Pak. J. Statist.
2010, Vol. 26(1) 195-202

SOME IMPROVED ESTIMATORS IN MULTIPHASE SAMPLING

Muhammad Hanif[1], **Muhammad Qaiser Shahbaz**[2], and **Zahoor Ahmad**[3]

[1] Lahore University of Management Sciences, Lahore, Pakistan
Email: hanif@lums.edu.pk
[2] Department of Mathematics, COMSATS Institute of IT, Lahore, Pakistan
Email: qshahbaz@gmail.com
[3] Department of Statistics, University of Gujrat, Gujrat, Pakistan
Email: zahoorstat@yahoo.com

ABSTRACT

Some improved estimators of population mean has been proposed in two phase and multiphase sampling using information on two and several auxiliary variables. The minimum variance of the proposed estimators has been obtained. Comparison of the proposed estimators has been done with some available estimators of two phase sampling that utilizes information of two and several auxiliary variables.

KEYWORDS

Regression estimators, correlation coefficients, two phase sampling, optimum variance.

1 INTRODUCTION

The use of auxiliary information has always been a source of improvement in estimation of certain population characteristics. The auxiliary variables that have string relationship with the estimand variable always improve the precision of the estimates resulting in smaller standard error of the estimate. The use of auxiliary variables in survey sampling has very old history. Several regression type estimators are available in literature that uses the auxiliary information to increase the precision of the estimates. The historical use of the auxiliary variables can be found in the regression estimator given by Hansen et al (1953) as:

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

where β is the regression coefficient of Y on X. The variance of classical regression estimator is:

$$Var(\bar{y}_{lr}) = \theta S_y^2 \left(1-\rho_{yx}^2\right); \tag{1.1}$$

where $\theta = n^{-1} - N^{-1}$ and ρ_{yx} is the correlation coefficient between X and Y. The use of several auxiliary variable in the context of regression estimator has been discussed by Ahmad (2008). The multiple regression estimator has the general form:

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

where β is the regression coefficient of Y on X. The variance of classical regression estimator is:

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

where $\rho_{y.\mathbf{x}}^2$ is the squared multiple correlation coefficient between Y and the combined effect of all the auxiliary variables. The regression estimator has been effectively used in two phase and multiphase sampling. The estimator of mean in two phase sampling has been discussed by Hansen et al (1953) and is given as:

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

where $\bar{x}_1$ and $\bar{x}_2$ are first phase and second phase mean of auxiliary variable X based upon the samples of sizes $n_1$and n_2 respectively; $\bar{y}_2$ is mean of Y for the second phase sample of size n_2. The variance of regression estimator for two phase sampling is given as:

$$Var\left(\bar{y}_{lr(2)}\right) = S_y^2\left\{\theta_2\left(1-\rho_{yx}^2\right) + \theta_1\rho_{yx}^2\right\}; \tag{1.3}$$

where $\theta_h = n_h^{-1} - N^{-1}$ and n_h is sample size at hth phase. A slightly modified regression type estimator has been proposed by Sahoo et al (1993) by using information of two auxiliary variables. The proposed estimator is:

$$\bar{y}_{ssm} = \bar{y}_2 + \beta_1(\bar{x}_1 - \bar{x}_2) + \beta_2(\bar{W} - \bar{w}_1);$$

The variance of estimator given by Sahoo et al (1993) is:

$$Var(\bar{y}_{ssm}) = S_y^2\left\{\theta_2\left(1-\rho_{yx}^2\right) + \theta_1\left(\rho_{yx}^2 - \rho_{yw}^2\right)\right\}; \tag{1.4}$$

where ρ_{yw}^2 is squared correlation coefficient between Y and W. Kiregyera (1984) has proposed the following regression-in-regression estimator:

$$\bar{y}_K = \bar{y}_2 + \beta_{yx}\left[(\bar{x}_1 - \bar{x}_2) + \beta_{wx}(\bar{W} - \bar{w}_1)\right];$$

The variance of $\bar{y}_K$ is:

$$Var(\bar{y}_K) = S_y^2\left[\theta_2\left(1-\rho_{yx}^2\right)+\theta_1\left(\rho_{yx}^2+\rho_{yx}^2\rho_{wx}^2-2\rho_{yx}\rho_{yw}\rho_{wx}\right)\right]; \tag{1.5}$$

Roy (2003) has given the following regression type estimator:

$$\bar{y}_R = \bar{y}_2+\alpha\left[\bar{x}_1+\beta\left(\bar{W}-\bar{w}_1\right)-\left\{\bar{x}_2+\gamma(\bar{W}-\bar{w}_2)\right\}\right];$$

The variance of estimator given by Roy (2003) is:

$$Var(\bar{y}_R) = S_y^2\left[\theta_2\left(1-\rho_{y.wx}^2\right)+\theta_1\rho_{yx.w}^2\left(1-\rho_{yw}^2\right)\right]; \tag{1.6}$$

where $\rho_{y.w\dot{x}}^2$ is the squared multiple correlation between Y and combined effects of X and W, $\rho_{yx.w}^2$ is the squared partial correlation between Y and X keeping W at a constant level and ρ_{yw}^2 is the squared correlation between Y and W. Another regression type estimator has been proposed by Samiuddin and Hanif (2007). The estimator is given as:

$$\bar{y}_{sh(2)} = \bar{y}_2+\alpha\left(\bar{x}_1-\bar{x}_2\right)+\beta\left(\bar{W}-\bar{w}_1\right)+\gamma(\bar{W}-\bar{w}_2); \tag{1.7}$$

