Journal of Scientific Research
Banaras Hindu University, Varanasi
Vol. 57, 2013 : 147-153
ISSN : 0447-9483

A Generalized Chain Ratio in Regression Estimator for Population Mean Using two Auxiliary Characters in Sample Survey

B.B. Khare[#], U. Srivastava[*] and Kamlesh Kumar[#]
[#]Department of Statistics, Banaras Hindu University, Varanasi, India
[*]Statistics Section, M.M.V., Banaras Hindu University, Varanasi, India

Abstract

A generalized chain ratio in regression estimator for population mean using two auxiliary characters has been proposed and its properties have been studied. A comparative study of the proposed estimator has been made with the relevant estimators. An effective range for α has been obtained for which the proposed estimator is more efficient than relevant estimators. For optimum value of α , the proposed estimator is found to be more efficient than the relevant estimators which is supported by the empirical study.

Key words: Bias, mean square error, two phase sampling, auxiliary characters, chained estimators.

1. Introduction

The information on auxiliary character is used to increase the efficiency of the estimators. Such information is generally used in ratio, product and regression type estimators for the estimation of population mean of study character y . Several research works have been done in developing ratio, product and regression type estimators by using an auxiliary character. Using the information on two auxiliary characters x and z, in which the population mean of first is unknown and second is known, chain ratio type and regression type estimators for population mean of study character y have been proposed by Chand (1975) and Kiregyera (1980,84). Further, Srivastava et al. (1990) proposed generalized chain ratio estimator for population mean of study character.

In the present paper, we have proposed a generalized chain ratio in regression estimator for population mean using auxiliary characters. We obtained the expressions for bias and mean square error of the proposed estimator. A comparative study of the proposed estimator is carried out with the relevant estimators. An empirical study is given to show the performance of the proposed estimator.

2. The estimators

Let $\overline{Y}$, $\overline{X}$ and $\overline{Z}$ denote the population mean of study character y, auxiliary character x and additional auxiliary character z having jth values Y_j, X_j and Z_j: $j = 1,2,3...N$.

In the case when population mean of auxiliary character is not known, we draw a large preliminary sample of size $n'(< N)$ from population of size N by using SRSWOR scheme of sampling and estimate the population mean $\overline{X}$ by first phase sample mean $\overline{x}'$ based on n' units. Further, we draw a sub-sample of size $n(< n')$ from large preliminary sample of size n' and compute $\overline{y}$ and $\overline{x}$ which are the sub-sample means based on n units. In such case, the double sampling ratio and regression estimators are defined by

$$T_1 = \overline{y}\,\frac{\overline{x}'}{\overline{x}} \tag{2.1}$$

and

$$T_{l1} = \overline{y} + b_{yx}(\overline{x}' - \overline{x}), \tag{2.2}$$

where $b_{yx} = \dfrac{\hat{S}_{yx}}{\hat{S}_x^2}$, $\overline{x}' = \dfrac{1}{n'}\sum_{j=1}^{n'} x_j$, $\hat{S}_{yx}$ and $\hat{S}_x^2$ denote the estimates of S_{yx} and S_x^2 based on n units.

Further, Srivastava (1970) has proposed a generalized two phase sampling estimator for estimating population mean $\overline{Y}$ using information on auxiliary character x, which is given as follows:

$$T_2 = \overline{y}\left(\frac{\overline{x}'}{\overline{x}}\right)^{\alpha_1}, \tag{2.3}$$

where α_1 is unknown constant.

