Research Article
A New Regression Estimator Based on Two Auxiliary Variables

[1]Saman Shahbaz, [2]Muhammad Zubair, [1]Nadeem Shafique Butt and [1]Muhammad Ismail
[1]Department of Statistics, COMSATS Institute of Information Technology, Lahore, Punjab, Pakistan
[2]Department of Statistics, University of Sargodha, Sargodha, Punjab, Pakistan

Abstract: A multiple regression estimator has been developed by using information on "k" auxiliary variables. The mean square error has been obtained of the Multiple Regression estimator and comparison has been made with existing estimators for estimation of population characteristic.

Keywords: Auxiliary variables, mean square error, regression estimator

INTRODUCTION

Auxiliary information has been very vital in improving the efficiency of estimates in surveys. Many estimators have been proposed from time to time that utilizes the information of auxiliary variables while estimation of population mean or total. Some notable references using of auxiliary variables is Rao (1986), Khare and Srivastava (1993, 1995), Tabasum and Khan (2004, 2006), Khare and Srivastava (2010) and Ismail *et al.* (2011).

To set up the seen suppose information about variables of interest Y is available for the sample. Also the information about a set of auxiliary variables, say X_j for j = 1, 2, …, k is available for the complete population units. The information about these auxiliary variables proves very helpful in improving efficiency of the estimates. Number of estimators are available in literature that utilizes the information about these auxiliary variables. The simplest one and perhaps the most commonly used estimator in literature that uses the information of a single auxiliary variable is the popular ratio estimator given as:

$$\bar{y}_r = \frac{\bar{y}}{\bar{x}}\bar{X} \tag{1}$$

This estimator is biased. The approximate mean square error of estimator (1) is given as:

$$Var(\bar{y}) \approx \frac{(N-n)}{Nn}\left[S_Y^2 - 2R\rho_{xy}S_YS_X + R^2S_X^2\right] \tag{2}$$

where, notations have there usual meanings. The estimator given in (1) enjoys the property that it is the best linear unbiased estimator of population total under the linear stochastic model of the form:

$$y_i = \beta_i x_i + \varepsilon_i$$

with usual assumptions. Some modifications of the ratio estimator have also been proposed in literature from time to time. The most reputed of these modifications is the estimator proposed by Hartley and Ross (1954) for estimation of population ratio of the variable Y to variable X. This estimator is given as:

$$r_{HR} = \bar{r} + \frac{n(N-1)}{N(n-1)} \cdot \frac{(\bar{y} - \bar{r}\bar{x})}{\bar{X}} \tag{3}$$

The approximate variance of (3) is given as:

$$Var(r_{HR}) \approx \frac{1}{n\bar{X}^2}\left[S_y^2 - 2\bar{R}\rho_{xy}S_yS_x + \bar{R}^2S_x^2\right] \tag{4}$$

A ratio estimator has also been given by Olkin (1958) that uses information of "k" auxiliary variables. The estimator is given as:

$$\hat{Y}_{MR} = W_1\frac{\bar{y}}{\bar{x}_1}X_1 + W_2\frac{\bar{y}}{\bar{x}_2}X_2 + \ldots\ldots + W_k\frac{\bar{y}}{\bar{x}_k}X_k = \sum_{j=1}^{k}W_j\frac{\bar{y}}{\bar{x}_j}X_j = \sum_{j=1}^{k}W_j\hat{Y}_{R_j} \tag{5}$$

An adoption from the ratio estimator is the regression estimator that has the simple form:

$$\bar{y}_{reg} = \bar{y} + \beta(\bar{X} - \bar{x}) \tag{6}$$

If, in (7), β is taken as the population regression coefficient then the estimator is unbiased with minimum variance given as:

$$Var(\bar{y}_{reg}) = \frac{N-n}{Nn}S_y^2(1-\rho^2) \tag{7}$$

Many modifications have been given for the regression estimator from time to time but they are not much popular. These estimators have also been widely used in two phase sampling where complete information is not available for the auxiliary variable. In that case the efficiency of the estimator is somewhat lower as compared to that for single phase sampling.

THE NEW PROPOSED ESTIMATOR

In this section we have developed the new estimator for estimation of population mean. This estimator is a simple extension of the regression estimator by using information on "*k*" auxiliary variables. The proposed estimator is given as:

$$\bar{y}_{lrm} = W\bar{y}_{lr1} + (1-W)\bar{y}_{lr2} \;;\;\; \bar{y}_{lrj} = \bar{y} + \beta_j \left(\bar{X}_j - \bar{x}_j \right) \quad (8)$$

The variance of (8) is:

$$V = V\left(\bar{y}_{lrm}\right) = W^2 V_{11} + (1-W)^2 V_{22} + 2W(1-W)V_{12} \quad (9)$$

where,

$$V_{jj} = \frac{N-n}{Nn} S_Y^2 \left(1-\rho_{YX_j}^2\right) \;;\; V_{12} = \frac{N-n}{Nn} S_Y^2 \left(1-\rho_{YX_1}^2 - \rho_{YX_1}^2 + \rho_{YX_1}\rho_{YX_1}\rho_{X_1X_2}\right)$$

The value of *W* that minimizes (8) is obtained by differentiating (8) and setting the resulting derivative to zero.
Now:

$$\frac{\partial V}{\partial W} = 2WV_{11} - 2(1-W)V_{22} + 2(1-2W)V_{12}$$

Also $\partial V / \partial W = 0$ gives:

$$W = \frac{V_{22} - V_{12}}{V_{11} + V_{12} - 2V_{12}} \quad (10)$$

Using the value of *W* given in (10) in (8), the minimum variance of the estimator is given as:

$$V\left(\bar{y}_{lrm}\right) = \frac{V_{11}V_{22} - V_{12}^2}{V_{11} + V_{22} - 2V_{12}}$$

This estimator seems to have relatively better performance as compared to the conventional regression estimator based upon two auxiliary variables.

REFERENCES

Hartley, H.O. and A. Ross, 1954. Unbiased ratio estimates. Nature, 174: 270-271.

Ismail, M., M.Q. Shahbaz and M. Hanif, 2011. A general class of estimator of population mean in the presence of non-response. Pak. J. Stat., 27(4): 467-476.

Khare, B.B. and S. Srivastava, 1993. Estimation of population mean using auxiliary character in presence of non-response. Nat. Acad. Sci. Lett. India, 16: 111-114.

Khare, B.B. and S. Srivastava, 1995. Study of conventional and alternative two-phase sampling ratio, product and regression estimators in presence of non-response. Proc. Indian Nat. Sci. Acad., 65: 195-203.

Khare, B.B. and S. Srivastava, 2010. Generalized two phase estimators for the population mean in the presence of non response. Aligarh J. Stat., 30: 39-54.

Olkin, I., 1958. Multivariate ratio estimation for finite populations. Biometrika, 45: 154-165.

Rao, P.S.R.S., 1986. Ratio estimation with sub sampling the non respondents. Surv. Methodol., 12(2): 217-230.

Tabasum, R. and I.A. Khan, 2004. Double sampling for ratio estimation with non-response. J. Ind. Soc. Agric. Stat., 58(3): 300-306.

Tabasum, R. and I.A. Khan, 2006. Double sampling ratio estimator for the population mean in presence of non response. Assam. Stat. Rev., 20(1): 73-83.