A FAMILY OF ESTIMATORS FOR ESTIMATING POPULATION MEAN IN STRATIFIED SAMPLING UNDER NON-RESPONSE

MANOJ K. CHAUDHARY, RAJESH SINGH, RAKESH K. SHUKLA, MUKESH KUMAR, FLORENTIN SMARANDACHE

Abstract

Khoshnevisan et al. (2007) proposed a general family of estimators for population mean using known value of some population parameters in simple random sampling. The objective of this paper is to propose a family of combined-type estimators in stratified random sampling adapting the family of estimators proposed by Khoshnevisan et al. (2007) under non-response. The properties of proposed family have been discussed. We have also obtained the expressions for optimum sample sizes of the strata in respect to cost of the survey. Results are also supported by numerical analysis.

1. Introduction

There are several authors who have suggested estimators using some known population parameters of an auxiliary variable. Upadhyaya and Singh (1999) and Singh et al. (2007) have suggested the class of estimators in simple random sampling. Kadilar and Cingi (2003) adapted Upadhyaya and Singh (1999) estimator in stratified random sampling. Singh et al. (2008) suggested class of estimators using power transformation based on the estimators developed by Kadilar and Cingi (2003). Kadilar and Cingi (2005), Shabbir and Gupta (2005, 06) and Singh and Vishwakarma (2008) have suggested new ratio estimators in stratified sampling to improve the efficiency of the estimators.

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

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

where $a \neq 0$ and b are either real numbers or functions of known parameters of auxiliary variable X. Koyuncu and Kadilar (2008, 09) have proposed family of combined-type estimators for estimating population mean in stratified random sampling by adapting the estimator of Khoshnevisan et al. (2007). These authors assumed that there is complete response from all the sample units. It is fact in most of the surveys that information is usually not obtained from all the sample units even after callbacks. The method of sub-sampling the non-respondents proposed by Hansen and Hurwitz (1946) can be applied in order to adjust the non-response in a mail survey.

In the next sections, we have tried to propose a family of combined-type estimators considering the above family of estimators in stratified random sampling under non-response. We have discussed the properties of proposed family of estimators. We have also derived the expressions for optimum sample sizes of the strata in respect to cost of the survey.

2. Sampling Strategies and Estimation Procedure

Let us consider a population consisting of N units divided into k strata. Let the size of i^{th} stratum is N_i, ($i = 1,2,..................,k$). We decide to select a sample of size n from the entire population in such a way that n_i units are selected from the N_i units in the i^{th} stratum. Thus, we have $\sum_{i=1}^{k} n_i = n$. Let Y and X be the study and auxiliary characteristics respectively with respective population mean $\overline{Y}$ and $\overline{X}$. It is considered that the non-response is detected on study variable Y only and auxiliary variable X is free from non-response.

Let $\overline{y}_i^*$ be the unbiased estimator of population mean $\overline{Y}_i$ for the i^{th} stratum, given by

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

where $\overline{y}_{ni1}$ and $\overline{y}_{ui2}$ are the means based on n_{i1} units of response group and u_{i2} units of sub-sample of non-response group respectively in the sample for the i^{th} stratum. $\overline{x}_i$ be the unbiased estimator of population mean $\overline{X}_i$, based on n_i sample units in the i^{th} stratum.

Using Hansen-Hurwitz technique, an unbiased estimator of population mean $\overline{Y}$ is given by

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

and the variance of the estimator is given by the following expression

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

where S_{yi}^2 and S_{yi2}^2 are respectively the mean-square errors 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 Proposed Estimators

Motivated by Khoshnevisan et al. (2007), we propose a family of combined-type estimators of population mean $\bar{Y}$, given by

$$T_C = \bar{y}_{st}^{*}\left[\frac{a\bar{X}+b}{\alpha(a\bar{x}_{st}+b)+(1-\alpha)(a\bar{X}+b)}\right]^{g} \tag{2.1.1}$$

where $\bar{x}_{st} = \sum_{i=1}^{k} p_i \bar{x}_i$ (unbiased for $\bar{X}$)

and $\bar{X} = \sum_{i=1}^{k} p_i \bar{X}_i$.

