Middle-East Journal of Scientific Research 6 (2): 198-200, 2010
ISSN 1990-9233

Estimation of Population Mean in Two Phase Sampling

[1]Nadeem Shafique Butt, [1]Shahid Kamal and [2]Muhammad Qaiser Shahbaz

[1]College of Statistical and Actuarial Sciences University of the Punjab, Lahore, Pakistan
[2]Department of Mathematics, COMSATS Institute of IT, Lahore, Pakistan

Abstract: A new estimator for population mean has been proposed in two phase sampling by using information of multiple auxiliary variables. The minimum variance of the proposed estimator has been obtained. Comparison has also been made with some available estimators of two phase sampling.

Key words: Two phase sampling · Multiple auxiliary variables · Minimum variance

INTRODUCTION

The auxiliary information has always been a source of improvement in estimation of certain population characteristics. Several estimators have been developed in single and two phase sampling which utilizes information on auxiliary variables as well as auxiliary attributes. The classical estimators which use information on auxiliary variables are the ratio and regression estimators as given in Hansen, *et al.* [1]. The classical regression estimator of population mean is given as:

$$\bar{y}_{lr} = \bar{y} + \beta(\bar{X} - \bar{x}) \tag{1.1}$$

The value of β for which the variance of (1.1) is minimum is $\beta = S_{xy}/S^2_x$. The minimum variance of (1.1) is given as:

$$Var(\bar{y}_{lr}) = \theta S_y^2\left(1-\rho_{yx}^2\right) \tag{1.2}$$

Where

$\theta = n^{-1} - N^{-1}$ and ρ_{yx} is the correlation coefficient between X and Y. The estimator (1.1) in case of several auxiliary variables has been discussed by number of statisticians and the estimator in this case is given as:

$$\bar{y}_{mlr} = \bar{y} + \beta^{/}\left(\bar{\mathbf{X}} - \bar{\mathbf{x}}\right); \tag{1.3}$$

Where:

$\bar{\mathbf{x}}$ is vector of sample means for auxiliary variables. The variance of (1.3); reported by Ahmad [2] among others; is given as:

$$Var(\bar{y}_{mlr}) = \theta S_{\bar{y}}^2\left(1-\rho_{\bar{y}.\mathbf{x}}^2\right); \tag{1.4}$$

Where:

$\rho^2_{\bar{y}.\mathbf{x}}$ is the squared multiple correlation coefficient between Y and x. The classical regression estimator for two phase sampling is given by Hansen, *et al.* [1] as:

$$\bar{y}_{lr(2)} = \bar{y}_2 + \beta(\bar{x}_1 - \bar{x}_2); \tag{1.5}$$

Where:

$\bar{x}_1$ and $\bar{x}_2$ are first phase and second phase means of auxiliary variable *X* and $\bar{y}_2$ is second phase mean of *Y*. The variance of (1.5) is given as:

$$Var\left(\bar{y}_{lr(2)}\right) = S_{\bar{y}}^2\left\{\theta_2\left(1-\rho_{yx}^2\right)+\theta_1\rho_{yx}^2\right\}; \tag{1.6}$$

Where:

$\theta_h = n^{-1}{}_h - N^{-1}{}_h$ and n_h is sample size at h^{th} phase. Ahmed [2] has extended the (1.6) the case of several variables. Sahoo, *et al.* [3] has proposed the regression type estimator using information of two auxiliary variables. The estimator proposed by Sahoo, *et al.* [3] is given as:

$$\bar{y}_{ssm} = \bar{y}_2 + \beta_1(\bar{x}_1 - \bar{x}_2) + \beta_2\left(\bar{Z} - \bar{z}\right) \tag{1.7}$$

The variance of (1.7) is:

$$Var(\bar{y}_{ssm}) = S_{\bar{y}}^2\left\{\theta_2\left(1-\rho_{yx}^2\right)+\theta_1\left(\rho_{yx}^2-\rho_{yz}^2\right)\right\} \tag{1.8}$$

Corresponding Author: Nadeem Shafique Butt, College of Statistical and Actuarial Sciences University of the Punjab, Lahore, Pakistan. E-mail: nadeemshafique@hotmail.com & kamal_shahid@hotmail.com.

Where:

ρ_{yz}^2 is squared correlation coefficient between *Y* and *Z*.

