Research Journal of Mathematical and Statistical Sciences ISSN 2320–6047
Vol. **3(1),** 6-9, January **(2015)** *Res. J. Mathematical and Statistical Sci.*

A Class of Chain Type Estimators for Population Mean using two Auxiliary Variables in the Presence of Non-Response

N. Kumar and Ashish Sharma
Udai Pratap Autonomous College, Department of Statistics, Varanasi, INDIA

Available online at: www.isca.in, www.isca.me
Received 25th September 2014, revised 3rd December 2014, accepted 10th January 2015

Abstract

This paper discusses the problem of estimation of population mean in the case of non-response in double sampling. Here we obtained the mean square error and relative bias of the suggested class of estimators. The comparison of the suggested class of estimators to other relevant estimators is also derived. The suggested class of estimators found to be more efficient as compare to another estimators obtained by applying suitable values of some constant. An empirical study is also carried out to show the performance of the suggested estimators.

Keywords: Study variable, auxiliary variables, relative bias, Non-response, mean square error, double sampling.

Introduction

Some times during sample survey we find that we not get the full information in the sample because the units we select in the sample suffering from the problem of non-response. When this type of situation arises we get misleading estimate it occurs when the units which are responding is different from the units which are not responding because the estimate can be biased. The problem of non-response was first considered by Hansen and Hurwitz[1], Rao[2,3], Khare and Srivastava[4,5]. Khare and Srivastava[6], Okafor and Lee[7], Singh and Kumar[8], Tabasum and Khan[9,10] have proposed the ratio and regression estimator in the presence of non-response when the population mean of auxiliary variable is not known.

This paper suggested the class of chain type estimators under non-response in case of double sampling. The expressions of mean square error (MSE) and relative bias (RB) of the suggested class of estimators are also given. Comparison and an empirical study of the suggested class of estimators are also given to show the performance of the suggested estimator.

The Estimators

Let y be the main variable with population mean $\bar{Y}$ and x be the auxiliary variable with population mean $\bar{X}$. z be the additional auxiliary variable having population mean $\bar{Z}$. Here we find that the population mean $(\bar{X})$ of the auxiliary variable (x) is unknown but the population mean $(\bar{Z})$ of the additional auxiliary variable (z) (nearly related to x) is known and less correlated to the (y) in comparison to the (x). Using the technique of (SRSWOR) simple random sampling without replacement we take a preliminary sample of size n' from the population with size N. Now we estimate the population mean $(\bar{X})$ with the help of additional auxiliary variable (z) with known population mean $\bar{Z}$ and n' observation on x and z. Here we again select a sub sample of size n ($<n'$) from n' by (SRSWOR) simple random sampling without replacement. Here we find that n_1 units respond on the study variable y and n_2 units do not respond in the sample of size n for the study variable y. Here N is assumed to be the combination of N_1 responding and N_2 non-responding units. In this situation we use the technique of sub sampling from non-respondent of Hansen and Hurwitz[1]. Again we take a sub-sample of size $r(r = n_2 k^{-1}, k > 1)$ units is selected from n_2 non-respondent units using the SRSWOR sampling scheme by making extra effort. Now we have $(n_1 + r)$ responding units on the study variable y.

The estimator for the population mean $\bar{Y}$ having $(n_1 + r)$ observation on study variable y given by Hansen and Hurwitz as follows

$$\bar{y}^* = \frac{n_1}{n}\bar{y}_1 + \frac{n_2}{n}\bar{y}_2', \tag{1}$$

where $\bar{y}_1$ are the sample mean of y on n_1 units and $\bar{y}_2'$ are the sample means of y based on r units. The variance of the unbiased estimator $\bar{y}^*$ is given as follows

$$V(\bar{y}^*) = \frac{1-f}{n}S_y^2 + \frac{W_2(k-1)}{n}S_{y(2)}^2, \tag{2}$$

where $f = \frac{n}{N}$, $W_2 = \frac{N_2}{N}$,

$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$ denotes the population mean squares of Y for the entire population and $S_{y(2)}^2 = \frac{1}{N_2 - 1}\sum_{i=1}^{N_2}(Y_{i(2)} - \bar{Y}_2)^2$ denotes the population mean squares of Y for non-responding part of the population.

