A Class Of Separate-Type Estimators For Population Mean In Stratified Sampling Using Known Parameters Under Non-Response

[1]Manoj K. Chaudhary, [2]Rajesh Singh , and [3]Mukesh Kumar, [4]Rakesh K. Shukla and [5]Florentin Smarandache

[1,2,3,4]Department of Statistics, Banaras Hindu University

Varanasi-221005, INDIA

[5]Chair of Department of Mathematics, University of New Mexico, Gallup, USA

[1]ritamanoj15@gmail.com,[2]rsinghstat@gmail.com,[3]Mukesh.stat@gamil.com,

[4]rimgyanu@gamil.com [5]smarand@unm.edu

ABSTRACT

The objective of the present paper is to propose a family of separate-type estimators of population mean in stratified random sampling in presence of non-response based on the family of estimators proposed by Khoshnevisan et al. (2007). Under simple random sampling without replacement (SRSWOR) the expressions of bias and mean square error (MSE) up to the first order of approximation are derived. The comparative study of the family with respect to usual estimator has been discussed. The expressions for optimum sample sizes of the strata in respect to cost of the survey have also been derived. An empirical study is carried out to shoe the properties of the estimators.

2000 Mathematics Subject Classification: 62D05

Key words and phrases : Stratified sampling, separate-type estimator, non-response, auxiliary information, mean square error.

1. INTRODUCTION

Upadhyaya and Singh (1999) have suggested the class of estimators in simple random sampling using some known population parameter(s) of an auxiliary variable. These estimators have been extended by Kadilar and Cingi (2003) for stratified random sampling. In an attempt to improve the estimators, Kadilar and Cingi (2005), Shabbir and Gupta (2005, 2006) and Singh and Vishwakarma (2008) have suggested new ratio estimators in stratified random sampling. Using power transformation Singh et al. (2008) have suggested a class of estimators adapting the estimators developed by Kadilar and Cingi (2003). Koyuncu and Kadilar (2008, 2009) have proposed a family of combined-type estimators in stratified random sampling based on the family of estimators proposed by Khoshnevisan et al. (2007). Singh et al. (2008) suggested some exponential ratio type estimators in stratified random sampling. Recently Koyuncu and Kadilar (2010) have suggested a family of estimators in stratified random sampling following Diana (1993) and Kadilar and Cingi (2003).

Let Y and X be the study and auxiliary variables respectively, with respective population means $\overline{Y}$ and $\overline{X}$. Khoshnevisan et al. (2007) have proposed a family of estimators for population mean using known values of some population parameters in simple random sampling (SRS) given by

$$t = \overline{y}\left[\frac{a\overline{X}+b}{\alpha(a\overline{x}+b)+(1-\alpha)(a\overline{X}+b)}\right]^{g} \qquad (1.1)$$

where $a \neq 0$, b are either real numbers or functions of known parameters of the auxiliary variable X.

In this paper, we have proposed a family of separate-type estimators of population mean in stratified random sampling in presence of non-response on study variable adapting the above family of estimators. The properties of the proposed family of estimators in comparison with usual estimators have been discussed. The expressions for optimum sample sizes of the strata with respect to cost of the survey have been obtained.

2. SAMPLING STRATEGY AND ESTIMATION PROCEDURES

Let us consider a finite population of size, N , is divided into k strata. Let N_i be the size of i^{th} stratum $(i = 1,2,...,k)$ and a sample of size n_i is drawn from the i^{th} stratum using SRSWOR scheme such that $\sum_{i=1}^{k} n_i = n$. It is assumed that the non-response is detected on study variable Y only and auxiliary variable X is free from non-response.

Let $\bar{y}_i^*$ and $\bar{x}_i$ are the unbiased estimators of population means $\bar{Y}_i$ and $\bar{X}_i$ respectively, for the i^{th} stratum, given as

$$\bar{y}_i^* = \frac{n_{i1}\bar{y}_{ni1} + n_{i2}\bar{y}_{ui2}}{n_i} \tag{2.1}$$

where $\bar{y}_{ni1}$ and $\bar{y}_{ui2}$ are the means based on n_{i1} response units and u_{i2} non-response units of sub sample selected from n_{i2} non-response units respectively. $\bar{x}_i$ be the sample mean based on n_i units.

