A Family of Estimators of a Sensitive Variable Using Auxiliary Information in Stratified Random Sampling

Nadia Mushtaq
National College of Business Administration and Economics
Lahore. Pakistan
dia_qau@yahoo.com

Muhammad Noor-ul-Amin
COMSATS Institute of Information and Technology
nooramin.stats@gmail.com

Muhammad Hanif
National College of Business Administration and Economics
Lahore. Pakistan
drhanif@ncbae.edu.pk

Abstract

In this article, a combined general family of estimators is proposed for estimating finite population mean of a sensitive variable in stratified random sampling with non-sensitive auxiliary variable based on randomized response technique. Under stratified random sampling without replacement scheme, the expression of bias and mean square error (MSE) up to the first-order approximations are derived. Theoretical and empirical results through a simulation study show that the proposed class of estimators is more efficient than the existing estimators, i.e., usual stratified random sample mean estimator, Sousa et al (2014) ratio and regression estimator of the sensitive variable in stratified sampling.

Keywords: Stratified Random Sampling, Sensitive Variable, Randomized Response Technique.

1. Introduction

It is common practice in sample survey related to agriculture, market, industries, and social research, and so forth that usually more than one characteristic is observed from each sampled unit of population. Stratified random sampling is more suitable than other survey designs used for obtaining information from heterogeneous population for reasons of economy and efficiency. And the problem of estimation of the population parameters of a sensitive quantitative variable is well known in survey sampling. In this study, the main goal is to propose a combined general family of estimators for estimating the finite population mean of a sensitive variable in stratified random sampling with non-sensitive auxiliary variable based on randomized response technique.

Many authors have discussed ratio and regression estimators when both Y and X are directly observable. These include Kadilar and Cingi (2003), Shabbir and Gupta (2005), and Nangsue (2009). Gupta and shabbir (2008) have suggested a general class of ratio estimators when the population parameters of the auxiliary variable are known. These estimators have also been extended by kadilar and Cingi (2003) to stratified random sampling scheme. Koyuncu and Kadilar (2010) have suggested a family of estimators in stratified random sampling following Kadilar and Cingi (2003). Sousa et al. (2010) and Gupta et al. (2012) have introduced ratio and regression mean estimators for a sensitive variable and Sousa et al. (2014) have suggested mean estimation of a sensitive variable using auxiliary information in stratified random sampling.

This paper suggests a combined general family of estimators of population mean of a sensitive variable using non-sensitive auxiliary information, using RRT methodology (Warner 1965; Gupta et al. 2002 and 2010) in stratified random sampling. Under stratified random sampling without replacement scheme, the expression of bias and mean square error (MSE) up to the first-order approximations are derived. Theoretical and empirical results through a simulation results support the reliability of the present study.

2. Terminology

We denote the finite population $U=(U_1,U_2,...,U_N)$. "Consider stratifies random samples (Cochran, 1977), selected from U with sampling rate $f=n/N$. The study population is divided into L strata with strata size N_h, such that $\sum_{h=1}^{L} N_h = N \ (h=1,...,L)$. Let Y be the study variable, a sensitive variable which cannot be observed directly due to respondent bias. Let X be non-sensitive auxiliary variable which is correlated with Y. Let S be scrambling variable independent of Y and X.

Consider a stratified random sample of size n be drawn from U such that the sample size in the h^{th} stratum is n_h and $\sum_{h=1}^{L} n_h = n$. Let y_{hi} and x_{hi} respectively be the values of the study variable Y and the auxiliary variable X in the h^{th} stratum. Let $\bar{y}_{st}=\sum_{h=1}^{L} W_h \bar{y}_h$, $\bar{x}_{st}=\sum_{h=1}^{L} W_h \bar{x}_h$ and $\bar{z}_{st}=\sum_{h=1}^{L} W_h \bar{z}_h$ be the stratified sample means,
where $\bar{y}_h=\frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$, $\bar{x}_h=\frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}$ and $\bar{z}_h=\frac{1}{n_h}\sum_{i=1}^{n_h} z_{hi}$ are the stratum sample means corresponding to population stratum means $\bar{Y}_h=E(Y_h)$, $\bar{X}_h=E(X_h)$ and $\bar{Z}_h=E(Z_h)$ and $W_h=N_h/N$ are the known stratum weights".