The mean square errors of the estimators T_1, T_2 and T_{l1} are given as follows:

$$MSE\,(T_1) = V(\overline{y}) + \overline{Y}^2\left(\frac{1}{n} - \frac{1}{n'}\right)\left(C_x^2 - 2\rho_{yx} C_y C_x\right), \tag{2.4}$$

$$MSE\,(T_2)_{\min} = V(\overline{y}) - \overline{Y}^2\left(\frac{1}{n} - \frac{1}{n'}\right)\rho_{yx}^2 C_y^2 \tag{2.5}$$

and

$$MSE(T_{l1}) = V(\bar{y}) - \bar{Y}^2\left(\frac{1}{n} - \frac{1}{n'}\right)\rho_{yx}^2 C_y^2, \tag{2.6}$$

where $V(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2$

In the case when the population mean $(\bar{X})$ of an auxiliary character x is not known but the population mean $(\bar{Z})$ of an additional auxiliary character z is known. It is suggested to take a large preliminary sample of size $n'(< N)$ from population of size N by using simple random sampling without replacement scheme of sampling and estimate the population mean $\bar{X}$ by $\hat{\bar{X}} = \bar{x}'\frac{\bar{Z}}{\bar{z}'}$ which is more efficient in comparison to $\bar{x}'$ if $\rho_{xz} > \frac{1}{2}\frac{C_z}{C_x}$, where $\bar{x}'$ and $\bar{z}'$ are the preliminary sample means based on n' units. Further, a sub-sample of size $n(< n')$ is selected from large preliminary sample of size n' and compute $\bar{y}$ and $\bar{x}$ based on n units. In this case, Chand (1975) and Kiregyera (1984) proposed chain ratio type and ratio in regression estimators, which are given as follows:

$$T_3 = \bar{y}\frac{\bar{x}'}{\bar{x}}\frac{\bar{Z}}{\bar{z}'} \quad \text{and} \quad T_{l2} = \bar{y} + b_{yx}\left[\bar{x}'\frac{\bar{Z}}{\bar{z}'} - \bar{x}\right]. \tag{2.7}$$

Now, we propose generalized chain ratio in regression estimator for population mean by using auxiliary characters which is given as follows:

$$T_{l3} = \bar{y} + b_{yx}\left[\bar{x}'\left(\frac{\bar{Z}}{\bar{z}'}\right)^{\alpha} - \bar{x}\right], \tag{2.8}$$

where α is unknown constant which is determined later.

3. Bias and mean square error of the estimator T_{l3}

Using the large sample approximation, the expressions for bias and mean square error of the estimator T_{l3} up to the terms of order (n^{-1}) are given as follows:

$$Bias(T_{l3}) = \theta\left[-\mu_{14} + \mu_{15} + \mu_{24} - \mu_{25} - \alpha\mu_{34} + \alpha\mu_{35} + \frac{f'}{n'}\left\{\frac{\alpha(\alpha+1)}{2}C_z^2 - \alpha\rho_{xz}C_xC_z\right\}\right], \quad (3.1)$$

and

$$MSE(T_{l3}) = MSE(T_{l1}) + \bar{Y}^2\frac{f'}{n'}\left[\alpha^2\frac{\rho_{yx}^2C_y^2C_z^2}{C_x^2} - 2\alpha\frac{\rho_{yx}\rho_{yz}C_y^2C_z}{C_x}\right], \quad (3.2)$$

where

$$\theta = \bar{X}\frac{S_{yx}}{S_x^2},\ \mu_{14} = Cov(\bar{x}, \hat{S}_{yx}),\ \mu_{15} = Cov(\bar{x}, \hat{S}_x^2),\ \mu_{24} = Cov(\bar{x}', \hat{S}_{yx}),$$

$$\mu_{25} = Cov(\bar{x}', \hat{S}_x^2),\ \mu_{34} = Cov(\bar{z}', \hat{S}_{yx}),\ \mu_{35} = Cov(\bar{z}', \hat{S}_x^2),$$

$$f' = 1 - \frac{n'}{N},\ C_y = \frac{S_y}{\bar{Y}},\ C_x = \frac{S_x}{\bar{X}},\ C_z = \frac{S_z}{\bar{Z}}\ ;\ S_y^2 = \frac{1}{N-1}\sum_{j=1}^{N}(Y_j - \bar{Y})^2,$$

$$S_x^2 = \frac{1}{N-1}\sum_{j=1}^{N}(X_j - \bar{X})^2, \quad S_z^2 = \frac{1}{N-1}\sum_{j=1}^{N}(Z_j - \bar{Z})^2$$

ρ_{yx} and ρ_{yz} are the correlation coefficients between (y, x) and (y, z).