Therefore an unbiased estimator of population mean $\bar{Y}$ is given by

$$\bar{y}_{st}^* = \sum_{i=1}^{k} p_i \bar{y}_i^* \tag{2.2}$$

and variance of the estimator is expressed as

$$V(\bar{y}_{st}^*) = \sum_{i=1}^{k}\left(\frac{1}{n} - \frac{1}{N}\right)p_i^2 S_{yi}^2 + \sum_{i=1}^{k}\frac{(k_i - 1)}{n_i} W_{i2} p_i^2 S_{yi2}^2 \tag{2.3}$$

where S_{yi}^2 and S_{yi2}^2 are respectively the mean-squares of entire group and non-response group of study variable in the population for the i^{th} stratum. $k_i = \frac{n_{i2}}{u_{i2}}$, $p_i = \frac{N_i}{N}$ and $W_{i2} =$ non-response rate of the i^{th} stratum in the population $= \frac{N_{i2}}{N_i}$.

2.1 SUGGESTED FAMILY OF ESTIMATORS

Adapting the idea of Khoshnevisan et al. (2007), we propose a family of separate-type estimators of population mean $\overline{Y}$, given by

$$T_S = \sum_{i=1}^{k} p_i T_i^* \tag{2.4}$$

where $$T_i^* = \overline{y}_i^* \left[\frac{a\overline{X}_i + b}{\alpha(a\overline{x}_i + b) + (1-\alpha)(a\overline{X}_i + b)} \right]^g \tag{2.5}$$

Obviously, T_S is biased for $\overline{Y}$. Therefore, bias and MSE of T_S can be obtained on using large sample approximations. Let

$$\overline{y}_i^* = \overline{Y}_i(1+e_0) \; ; \; \overline{x}_i = \overline{X}_i(1+e_1)$$

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

$$E(e_0^2) = \frac{V(\overline{y}_i^*)}{\overline{Y}_i^2} = f_i C_{Yi}^2 + \frac{(k_i - 1)W_{i2} S_{Yi2}^2}{n_i \overline{Y}_i^2},$$

$$E(e_1^2) = \frac{V(\overline{x}_i)}{\overline{X}_i^2} = f_i C_{Xi}^2,$$

$$E(e_0 e_1) = \frac{Cov(\overline{y}_i^*, \overline{x}_i)}{\overline{Y}_i \overline{X}_i^2} = f_i \rho_i C_{Yi} C_{Xi},$$

where $f_i = \dfrac{N_i - n_i}{N_i n_i}$, $C_{Yi}^2 = \dfrac{S_{Yi}^2}{\overline{Y}_i^2}$, $C_{Xi}^2 = \dfrac{S_{Xi}^2}{\overline{X}_i^2}$, S_{Xi}^2 be the mean-square of entire group of auxiliary variable in the population for the i^{th} stratum and ρ_i is the correlation coefficient between Y and X in the i^{th} stratum.

Expressing the estimator T_S in terms of e_0 and e_1, we get

$$T_S = \sum_{i=1}^{k} p_i \overline{Y}_i (1+e_0)[1 + \alpha\lambda_i e_1]^{-g} \tag{2.6}$$

where $\lambda_i = \dfrac{a\overline{X}_i}{a\overline{X}_i + b}$.

Suppose $|\alpha\lambda_i e_1| < 1$ so that $[1+\alpha\lambda_i e_1]^{-g}$ is expandable. Expanding the right hand side of the equation (2.6) up to the first order of approximation, we obtain

$$\left(T_S - \overline{Y}\right) = \sum_{i=1}^{k} \overline{Y}_i \left[e_0 - g\alpha\lambda_i e_1 + \frac{g(g+1)}{2}\alpha^2 \lambda_i^2 e_1^2 - g\alpha\lambda_i e_0 e_1 \right] \tag{2.7}$$

Taking expectations of both sides of (2.7), we get the bias of T_S up to the first order of approximation, as

$$B(T_S) = \sum_{i=1}^{k} p_i f_i \overline{Y}_i \left[\frac{g(g+1)}{2}\alpha^2 \lambda_i^2 C_{Xi}^2 - \alpha\lambda_i g\rho_i C_{Yi} C_{Xi} \right] \tag{2.8}$$