To estimate $\bar{Y}=\sum_{i=1}^{N_h} W_h \bar{Y}_h$, we estimate that $\bar{X}=\sum_{i=1}^{N_h} W_h \bar{X}_h$ is known. Let $\bar{Z}=\sum_{i=1}^{N_h} W_h \bar{Z}_h$ be the population mean for the scrambled variable Z. The respondent is asked to report a scrambled response for Y given by $Z=Y+S$ but is asked to provide a true response for X.

To discuss the properties of the different estimators,

$e_{0st}=\frac{\bar{z}_{st}-\bar{Z}}{\bar{Z}}$ and $e_{1st}=\frac{\bar{x}_{st}-\bar{X}}{\bar{X}}$, $e_{2st}=\frac{s_{xst}^2-S_{xst}^2}{S_{xst}^2}$ and $e_{3st}=\frac{s_{zxst}^2-S_{zxst}^2}{S_{zxst}^2}$. Such that $E\left(e_{ist}\right)=0 \quad \left(i=0,1,2,3\right)$.

$E\left(e_{0st}^2\right)=\frac{\sum_{h=1}^{L}W_h^2\gamma_h S_{zh}^2}{\bar{Z}^2}$, $E\left(e_{1st}^2\right)=\frac{\sum_{h=1}^{L}W_h^2\gamma_h S_{xh}^2}{\bar{X}^2}$, $E\left(e_{0st}e_{1st}\right)=\frac{\sum_{h=1}^{L}W_h^2\gamma_h S_{zxh}}{\bar{Z}_h\bar{X}_h}$,

$E\left(e_{1st}e_{2st}\right)=\frac{1}{\bar{X}}\sum_{h=1}^{L}W_h^2\gamma_h\frac{\mu_{03h}}{\mu_{02h}}$

And $E\left(e_{1st}e_{3st}\right)=\frac{1}{\bar{X}}\sum_{h=1}^{L}W_h^2\gamma_h\frac{\mu_{12h}}{\mu_{11h}}$, $\gamma_h=\left(\frac{1}{n_h}-\frac{1}{N_h}\right)$ and $\mu_{rsh}=\frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(z_{hi}-\bar{Z}_h\right)^r\left(x_{hi}-\bar{X}_h\right)^s$.

For a stratified random sample the usual combined sample mean of the sensitive variable, ignoring auxiliary information, is given by

$$t_{Yst}=\bar{z}_{st} \tag{2.1}$$

Which is unbiased estimator of population mean $\bar{Y}$.

The MSE of t_{Yst} is given by

$$MSE\left(t_{Yst}\right)=\sum_{h=1}^{L}W_h^2\gamma_h\left(S_{yh}^2+S_{sh}^2\right) \tag{2.2}$$

Where $\gamma_h=\left(\frac{1}{n_h}-\frac{1}{N_h}\right)$, $S_{yh}^2=\frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(y_{hi}-\bar{Y}_h\right)^2$ and $S_{sh}^2=\frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(s_{hi}-\bar{S}_h\right)^2$.

A combined stratified ratio estimator by souse et al. (2014) given as:

$$t_{Rst}=\bar{z}_{st}\left(\frac{\bar{X}}{\bar{x}_{st}}\right) \tag{2.3}$$

$$MSE\left(t_{Rst}\right)\cong\bar{Y}^2\sum_{h=1}^{L}W_h^2\gamma_h\left\{C_{zh}^2+C_{xh}^2-2C_{zxh}\right\} \tag{2.4}$$

And combined stratified regression estimator given as

$$t_{\mathrm{Re}gst}=\bar{z}_{st}+\hat{\beta}_c\left(\bar{X}-\bar{x}_{st}\right) \tag{2.5}$$

Where $\hat{\beta}_c = \dfrac{\sum_{h=1}^{L} W_h^2 \gamma_h s_{zxh}}{\sum_{h=1}^{L} W_h^2 \gamma_h s_{xh}^2}$

