Jour. Ind. Soc. Ag. Statistics
54(3), 2001 : 370-375

A Generalised Chain Estimator for Finite Population Mean in Two Phase Sampling

Arun K. Singh, Housila P. Singh[1] and Lakshmi N. Upadhyaya[2]
Nagaland University, Medziphema–797 106
(Received : May, 2001)

SUMMARY

This paper proposes a generalized class of chain estimators for finite population mean using two auxiliary variates in two phase sampling and analyses its properties.

Key words : Auxiliary character, Study character, Two phase sampling, Mean squared error.

1. Introduction

Consider a finite population $U = \{U_1, U_2, ..., U_i, ..., U_N\}$ Let y and x be the study and auxiliary variables, taking values y_i and x_i respectively for the i^{th} unit. When the two variables are strongly related but the population mean $\overline{X}$ of x is not known, we seek to estimate the population mean $\overline{Y}$ of y from a sample s_n, obtained through a two phase selection. Allowing simple random sampling without replacement scheme in each phase, the double sampling scheme will be as follows

(a) The first phase sample s'_n $(s'_n \subset U)$ of fixed size n' is drawn to observe only x in order to obtain a good estimate of $\overline{X}$.

(b) Given s'_n, the second phase sample s_n $(s_n \subset s'_n)$ of fixed size n is drawn to observe y only.

Sometimes even if $\overline{X}$ is unknown, information on a second auxiliary variable z closely related to x but compared to x remotely related to y (i. e. $\rho_{yx} > \rho_{yz}$) is readily available. This type of situation has been briefly

1. School of Studies in Statistics, Vikram University, Ujjain – 456 010, M. P.
2. Indian School of Mines, Dhanbad – 826 004, Jharkhand

discussed by, among others, Chand [1], Kiregyera ([3], [4]) and Sahoo and Sahoo [6]. Let $\overline{Z}$, be the population mean of second auxiliary variable z. Let $\overline{y} = \sum_{i=1}^{n} y_i \Big/ n$, $\overline{x} = \sum_{i=1}^{n} x_i \Big/ n$ be the unbiased estimators of $\overline{Y}$ and $\overline{X}$, the population mean of y and x, respectively, based on s_n and let $\overline{x}' = \sum_{i=1}^{n'} x_i \Big/ n'$ and $\overline{z}' = \sum_{i=1}^{n'} z_i \Big/ n'$ be the unbiased estimators of population means $\overline{Y}$ and $\overline{Z}$ respectively based on s_n'.

By analogy, if the correlation x and z is highly positive, $\left(\frac{\overline{x}'}{\overline{z}'}\right)\overline{Z}$ will estimate $\overline{X}$ more precisely than $\overline{x}'$. Accordingly, Chand [1] has chained $\left(\frac{\overline{x}'}{\overline{z}'}\right)\overline{Z}$ into $\frac{\overline{y}}{\overline{x}}$ and developed a chain ratio-type estimator

$$\overline{y}_1 = \left(\frac{\overline{y}}{\overline{x}}\right)\left(\frac{\overline{x}'}{\overline{z}'}\right)\overline{Z} \tag{1.1}$$

By chaining a regression estimator $\overline{x}' + b'_{xz}\left(\overline{Z} - \overline{z}'\right)$ of $\overline{X}$ into $\frac{\overline{y}}{\overline{x}}$, Kiregyra [3] derived a chain ratio-to-regression estimator

$$\overline{y}_2 = \left(\frac{\overline{y}}{\overline{x}}\right)\left[\overline{x}' + b'_{xz}\left(\overline{Z} - \overline{z}\right)\right] \tag{1.2}$$

where $b'_{xz} = \sum_{i=1}^{n'} (x_i - \overline{x})(z_i - \overline{z}') \Big/ \sum_{i=1}^{n'} (z_i - \overline{z}')^2$ is the estimate of population regression coefficient of x on z. Kiregyera [4] also extended this formulation to obtain a ratio-in-regression estimator

$$\overline{y}_3 = \overline{y} + b_{yx}\left[\left(\frac{\overline{x}'}{\overline{z}'}\right)\overline{Z} - \overline{x}\right] \tag{1.3}$$

where $b_{yx} = \sum_{i=1}^{n} (y_i - \overline{y})(x_i - \overline{x}) \Big/ \sum_{i=1}^{n} (x_i - \overline{x})^2$ is the estimate of population regression coefficient of y on x.

Motivated by Das and Tripathi [2] we have suggested a class of chain regression estimators of $\overline{Y}$ and discussed its properties.

