}_2\left(\frac{P_1}{p_{1(1)}}\right)\left(\frac{P_2}{p_{2(1)}}\right)\cdots\left(\frac{P_m}{p_{m(1)}}\right)\left(\frac{p_{(m+1)(1)}}{p_{(m+1)(2)}}\right)\cdots\left(\frac{p_{k(1)}}{p_{k(2)}}\right)$$

or

$$t_{2(2)}=\bar{y}_2\prod_{j=1}^{m}\left(P_j/p_{j(1)}\right)\prod_{h=m+1}^{k}\left(p_{h(1)}/p_{h(2)}\right) \tag{4.1}$$

The mean square error of above estimator can be expanded as:

$$MSE\left(t_{2(2)}\right)\approx E\left(\begin{aligned}&\bar{e}_{y_2}-\sum_{j=1}^{m}\frac{\bar{Y}}{P_j}\bar{e}_{\tau_{j(1)}}\\&-\sum_{j=m+1}^{k}\frac{\bar{Y}}{P_j}\left(\bar{e}_{\tau_{j(2)}}-\bar{e}_{\tau_{j(1)}}\right)\end{aligned}\right)^2.$$

Using the expectations from (2.2), the mean square error of (4.1) is:

$$\begin{aligned} MSE\left(t_{2(2)}\right) \approx \bar{Y}^2 \Bigg[&\theta_2 \left\{ C_y^2 + \sum_{j=m+1}^{k} C_{\tau_j}^2 - 2 \sum_{j=m+1}^{k} C_y C_{\tau_j} \rho_{Pb_j} + 2 \sum_{j \neq \psi = m+1}^{k} C_{\tau_j} C_{\tau_\psi} Q_{j\psi} \right\} \\ &\theta_1 \Bigg\{ \left(\sum_{j=1}^{m} C_{\tau_j}^2 - \sum_{j=m+1}^{k} C_{\tau_j}^2 \right) - 2\left(\sum_{j=1}^{m} C_y C_{\tau_j} \rho_{Pb_j} - \sum_{j=m+1}^{k} C_y C_{\tau_j} \rho_{Pb_j} \right) \\ &+ 2\left(\sum_{j \neq \psi = 1}^{m} C_{\tau_j} C_{\tau_\psi} Q_{j\psi} - \sum_{j \neq \psi = m+1}^{k} C_{\tau_j} C_{\tau_\psi} Q_{j\psi} \right) \Bigg\} \Bigg] \end{aligned} \tag{4.2}$$

The second estimator that we propose for partial information case is:

$$t_{3(2)} = \bar{y}_2 \left(\frac{P_1}{p_{1(2)}}\right)\left(\frac{P_2}{p_{2(2)}}\right) \ldots \left(\frac{P_m}{p_{m(2)}}\right)\left(\frac{p_{(m+)(1)}}{p_{(m+)(2)}}\right) \ldots \left(\frac{p_{k(1)}}{p_{k(2)}}\right) \text{ or } t_{2(2)} = \bar{y}_2 \prod_{j=1}^{m} \left(P_j / p_{j(2)}\right) \prod_{h=m+1}^{k} \left(p_{h(1)} / p_{h(2)}\right) \tag{4.3}$$

The mean square error of (4.3) can be expanded as:

$$MSE\left(t_{3(2)}\right) \approx E\left(\bar{e}_{y_2} - \sum_{j=1}^{m} \frac{\bar{Y}}{P_j} \bar{e}_{\tau_{j(2)}} - \sum_{j=m+1}^{k} \frac{\bar{Y}}{P_j} \left(\bar{e}_{\tau_{j(2)}} - \bar{e}_{\tau_{j(1)}} \right) \right)^2 .$$