And $MSE\left(\hat{\mu}_{\mathrm{Re}gst}\right) \cong \bar{Y}^2 \sum_{h=1}^{L} W_h^2 \gamma_h C_{zh}^2 \left(1-\rho_c^2\right)$ (2.6)

Where $\rho_c = \dfrac{\sum_{h=1}^{L} W_h^2 \gamma_h s_{zxh}}{\sqrt{\sum_{h=1}^{L} W_h^2 \gamma_h C_{zh}^2}\sqrt{\sum_{h=1}^{L} W_h^2 \gamma_h C_{xh}^2}}$.

3. Proposed a combined general family of Estimators

Motivated by Yadav et. al. (2015) and Sousa et al. (2014) suggested the following combined general family of estimators,

$$t_{Sist} = \left[k_1 \bar{z}_{st} + k_2\left(\bar{X} - \bar{x}_{st}\right)\right]\left[\alpha\left(\frac{a_{st}\bar{X} + b_{st}}{a_{st}\bar{x}_{st} + b_{st}}\right) + \left(1-\alpha\right)\exp\left(\frac{a_{st}\left(\bar{X} - \bar{x}_{st}\right)}{a_{st}\left(\bar{X} + \bar{x}_{st}\right) + 2b_{st}}\right)\right] \quad (3.1)$$

Where k_1 and k_2 are weights whose values are to be determined, $\alpha = 0$ or 1, a_{st} and b_{st} are the real numbers or known parameters of the auxiliary variable such as

$\psi_1 = \sum_{h=1}^{L} W_h C_{xh}$ and $\psi_2 = \sum_{h=1}^{L} W_h \beta_{2xh}$ where $\beta_{2xh} = \dfrac{E\left(x_h - \bar{X}_h\right)^4}{\left\{E\left(x_h - \bar{X}_h\right)^2\right\}^2}$.

From t_{Sist} for $\alpha = 0$, we obtain the following estimator

$t_{S0st} = \left[k_1 \bar{z}_{st} + k_2\left(\bar{X} - \bar{x}_{st}\right)\right]\left[\exp\left(\dfrac{\left(\bar{X} - \bar{x}_{st}\right)}{\left(\bar{X} + \bar{x}_{st}\right)}\right)\right]$ where $a_{st} = 1, b_{st} = 0$.

$t_{S1st} = \left[k_1 \bar{z}_{st} + k_2\left(\bar{X} - \bar{x}_{st}\right)\right]\left[\exp\left(\dfrac{\left(\bar{X} - \bar{x}_{st}\right)}{\left(\bar{X} + \bar{x}_{st}\right) + 2\psi_1}\right)\right]$ where $a_{st} = 1, b_{st} = \psi_1$.

$t_{S2st} = \left[k_1 \bar{z}_{st} + k_2\left(\bar{X} - \bar{x}_{st}\right)\right]\left[\exp\left(\dfrac{\left(\bar{X} - \bar{x}_{st}\right)}{\left(\bar{X} + \bar{x}_{st}\right) + 2\psi_2}\right)\right]$ where $a_{st} = 1, b_{st} = \psi_2$.

$t_{S3st} = \left[k_1 \bar{z}_{st} + k_2\left(\bar{X} - \bar{x}_{st}\right)\right]\left[\exp\left(\dfrac{\psi_1\left(\bar{X} - \bar{x}_{st}\right)}{\psi_1\left(\bar{X} + \bar{x}_{st}\right) + 2\psi_2}\right)\right]$ where $a_{st} = \psi_1, b_{st} = \psi_2$.

$t_{S4st} = \left[k_1 \bar{z}_{st} + k_2\left(\bar{X} - \bar{x}_{st}\right)\right]\left[\exp\left(\dfrac{\psi_2\left(\bar{X} - \bar{x}_{st}\right)}{\psi_2\left(\bar{X} + \bar{x}_{st}\right) + 2\psi_1}\right)\right]$ where $a_{st} = \psi_2, b_{st} = \psi_1$.

