Manoj K. Chaudhary, Sachin Malik, Rajesh Singh

1Department of Statistics, Banaras Hindu University

Varanasi-221005, India

Florentin Smarandache

University of New Mexico, Gallup, USA

Use of Auxiliary Information for Estimating Population Mean in Systematic Sampling under Non-Response

Published in:
Rajesh Singh, Florentin Smarandache (Editors)
ON IMPROVEMENT IN ESTIMATING POPULATION PARAMETER(S) USING AUXILIARY INFORMATION
Educational Publishing (Columbus) & Journal of Matter Regularity (Beijing), USA - China, 2013
ISBN: 978-1-59973-230-5
pp. 5 - 16

Abstract

In this paper we have adapted Singh and Shukla (1987) estimator in systematic sampling using auxiliary information in the presence of non-response. The properties of the suggested family have been discussed. Expressions for the bias and mean square error (MSE) of the suggested family have been derived. The comparative study of the optimum estimator of the family with ratio, product, dual to ratio and sample mean estimators in systematic sampling under non-response has also been done. One numerical illustration is carried out to verify the theoretical results.

Keywords: Auxiliary variable, systematic sampling, factor-type estimator, mean square error, non-response.

1. Introduction

There are some natural populations like forests etc., where it is not possible to apply easily the simple random sampling or other sampling schemes for estimating the population characteristics. In such situations, one can easily implement the method of systematic sampling for selecting a sample from the population. In this sampling scheme, only the first unit is selected at random, the rest being automatically selected according to a predetermined pattern. Systematic sampling has been considered in detail by Madow and Madow (1944), Cochran (1946) and Lahiri (1954). The application of systematic sampling to forest surveys has been illustrated by Hasel (1942), Finney (1948) and Nair and Bhargava (1951).

The use of auxiliary information has been permeated the important role to improve the efficiency of the estimators in systematic sampling. Kushwaha and Singh (1989) suggested a class of almost unbiased ratio and product type estimators for estimating the population mean using jack-knife technique initiated by Quenouille (1956). Later Banarasi et al. (1993), Singh and Singh (1998), Singh et al. (2012), Singh et al. (2012) and Singh and Solanki (2012) have made an attempt to improve the estimators of population mean using auxiliary information in systematic sampling.

The problem of non-response is very common in surveys and consequently the estimators may produce bias results. Hansen and Hurwitz (1946) considered the problem of estimation of population mean under non-response. They proposed a sampling plan that involves taking a subsample of non-respondents after the first mail attempt and then enumerating the subsample by personal interview. El-Badry (1956) extended Hansen and Hurwitz (1946) technique. Hansen and Hurwitz (1946) technique in simple random sampling is described as: From a population $U = (U_1, U_2, ---, U_N)$, a large first phase sample of size n' is selected by simple random sampling without replacement (SRSWOR). A smaller second phase sample of size n is selected from n' by SRSWOR. Non-response occurs on the second phase of size n in which n_1 units respond and n_2 units do not. From the n_2 non-respondents, by SRSWOR a sample of $r = n_2/ k; k > 1$units is selected. It is assumed that all the r units respond this time round. (see Singh and Kumar (20009)). Several authors such as Cochran (1977), Sodipo and Obisesan (2007), Rao (1987), Khare and Srivastava (1997) and Okafor and Lee (2000) have studied the problem of non-response under SRS.

In the sequence of improving the estimator, Singh and Shukla (1987) proposed a family of factor-type estimators for estimating the population mean in simple random sampling using an auxiliary variable, as

$$T_\alpha = \bar{y}\left[\frac{(A+C)\bar{X} + fB\bar{x}}{(A+fB)\bar{X} + C\bar{x}}\right] \tag{1.1}$$

where $\bar{y}$ and $\bar{x}$ are the sample means of the population means $\bar{Y}$ and $\bar{X}$ respectively. A, B and C are the functions of α, which is a scalar and chosen so as the MSE of the estimator T_α is minimum.