Obviously, T_C is biased. The bias and MSE can be obtained on using large sample approximations:

$\bar{y}_{st}^{*} = \bar{Y}(1+e_0)$; $\bar{x}_{st} = \bar{X}(1+e_1)$

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

$$E(e_0^2) = \frac{V\left(\bar{y}_{st}^{*}\right)}{\bar{Y}^2} = \frac{1}{\bar{Y}^2}\sum_{i=1}^{k} p_i^2\left[f_i S_{Yi}^2 + \frac{(k_i-1)}{n_i} W_{i2} S_{Yi2}^2\right]$$

$$E(e_1^2) = \frac{V(\bar{x}_{st})}{\bar{X}^2} = \frac{1}{\bar{X}^2}\sum_{i=i}^{k} p_i^2 f_i S_{Xi}^2$$

$$E(e_0 e_1) = \frac{Cov\left(\bar{y}_{st}^{*}, \bar{x}_{st}\right)}{\bar{Y}\bar{X}} = \frac{1}{\bar{Y}\bar{X}}\sum_{i=1}^{k} p_i^2 f_i \rho_i S_{Yi} S_{Xi}$$

where $f_i = \frac{N_i - n_i}{N_i n_i}$, S_{Xi}^2 be the mean-square error 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 T_C in terms of e_i $(i=0,1)$, we can write (2.1.1) as

$$T_C = \overline{Y}(1+e_0)[1+\alpha\lambda e_1]^{-g} \tag{2.1.2}$$

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

Suppose $|\alpha\lambda e_1|$< 1 so that $[1+\alpha\lambda e_1]^{-g}$ is expandable. Expanding the right hand side of (2.1.2) up to the first order of approximation, we obtain

$$\left(T_C - \overline{Y}\right) = \overline{Y}\left[e_0 - g\alpha\lambda e_1 + \frac{g(g+1)}{2}\alpha^2\lambda^2 e_1^2 - g\alpha\lambda e_0 e_1\right] \tag{2.1.3}$$

Taking expectation of both sides in (2.1.3), we get the bias of the estimator T_C as

$$B(T_C) = \frac{1}{\overline{Y}}\sum_{i=1}^{k} f_i p_i^2\left[\frac{g(g+1)}{2}\alpha^2\lambda^2 R^2 S_{Xi}^2 - \alpha\lambda g R \rho_i S_{Yi} S_{Xi}\right] \tag{2.1.4}$$

Squaring both sides of (2.1.3) and then taking expectation, we get the MSE of the estimator T_C , up to the first order approximation, as

$$MSE(T_C) = \sum_{i=1}^{k} f_i p_i^2\left[S_{Yi}^2 + \alpha^2\lambda^2 g^2 R^2 S_{Xi}^2 - 2\alpha\lambda g R \rho_i S_{Yi} S_{Xi}\right] + \sum_{i=1}^{k} p_i^2 \frac{(k_i - 1)}{n_i} W_{i2} S_{Yi2}^2 \tag{2.1.5}$$

Optimum choice of α

On minimizing $MSE(T_C)$ w.r.t. α **,** we get the optimum value of α as

$$\frac{\partial MSE(T_C)}{\partial\alpha} = 2\alpha\lambda^2 g^2 R^2 \sum_{i=1}^{k} f_i p_i^2 S_{Xi}^2 - 2\lambda g R \sum_{i=1}^{k} f_i p_i^2 \rho S_{Yi} S_{Xi} = 0$$

$$\Rightarrow \alpha_{(opt)} = \frac{\sum_{i=1}^{k} f_i p_i^2 \rho_i S_{Yi} S_{Xi}}{\lambda g R \sum_{i=1}^{k} f_i p_i^2 S_{Xi}^2} \tag{2.1.6}$$

Thus $\alpha_{(opt)}$ is the value of α at which $MSE(T_C)$ would attain its minimum.

3. 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

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

Now, we consider the average cost per stratum

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

Thus the total cost over all the strata is given by

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

$$= \sum_{i=1}^{k} n_i\left[C_{i0} + C_{i1}W_{i1} + C_{i2}\frac{W_{i2}}{k_i}\right] \quad (3.1)$$

Let us consider the function

$$\phi = MSE(T_C) + \mu C_0 \quad (3.2)$$

where μ is Lagrangian multiplier. Differentiating the equation (3.2) with respect to n_i and k_i separately and equating to zero, we get the following normal equations.

$$\frac{\partial\phi}{\partial n_i} = -\frac{p_i^2}{n_i^2}\left[S_{Yi}^2 + \alpha^2\lambda^2 g^2 R^2 S_{Xi}^2 - 2\alpha\lambda g R\rho_i S_{Yi}S_{Xi}\right] - \frac{p_i^2}{n_i^2}(k_i - 1)W_{i2}S_{Yi2}^2$$
$$+ \mu\left[C_{i0} + C_{i1}W_{i1} + C_{i2}\frac{W_{i2}}{k_i}\right] = 0 \quad (3.3)$$

$$\frac{\partial\phi}{\partial k_i} = \frac{p_i^2 W_{i2}S_{Yi2}^2}{n_i} - \mu n_i C_{i2}\frac{W_{i2}}{k_i^2} = 0 \quad (3.4)$$