Jhajj, *et al.* [4] have proposed a family of estimators in single and two phase sampling using information on auxiliary attributes. The variance of the proposed family depends upon the point bi-serial correlation coefficient. Samiuddin and Hanif [5] have also proposed several estimators in single and two phase sampling. A regression-in-ratio estimator proposed by Samiuddin and Hanif [5] is:

$$\bar{y}_{sh(2)} = \left[\bar{y}_2 + \beta_{yz}(\bar{z}_1 - \bar{z}_2)\right]\frac{\bar{X}}{\bar{x}_2}. \qquad (1.9)$$

The variance of (1.9) is:

$$MSE\left(\bar{y}_{sh(2)}\right) \approx \bar{Y}^2\left[\theta_2\left\{C_y^2\left(1-\rho_{xy}^2\right)+\left(C_x - C_y\,\rho_{xy}\right)^2\right\}\right.$$

$$\left.+(\theta_2-\theta_1)\left\{C_x^2\,\rho_{xz}^2 - \left(C_y\rho_{yz} - C_x\rho_{xz}\right)^2\right\}\right]. \qquad (1.10)$$

In this paper we have proposed a modified regression type estimator using information on several auxiliary variables.

Notations: In this section we define the notations used for the development of the estimator and its variance. Let **w** be a vector of auxiliary variables with covariance matrix Sw, *X* be another auxiliary variable and Y be the variable of interest.

Let s_{xw} be the vector of covariances between *X* and w, s_{yw} be the vector of covariances between *Y* and w. Using these notations we define $\alpha=S^{-1}{}_w s_{xw}$ as vector of regression coefficients between *X* and w, and $\gamma = S^{-1}{}_w s_{yw}$ as vector of regression coefficients between *Y* and w. We also define $\beta_{yx.w} = S_{xy.w}/S_{2.w}$ as partial regression coefficient between *Y* and *X* keeping the w at constant level. Also $S_{yx.\mathbf{w}} = S_{yx} - \mathbf{s}_{x\mathbf{w}}^{/}\mathbf{S}_{\mathbf{w}}^{-1}\mathbf{s}_{x\mathbf{w}}$ is partial covariance between *Y* and *X* after removing the effect of w, $S_{yx.\mathbf{w}} = S_{yx} - \mathbf{s}_{x\mathbf{w}}^{/}\mathbf{S}_{\mathbf{w}}^{-1}\mathbf{s}_{x\mathbf{w}}$ is the partial variance of *Y* and $S_{x.\mathbf{w}}^2 = S_x^2 - \mathbf{s}_{x\mathbf{w}}^{/}\mathbf{S}_{\mathbf{w}}^{-1}\mathbf{s}_{x\mathbf{w}}$ is the partial variance of *X*. We also define $\rho_{yx.\mathbf{w}}^2 = S_{yx.\mathbf{w}}^2 \Big/ \left(S_{x.\mathbf{w}}^2 S_{y.\mathbf{w}}^2\right)$ as partial correlation coefficient between *Y* and *X* after removing the effect of w, $\rho_{x.w}^2$ as squared multiple correlation coefficient between *Y* and combined effects of *X* and w, $\rho_{x.w}^2$ as squared multiple correlation coefficient between *Y* and combined effects of w.

Using the above notations we proposed the new estimators in the section 3.

The Proposed Estimator: We propose following unbiased estimator of population mean in two phase sampling using information of several auxiliary variables:

$$t_{nss} = \bar{y}_2 + k\left[\bar{x}_1 + \mathbf{a}^{/}(\bar{\mathbf{w}} - \bar{\mathbf{w}}_1) - \left\{\bar{x}_2 + \boldsymbol{ß}^{/}(\bar{\mathbf{w}} - \bar{\mathbf{w}}_2)\right\}\right] \qquad (3.1)$$

Using $\bar{y}_2 = \bar{Y} + \bar{e}_{y_2}, \bar{x}_1 = \bar{X} + \bar{e}_{x_1}, \bar{x}_2 = \bar{X} + \bar{e}_{x_2}, \bar{\mathbf{w}}_1 = \bar{\mathbf{w}} + \bar{\mathbf{e}}_{w_1}$ *and* $\bar{\mathbf{w}}_2 = \bar{\mathbf{w}} + \bar{\mathbf{e}}_{w_2}$

in (2.1) we have:

$$t_{nss} - \bar{Y} = \bar{e}_{y_2} + k\left[\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right) - \mathbf{a}^{/}\bar{\mathbf{e}}_{w_1} + \boldsymbol{ß}^{/}\bar{\mathbf{e}}_{w_2}^{/}\right]$$