In the same manner for estimating the population mean $\bar{X}$ of the auxiliary variable x, the estimator $\bar{x}^*$ is given as follows

$$\bar{x}^* = \frac{n_1}{n}\bar{x}_1 + \frac{n_2}{n}\bar{x}_2' ,$$

where $\bar{x}_1$ are the sample means of auxiliary variable x on n_1 units and $\bar{x}_2'$ are the sample means of auxiliary variable x on r units. We also have

$V(\bar{x}^*) = \frac{1-f}{n}S_x^2 + \frac{w_2(k-1)}{n}S_{x(2)}^2$ where; $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$ denotes the population mean square of x for the entire population and $S_{x(2)}^2 = \frac{1}{N_2 - 1}\sum_{i=1}^{N_2}(X_{i(2)} - \bar{X}_{(2)})^2$ denotes the population mean square of x for non-responding part of the population.

Khare and Srivastava[5] and Tabasum and Khan (2004) have suggested the estimator when the population mean of auxiliary variable is unknown given by

$$t_{h_1}^* = \bar{y}^*(\bar{x}'/\bar{x}^*) .$$

In this situation Kumar et al.[8] suggested the following estimator

$$t_{h_3}^* = (1-a)\bar{y}^* + a\bar{y}^*(\bar{x}'/\bar{x}^*)^{\alpha} ,$$

Where a and α are constants.

The chain ratio estimator for population mean using two auxiliary variables in presence of non-response suggested by Khare et al[6] is given by

$$t_{h_2}^* = \bar{y}^* \frac{\bar{x}'}{\bar{x}^*}\frac{\bar{Z}}{\bar{z}'} .$$

Now, we suggested a class of chain type estimator under double sampling using auxiliary variables for population mean in the presence of non-response which is given as

$$T_H^* = (1-a)\bar{y}^* + a\bar{y}^*(\bar{x}'/\bar{x}^*)^{\alpha_1}(\bar{Z}/\bar{z}')^{\alpha_2} , \quad (3)$$

where a, α_1 and α_2 are some constants, $\bar{x}'$ is the sample mean based on preliminary sample of size n' and $\bar{z}'$ is the sample mean based on preliminary sample of size n' .

Mean Square Error (MSE) and Relative Bias (RB)

$$RB(T_H^*) = a\alpha_2\left(\frac{1}{n'} - \frac{1}{N}\right)\left[\frac{(\alpha_2+1)}{2}C_z^2 - C_{yz}\right] + a\alpha_1\frac{W_2}{n}(k-1)\left[\frac{(\alpha_1+1)}{2}C_{x(2)}^2 - C_{xy(2)}\right]$$

$$+ a\alpha_1\left(\frac{1}{n} - \frac{1}{n'}\right)\left[\frac{(\alpha_1+1)}{2}C_x^2 - C_{xy}\right] \quad (4)$$

$$MSE(T_H^*) = V(\bar{y}^*) + \bar{Y}^2\left[a^2\alpha_1^2\left\{\left(\frac{1}{n} - \frac{1}{n'}\right)C_x^2 + \frac{w_2}{n}(k-1)C_{x(2)}^2\right\} - 2a\alpha_1\left\{\left(\frac{1}{n} - \frac{1}{n'}\right)C_{yx} + \frac{w_2}{n}(k-1)C_{yx(2)}\right\}\right]$$

$$+ \bar{Y}^2\left[a\alpha_2\left(\frac{1}{n'} - \frac{1}{N}\right)\left\{a\alpha_2 C_z^2 - 2C_{yz}\right\}\right] \quad (5)$$

$$= V(\bar{y}^*) + \bar{Y}^2\left[a^2\alpha_1^2 A - 2a\alpha_1 B + a\alpha_2\left(\frac{1}{n'} - \frac{1}{N}\right)\left\{a\alpha_2 C_z^2 - 2C_{yz}\right\}\right] \quad (6)$$

Table-1
Estimators

Estimators	a	α_1	α_2
(i) Usual unbiased estimator = $\bar{y}^*$	1	0	0
(ii) Ratio estimator: $t_{h_1}^* = \bar{y}^*(\bar{x}'/\bar{x}^*)$	1	1	0
(iii) The estimator: $t_{h_2}^* = \bar{y}^*(\bar{x}'/\bar{x}^*)(\bar{Z}/\bar{z}')$	1	1	1
(iv)The estimator: $t_{h_3}^* = (1-a)\bar{y}^* + a\bar{y}^*(\bar{x}'/\bar{x}^*)^{\alpha_1}$	a	α_1	0

where, $A = \left(\frac{1}{n}-\frac{1}{n'}\right)C_x^2 + \frac{W_2(k-1)}{n}C_{x(2)}^2$,

