VARIOUS METHODS OF ESTIMATOR CONSTRUCTION IN MULTIPHASE SAMPLING

Saman Hanif and Muhammad Qaiser Shahbaz
Department of Statistics, Faculty of Sciences, King Abdul Aziz University, Jeddah, Saudi Arabia
saman.hanif@hotmail.com, qshahbaz@gmail.com

ABSTRACT: In this short paper we have shown that in multiphase sampling the regression type estimator is based upon two phases only. We have proved the result by considering various construction methods.

Key Words: *Auxiliary Information, Three and Four Phase Sampling*

1. INTRODUCTION

Often we deal with situations where information on sampling frame is not available and so conventional methods are no more useful. In these situations the multiphase sampling can be used as an alternate. The situation can be set up by considering that a finite population has N differentiable units but information about sampling frame is not available and hence the popular designs like simple random sampling, stratified random sampling and others are not applicable. The multiphase sampling design is an alternative design that can be used in such situations.

The notations of multiphase sampling are different from conventional sampling designs in that sample is selected in phases and at each phase information is collected about some auxiliary variables that can increase the efficiency of estimation of variable of interest. The information from main variable of interest is collected at the last phase only. The main theme is given by presenting common notations. Let n_j; $j = 1, 2, ..., p$; be the size of sample drawn from a population of size N at j–th phase, $\bar{x}_j$; $j = 1, 2, ..., p$; be the mean of an auxiliary variable based upon the sample at j–th phase; that is $\bar{x}_j = n_j^{-1}\sum_{i=1}^{n_j} x_{ij}$ and $\bar{y}_p$ be the mean of main variable of interest based upon the last phase sample. Some other notations are described below:

$$\left.\begin{aligned}&\theta_j = 1/n_j - 1/N\ ; \bar{x}_j = \bar{X} + \bar{e}_{x_j},\ j = 1, 2, ..., p;\\ &\bar{y}_q = \bar{Y} + \bar{e}_{y_q}; E\left(\bar{e}_{y_q}^2\right) = \theta_j S_y^2\ ;\ E\left(\bar{e}_{x_j}^2\right) = \theta_j S_x^2;\\ &E\left(\bar{e}_{y_q}\bar{e}_{x_j}\right) = \theta_j S_{xy};\\ &E\left(\bar{e}_{x_h} - \bar{e}_{x_j}\right)^2 = \left(\theta_j - \theta_h\right)S_x^2\ ;\ h < j\end{aligned}\right\} \quad (1.1)$$

The description given in [2] and [5] will be followed about the quantities $\bar{e}_{y_q}$ and $\bar{e}_{x_j}$ and we assume that

these quantities are very small. Several two phase estimators have been proposed from time to time. The following estimator is proposed by [5] for two phase sampling using two auxiliary variables:

$$t_1 = \left[\bar{y}_2 + b_{yx}\left(\bar{x}_1 - \bar{x}_2\right)\right]\frac{\bar{z}_1}{\bar{z}_2} \quad (1.2)$$

Another estimator proposed by [5] is:

$$t_2 = \left[\bar{y}_2 + b_{y1}\left(\bar{z}_1 - \bar{z}_2\right)\right]\frac{\bar{X}}{\bar{x}_2} \quad (1.3)$$

Two chain–ratio type estimators have been proposed by [1] and are given as:

$$t_3 = y_2\frac{\bar{z}_1}{\bar{z}_2}\cdot\frac{\bar{X}}{\bar{x}_1}\ \ and\ \ t_4 = y_2\frac{\bar{z}_2}{\bar{z}_1}\cdot\frac{\bar{x}_1}{\bar{X}} \quad (1.4)$$

In [4] three estimators have been proposed for use in two phase sampling. These estimators are ratio–in–regression and regression–in–ratio type estimators. The first estimator proposed in [4] is a regression–in–ratio type estimator and is given as:

$$t_5 = \frac{\bar{y}_2}{\bar{z}_2}\left[\bar{z}_1 + b_{zx}\left(\bar{X} - \bar{x}_1\right)\right] \quad (1.5)$$

Another estimator proposed in [4] is a ratio–in–regression and is given as:

$$t_6 = \bar{y}_2 + b_{yz}\left(\frac{\bar{z}_1}{\bar{x}_1}\bar{X} - \bar{z}_2\right) \quad (1.6)$$

A third estimator proposed in [4] is a regression–in–regression type estimator and is given as:

$$t_7 = \bar{y}_2 + b_{yz}\left\{\left(\bar{z}_1 - \bar{z}_2\right) - b_{zx}\left(\bar{x}_1 - \bar{X}\right)\right\} \quad (1.7)$$

Several other estimators have been proposed in literature. Some other notable references are [6], [7], [8], [9] and [10] among many others.