Where,

$$A = (\alpha-1)(\alpha-2),\ B = (\alpha-1)(\alpha-4),$$

$$C = (\alpha-2)(\alpha-3)(\alpha-4);\ \alpha > 0 \text{ and}$$

$$f = \frac{n}{N}.$$

Remark 1 : If we take α = 1, 2, 3 and 4, the resulting estimators will be ratio, product, dual to ratio and sample mean estimators of population mean in simple random sampling respectively (for details see Singh and Shukla (1987)).

In this paper, we have proposed a family of factor-type estimators for estimating the population mean in systematic sampling in the presence of non-response adapting Singh and Shukla (1987) estimator. The properties of the proposed family have been discussed with the help of empirical study.

2. Sampling Strategy and Estimation Procedure

Let us assume that a population consists of N units numbered from 1 to N in some order. If $N = nk$, where k is a positive integer, then there will be k possible samples each consisting of n units. We select a sample at random and collect the information from the units of the selected sample. Let n_1 units in the sample responded and n_2 units did not respond, so that $n_1 + n_2 = n$. The n_1 units may be regarded as a sample from the response class and n_2 units as a sample from the non-response class belonging to the population. Let us assume that N_1 and N_2 be the number of units in the response class and non-response class respectively in the population. Obviously, N_1 and N_2 are not known but their unbiased estimates can be obtained from the sample as

$$\hat{N}_1 = n_1 N / n \,; \;\; \hat{N}_2 = n_2 N / n \,.$$

Further, using Hansen and Hurwitz (1946) technique we select a sub-sample of size h_2 from the n_2 non-respondent units such that $n_2 = h_2 L$ ($L > 1$) and gather the information on all the units selected in the sub-sample (for details on Hansen and Hurwitz (1946) technique see Singh and Kumar (2009)).

Let Y and X be the study and auxiliary variables with respective population means $\overline{Y}$ and $\overline{X}$. Let $y_{ij}\left(x_{ij}\right)$ be the observation on the j^{th} unit in the i^{th} systematic sample under study (auxiliary) variable ($i = 1...k : j = 1...n$).Let us consider the situation in which non-response is observed on study variable and auxiliary variable is free from non-response. The Hansen-Hurwitz (1946) estimator of population mean $\overline{Y}$ and sample mean estimator of $\overline{X}$ based on a systematic sample of size n, are respectively given by

$$\overline{y}^* = \frac{n_1 \overline{y}_{n1} + n_2 \overline{y}_{h_2}}{n}$$

and $\overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$

where $\overline{y}_{n1}$ and $\overline{y}_{h_2}$ are respectively the means based on n_1 respondent units and h_2 non-respondent units. Obviously, $\overline{y}^*$ and $\overline{x}$ are unbiased estimators of $\overline{Y}$ and $\overline{X}$ respectively. The respective variances of $\overline{y}^*$ and $\overline{x}$ are expressed as

$$V\left(\overline{y}^*\right) = \frac{N-1}{nN}\left\{1 + (n-1)\rho_Y\right\}S_Y^2 + \frac{L-1}{n} W_2 S_{Y2}^2 \qquad (2.1)$$

and

$$V\left(\overline{x}\right) = \frac{N-1}{nN}\left\{1 + (n-1)\rho_X\right\}S_X^2 \qquad (2.2)$$

where ρ_Y and ρ_X are the correlation coefficients between a pair of units within the systematic sample for the study and auxiliary variables respectively. S_Y^2 and S_X^2 are respectively the mean squares of the entire group for study and auxiliary variables. S_{Y2}^2 be the population mean square of non-response group under study variable and W_2 is the non-response rate in the population.