$$\begin{aligned} MSE\left(t_{3(2)}\right) \approx \bar{Y}^2 \Bigg[&\theta_2 \left\{ C_y^2 + \sum_{j=1}^{m} C_{\tau_j}^2 - 2 \sum_{j=1}^{m} C_y C_{\tau_j} \rho_{Pb_j} + 2 \sum_{j \neq \psi = 1}^{m} C_{\tau_j} C_{\tau_\psi} Q_{j\psi} \right\} \\ &+ \theta_3 \left\{ \sum_{j=m+1}^{k} C_{\tau_j}^2 - 2 \sum_{j=m+1}^{k} C_y C_{\tau_j} \rho_{Pb_j} + 2 \sum_{\substack{j=m+1 \\ \psi \geq m+1}} C_{\tau_j} C_{\tau_\psi} Q_{j\psi} \right. \\ &\left. 2 \sum_{j \neq \psi = m+1}^{k} C_{\tau_j} C_{\tau_\psi} Q_{j\psi} \right\} \Bigg] \end{aligned} \tag{4.4}$$

If P_j for j =, m+1, m+2,…k) is also known then $\theta_2 = \theta_3 = \theta$; then (4.4) reduces to (3.2).

Estimator for no information case: We propose the following estimator when information on all the auxiliary attributes is unavailable:

$$t_{4(2)} = \bar{y}_2 \left(\frac{p_{1(1)}}{p_{1(2)}}\right)\left(\frac{p_{2(1)}}{p_{2(2)}}\right) \ldots \left(\frac{p_{k(1)}}{p_{k(2)}}\right) = \bar{y}_2 \prod_{j=1}^{k} \left(p_{j(1)} / p_{j(2)}\right) \tag{4.5}$$

The mean square error of (4.5) is given as:

$$MSE\left(t_{4(2)}\right) \approx E\left(\bar{e}_{y_2} - \sum_{j=1}^{k} \frac{\bar{Y}}{P_j} \left(\bar{e}_{\tau_{j(2)}} - \bar{e}_{\tau_{j(1)}} \right) \right)^2$$

Using the expectations from (3.2), the mean square error of (4.5) is given as:

$$MSE\left(t_{4(2)}\right) \approx \bar{Y}^2 \left[\theta_2 C_y^2 + \theta_3 \left\{ \sum_{j=1}^{k} C_{\tau_j}^2 - 2 \sum_{j=1}^{k} C_y C_{\tau_j} \rho_{Pb_j} + 2 \sum_{j \neq \psi = 1}^{k} C_{\tau_j} C_{\tau_\psi} Q_{j\psi} \right\} \right] \tag{4.6}$$

The estimate of (4.6) can be easily written by replacing population parameters with sample estimates.

THE SHRINKAGE ESTIMATORS

We have proposed four estimators for single and two phase sampling using information on multiple auxiliary attributes. The shrinkage version of all the above estimators can be obtained by using the method proposed by Shahbaz and Hanif [5]. For example the shrinkage estimator for single phase sampling is:

$$t_{1(1)s} = d_0\left[\bar{y}\prod_{j=1}^{k}\left(P_j/p_j\right)\right] = d_0 t_{1(1)} \quad (5.1)$$

Now using the method of Shahbaz and Hanif [5], the mean square error of (5.1) is:

$$MSE\left(t_{1(1)s}\right) \approx MSE\left(t_{1(1)}\right)\left[1+\bar{Y}^{-2}MSE\left(t_{1(1)}\right)\right]^{-1} \quad (5.2)$$

Similarly, the mean square errors of shrinkage version of other estimators can be directly written.

REFERENCES

1. Hanif, M., I. Haq and M.Q. Shahbaz, 2009. On a new family of estimators using multiple auxiliary attributes. To Appear in World Applied Sciences Journal.
2. Haq, I., 2008. A chain ratio type estimator using two auxiliary attributes. In the Proc. of 4th Int. Conf. on Statistical Sciences, 15: 69-72.
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. Stat., 22 (1): 43-50.
4. Naik, V.D. and P.C. Gupta, 1996. A note on estimation of mean with known population of an auxiliary character. J. Indian Soc. Agri. Stat., 48 (2): 151-158.
5. Shahbaz, M.Q. and M. Hanif, 2009. A general Shrinkage estimator in Survey Sampling. World App. Sci. J., 7 (5): 593-596.