2. MAIN RESULTS

In this section we present main results for estimation in multiphase sampling. These results have been given by considering two different regression type estimators in multiphase sampling design.

The first estimator is given as:

$$\bar{y}' = \bar{y}_q + \sum_{j=1}^{q-1}\beta_j\left(\bar{x}_j - \bar{x}_{j+1}\right) \quad (2.1)$$

Using the notations given in (1.1), the estimator (2.1) can be written as:

$$\bar{y}' = \left(\bar{Y} + \bar{e}_{y_q}\right) + \sum_{j=1}^{q-1}\beta_j\left\{\left(\bar{X} + \bar{e}_{x_j}\right) - \left(\bar{X} + \bar{e}_{x_{j+1}}\right)\right\}$$

$$\bar{y}' - \bar{Y} = \bar{e}_{y_q} + \sum_{j=1}^{q-1}\beta_j\left(\bar{e}_{x_j} - \bar{e}_{x_{j+1}}\right)$$

Squaring and applying the expectation, the mean square error; by using (1.1); of (2.1) is given as:

$$S = MSE\left(\bar{y}'\right) = E\left(\bar{y}' - \bar{Y}\right)^2$$
$$= \theta_q S_y^2 + \sum_{j=1}^{q-1}\beta_j^2\left(\theta_{j+1}-\theta_j\right)S_x^2$$
$$-2\sum_{j=1}^{q-1}\beta_j\left(\theta_{j+1}-\theta_j\right)S_{xy} \qquad (2.2)$$

The optimum values of $\beta_j's$ are obtained by partially differentiating (2.2) with respect to these unknowns. These derivatives are:

$$\frac{\partial S}{\partial \beta_j} = 2\left(\theta_{j+1}-\theta_j\right)\beta_j S_x^2 - 2\left(\theta_{j+1}-\theta_j\right)S_{xy} \;;\; j = 1,2,\cdots,(q-1) \qquad (2.3)$$

Now, setting above equations equal to zero we get $\beta_j = \beta = S_{xy}/S_x^2$. Using these values in (2.1), the estimator become:

$$\bar{y}' = \bar{y}_q + \beta\left(\bar{x}_1 - \bar{x}_q\right) \qquad (2.4)$$

Again using the value of $\beta_j's$ in (2.2), the mean square error of (2.4) is:

$$S = MSE\left(\bar{y}'\right) = \theta_q S_y^2 + \beta^2\left(\theta_q - \theta_1\right)S_x^2$$
$$-2\beta\left(\theta_q-\theta_1\right)S_{xy}$$
$$= \theta_q S_y^2\left(1-\rho^2\right)+\theta_1 S_y^2 \qquad (2.5)$$

From (2.4) and (2.5), we can see that the estimator and its mean square error is based upon only on two phases which are first and last phase.

We again consider another formation of the estimator for use in multiphase sampling. This formation is given as:

$$\bar{y}' = \bar{y}_q + \sum_{j=1}^{q-1}\beta_j\left(\bar{x}_j - \bar{x}_q\right) \qquad (2.6)$$

Using (1.1) in (2.6), the estimator can be written as:

$$\bar{y}' = \left(\bar{Y} + \bar{e}_{y_q}\right) + \sum_{j=1}^{q-1}\beta_j\left\{\left(\bar{X}+\bar{e}_{x_j}\right)-\left(\bar{X}+\bar{e}_{x_q}\right)\right\}$$
$$\bar{y}' - \bar{Y} = \bar{e}_{y_q} + \sum_{j=1}^{q-1}\beta_j\left(\bar{e}_{x_j}-\bar{e}_{x_q}\right)$$

Squaring and applying expectation, the mean square error of (2.6) is:

$$S = MSE\left(\bar{y}'\right) = E\left(\bar{y}'-\bar{Y}\right)^2$$
$$= \theta_q S_y^2 + \sum_{j=1}^{q-1}\beta_j^2\left(\theta_q-\theta_j\right)S_x^2 - \sum_{j=1}^{q-1}\beta_j\left(\theta_q-\theta_j\right)S_{xy}$$
$$+2\sum_{j=1}^{q-2}\sum_{\substack{h=1\\h>j}}^{q-1}\beta_j\beta_h\left(\theta_q-\theta_h\right)S_x^2 \qquad (2.7)$$