From t_{Sist} for $\alpha = 1$, we obtain following estimators

$$t_{S5st}=\left[k_1\bar{z}_{st}+k_2\left(\bar{X}-\bar{x}_{st}\right)\right]\left[\left(\frac{\bar{X}}{\bar{x}_{st}}\right)\right] \text{ where } a_{st}=1, b_{st}=0.$$

$$t_{S6st}=\left[k_1\bar{z}_{st}+k_2\left(\bar{X}-\bar{x}_{st}\right)\right]\left[\left(\frac{\bar{X}+\psi_1}{\bar{x}_{st}+\psi_1}\right)\right] \text{ where } a_{st}=1, b_{st}=\psi_1.$$

$$t_{S7st}=\left[k_1\bar{z}_{st}+k_2\left(\bar{X}-\bar{x}_{st}\right)\right]\left[\left(\frac{\bar{X}+\psi_2}{\bar{x}_{st}+\psi_2}\right)\right] \text{ where } a_{st}=1, b_{st}=\psi_2.$$

$$t_{S8st}=\left[k_1\bar{z}_{st}+k_2\left(\bar{X}-\bar{x}_{st}\right)\right]\left[\left(\frac{\psi_1\bar{X}+\psi_2}{\psi_1\bar{x}_{st}+\psi_2}\right)\right] \text{ where } a_{st}=\psi_1, b_{st}=\psi_2.$$

$$t_{S9st}=\left[k_1\bar{z}_{st}+k_2\left(\bar{X}-\bar{x}_{st}\right)\right]\left[\left(\frac{\psi_2\bar{X}+\psi_1}{\psi_2\bar{x}_{st}+\psi_1}\right)\right] \text{ where } a_{st}=\psi_2, b_{st}=\psi_1.$$

Expressing (3.1) in terms of $e's$ and retaining terms in $e's$ up to first order approximation, we have

$$t_{Sist}\cong\left[k_1\bar{Z}\left(1+e_{0st}\right)-k_2\bar{X}e_{1st}\right]\left[\alpha\left(1+ge_{1st}\right)^{-1}+\left(1-\alpha\right)\exp\left\{\frac{-1}{2}ge_{1st}\left(1+\frac{1}{2}e_{1st}\right)^{-1}\right\}\right] \quad (3.2)$$

Where $g=\dfrac{a_{st}\bar{X}}{a_{st}\bar{X}+b_{st}}$

$$t_{Sist}-\bar{Y}\cong\left(k_1-1\right)\bar{Y}+k_1\bar{Y}\left[e_{0st}-\frac{1}{2}g\left(1+\alpha\right)e_{1st}+\frac{1}{8}g^2\left(3+5\alpha\right)e_{1st}^2-\frac{1}{2}g\left(1+\alpha\right)e_{0st}e_{1st}\right]$$

Using (3.3), the $Bias$ and MSE of t_{Sist}, are given by

$$Bias\left(t_{Sist}\right)\cong\left(k_1-1\right)\bar{Y}+k_1\bar{Y}\sum_{h=1}^{L}W_h^2\gamma_h\left(\frac{1}{8}g^2\left(3+5\alpha\right)C_{xh}^2-\frac{1}{2}g\left(1+\alpha\right)C_{zxh}\right)$$
$$-\frac{1}{2}k_2\bar{X}\sum_{h=1}^{L}W_h^2\gamma_h g\left(1+\alpha\right)C_{xh}^2. \quad (3.4)$$

$$-k_2\bar{X}\left[e_{1st}-\frac{1}{2}g\left(1+\alpha\right)e_{1st}^2\right] \quad (3.3)$$