Squaring and applying expectation, the variance of (3.1) is given as:

$$S = Var(t_{nss}) = \theta_2 s_y^2 + k^2\left[(\theta_2-\theta_1)s_x^2 + \theta_1\mathbf{a}^{/}\mathbf{S}_{\mathbf{W}}\mathbf{a} + \theta_2\boldsymbol{ß}^{/}\right.$$

$$\left.\mathbf{S}_{\mathbf{W}}\boldsymbol{ß} + 2(\theta_1-\theta_2)\boldsymbol{ß}^{/}\mathbf{s}_{x\mathbf{W}} - 2\theta_1\mathbf{a}^{/}\mathbf{S}_{\mathbf{W}}\boldsymbol{ß}\right] +$$

$$2k\left[(\theta_1-\theta_2)s_{yx} - \theta_1\mathbf{a}^{/}\mathbf{s}_{yw} + \theta_2\boldsymbol{ß}^{/}\mathbf{s}_{y\mathbf{W}}\right] \qquad (3.2)$$

The optimum values of α, β and *k* are obtained by minimizing (3.2). These values are obtained by solving following three equations, obtained by partially differentiating (3.2) and setting the derivative to zero.

$$2k\left[(\theta_2-\theta_1)s_x^2 + \theta_1\mathbf{a}^{/}\mathbf{S}_{\mathbf{W}}\mathbf{a} + \theta_2\boldsymbol{ß}^{/}\mathbf{S}_{\mathbf{W}}\boldsymbol{ß} + 2(\theta_1-\theta_2)\boldsymbol{ß}^{/}\mathbf{s}_{x\mathbf{W}} - 2\theta_1\mathbf{a}^{/}\mathbf{S}_{\mathbf{W}}\boldsymbol{ß}\right]$$

$$+2\left[(\theta_1-\theta_2)s_{yx} - \theta_1\mathbf{a}^{/}\mathbf{s}_{y\mathbf{W}} + \theta_2\boldsymbol{ß}^{/}\mathbf{s}_{y\mathbf{W}}\right] = 0 \qquad (I)$$

$$k\mathbf{S}_{\mathbf{W}}(\mathbf{a}-\boldsymbol{ß}) - \mathbf{s}_{y\mathbf{W}} = 0 \qquad (ii)$$

$$k\mathbf{S}_w(\theta_2\boldsymbol{ß} - \theta_1\mathbf{a}) - k(\theta_2-\theta_1)\mathbf{s}_{x\mathbf{W}} + \theta_2\mathbf{s}_{y\mathbf{W}} = \mathbf{0} \qquad (iii)$$

Solving the above equations simultaneously, the optimum values of α,β and *k* are:

$$\mathbf{a} = S_{\mathbf{W}}^{-1}s_{x\mathbf{W}},\ \boldsymbol{ß} = \mathbf{a} - k^{-1}\gamma \text{ and } k = \beta_{yx.\mathbf{W}} = S_{yx.\mathbf{W}}\Big/S_{x.\mathbf{W}}^2$$

Using the optimum values in (3.2) and simplifying, the variance of proposed estimator is:

$$Var\left(t_{nss}\right) = s_{y.\mathbf{W}}^2\left[\theta_2\left(1-\rho_{xy.\mathbf{W}}^2\right) + \theta_1\rho_{xy.\mathbf{W}}^2\right] \qquad (3.3)$$

Further, by using the fact that $s^2_{y.w} = S^2_y(1 - \rho^2_{xy.w})$ and utilizing the relationship that $1 - \rho^2_{y.xw} = (1 - \rho^2_{y.w})(1 - \rho^2_{yx.w})$ the variance of proposed estimator can be written as:

$$Var(t_{nss}) = s^2_{\bar{y}}\left\{\theta_2\left(1-\rho^2_{\bar{y}.x\mathbf{w}}\right)+\theta_1\rho^2_{\bar{x}y.\mathbf{w}}\left(1-\rho^2_{\bar{y}.\mathbf{w}}\right)\right\} \quad (3.4)$$

From (3.4) we can see that the variance of (3.1) depends upon the squared multiple and partial correlation coefficients. The estimator and its variance for multiphase sampling can be analogously written from (3.1) and (3.4). Specifically if a sample of size n_h is taken at h^{th} phase and a sample of n_q is taken at q_{th} phase with $n_q<n_h$, the estimator of the population mean is:

$$t_{nss} = \bar{y}2 + k\left[\bar{x}_h + \mathbf{a}'\left(\bar{\mathbf{w}} - \bar{\mathbf{w}}_h\right) - \left\{\bar{x}_q + \mathbf{ß}'\left(\bar{\mathbf{w}} - \bar{\mathbf{w}}_q\right)\right\}\right] \quad (3.5)$$