The optimum value of α for which $MSE(T_{l3})$ is minimum, given by

$$\alpha_{opt} = \frac{\rho_{yz}}{\rho_{yx}}\frac{C_x}{C_z} = Q_1 \quad (3.3)$$

and the minimum mean square error of the estimator T_{l3} is given by

$$MSE\,(T_{l3})_{\min} = MSE\,(T_{l1}) - \bar{Y}^2\frac{f'}{n'}\rho_{yz}^2C_y^2. \quad (3.4)$$

The optimum value of α may be obtained from past data. If past data is not available then one may estimate it on the basis of sample observations without having any loss in the efficiency of the estimator [Reddy (1978)]. If we estimate the optimum value of the constant by using the sample value, the minimum value of the mean square error of the estimator up to the terms of order (n^{-1}) are unchanged [Srivastava and Jhajj (1983)].

The mean square errors of the estimators (T_3) and (T_{l2}) are given by

$$MSE(T_3) = MSE(T_1) + \bar{Y}^2 \frac{f'}{n'}(C_z^2 - 2\rho_{yz}C_yC_z) \tag{3.5}$$

and

$$MSE(T_{l2}) = MSE(T_{l1}) + \bar{Y}^2 \frac{f'}{n'}\left[\frac{\rho_{yx}^2 C_y^2 C_z^2}{C_x^2} - 2\frac{\rho_{yx}\rho_{yz}C_y^2 C_z}{C_x}\right], \tag{3.6}$$

4. Some special cases of the proposed estimator

(i) If $\alpha = 0$ then T_{l3} reduces to $T_{l1} = \bar{y} + b_{yx}(\bar{x}' - \bar{x})$. (4.1)

(ii) If $\alpha = 1$ then T_{l3} reduces to $T_{l2} = \bar{y} + b_{yx}\left[\bar{x}'\frac{\bar{Z}}{\bar{z}'} - \bar{x}\right]$. (4.2)

(iii) If $\alpha = -1$ then T_{l3} reduces to $T'_{l2} = \bar{y} + b_{yx}\left[\bar{x}'\frac{\bar{z}'}{\bar{Z}} - \bar{x}\right]$. (4.3)

5. Comparison of the proposed estimator T_{l3} with T_1, T_2, T_3, T_{l1}, T_{l2}

$$MSE(T_{l3}) < MSE(T_1) \text{ if } \frac{B - \sqrt{B^2 + AC}}{A} < \alpha < \frac{B + \sqrt{B^2 + AC}}{A} \tag{5.1}$$

$$MSE(T_{l3}) < MSE(T_2)_{\min} \text{ if } 0 < \alpha < 2\frac{B}{A} \tag{5.2}$$

$$MSE(T_{l3}) < MSE(T_3) \text{ if } \frac{B - \sqrt{B^2 + AC'}}{A} < \alpha < \frac{B + \sqrt{B^2 + AC'}}{A} \tag{5.3}$$

$$MSE(T_{l3}) < MSE(T_{l1}) \text{ if } 0 < \alpha < 2\frac{B}{A} \tag{5.4}$$

$$MSE(T_{l3}) < MSE(T_{l2}) \text{ if } 1 < \alpha < 2\frac{B}{A} - 1 \tag{5.5}$$

where $A = \frac{f'}{n'}\frac{\rho_{yx}^2 C_y^2 C_z^2}{C_x^2}$, $B = \frac{f'}{n'}\frac{\rho_{yx}\rho_{yz}C_y^2 C_z}{C_x}$,

$$C = \left(\frac{1}{n} - \frac{1}{n'}\right)\left\{C_x^2 + \rho_{yx}^2 C_y^2 - 2\rho_{yx}C_yC_x\right\},$$

and $C' = \left[\left(\frac{1}{n} - \frac{1}{n'}\right)\left\{C_x^2 + \rho_{yx}^2 C_y^2 - 2\rho_{yx}C_yC_x\right\} + \frac{f'}{n'}\left(C_z^2 - 2\rho_{yz}C_yC_z\right)\right]$