and

$$MSE\left(t_{Sist}\right)\cong\bar{Y}^2\left[\left(k_1-1\right)^2+k_1^2\sum_{h=1}^{L}W_h^2\gamma_h\left\{C_{zh}^2+\frac{1}{4}g^2C_{xh}^2\left(\alpha^2+7\alpha+4\right)-2gC_{zxh}\left(1+\alpha\right)\right\}\right.$$
$$-2k_1\sum_{h=1}^{L}W_h^2\gamma_h\left\{\frac{1}{8}g^2\left(5\alpha+3\right)C_{xh}^2-\frac{1}{2}g\left(1+\alpha\right)C_{zxh}\right\}+k_2^2\frac{\bar{X}^2}{\bar{Y}^2}\sum_{h=1}^{L}W_h^2\gamma_hC_{xh}^2$$
$$\left.-2k_2\frac{\bar{X}}{\bar{Y}}\frac{1}{2}g\sum_{h=1}^{L}W_h^2\gamma_h\left(1+\alpha\right)C_{xh}^2-2k_1k_2\frac{\bar{X}}{\bar{Y}}\sum_{h=1}^{L}W_h^2\gamma_h\left(C_{zxh}-g\left(1+\alpha\right)C_{xh}^2\right)\right] \quad (3.5)$$

And optimum values of k_1 and k_2, respectively, are found as,

$$k_{1(opt)} = \frac{1-\frac{1}{8}g^2\left(4\alpha^2+3\alpha+1\right)\sum_{h=1}^{L}W_h^2\gamma_h C_{xh}^2}{1+\left\{\sum_{h=1}^{L}W_h^2\gamma_h C_{zh}^2\left(1-\rho_{zxh}^2\right)-g^2\frac{1}{4}\left(\alpha+3\alpha^2\right)\sum_{h=1}^{L}W_h^2\gamma_h C_{xh}^2\right\}}$$

and

$$k_{2(opt)} = \frac{\bar{Y}}{\bar{X}}\left\{\frac{1}{2}g\left(1+\alpha\right)+k_{1(opt)}\left(\frac{C_{zxh}}{C_{xh}^2}-g\left(1+\alpha\right)\right)\right\}$$

Substituting these optimum values in (3.5), the minimum MSE of t_{Sist} is given by

$$MSE\left(t_{Sist}\right)_{\min} \cong \bar{Y}^2\left[1-\frac{1}{4}g^2\left(1+\alpha\right)^2\sum_{h=1}^{L}W_h^2\gamma_h C_{xh}^2-\frac{\left\{1-\frac{1}{8}g^2\left(4\alpha^2+3\alpha+1\right)\sum_{h=1}^{L}W_h^2\gamma_h C_{xh}^2\right\}^2}{\left\{1+\left\{\sum_{h=1}^{L}W_h^2\gamma_h C_{zh}^2\left(1-\rho_{zxh}^2\right)-g^2\frac{1}{4}\left(\alpha+3\alpha^2\right)\sum_{h=1}^{L}W_h^2\gamma_h C_{xh}^2\right\}\right\}}\right] \quad (3.6)$$

By using (3.6), for different values of a_{st}, b_{st} and $\alpha = 0$ or $\alpha = 1$, we can get the minimum MSE_s of t_{Sist} $\left(i = 0,1,2,...,9\right)$.

4. Simulation Study

For simulation study, we use "two bivariate normal populations with different covariance matrices to represent the distribution of $\left(Y, X\right)$. The scrambling variable S is taken to be normal distribution with mean equal to zero and standard deviation equal to 10% of the standard deviation of X. The reported scrambled responses on Y is given by $Z = Y + S$. The simulated populations have theoretical mean of $\left[Y, X\right]$ as $\mu = \left[5,5\right]$ and covariance matrices as given below.

Population 1:

$$N = 1000$$

$$\Sigma = \begin{bmatrix} 9 & 3.2 \\ 3.2 & 4 \end{bmatrix}, \quad \rho_{XY} = 0.5333$$

Population 2:

$$N = 1000$$

$$\Sigma = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}, \rho_{XY} = 0.9487$$

For each population we considered four sample sizes: n=30, 60, 150 and 300. The population is divided in two strata according to a certain criteria set for the auxiliary variable. The sample size from each stratum is based on Neyman allocation".

Table 1 and 2 gives the empirical and theoretical MSE's for the various estimators based on 1st order approximation. We estimate the empirical MSE using 5000 samples of size n and considering the average of all the observed values. We use the following expression to find the percent relative efficiency (PRE) of study estimators as compared to the ordinary sample mean:

$$PRE = \frac{MSE\left(t_{Yst}\right)}{MSE\left(t_{\beta}\right)} \times 100$$

where $\beta = R_{st}, \mathrm{Re}\,g_{st}, S0_{st}, S1_{st}, S2_{st}, S3_{st}, S4_{st}, S5_{st}, S6_{st}, S7_{st}, S8_{st}, S9_{st}$.