Assuming population mean $\overline{X}$ of auxiliary variable is known, the usual ratio, product and dual to ratio estimators based on a systematic sample under non-response are respectively given by

$$\overline{y}_R^* = \frac{\overline{y}^*}{\overline{x}}\overline{X}, \tag{2.3}$$

$$\overline{y}_P^* = \frac{\overline{y}^*\overline{x}}{\overline{X}} \tag{2.4}$$

and $$\overline{y}_D^* = \overline{y}^* \frac{\left(N\overline{X} - n\overline{x}\right)}{(N-n)\overline{X}}. \tag{2.5}$$

Obviously, all the above estimators $\overline{y}_R^*$, $\overline{y}_P^*$ and $\overline{y}_D^*$ are biased. To derive the biases and mean square errors (MSE) of the estimators $\overline{y}_R^*$, $\overline{y}_P^*$ and $\overline{y}_D^*$ under large sample approximation, let

$$\overline{y}^* = \overline{Y}(1+e_0)$$

$$\overline{x} = \overline{X}(1+e_1)$$

such that $E(e_0) = E(e_1) = 0$,

$$E\left(e_0^2\right) = \frac{V\left(\overline{y}^*\right)}{\overline{Y}^2} = \frac{N-1}{nN}\{1+(n-1)\rho_Y\}C_Y^2 + \frac{L-1}{n}W_2\frac{S_{Y2}^2}{\overline{Y}^2}, \tag{2.6}$$

$$E\left(e_1^2\right) = \frac{V\left(\overline{x}\right)}{\overline{X}^2} = \frac{N-1}{nN}\{1+(n-1)\rho_X\}C_X^2 \tag{2.7}$$

and

$$E(e_0 e_1) = \frac{Cov\left(\bar{y}^*, \bar{x}\right)}{\bar{Y}\bar{X}} = \frac{N-1}{nN}\{1+(n-1)\rho_Y\}^{1/2}\{1+(n-1)\rho_X\}^{1/2}\rho C_Y C_X \qquad (2.8)$$

where C_Y and C_X are the coefficients of variation of study and auxiliary variables respectively in the population (for proof see Singh and Singh(1998) and Singh (2003, pg. no. 138)).

The biases and MSE's of the estimators $\bar{y}_R^*$, $\bar{y}_P^*$ and $\bar{y}_D^*$ up to the first order of approximation using (2.6-2.8), are respectively given by

$$B\left(\bar{y}_R^*\right) = \frac{N-1}{nN}\bar{Y}\{1+(n-1)\rho_X\}\left(1-K\rho^*\right)C_X^2, \qquad (2.9)$$

$$MSE\left(\bar{y}_R^*\right) = \frac{N-1}{nN}\bar{Y}^2\{1+(n-1)\rho_X\}\left[\rho^{*2}C_Y^2+\left(1-2K\rho^*\right)C_X^2\right] + \frac{L-1}{n}W_2 S_{Y2}^2, \qquad (2.10)$$

$$B\left(\bar{y}_P^*\right) = \frac{N-1}{nN}\bar{Y}\{1+(n-1)\rho_X\}K\rho^* C_X^2, \qquad (2.11)$$

$$MSE\left(\bar{y}_P^*\right) = \frac{N-1}{nN}\bar{Y}^2\{1+(n-1)\rho_X\}\left[\rho^{*2}C_Y^2+\left(1+2K\rho^*\right)C_X^2\right] + \frac{L-1}{n}W_2 S_{Y2}^2, \qquad (2.12)$$

$$B\left(\bar{y}_D^*\right) = \frac{N-1}{nN}\bar{Y}\{1+(n-1)\rho_X\}\left[-\rho^* K\right]C_X^2, \qquad (2.13)$$

$$MSE\left(\bar{y}_D^*\right) = \frac{N-1}{nN}\bar{Y}^2\{1+(n-1)\rho_X\}\left[\rho^{*2}C_Y^2+\left(\frac{f}{1-f}\right)\left\{\left(\frac{f}{1-f}\right)-2\rho^* K\right\}C_X^2\right]$$

$$+ \frac{(L-1)}{n}W_2 S_{Y2}^2 \qquad (2.14)$$

where,

$$\rho^* = \{1+(n-1)\rho_Y\}^{1/2} \Big/ \{1+(n-1)\rho_X\}^{1/2} \quad \text{and} \quad K = \rho\frac{C_Y}{C_X}.$$

(for details of proof refer to Singh et al.(2012)).