The variance of (3.5) can be written from (3.4) as:

$$Var(t_{nss}) = s^2_{\bar{y}}\left\{\theta_q\left(1-\rho^2_{\bar{y}.x\mathbf{w}}\right)+\theta_h\rho^2_{\bar{x}y.\mathbf{w}}\left(1-\rho^2_{\bar{y}.\mathbf{w}}\right)\right\} \quad (3.6)$$

For practical applicability, the proposed estimator can be easily modified by using the sample estimates in place of population parameters. The consistent estimate of population mean can be straight-away written as:

$$t_{nss} = \bar{y}2 + b_{yx.\mathbf{w}}(\bar{x}1 - \bar{x}2) + b_{yx.\mathbf{w}}\mathbf{b}_{x\mathbf{w}}(\bar{\mathbf{w}}2 - \bar{\mathbf{w}}1) + \mathbf{b}'_{y\mathbf{w}}(\bar{\mathbf{w}} - \bar{\mathbf{w}}2) \quad (3.7)$$

The estimated standard error of (3.1) is given as:

$$S.E(t_{nss}) = s_y\sqrt{\theta_2\left(1-r^2_{\bar{y}.x\mathbf{w}}\right)+\theta_1 r^2_{\bar{y}x.\mathbf{w}}\left(1-r^2_{\bar{y}.\mathbf{w}}\right)} \quad (3.8)$$

Using (3.7) and (3.8), the confidence interval for true population mean can be constructed.

Comparison with Available Estimators: Ahmed [2] has proposed various estimators for two phase and multiphase sampling using information son several auxiliary variables. We have compared the estimator (3.1) with following estimator given in Ahmed [2]:

$$t_1 = y2 + \sum_{i=1}^{r}\alpha_i\left(\bar{W}_i - \bar{w}_i1\right) + \sum_{i=1}^{r}\beta_i\left(\bar{W}_i - \bar{w}_i2\right) + \sum_{i=r+1}^{p}\beta_i\left(\bar{w}_i1 - \bar{w}_i2\right)$$

The variance of above estimator is:

$$Var(t_1) = S^2_{\bar{y}}\left[\theta_2\left(1-\rho^2_{\bar{y}.\mathbf{w}}\right)+\theta_1\left(\rho^2_{\bar{y}.\mathbf{w}} - \rho^2_{y.\mathbf{w}_1}\right)\right] \quad (4.1)$$

Where:

$\rho^2_{y.w}$ is squared multiple correlation between Y and combined effect of all auxiliary variables and $\rho^2_{y.w1}$ is the squared multiple correlation between Y and first r auxiliary variables. Now comparing (3.4) with (4.1) gives:

$$Var(t_1) - Var(t_{nss}) = \theta_2\left(\rho^2_{\bar{y}.x\mathbf{w}} - \rho^2_{\bar{y}.\mathbf{w}}\right) + \theta_1\left[\rho^2_{\bar{y}.\mathbf{w}}\left(1-\rho^2_{\bar{y}x.\mathbf{w}}\right) - \rho^2_{\bar{y}x.\mathbf{w}} - \rho^2_{y.\mathbf{w}_1}\right] > 0 \quad (4.2)$$

From (4.2) we can readily see that the proposed estimator performs well as compared with the estimator proposed by Ahmed [2].

REFERENCES

1. Hansen, M.H., Hurwitz, W.N., and W.G. Madow, 1953. Sample Survey Methods and Theory (Vol. II): John Wiley.
2. Ahmed, Z., 2008. Generalized Ratio and Regression Estimators in Muliphase Sampling. Unpublished PhD thesis.
3. Sahoo, J., L.N. Sahoo and S. Mohanty, 1993. A regression approach to estimation in two phase sampling using two auxiliary variables. Current Sciences, 65(1): 73-75.
4. Jhajj, H.S., M.K. Sharma and L.K. Grover, 2006. A familty of estimators of population mean using information of auxiliary attribute. Pak. J. Stat., 22(1): 43-50.
5. Samiuddin, M. and M. Hanif, 2007. Estimation of population mean in single and two phase sampling with or without additional information. Pak. J. Stat., 23(2): 99-118.