Table 2 and 3 below gives the empirical and theoretical MSE's and PRE for the competing estimators. The results of the MSE's and PRE show that the proposed a combined general family of estimators performs better than the existing estimators. The use of auxiliary information provides a gain for a stratified random sample. And the proposed estimators get more efficient as ρ_{XY} increases.

5. Numerical Example

In this study, we use the data set earlier considered by Sousa et al. (2014). In this data, "the variable of interest Y is the purchase orders in 2010 and the auxiliary variable X is the enterprises of turnover. So we consider three strata: the first is enterprises with less than 10 million of turnover, the second between 10 and less than 30 million of turnover, and third with 30 million or more of turnover.

The scrambling variable S is taken to be normal distribution with mean equal to zero and standard deviation equal to 10% of the standard deviation of X. The reported scrambled responses on Y is given by $Z = Y + S$ (the purchase order value plus a random quantity).

Sampling information: $N = 1698$, $\rho_{XY} = 0.9368$, $\beta_{XY} = 0.8284$, $\mu_Y = 14.44$, $\mu_X = 17.97$, $\sigma_Y = 22.39$, $\sigma_X = 25.31$. The variable X and Y are expressed in millions of Euros. We test our stratified sample estimators with random sample of sizes $n = 100, 250$ and 500. The sample size of each stratum is allocated proportionally to the dimension of strata population".

Stratum	N	ρ_{XY}	μ_Y	σ_Y	μ_X	σ_X
1	979	0.7802	2.15	2.46	3.12	2.68
2	362	0.7952	16.67	6.86	20.31	6.02
3	357	0.8408	45.88	30.21	56.33	30.18

Table 3 presents the empirical and theoretical results of MSE estimates and PRE of the various estimators in the stratified sample. We estimate the empirical MSE using 5000 samples of size n selected from the population.

According to the MSE and PRE results in table 3, the proposed a combined general family of estimators is considerably better than the existing estimators i.e., usual stratified random sample mean estimator, Sousa et al (2014) ratio and regression estimator of the sensitive variable in stratified sampling.

6. Conclusion

We can conclude from this study, the mean estimation of a sensitive variable by using non-sensitive variable can be improved in stratified random sampling based on randomized response technique. Our simulation study and numerical example show that the proposed a combined family of estimators are more efficient than the existing estimators i.e., usual stratified random sample mean estimator, Sousa et al (2014) ratio and regression estimator of the sensitive variable in stratified sampling. Also there is no additional loss of privacy as compared to what it is for an additively scrambled RRT model.

References

1. Cochran, W.G. Sampling Techniques, 3rd ed., Wiley Eastern Ltd., New York, 1977.
2. Gupta, S. N., Gupta, B. C. and Singh, S. Estimation of sensitivity level of personal interview survey questions, *Journal of Statistical Planning and Inference*, Vol. 100, 2002, pp. 239–247.
3. Gupta, S and Shabbir, J. On improvement in estimating the population mean in simple random sampling, *Journal of Applied Statistics*, Vol. 35(5), 2008, pp. 559–566.
4. Gupta, S., Shabbir, J. and Sehra, S. Mean and sensitivity estimation in optional randomized response models, *Journal of Statistical Planning and Inference*, Vol. 140(10), 2010, pp. 2870–2874.
5. Gupta, S., Shabbir, J., Sousa, R., and Real, P. C. Estimation of the mean of a sensitive variable in the presence of auxiliary information, *Communications in Statistics-Theory and Methods*, Vol. 41, 2012, pp.2394–2404.
6. Kadilar, C. and Cingi, H. Ratio estimator in stratified sampling, *Biometrical Journal*, Vol. 45, 2003, pp. 218–225.
7. Koyuncu, N. and Kadılar, C. On the family of estimators of population mean in stratified random sampling, *Pakistan Journal of Statistics*, Vol. 26(2), 2010, pp. 427–443.
8. Nangsue, N. Adjusted Ratio and Regression Type Estimators for Estimation of Population Mean when some Observations are missing, World Academy of Science, *Engineering and Technology*, Vol. 53, 2009, pp. 787–790.
9. Shabbir, J. and Gupta, S. Improved ratio estimators in stratified sampling, *American Journal of Mathematical and Management Sciences*, Vol. 25, 2005, pp. 293–311.