The regression estimator based on a systematic sample under non-response is given by

$$\bar{y}_{lr}^{*} = \bar{y}^{*} + b(\bar{X} - \bar{x}) \tag{2.15}$$

MSE of the estimator $\bar{y}_{lr}^{*}$ is given by

$$\mathrm{MSE}(\bar{y}_{lr}^{*}) = \frac{N-1}{nN}\bar{Y}^{2}\{1+(n-1)\rho_X\}\left[C_Y^2 - K^2 C_X^2\right]\rho^{*2} + \frac{(L-1)}{n}W_2 S_{Y2}^2 \tag{2.16}$$

3. Adapted Family of Estimators

Adapting the estimator proposed by Singh and Shukla (1987), a family of factor-type estimators of population mean in systematic sampling under non-response is written as

$$T_\alpha^{*} = \bar{y}^{*}\left[\frac{(A+C)\bar{X} + fB\bar{x}}{(A+fB)\bar{X} + C\bar{x}}\right]. \tag{3.1}$$

The constants A, B, C, and f are same as defined in (1.1).

It can easily be seen that the proposed family generates the non-response versions of some well known estimators of population mean in systematic sampling on putting different choices of α. For example, if we take α = 1, 2, 3 and 4, the resulting estimators will be ratio, product, dual to ratio and sample mean estimators of population mean in systematic sampling under non-response respectively.

3.1 Properties of T_α^{*}

Obviously, the proposed family is biased for the population mean $\bar{Y}$. In order to find the bias and MSE of T_α^{*}, we use large sample approximations. Expressing the equation (3.1) in terms of e_i's $(i = 0,1)$ we have

$$T_\alpha^{*} = \frac{\bar{Y}(1+e_0)(1+De_1)^{-1}\left[(A+C)+fB(1+e_1)\right]}{A+fB+C} \tag{3.2}$$

where $D = \dfrac{C}{A+fB+C}$.

Since $|D| < 1$ and $|e_i| < 1$, neglecting the terms of e_i's $(i = 0,1)$ having power greater than two, the equation (3.2) can be written as

$$T_\alpha^* - \bar{Y} = \frac{\bar{Y}}{A + fB + C}\Big[(A+C)\left\{e_0 - De_1 + D^2 e_1^2 - De_0 e_1\right\}$$

$$+ fB\left\{e_0 - (D-1)e_1 + D(D-1)e_1^2 - (D-1)e_0 e_1\right\}\Big]. \tag{3.3}$$

Taking expectation of both sides of the equation (3.3), we get

$$E\left[T_\alpha^* - \bar{Y}\right] = \frac{\bar{Y}(C - fB)}{A + fB + C}\left[\frac{C}{A + fB + C}E\left(e_1^2\right) - E\left(e_0 e_1\right)\right].$$

Let $\phi_1(\alpha) = \dfrac{fB}{A + fB + C}$ and $\phi_2(\alpha) = \dfrac{C}{A + fB + C}$ then

$$\phi(\alpha) = \phi_2(\alpha) \text{ - } \phi_1(\alpha) = \frac{C - fB}{A + fB + C}.$$

Thus, we have

$$E\left[T_\alpha^* - \bar{Y}\right] = \bar{Y}\phi(\alpha)\left[\phi_2(\alpha)E\left(e_1^2\right) - E\left(e_0 e_1\right)\right]. \tag{3.4}$$