From the equations (3.3) and (3.4) respectively, we have

$$n_i = \frac{p_i\sqrt{S_{Yi}^2 + \alpha^2\lambda^2 g^2 R^2 S_{Xi}^2 - 2\alpha\lambda g R\rho_i S_{Yi}S_{Xi} + (k_i - 1)W_{i2}S_{Yi2}^2}}{\sqrt{\mu}\sqrt{C_{i0} + C_{i1}W_{i1} + C_{i2}\frac{W_{i2}}{k_i}}} \quad (3.5)$$

and

$$\sqrt{\mu} = \frac{k_i p_i S_{Yi2}}{n_i\sqrt{C_{i2}}} \quad (3.6)$$

Putting the value of the $\sqrt{\mu}$ from equation (3.6) into the equation (3.5), we get

$$k_{i(opt)} = \frac{\sqrt{C_{i2}}B_i}{S_{Yi2}A_i} \quad (3.7)$$

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

and $B_i = \sqrt{S_{Yi}^2 + \alpha^2\lambda^2 g^2 R^2 S_{Xi}^2 - 2\alpha\lambda g R\rho_i S_{Yi}S_{Xi} - W_{i2}S_{Yi2}^2}$

Substituting $k_{i(opt)}$ from equation (3.7) into equation (3.5), n_i can be expressed as

$$n_i = \frac{p_i\sqrt{B_i^2 + \frac{\left(\sqrt{C_{i2}}B_i W_{i2} S_{Yi2}\right)}{A_i}}}{\sqrt{\mu}\sqrt{A_i^2 + \frac{\sqrt{C_{i2}}A_i W_{i2} S_{Yi2}}{B_i}}} \tag{3.8}$$

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 (3.7) and (3.8) respectively into equation (3.1)

$$\sqrt{\mu} = \frac{1}{C_0}\sum_{i=1}^{k} p_i\left(A_i B_i + \sqrt{C_{i2}}W_{i2} S_{Yi2}\right) \tag{3.9}$$

Now we can express n_i in terms of total cost C_0

$$n_{i(opt)} = \frac{C_0}{\sum_{i=1}^{k} p_i\left(A_i B_i + \sqrt{C_{i2}}W_{i2} S_{Yi2}\right)} \frac{p_i\sqrt{B_i^2 + \frac{\left(\sqrt{C_{i2}}B_i W_{i2} S_{Yi2}\right)}{A_i}}}{\sqrt{A_i^2 + \frac{\sqrt{C_{i2}}A_i W_{i2} S_{Yi2}}{B_i}}} \tag{3.10}$$

Thus $n_{i(opt)}$ can be obtained by equation (3.10) by putting different values of W_{i2} and k_i.

4. Numerical Analysis

For numerical analysis we have used data considered by Koyuncu and Kadilar (2008). 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). Details are given below:

Table No.4.1: Stratum means, Mean Square Errors and Correlation Coefficients S_{Yi2}

Stratum No.	N_i	n_i	$\bar{Y}_i$	$\bar{X}_i$	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 No.4.2: % Relative efficiency (R.E.) of T_C w.r.t. to $\bar{y}_{st}^{*}$ at $\alpha_{(opt)}$, $a=1, b=1$

W_{i2}	k_i	$R.E.(T_C)$
0.1	2.0	914.25
	2.5	834.05
	3.0	768.23
	3.5	713.25
0.2	2.0	768.23
	2.5	666.62
	3.0	591.84
	3.5	534.49
0.3	2.0	666.62
	2.5	561.39
	3.0	489.12
	3.5	436.42
0.4	2.0	591.84
	2.5	489.12
	3.0	421.89
	3.5	374.47

5. Conclusion

We have proposed a family of estimators in stratified sampling using an auxiliary variable in the presence of non-response on study variable. We have also derived the expressions for optimum sample sizes in respect to cost of the survey. Table 4.2 reveals that the proposed estimator T_C has greater precision than the usual estimator $\bar{y}_{st}^{*}$ under non-response.

References

1. Hansen, M. H., Hurwitz, W. N. (1946): The problem of non-response in sample surveys. Journal of American Statistical Association, 41, 517-529.
2. Kadilar, C., Cingi, H. (2005): A new estimator in stratified random sampling. Communication in Statistics Theory and Methods, 34, 597-602.
3. Kadilar, C., Cingi, H. (2003): Ratio estimator in stratified sampling. Biometrical Journal, 45, 218-225.
4. Khoshnevisan, M., Singh, R., Chauhan, P., Sawan, N., 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.

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

6. 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.

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

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

9. 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.

10. Singh, H., P., 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.

11. Singh, R., Cauhan, P., Sawan, N. and Smarandache, F. (2007): Auxiliary information and a priori values in construction of improved estimators. Renaissance High press.

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

Published in "Pakistani Journal of Statistics and Operational Research", Vol. V, No. 1, pp. 47-54, 2009.