10. Sousa, R., Shabbir, J., Real, P.C., Gupta, S. Ratio Estimation of the Mean of a Sensitive Variable in the Presence of Auxiliary Information, *Journal of Statistical Theory and Practice*, Vol. 4(3), 2010, pp. 495–507.

11. Sousa, R. Gupta, S. Shabbir, J. and Real, P.C. Improved Mean Estimation of a Sensitive Variable Using Auxiliary Information in Stratified Sampling. *Journal of Statistics & Management System*, Vol. 17, 2014, No. 5 &6, pp. 503-518.

12. Warner, S. L. Randomized Response: A survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association*, Vol. 76, 1965, pp. 916–923.

13. Yadav, S.K. Kadilar, C. Shabbir, J. Gupta, S. Improved family of estimators of population variance in simple random sampling, *Journal of Statistical Theory and Practice*, 9(2015) 219-226.

Table 1: Empirical and Theoretical MSE, PRE for the estimators relative to RRT mean estimator in stratified random sampling for Population 1.

N	N_h	ρ_{XYh}	n	$Estimation$	MSE $Estimation$		PRE
					$Empirical$	$Theoretical$	
1000	$N_1 = 550$	$\rho_{XY1} = 0.5397$	60	$t_{Y_{st}}$	0.1414	0.1428	100
	$N_2 = 450$	$\rho_{XY2} = 0.5410$		t_{Rst}	0.1049	0.1057	135.09
				$t_{\mathrm{Re}gst}$	0.1039	0.1044	136.78
				t_{S0st}	0.1026 `	0.1030	138.64
				t_{S1st}	0.1022	0.1019	140.13
				t_{S2st}	0.1007	0.1008	141.67
				t_{S3st}	0.1004	0.1004	142.23
				t_{S4st}	0.1012	0.1010	141.38
				t_{S5st}	0.0993	0.0998	143.08
				t_{S6st}	0.0995	0.1010	141.38
				t_{S7st}	0.0923	0.1008	141.66
				t_{S8st}	0.1020	0.1017	140.41
				t_{S9st}	0.1029	0.1025	139.31
			150	$t_{Y_{st}}$	0.0501	0.0497	100
				t_{Rst}	0.0383	0.0381	130.44
				$t_{\mathrm{Re}gst}$	0.0408	0.0377	131.83
				t_{S0st}	0.0335	0.0329	151.06
				t_{S1st}	0.0341	0.0330	150.61
				t_{S2st}	0.0334	0.0338	147.04
				t_{S3st}	0.0340	0.0339	146.61
				t_{S4st}	0.0310	0.0315	157.78
				t_{S5st}	0.0340	0.0336	147.92
				t_{S6st}	0.0343	0.0337	147.47
				t_{S7st}	0.0365	0.0345	144.06
				t_{S8st}	0.0359	0.0341	145.75
				t_{S9st}	0.0366	0.0351	141.59
			300	$t_{Y_{st}}$	0.0202	0.0207	100

Table 1: Continued

N	N_h	ρ_{XYh}	n	$Estimation$	MSE $Estimation$ $Empirical$	$Theoretical$	PRE
				t_{Rst}	0.0167	0.0165	125.45
				$t_{\mathrm{Re}\,gst}$	0.0159	0.0157	131.84
				t_{S0st}	0.0154	0.0153	135.29
				t_{S1st}	0.0150	0.0146	141.78
				t_{S2st}	0.0152	0.0151	137.08
				t_{S3st}	0.0153	0.0148	139.87
				t_{S4st}	0.0150	0.0149	138.92
				t_{S5st}	0.0144	0.0150	138.00
				t_{S6st}	0.0139	0.0147	140.82
				t_{S7st}	0.0139	0.0141	146.81
				t_{S8st}	0.0139	0.0145	142.76
				t_{S9st}	0.0142	0.0146	141.78

Table 2: Empirical and Theoretical MSE, PRE for the estimators relative to RRT mean estimator in stratified random sampling for Population 2.