Putting the values of $E\left(e_1^2\right)$ and $E\left(e_0 e_1\right)$ from equations (2.7) and (2.8) into the equation (3.4), we get the bias of T_α^* as

$$B\left(T_\alpha^*\right) = \phi(\alpha)\frac{N-1}{nN}\bar{Y}\left\{1 + (n-1)\rho_X\right\}\left[\phi_2(\alpha) - \rho^* K\right]C_X^2. \tag{3.5}$$

Squaring both the sides of the equation (3.3) and then taking expectation, we get

$$E\left[T_\alpha^* - \bar{Y}\right]^2 = \bar{Y}^2\left[E\left(e_0^2\right) + \phi^2(\alpha)E\left(e_1^2\right) - 2\phi(\alpha)E\left(e_0 e_1\right)\right]. \tag{3.6}$$

Substituting the values of $E\left(e_0^2\right)$, $E\left(e_1^2\right)$ and $E\left(e_0 e_1\right)$ from the respective equations (2.6), (2.7) and (2.8) into the equation (3.6), we get the MSE of T_α^* as

$$MSE\left(T_\alpha^*\right) = \frac{N-1}{nN}\bar{Y}^2\left\{1 + (n-1)\rho_X\right\}\left[\rho^{*2} C_Y^2 + \left\{\phi^2(\alpha) - 2\phi(\alpha)\rho^* K\right\}C_X^2\right]$$

$$+ \frac{(L-1)}{n}W_2 S_{Y2}^2. \tag{3.7}$$

3.2 Optimum Choice of α

In order to obtain the optimum choice of α, we differentiate the equation (3.7) with respect to α and equating the derivative to zero, we get the normal equation as

$$\frac{N-1}{nN}\bar{Y}^2\{1+(n-1)\rho_X\}[2\phi(\alpha)\phi'(\alpha)-2\phi'(\alpha)\rho^* K]C_X^2 = 0 \tag{3.8}$$

where $\phi'(\alpha)$ is the first derivative of $\phi(\alpha)$ with respect to α.

Now from equation (3.8), we get

$$\phi(\alpha) = \rho^* K \tag{3.9}$$

which is the cubic equation in α. Thus α has three real roots for which the MSE of proposed family would attain its minimum.

Putting the value of $\phi(\alpha)$ from equation (3.9) into equation (3.7), we get

$$MSE(T_\alpha^*)_{\min} = \frac{N-1}{nN}\bar{Y}^2\{1+(n-1)\rho_X\}[C_Y^2 - K^2 C_X^2]\rho^{*2} + \frac{(L-1)}{n}W_2 S_{Y2}^2 \tag{3.10}$$

which is the MSE of the usual regression estimator of population mean in systematic sampling under non-response.

4. Empirical Study

In the support of theoretical results, we have considered the data given in Murthy (1967, p. 131-132). These data are related to the length and timber volume for ten blocks of the blacks mountain experimental forest. The value of intraclass correlation coefficients ρ_X and ρ_Y have been given approximately equal by Murthy (1967, p. 149) and Kushwaha and Singh (1989) for the systematic sample of size 16 by enumerating all possible systematic samples after arranging the data in ascending order of strip length. The particulars of the population are given below:

$N = 176$, $n = 16$, $\bar{Y} = 282.6136$, $\bar{X} = 6.9943$,

$S_Y^2 = 24114.6700$, $S_X^2 = 8.7600$, $\rho = 0.8710$,

$$S_{Y2}^{2} = \frac{3}{4} S_{Y}^{2} = 18086.0025.$$

Table 1 depicts the MSE's and variance of the estimators of proposed family with respect to non-response rate (W_2).

Table 1: MSE and Variance of the Estimators for $L = 2$.