Squaring both side of equation (2.7) and taking expectations on both sides of this equation, we get the MSE(T_S) to the first order of approximation as given below:

$$MSE(T_S) = \sum_{i=1}^{k} p_i^2 \left[f_i \overline{Y}_i^2 \left(C_{Yi}^2 + \alpha^2 \lambda_i^2 g^2 C_{Xi}^2 - 2\alpha\lambda_i g\rho_i C_{Yi} C_{Xi} \right) + \frac{(k_i - 1)}{n_i} W_{i2} S_{Yi2}^2 \right] \tag{2.9}$$

2.2 SOME SPECIAL CASES

Case 1: If we put $\alpha = 1$**,** $a = 1, b = 0$ and $g = 1$ in equation (2.4)**,** we get

$$T_S = \sum_{i=1}^{k} p_i \overline{y}_i^* \frac{\overline{X}_i}{\overline{x}_i} \tag{2.10}$$

which is separate ratio estimator of population mean $\overline{Y}$ under non-response**.**

Case 2: If $\alpha = 1$**,** $a = 1, b = 0$ and $g = -1$**,** the equation (2.4) gives

$$T_S = \sum_{i=1}^{k} p_i \overline{y}_i^* \frac{\overline{x}_i}{\overline{X}_i} \tag{2.11}$$

which is separate product estimator of population mean $\overline{Y}$ under non-response**.**

Case 3: If we take $\alpha = 0$**,** $a = 0, b = 0$ and $g = 0$**,** the equation (2.4) provides

$$T_S = \sum_{i=1}^{k} p_i \overline{y}_i^* \tag{2.12}$$

which is the usual estimator of population mean $\overline{Y}$ under non-response**.**

Similarly, we can obtain the various existing estimators of the family under non-response on different choices of α , a , b and g [See Khoshnevisan et al. (2007)].

2.3 OPTIMUM CHOICE OF α

In order to obtain the optimum α we minimize $MSE(T_S)$ with respect to α. Differentiating $MSE(T_S)$ with respect to α and equating the derivative to zero, we get the normal equation

$$\frac{\partial MSE(T_S)}{\partial \alpha} = \sum_{i=1}^{k} p_i^2 f_i \overline{Y}_i^2 \left[2\alpha\lambda_i^2 C_{Xi}^2 - 2\alpha\lambda_i g\rho_i C_{Yi} C_{Xi}\right] = 0 \tag{2.13}$$

$$\Rightarrow \alpha_{(opt)} = \frac{\sum_{i=1}^{k} p_i^2 f_i \overline{Y}_i^2 \rho_i C_{Yi} C_{Xi}}{\lambda_i g \sum_{i=1}^{k} p_i^2 f_i \overline{Y}_i^2 C_{Xi}^2} \tag{2.14}$$

Thus the equation (2.14) provides the value of α at which $MSE(T_S)$ would be minimum.

2.4 OPTIMUM n_i WITH RESPECT TO COST OF THE SURVEY

Let C_{i0} be the cost per unit of selecting n_i units, C_{i1} be the cost per unit in enumerating n_{i1} units and C_{i2} be the cost per unit of enumerating u_{i2} units. Then the total cost for the i^{th} stratum is given by

$$C_i = C_{i0} n_i + C_{i1} n_{i1} + C_{i2} u_{i2} \qquad \forall \; i = 1,2,...,k \tag{2.15}$$

Now, we consider the average cost per stratum

$$E(C_i) = n_i \left[C_{i0} + C_{i1} W_{i1} + C_{i2} \frac{W_{i2}}{k_i} \right] \tag{2.16}$$

Thus the total cost over all the strata is given by

$$C_0 = \sum_{i=1}^{k} E(C_i)$$

$$= \sum_{i=1}^{k} n_i \left[C_{i0} + C_{i1} W_{i1} + C_{i2} \frac{W_{i2}}{k_i} \right] \tag{2.17}$$