2. The Proposed Class of Estimators and its Properties

Following Das and Tripathi [2], we suggest a class of ratio-in-regression type estimators for population mean $\overline{Y}$ as

$$\overline{y}_t = \overline{y} + b_{yx}\left[\frac{\left\{\overline{x}' - t_1\left(\overline{z}' - \overline{Z}\right)\right\}}{\left\{\overline{z}' - t_2\left(\overline{z}' - \overline{Z}\right)\right\}^{\alpha}}\left(\overline{Z}\right)^{\alpha} - \overline{x}\right] \tag{2.1}$$

where α is a suitably chosen constant and t_1 and t_2 are suitably chosen statistics such that their means exist (which may in particular be constant).

The mean squared error of $\overline{y}_t$, to the first degree of approximation, is given by

$$\mathrm{MSE}\left(\overline{y}_t\right) = \overline{Y}^2\left[\lambda C_y^2 - (\lambda - \lambda')C^2 C_x^2 + \lambda' C C_z^2 \gamma\left(\gamma C - 2C^*\right)\right], \tag{2.2}$$

where $C = \rho_{yx}\dfrac{C_y}{C_x}$, $C^* = \rho_{yz}\dfrac{C_y}{C_z}$, $\rho_{yx} = \dfrac{S_{yx}}{\left(S_y S_x\right)}$, $\rho_{yz} = \dfrac{S_{yz}}{\left(S_y S_z\right)}$

$$S_{yv} = \sum_{i=1}^{N}\left(y_i - \overline{Y}\right)\left(v_i - \overline{V}\right);\; v = x, z$$

$$S_u^2 = \sum_{i=1}^{N}\left(u_i - \overline{U}\right)^2 \Big/ (N-1);\; u = x, y, z$$

$$C_u = \frac{S_u}{\overline{U}};\; u = x, y, z \qquad \lambda = \frac{(N-n)}{Nn} \qquad \lambda' = \frac{(N-n')}{Nn'}$$

$$\gamma = \left[R\left(E_0 t_1\right) + \alpha\left(1 - \left(E_0 t_2\right)\right)\right] \qquad R = \frac{\overline{Z}}{\overline{X}}$$

$$Et_i = \left(E_0 t_i\right) + 0\left(n^{-q_i}\right),\; q_i > 0,\; i = 1, 2$$

and $\left(E_0 t_i\right)$ is a constant (parameter) not depending on n.

The $\mathrm{MSE}\left(\overline{y}_t\right)$ at (2.2) is minimised for

$$\gamma = \left(\frac{C^*}{C}\right) \tag{2.3}$$

Thus the resulting (minimum) MSE of $\bar{y}_t$ is given by

$$\min.\,MSE(\bar{y}_t) = \bar{Y}^2\left[\lambda C_y^2 - (\lambda - \lambda')C^2C_x^2 - \lambda' C_z^2 C^{*2}\right] \tag{2.4}$$

A large number of estimators may be identified as particular cases of the suggested class of estimators $\bar{y}_t$. Few examples are

$$\bar{y}_t^{(1)} = \bar{y} + b_{yx}\left\{\bar{x}'\left(\frac{\bar{Z}}{\bar{z}'}\right) - \bar{x}\right\}$$

$$\bar{y}_t^{(2)} = \bar{y} + b_{yx}\left[\bar{x}'\left\{2 - \left(\frac{\bar{Z}}{\bar{z}'}\right)^{\alpha_1}\right\} - \bar{x}\right]$$

$$\bar{y}_t^{(3)} = \bar{y} + b_{yx}\left\{\bar{x}'\left(\frac{\bar{Z}}{\left\{\bar{Z} + \alpha_1\left(\bar{z}' - \bar{Z}\right)\right\}}\right) - \bar{x}\right\}$$

$$\bar{y}_t^{(4)} = \bar{y} + b_{yx}\left[\left\{\alpha_1\,\bar{x}' + (1 - \alpha_1)\bar{x}'\left(\frac{\bar{Z}}{\bar{z}'}\right)^{\alpha_2} - \bar{x}\right\}\right]$$

etc. where α_1 and α_2 are suitably chosen constants.