$B = \left(\frac{1}{n}-\frac{1}{n'}\right)C_{yx} + \frac{W_2(k-1)}{n}C_{yx(2)}$,

$$C_x^2=\frac{S_x^2}{\bar{X}^2}, C_{x(2)}^2=\frac{S_{x(2)}^2}{\bar{X}^2}, C_{yx}=\rho_{yx}C_yC_x, C_{yx(2)}=\rho_{yx(2)}C_{y(2)}C_{x(2)},$$

$$C_z^2=\frac{S_z^2}{\bar{Z}^2}, C_{yz}=\rho_{yz}C_yC_z,\ S_x^2=\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\bar{X})^2$$

$$S_{x(2)}^2=\frac{1}{N_2-1}\sum_{i=1}^{N_2}(X_{i(2)}-\bar{X}_{(2)})^2\quad S_z^2=\frac{1}{N-1}\sum_{i=1}^{N}(Z_i-\bar{Z})^2$$

The MSE of T_H^* in (6) is minimized when

$$\alpha_1 = \frac{B}{aA} \tag{7}$$

$$\alpha_2 = \frac{C_{yz}}{aC_z^2} = \frac{\rho_{yz}C_y}{aC_z} \tag{8}$$

Now putting the optimum values of α_1 and α_2 in equation (6) we have

Min MSE (T_H^*) =

$$V(\bar{y}^*) - \bar{Y}^2\left[\frac{B^2}{A} + \left(\frac{1}{n'}-\frac{1}{N}\right)\frac{C_{yz}^2}{C_z^2}\right], \tag{9}$$

Substituting the values of a, α_1 and α_2 into the MSE of T_H^* we get the MSE of $t_{h_i}^*$= 1, 2……..4 to the first degree of approximation as follow:

$$MSE(\bar{y}^*) = \left(\frac{1}{n}-\frac{1}{N}\right)S_y^2 + \frac{W_2}{n}(k-1)S_{y(2)}^2 \tag{10}$$

$$MSE(t_{h_1}^*) = V(\bar{y}^*) + \bar{Y}^2\left[\left(\frac{1}{n}-\frac{1}{n'}\right)C_x^2 + \frac{W_2}{n}(k-1)C_{x(2)}^2\right]$$

$$-2\bar{Y}^2\left[\left(\frac{1}{n}-\frac{1}{n'}\right)C_{yx} + \frac{W_2}{n}(k-1)C_{yx(2)}\right] \tag{11}$$

$$MSE(t_{h_2}^*) = V(\bar{y}^*) + \bar{Y}^2\left[\left(\frac{1}{n}-\frac{1}{n'}\right)C_x^2 + \frac{W_2}{n}(k-1)C_{x(2)}^2 - 2\left\{\left(\frac{1}{n}-\frac{1}{n'}\right)C_{yx} + \frac{W_2}{n}(k-1)C_{yx(2)}\right\}\right]$$

$$+\bar{Y}^2\left[\left(\frac{1}{n'}-\frac{1}{N}\right)\left\{C_z^2 - 2C_{yz}\right\}\right] \tag{12}$$

$$MSE(t_{h_3}^*) = V(\bar{y}^*) + \bar{Y}^2\left[a^2\alpha_1^2\left\{\left(\frac{1}{n}-\frac{1}{n'}\right)C_x^2 + \frac{W_2}{n}(k-1)C_{x(2)}^2\right\} - 2a\alpha_1^2\left\{\left(\frac{1}{n}-\frac{1}{n'}\right)C_{yx} + \frac{W_2}{n}(k-1)C_{yx(2)}\right\}\right] \tag{13}$$

Comparison of T_H^* with the other estimators (14)

$MSE(T_H^*) < MSE(\bar{y}^*)$ If $\frac{2B}{aA} < \alpha_1 < 0$ or $0 < \alpha_1 < \frac{2B}{aA}$

If $\frac{2C_{yz}}{aC_z^2} < \alpha_2 < 0$ or $0 < \alpha_2 < \frac{2C_{yz}}{aC_z^2}$

$MSE(T_H^*) < MSE(t_{h_1}^*)$ If

$\frac{2B}{aA} + \frac{1}{a} < \alpha_1 < -\frac{1}{a}$ or $-\frac{1}{a} < \alpha_1 < \frac{2B}{aA} + \frac{1}{a}$

If $0 < \alpha_2 < \frac{2C_{yz}}{aC_z^2}$ or $\frac{2C_{yz}}{aC_z^2} < \alpha_2 < 0$

$MSE(T_H^*) < MSE(t_{h_2}^*)$ If

$\frac{2B}{aA} - \frac{1}{a} < \alpha_1 < \frac{1}{a}$ or $\frac{1}{a} < \alpha_1 < \frac{2B}{aA} - \frac{1}{a}$

If $\frac{2C_{yz}}{aC_z^2} - \frac{1}{a} < \alpha_2 < \frac{1}{a}$ or $\frac{1}{a} < \alpha_2 < \frac{2C_{yz}}{aC_z^2} - \frac{1}{a}$