Let us consider the function

$$\phi = MSE(T_S) + \mu C_0 \tag{2.18}$$

where μ is Lagrangian multiplier. Differentiating the equation (2.18) with respect to n_i and k_i respectively and equating the derivatives to zero, we get the following normal equations:

$$\frac{\partial \phi}{\partial n_i} = -\frac{p_i^2}{n_i^2}\left[\overline{Y}_i^2\left(C_{Yi}^2 + \alpha^2\lambda_i^2 g^2 C_{Xi}^2 - 2\alpha\lambda_i g\rho_i C_{Yi}C_{Xi}\right) + (k_i - 1)W_{i2}S_{Yi2}^2\right] +$$

$$\mu\left(C_{i0} + C_{i1}W_{i1} + C_{i2}\frac{W_{i2}}{k_i}\right) = 0 \qquad (2.19)$$

$$\frac{\partial \phi}{\partial k_i} = \frac{p_i^2 W_{i2}S_{Yi2}^2}{n_i} - \mu n_i C_{i2}\frac{W_{i2}}{k_i^2} = 0 \qquad (2.20)$$

From the equations (2.19) and (2.20), we have

$$n_i = \frac{p_i\sqrt{\overline{Y}_i^2\left(C_{Yi}^2 + \alpha^2\lambda_i^2 g^2 C_{Xi}^2 - 2\alpha\lambda_i g\rho_i C_{Yi}C_{Xi}\right) + (k_i - 1)W_{i2}S_{Yi2}^2}}{\sqrt{\mu}\sqrt{C_{i0} + C_{i1}W_{i1} + C_{i2}\frac{W_{i2}}{k_i}}} \qquad (2.21)$$

and $\sqrt{\mu} = \dfrac{k_i p_i S_{Yi2}}{n_i\sqrt{C_{i2}}}$ (2.22)

Putting the value of the $\sqrt{\mu}$ from equation (2.22) into the equation (2.21), we get

$$k_{i(opt)} = \frac{\sqrt{C_{i2}}B_i}{S_{Yi2}A_i} \qquad (2.23)$$

where $A_i = \sqrt{C_{i0} + C_{i1}W_{i1}}$

and $B_i = \sqrt{\overline{Y}_i^2\left(C_{Yi}^2 + \alpha^2\lambda_i^2 g^2 C_{Xi}^2 - 2\alpha\lambda_i g\rho_i C_{Yi}C_{Xi}\right) - W_{i2}S_{Yi2}^2}$

On substituting $k_{i(opt)}$ into equation (2.21), n_i can be expressed as

$$n_i = \frac{p_i\sqrt{B_i^2 + \frac{\left(\sqrt{C_{i2}}B_iW_{i2}S_{Yi2}\right)}{A_i}}}{\sqrt{\mu}\sqrt{A_i^2 + \frac{\sqrt{C_{i2}}A_iW_{i2}S_{Yi2}}{B_i}}} \tag{2.24}$$

The $\sqrt{\mu}$ in terms of total cost C_0 can be obtained by putting the values of $k_{i(opt)}$ and n_i from equations (2.23) and (2.24) respectively into equation (2.17) as

$$\sqrt{\mu} = \frac{1}{C_0}\sum_{i=1}^{k} p_i\left(A_iB_i + \sqrt{C_{i2}}W_{i2}S_{Yi2}\right) \tag{2.25}$$

Thus the n_i can be expressed in terms of the total cost C_0 as

$$n_{i(opt)} = \frac{C_0}{\sum_{i=1}^{k} p_i\left(A_iB_i + \sqrt{C_{i2}}W_{i2}S_{Yi2}\right)} \frac{p_i\sqrt{B_i^2 + \frac{\left(\sqrt{C_{i2}}B_iW_{i2}S_{Yi2}\right)}{A_i}}}{\sqrt{A_i^2 + \frac{\sqrt{C_{i2}}A_iW_{i2}S_{Yi2}}{B_i}}} \tag{2.26}$$

The optimum values of n_i and k_i can be obtained by the expressions (2.26) and (2.23) respectively.

3. EMPIRICAL STUDY

In this section, we use the data set in Koyuncu and Kadilar (2009). The data concerning the number of teachers as study variable and the number of students as auxiliary variable in both primary and secondary school for 923 districts at 6 regions (as 1: Marmara, 2: Agean, 3: Mediterranean, 4: Central Anatolia, 5: Black Sea, 6: East and Southeast Anatolia) in Turkey in 2007 (Source: Ministry of Education Republic of Turkey).