3. Efficiency Comparisons

To compare the proposed estimator with $\bar{y}_i\ (i = 1, 2, 3)$ we write the MSE to the first degree of approximation, as

$$MSE(\bar{y}) = \lambda\bar{Y}^2C_y^2 \tag{3.1}$$

$$MSE(\bar{y}_1) = \bar{Y}^2\left[\lambda C_y^2 + (\lambda - \lambda')C_x^2(1 - 2C) + \lambda' C_z^2\left(1 - 2C^*\right)\right] \tag{3.2}$$

$$MSE(\bar{y}_2) = \bar{Y}^2\left[\lambda C_y^2 + (\lambda - \lambda')C_x^2(1 - 2C) + \lambda' C^{**}C_z^2\left(C^{**} - 2C^*\right)\right] \tag{3.3}$$

$$MSE(\bar{y}_3) = \bar{Y}^2\left[\lambda C_y^2 - (\lambda - \lambda')C^2C_x^2 + \lambda' CC_z^2\left(C - 2C^*\right)\right] \tag{3.4}$$

From (2.4), (3.1), (3.2), (3.3) and (3.4), it can be easily shown that the proposed class of estimators $\bar{y}_t$ is more efficient than usual unbiased estimator $\bar{y}$, Chand's [1] estimator $\bar{y}_1$, Kiregyera's ([3], [4]) estimators $\bar{y}_2$ and $\bar{y}_3$.

If the constant γ does not coincide with the exact optimum value $\left(\frac{C^*}{C}\right)$, then the suggested estimator $\bar{y}_t$ is more efficient than

(i) The usual unbiased estimator $\bar{y}$ if $\left[\gamma^2 - 2\gamma\left(\frac{C^*}{C}\right) - \frac{(\lambda - \lambda')}{\lambda'}\left(\frac{C_x^2}{C_z^2}\right)\right] < 0$

(ii) Chand's [1] chain ratio estimator $\bar{y}_1$ if

$$\left.\begin{array}{ll} \text{either} & \frac{(2C^* - 1)}{C} < \gamma < \frac{1}{C} \\ \text{or} & \frac{1}{C} < \gamma < \frac{(2C^* - 1)}{C} \end{array}\right\}$$

(iii) Kiregyera's [3] estimator $\bar{y}_2$ if

$$\left[\gamma^2 - 2\gamma\left(\frac{C^*}{C}\right) - \left\{\frac{(\lambda - \lambda')}{\lambda'}\frac{C_x^2}{C_z^2}(1 - C)^2 + C^{**}(C^{**} - 2C^*)\right\}\right] < 0$$

(iv) Kiregyera's [4] estimator $\bar{y}_3$ if

$$\left.\begin{array}{ll} \text{either} & 1 < \gamma < \left(\frac{2C^*}{C} - 1\right) \\ \text{or} & \left(\frac{2C^*}{C} - 1\right) < \gamma < 1 \end{array}\right\}$$

4. Empirical Study

To examine the efficiency of the suggested estimator over other estimators $\bar{y}, \bar{y}_1, \bar{y}_2$ and $\bar{y}_3$ we have considered the data earlier used by Chand [1].

y : bushels of corn harvested in 1964

x : acres under corn in 1964

z : acres of corn harvested for grain in 1959

$\rho_{yx} = 0.92, \rho_{yz} = 0.89, \rho_{xy} = 0.98, n = 60, n' = 120$ and N is very large. We have computed the relative efficiency of various estimators of $\bar{Y}$ with respect to $\bar{y}$ and displayed in Table 4.1.

Table 4.1 clearly indicates that the proposed estimator $\bar{y}_t$ is more efficient than rest of the estimators.

Table 4.1. Relative efficiency (%) of various estimators of $\overline{Y}$ with respect to $\bar{y}$

Estimator	$\bar{y}$	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_t$
$RE(.,\bar{y})\times 100$	100.00	371.90	525.24	224.79	553.25 (optimum $\gamma = 1.3696$)

ACKNOWLEDGEMENTS

Authors express their thanks to the referee for useful suggestions in improving the paper.

REFERENCES

[1] Chand, L. (1975). *Some ratio-type estimators based on two or more auxiliary variables.* Ph.D. Dissertation, Iowa State University, Ames, Iowa.

[2] Das, A.K. and Tripathi, T.P. (1979). A class of estimators for population mean when mean of an auxiliary character is known. *Math. Tech. Report No. 22/79,* ISI, Calcutta.

[3] Kiregyera, B. (1980). A chain ratio-type estimator of finite population double sampling using two auxiliary variables. *Metrika,* **27,** 217-223.

[4] Kiregyera, B. (1984). Regression type estimators using two auxiliary vaiables and the model of double sampling. *Metrika,* **31,** 215-226.

[5] Sahoo, J. and Sahoo, L.N. (1993). A class of estimators in two-phase sampling using two auxiliary variables. *Jour. Ind. Statist. Assoc.,* **31,** 107-114.