Differentiating (2.7) with respect to the unknown coefficients $\beta_j's$, we have:

$$\frac{\partial S}{\partial \beta_j} = 2\beta_j\left(\theta_j-\theta_j\right)S_x^2 - 2\left(\theta_q-\theta_j\right)S_{xy}$$
$$+2\sum_{\substack{h=1\\h\neq j}}^{q-2}\beta_h\left(\theta_q-\theta_h\right)S_x^2 \;;\; j=1,2,\cdots,q-2 \qquad (2.8)$$
$$= 2\beta_{q-1}\left(\theta_q-\theta_{q-1}\right)S_x^2 - 2\left(\theta_q-\theta_{q-1}\right)S_{xy}$$
$$+2\left(\theta-\theta_{q-1}\right)\sum_{h=1}^{q-2}\beta_h S_x^2 \qquad (2.9)$$

Now setting (2.9) equals to zero, we have:

$$\beta_{q-1}S_x^2 + \sum_{h=1}^{q-2}\beta_h S_x^2 = S_{xy} \Rightarrow \beta_{q-1} = S_{xy}/S_x^2 - \sum_{h=1}^{q-2}\beta_h \qquad (2.10)$$

Now by backward substitution in (2.8) we have:

$$\beta_j = S_{xy}/S_x^2 - \sum_{h=1}^{j-1}\beta_h \quad ; \quad j = q-2, q-1, \cdots, 2 \qquad (2.11)$$

Now using (2.10) in (2.8) we have $\beta_1 = \beta = S_{xy}/S_x^2$ and so $\beta_2 = \beta_3 = \cdots = \beta_{q-1} = 0$ and hence the estimator becomes:

$$\bar{y}' = \bar{y}_q + \beta\left(\bar{x}_1 - \bar{x}_q\right)$$

with mean square error:

$$S = MSE\left(\bar{y}'\right) = \theta_q S_y^2 + \beta^2\left(\theta_q-\theta_1\right)S_x^2 - 2\beta\left(\theta_q-\theta_1\right)S_{xy}$$
$$= \theta_q S_y^2\left(1-\rho^2\right)+\theta_1 S_y^2 \qquad (2.12)$$

From these developments we have deduced the following conclusion.

3. CONCLUSION

In this short article we have considered the estimation problem in multiphase sampling. For this we have developed two different versions of the regression type estimator. These two estimators are described in (2.1) and (2.6). These two estimators are optimized to obtain the value of unknown coefficients. The optimization in both these estimators has lead us to the same conclusion and based upon these findings we have concluded that the following regression type estimator can be used for estimation of population mean in multiphase sampling:

$$\bar{y}' = \bar{y}_q + \beta\left(\bar{x}_1 - \bar{x}_q\right)$$

with $MSE\left(\bar{y}'\right) = \theta_q S_y^2\left(1-\rho^2\right)+\theta_1 S_y^2$

from these we can conclude that the regression type estimator in two phase sampling can only be based upon the first and last phase.

REFERENCES:

1. Chand, L. (1975). *Some Ratio-type Estimators based on two or more Auxiliary Variables.* Unpublished Ph.D. Dissertation. Iowa State University, Iowa.
2. Cochran (1977). *Sampling techniques.* J. Wily New York

3. Kiregyera, B. (1980). A Chain Ratio-type Estimator in Finite Population Double Sampling using two Auxiliary Variables. *Metrika*. 27, 217-223.
4. Kiregyera, B. (1984). Regression-type Estimator using two Auxiliary Variables and Model of Double Sampling from Finite Populations. *Metrika*. 31, 215-226.
5. Mohanty, S. (1967). Combination of Regression and Ratio Estimate. *Jour. Ind. Statist. Asso*. 5,16-19.
6. Roy, D. C. (2003). A regression type estimates in two-phase sampling using two auxiliary variables. *Pak. J. Statist.* Vol. 19(3), 281-290.
7. Sahoo, J., Sahoo, L. N. and Mohanty, S. (1993). A Regression Approach to Estimation in Two-phase Sampling using two Auxiliary Variables. *Current Sciences*. 65, 1, 73-75.
8. Srivastava, S.K. (1970). A two phase sampling estimator in sample survey. *Austrl. J. Statist.* 12(1), 23-27.
9. Srivastava, S.K. (1980). A generalized two-phase sampling estimator. *J. Indian. Soc. Agri. Statist.* 33,33-46.
10. Srivastava, S.K. (1981). A generalized two-phase sampling estimator. *J. Ind. Soc. Agri. Statist.* 33(1), 38-46.