Table 1: Stratum means, Mean Squares and Correlation Coefficients

Stratum No.	N_i	n_i	$\overline{Y}_i$	$\overline{X}_i$	S_{Yi}	S_{Xi}	S_{XYi}	ρ_i	S_{Yi2}
1	127	31	703.74	20804.59	883.835	30486.751	25237153.52	.936	440
2	117	21	413.00	9211.79	644.922	15180.769	9747942.85	.996	200
3	103	29	573.17	14309.30	1033.467	27549.697	28294397.04	.994	400
4	170	38	424.66	9478.85	810.585	18218.931	14523885.53	.983	405
5	205	22	527.03	5569.95	403.654	8497.776	3393591.75	.989	180
6	201	39	393.84	12997.59	711.723	23094.141	15864573.97	.965	300

Table 2: Percent Relative Efficiency (P.R.E.) of T_S with respect to $\overline{y}^{*}_{st}$ at $\alpha_{(opt)}$ = 0.9317, $a=1$ and $b=1$

W_{i2}	k_i	$P.R.E.(T_S)$
0.1	2.0	1319.17
	2.5	1153.15
	3.0	1026.92
	3.5	927.72
0.2	2.0	1026.92
	2.5	847.70
	3.0	726.55
	3.5	639.18
0.3	2.0	847.70
	2.5	679.59
	3.0	573.20
	3.5	499.81
0.4	2.0	726.55
	2.5	573.20
	3.0	480.16
	3.5	417.69

4. CONCLUSION

In this paper, a class of separate-type estimators for estimating the population mean in stratified random sampling under non-response has been proposed and method of finding the optimum estimator of the family has also been discussed. We have derived the expressions for optimum sample sizes in respect to cost of the survey. From the Table 2, it is easily observed that the optimum estimator of the proposed class T_S provides better estimate than usual estimator $\bar{y}^*_{st}$ under non-response. It is also observed that the relative efficiency of T_S decreases with increase in the non-response rate W_{i2} and k_i.

REFERENCES

Diana, G. (1993) : A class of estimators of the population mean in stratified random sampling, *Statistica*, 53 (1), 59–66.

Hansen, M. H. and Hurvitz, W. N. (1946) : The problem of non-response in sample surveys, *Journal of American Statistical Association,* 41, 517-529.

Kadilar, C. and Cingi, H. (2003) : Ratio estimator in stratified sampling, *Biometrical Journal*, 45,218-225.

Kadilar, C. and Cingi, H. (2005) : A new estimator in stratified random sampling, *Communication in Statistics Theory and Methods, 34, 597-602*.

Khoshnevisan, M., Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. (2007) : A general family of estimators for estimating population mean using known value of some population parameter(s), *Far East Journal of Theoretical Statistics,* 22, 181-191.

Koyuncu, N. and Kadilar, C. (2008) : Ratio and product estimators in stratified random sampling, *Journal of Statistical Planning and Inference*, 139, 8, 2552-2558.

Koyuncu, N., Kadilar, C. (2009) : Family of estimators of population mean using two auxiliary variables in stratified random sampling, *Communication in Statistics Theory and Methods, 38:14, 2398-2417.*

Koyuncu, N. and Kadılar, C. (2010) : On the family of estimators of population mean in stratified random sampling, *Pakistan Journal of Statistics*, 26, 2, 427-443.

Shabbir, J. and Gupta, S. (2005) : Improved ratio estimators in stratified sampling, *American Journal of Mathematical and Management Sciences,* 25, 293-311.

Shabbir, J. and Gupta, S. (2006) : A new estimator of population mean in stratified sampling, *Communication in Statistics Theory and Methods,* 35, 1201-1209.

Singh, H., P., Tailor, R. Singh S. and Kim, J. M. (2008) : A modified estimator of population mean using power transformation, *Statistical papers,* Vol-49, No.1, 37-58.

Singh, H., P. and Vishwakarma, G. K. (2008) : A family of estimators of population mean using auxiliary information in stratified sampling, *Communication in Statistics Theory and Methods,* 37(7), 1038-1050.

Singh, R., Kumar, M., Singh, R. D., and Chaudhary, M.K. (2008): Exponential Ratio Type Estimators in Stratified Random Sampling. Presented in International Symposium on Optimisation and Statistics (I.S.O.S) at A.M.U., Aligarh, India, during 29-31 Dec 2008.

Upadhyaya, L. N. and Singh, H. P. (1999) : Use of transformed auxiliary variable in estimating the finite population mean, *Biometrical Journal*. 41, 627-636.