α	W_2			
	0.1	0.2	0.3	0.4
1 $(= \bar{y}_R^*)$	371.37	484.41	597.45	710.48
2 $(= \bar{y}_P^*)$	1908.81	2021.85	2134.89	2247.93
3$(= \bar{y}_D^*)$	1063.22	1176.26	1289.30	1402.33
4$(= \bar{y}^*)$	1140.69	1253.13	1366.17	1479.205
$\alpha_{opt} (= (T_\alpha^*)_{\min})$	270.67	383.71	496.75	609.78

5. Conclusion

In this paper, we have adapted Singh and Shukla (1987) estimator in systematic sampling in the presence of non-response using an auxiliary variable and obtained the optimum estimator of the proposed family. It is observed that the proposed family can generate the non-response versions of a number of estimators of population mean in systematic sampling on different choice of α. From Table 1, we observe that the proposed family under optimum condition has minimum MSE, which is equal to the MSE of the regression estimator (most of the class of estimators in sampling literature under optimum condition attains MSE equal to the MSE of the regression estimator). It is also seen that the MSE or variance of the estimators increases with increase in non response rate in the population.

References

1. Banarasi, Kushwaha, S.N.S. and Kushwaha, K.S. (1993): A class of ratio, product and difference (RPD) estimators in systematic sampling, *Microelectron. Reliab.*, 33, 4, 455–457.
2. Cochran, W. G. (1946): Relative accuracy of systematic and stratified random samples for a certain class of population, *AMS*, 17, 164-177.
3. Finney, D.J. (1948): Random and systematic sampling in timber surveys, *Forestry*, 22, 64-99.
4. Hansen, M. H. and Hurwitz, W. N. (1946) : The problem of non-response in sample surveys, *Jour. of The Amer. Stat. Assoc.*, 41, 517-529.
5. Hasel, A. A. (1942): Estimation of volume in timber stands by strip sampling, *AMS*, 13, 179-206.
6. Kushwaha, K. S. and Singh, H.P. (1989): Class of almost unbiased ratio and product estimators in systematic sampling, *Jour. Ind. Soc. Ag. Statistics*, 41, 2, 193–205.
7. Lahiri, D. B. (1954): On the question of bias of systematic sampling, *Proceedings of World Population Conference*, 6, 349-362.
8. Madow, W. G. and Madow, L.H. (1944): On the theory of systematic sampling, *I. Ann. Math. Statist.*, 15, 1-24.
9. Murthy, M.N. (1967): *Sampling Theory and Methods*. Statistical Publishing Society, Calcutta.
10. Nair, K. R. and Bhargava, R. P. (1951): Statistical sampling in timber surveys in India, *Forest Research Institute, Dehradun, Indian forest leaflet*, 153.
11. Quenouille, M. H. (1956): Notes on bias in estimation, *Biometrika*, 43, 353-360.
12. Singh, R and Singh, H. P. (1998): Almost unbiased ratio and product type- estimators in systematic sampling, *Questiio*, 22,3, 403-416.
13. Singh, R., Malik, S., Chaudhary, M.K., Verma, H. and Adewara, A. A. (2012) : A general family of ratio type- estimators in systematic sampling. Jour. Reliab. and Stat. Stud.,5(1), 73-82).
14. Singh, R., Malik, S., Singh, V. K. (2012) : An improved estimator in systematic sampling. Jour. Of Scie. Res., 56, 177-182.

15. Singh, H.P. and Kumar, S. (2009) : A general class of dss estimators of population ratio, product and mean in the presence of non-response based on the sub-sampling of the non-respondents. Pak J. Statist., 26(1), 203-238.

16. Singh, H.P. and Solanki, R. S. (2012) : An efficient class of estimators for the population mean using auxiliary information in systematic sampling. Jour. of Stat. Ther. and Pract., 6(2), 274-285.

17. Singh, S. (2003) : Advanced sampling theory with applications. Kluwer Academic Publishers.

18. Singh, V. K. and Shukla, D. (1987): One parameter family of factor-type ratio estimators, *Metron*, 45 (1-2), 273-283.