$MSE(T_H^*) < MSE(t_{h_3}^*)$ If

$0 < \alpha_2 < \frac{2C_{yz}}{aC_z^2}$ or $\frac{2C_{yz}}{aC_z^2} < \alpha_2 < 0$

Empirical study

This data based on physical growth of upper socio-economic group of 95 school going students of Varanasi under an ICMR study, Department of Pediatrics, B.H.U., during 1983-84 has been taken under study, Khare et al. The last 25% (i.e. 24 children) of units have been considered as non-responding units. The study variable(y), auxiliary variable (x) and the additional auxiliary variable (z) are taken as follow: y- the weight of the students in kg., x- the chest circumference of the students in cm., z- the skull circumference of the students in cm.

The values of the parameters are as follow:

$\bar{Y} = 21.9758, \bar{X} = 57.2116, \bar{Z} = 51.6084, C_y = 0.1905, C_x = 0.0705, C_z = 0.0322$

$C_{y(2)} = 0.1856, C_{x(2)} = 0.0752, \rho_{yx} = 0.8338, \rho_{yz} = 0.4274, \rho_{yx(2)} = 0.8426$

Table-2 Explain MSE and Relative efficiency (with respect to $\bar{y}^*$) of the estimators $\bar{y}^*$, $t_{h_1}^*, t_{h_2}^*, t_{h_3}^*, T_H^*$ for the fixed values of n', n and different values of k (N=95, n'=70 and n=50)

Table-2
MSE and Relative efficiency

Estimators	1/k			
	1/2	1/3	1/4	1/5
$\overline{y}^*$	100(0.250089)*	100(0.334144)	100(0.418199)	100(0.502254)
$t_{h_1}^*$	158(0.158406)	168(0.198868)	175(0.239329)	180(0.279790)
$t_{h_2}^*$	166(0.150769)	175(0.19123)	180(0.231692)	185(0.272153)
$t_{h_3}^*$	207(0.120994)	230(0.145441)	246(0.169853)	259(0.194252)
T_H^*	230(0.108959)	250(0.133406)	265(0.157818)	276(0.182216)

*Figures in parenthesis give the MSE (.) and outside give the Relative efficiency (RE)

Conclusion

After study the above table we conclude that for the fixed value of n', n and k=2, 3, 4, 5 the mean square error of the above estimators decreases as value of k^{-1} increases and the estimator T_H^* has getting the maximum relative efficiency with respect to $\overline{y}^*$. However, the estimators T_H^* are more efficient than $\overline{y}^*$ and correspondingly with $t_{h_1}^*$, $t_{h_2}^*$, $t_{h_3}^*$.

References

1. Hansen M.H. and Hurwitz W.N. The problem of non-response in sample surveys, *J. Amer. Statist. Assoc.* **41,** 517-529, **(1946)**
2. Rao, P.S.R.S., Ratio Estimation with sub sampling the non respondent, *Survey methodology,* **12(2),** 217-230, **(1986)**
3. Rao P.S.R.S., Ratio and Regression Estimates with sub sampling the non respondent, Paper presented at a special contributed session of the International Statistical Association Meetings, September, Tokyo, Japan,2-16, **(1987)**
4. Khare B.B. and Srivastava S., Estimation of population mean using auxiliary character in presence of non response, *Nat. Acad. Sc. Letters, India,* **16(3),** 111-114, **(1993)**
5. Khare B.B. and Srivastava S., Study of conventional and alternative two-phase sampling ratio, product and regression estimators in presence of non -response, Proc. *Nat. Acad .Sci. India,* **65(A) (II),** 195-203, **(1995)**
6. Khare B.B., Srivastava U. and Kumar K. Chain ratio estimator for the population mean in the presence of non-response, *Statistics in transition-new series,* **13(3),** 495-518, **(2012)**
7. Okafor F.C. and Lee H., Double sampling for ratio and regression estimation with sub sampling the non respondent, *Survey Methodology,* **26,** 183-188, **(2000)**
8. Singh H.P. and Kumar S., Estimation of mean in presence of non-response using two phase sampling scheme, Statistical Papers, DOI 10. 1007/s00362-008040-5**, (2008)**
9. Tabasum R. and Khan I.A. Double sampling for ratio estimation with non response, *J. Soc. Agril. Statist,* **58(3),** 300-306, **(2004)**
10. Tabasum R. and Khan I.A., Double sampling ratio estimator for the population mean in presence of non-response, *Assam Statist. Review,* **20(1),** 73-83, **(2006)**