6. Empirical study

Data Set I ($N = 82$, $n' = 43$, $n = 25$)

The data used by Srivastava et al. (1989), 82 children of age 3 months of Varanasi, India have consider for the present study,
y : mid arm circumference of the children,
x : chest circumference of the children ,
z : skull circumference of the children.

The values of the parameters of the y, x and z characters for the given data are given as follows:

$\bar{Y} = 11.90$, $\bar{Z} = 39.80$, $C_y^2 = 0.0052$, $C_x^2 = 0.0011$, $C_z^2 = 0.008$, $\rho_{yx} = 0.87$,
$\rho_{yz} = 0.86$

Data Set II ($N = 55$, $n' = 30$, $n = 18$)

The data used by Srivastava et al. (1989), 55 children of age 5 years of Varanasi, India have consider for the present study
y : weight of the children,
x : chest circumference of the children,
z : skull circumference of the children.

The values of the parameters of the y, x and z characters for the given data are given as follows:

$\bar{Y} = 17.08$, $\bar{Z} = 50.44$, $C_y^2 = 0.0161$, $C_x^2 = 0.0027$, $C_z^2 = 0.007$, $\rho_{yx} = 0.84$,
$\rho_{yz} = 0.51$

Table1

Relative efficiency of the estimators $\bar{y}$, T_1 , T_2 , T_3 , T_{l1}, T_{l2} and T_{l3} (in %) with respect to $\bar{y}$.

Estimators	Data Set 1	Data Set 2
$\bar{y}$	100.00 (0.02047)*	100.00 (0.17553)*
T_1	154.96 (0.01321)	144.79 (0.12123)
T_2	183.75 (0.01114)	172.27 (0.10189)

T_3	228.21 (0.00897)	160.76 (0.10919)
T_{l1}	183.75 (0.01114)	172.27 (0.10189)
T_{l2}	391.39 (0.00523)	209.31 (0.08386)
T_{l3}	400.59 (0.00511)	210.52 (0.08338)

*Figures in parenthesis give the $MSE(.)$.

From table 1, we observe that the proposed estimator T_{l3} is more efficient than the relevant estimators $\bar{y}$, T_1, T_2, T_{l1} and the chained estimators T_3 and T_{l2} proposed by Chand (1975) and Kiregyera (1984).

References

Chand, L. (1975): Some ratio-type estimators based on two or more auxiliary variables. Ph.D. thesis submitted to Iowa State University, Ames, IOWA.

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

Kiregyera, B. (1984): Regression-type estimator using two auxiliary variables and model of double sampling from finite population. *Metrika,* 31, 215-226.

Reddy, V.N. (1978): A study on the use of prior knowledge on certain population parameters in estimation. *Sankhya* C, 40, 29-37.

Srivastava, S.K. (1970): A two phase estimator in sampling surveys. *Aust. Jour. Statist.*, 12, 23-27.

Srivastava, S.K. and Jhajj, H.S. (1983): A class of estimators of the population mean using multi-auxiliary information. *Cal. Stat. Assoc. Bull.*, 32, 47-56.

Srivastava, S. Rani, Srivastava, S.R. and Khare, B.B. (1989): Chain ratio-type estimators for ratio of two population means using auxiliary characters. *Commun. Stat. –Theo. Meth.*, 18 (10), 3917-3926.

Srivastava S. Rani, Khare B.B. and Srivastava S.R. (1990): A generalised chain ratio estimator for mean of finite population. *Jour. Ind. Soc. Agri. Statistics*, 1, 108-117.