Samiuddin and Hanif (2007) has also proposed the following ratio type estimators using two auxiliary variables:

$$\bar{y}_{sh(2)} = \bar{y}_2\left(\frac{\bar{X}}{\bar{x}_1}\right)^{\alpha_1}\left(\frac{\bar{X}}{\bar{x}_2}\right)^{\alpha_2}\left(\frac{\bar{z}}{\bar{z}_2}\right)^{\alpha_3}; \tag{1.8}$$

Samiuddin and Hanif (2007) has shown that the variance of (1.7) and (1.8) is same as the variance of Roy's estimator given in (1.6). Mukerjee et al (1987) has also proposed a regression type estimator using two auxiliary variables. The estimator is:

$$\bar{y}_M = \bar{y}_2+\beta_{yx}\left(\bar{x}_1-\bar{x}_2\right)+\beta_{yw}\left(\bar{w}_1-\bar{w}_2\right);$$

The variance of $\bar{y}_M$ is given as:

$$Var(\bar{y}_M) = S_y^2\left[\theta_2-\left(\theta_2-\theta_1\right)\left(\rho_{yw}^2+\rho_{yx}^2-2\rho_{yx}\rho_{yw}\rho_{wx}\right)\right]; \tag{1.9}$$

Chand (1975) has proposed various ratio type estimators. One of the estimator proposed by Chand (1975) is given as:

$$\bar{y}_C = \bar{y}_2\frac{\bar{x}_2\bar{w}_1}{\bar{x}_1\bar{W}};$$

The variance of above estimator is of complex form. In this paper we have proposed a modified regression type estimator using information on several auxiliary variables.

with variance given as:

$$Var(\bar{y}_{sh1}) = S_y^2\left[\theta_h\left(1-\rho^2_{y.wx}\right)+\theta_m\rho^2_{yx.w}\left(1-\rho^2_{yw}\right)\right]; \tag{2.5}$$

A consistent estimator of population mean can be easily written from (2.1) as:

$$\bar{y}_{sh1} = \bar{y}_2 + b_{yx.w}(\bar{x}_1-\bar{x}_2)+b_{yw.x}(\bar{w}_1-\bar{w}_2)+b_{yw}(\bar{W}-\bar{w}_1); \tag{2.6}$$

with estimator of variance given as:

$$var(\bar{y}_{sh1}) = s_y^2\left[\theta_2\left(1-r^2_{y.wx}\right)+\theta_1 r^2_{yx.w}\left(1-r^2_{yw}\right)\right]; \tag{2.7}$$

The confidence interval for population mean can be easily constructed by using (2.6) and (2.7).

3 THE PROPOSED ESTIMATOR-2

We propose the following estimator of population mean by using information of several auxiliary variables:

$$\bar{y}_{sh2} = \bar{y}_2 + \alpha'(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)+\beta'\left(\bar{\mathbf{W}}-\bar{\mathbf{w}}_1\right); \tag{3.1}$$

where $\bar{\mathbf{x}}_1$ is vector of first phase mean of p auxiliary variables X, $\bar{\mathbf{x}}_2$ is vector of second phase mean of p auxiliary variables X, $\bar{\mathbf{w}}_1$ is vector of first phase mean of q auxiliary variables W and $\bar{\mathbf{W}}$ is the vector of population means for variable $W's$. The vectors α and β are the vectors of unknown parameters whose values are to be determined so that the variance of (3.1) is minimum. Now, using $\bar{y}_2 = \bar{Y}+\bar{e}_{y_2}, \bar{\mathbf{x}}_1 = \bar{\mathbf{X}}+\bar{\mathbf{e}}_{x_1}, \bar{\mathbf{x}}_2 = \bar{\mathbf{X}}+\bar{\mathbf{e}}_{x_2}, \bar{\mathbf{w}}_1 = \bar{\mathbf{W}}+\bar{\mathbf{e}}_{w_1}$ in (3.1) we have:

$$\bar{y}_{sh2}-\bar{Y} = \bar{e}_{y_2}+\alpha'(\bar{\mathbf{e}}_{x_1}-\bar{\mathbf{e}}_{x_2})-\beta'\bar{\mathbf{e}}_{w_1};$$

Squaring and applying expectation, the variance of $\bar{y}_{sh2}$ is given as:

$$\begin{aligned} Var(\bar{y}_{sh2}) &= \theta_2 S_y^2 + (\theta_2-\theta_1)\alpha'\mathbf{S}_x\alpha+\theta_1\beta'\mathbf{S}_w\beta \\ &\quad -2(\theta_2-\theta_1)\alpha'\mathbf{s}_{xy}-2\theta_1\beta'\mathbf{s}_{wy}. \end{aligned} \tag{3.2}$$

where $\mathbf{S}_x$ is the covariance matrix of $\mathbf{x}$, $\mathbf{S}_w$ is the covariance matrix of $\mathbf{w}$, $\mathbf{s}_{xy}$ is vector of covariances between Y and $\mathbf{x}$ and $\mathbf{s}_{wy}$ is vector of covariances between Y and $\mathbf{w}$. Partially differentiating (3.2) *w.r.t.*α and β and setting the derivatives to zero we obtain $\alpha = \mathbf{S}_x^{-1}\mathbf{s}_{xy}$ and $\beta = \mathbf{S}_w^{-1}\mathbf{s}_{wy}$. Further, by using the value of α and β in (3.2), the variance of (3.1) is:

$$Var(\bar{y}_{sh2}) = S_y^2\left[\theta_2\left(1-\rho^2_{y.\mathbf{x}}\right)+\theta_1\left(\rho^2_{y.\mathbf{x}}-\rho^2_{y.\mathbf{w}}\right)\right]; \tag{3.3}$$