N	N_h	ρ_{XYh}	n	$Estimation$	MSE $Estimation$ Empirical	Theoretical	PRE
1000	$N_1 = 550$	$\rho_{XY1} = 0.9522$	60	$t_{Y_{st}}$	0.0778	0.0776	100
	$N_2 = 450$	$\rho_{XY2} = 0.9478$		t_{Rst}	0.0163	0.0156	497.44
				$t_{\mathrm{Re}gst}$	0.0098	0.0101	768.32
				t_{S0st}	0.0081	0.0077	1007.79
				t_{S1st}	0.0081	0.0076	1021.05
				t_{S2st}	0.0079	0.0070	1108.57
				t_{S3st}	0.0076	0.0073	1063.01
				t_{S4st}	0.0080	0.0075	1034.66
				t_{S5st}	0.0084	0.0081	958.02
				t_{S6st}	0.0082	0.0079	982.27
				t_{S7st}	0.0088	0.0083	934.94
				t_{S8st}	0.0081	0.0080	970.00
				t_{S9st}	0.0081	0.0078	994.87
			150	$t_{Y_{st}}$	0.0275	0.0308	100
				t_{Rst}	0.0056	0.0057	540.35
				$t_{\mathrm{Re}gst}$	0.0049	0.0049	628.57
				t_{S0st}	0.0031	0.0031	993.55
				t_{S1st}	0.0031	0.0027	1141.74
				t_{S2st}	0.0031	0.0029	1062.07
				t_{S3st}	0.0029	0.0028	1100.00
				t_{S4st}	0.0030	0.0027	1141.74
				t_{S5st}	0.0031	0.0028	1100.0
				t_{S6st}	0.0029	0.0027	1141.74
				t_{S7st}	0.0028	0.0028	1100.00
				t_{S8st}	0.0031	0.0027	1141.74
				t_{S9st}	0.0031	0.0029	1062.07
			300	$t_{Y_{st}}$	0.0110	0.0112	100
				t_{Rst}	0.0021	0.0026	430.77
				$t_{\mathrm{Re}gst}$	0.0015	0.0015	746.67
				t_{S0st}	0.0013	0.0013	861.54
				t_{S1st}	0.0011	0.0012	933.33

Table 2: Continued

N	N_h	ρ_{XYh}	n	$Estimation$	MSE $Estimation$ $Empirical$	$Theoretical$	PRE
				t_{S2st}	0.0010	0.0010	1120.00
				t_{S3st}	0.0009	0.0009	1244.44
				t_{S4st}	0.0009	0.0010	1120.00
				t_{S5st}	0.0010	0.0010	1120.0
				t_{S6st}	0.0011	0.0010	1120.00
				t_{S7st}	0.0012	0.0010	1120.00
				t_{S8st}	0.0012	0.0011	1018.18
				t_{S9st}	0.0012	0.0010	1120.00

Table 3: Empirical and Theoretical MSE, PRE for the estimators relative to RRT mean estimator in stratified random sampling for Real data:

N	N_h	ρ_{XYh}	n	*Estimation*	*MSE Estimation* *Empirical*	*Theoretical*	*PRE*
1698	$N_1 = 979$	$\rho_{XY1} = 0.7802$	100	$t_{Y_{st}}$	2.0071	1.9876	100
	$N_2 = 362$	$\rho_{XY2} = 0.7952$		t_{Rst}	0.7202	0.7057	281.65
	$N_3 = 357$	$\rho_{XY3} = 0.8408$		$t_{\mathrm{Re}gst}$	0.6653	0.6883	288.77
				t_{S0st}	0.6215	0.6209	320.12
				t_{S1st}	0.5603	0.5532	359.29
				t_{S2st}	0.6054	0.6096	326.04
				t_{S3st}	0.6158	0.6155	322.92
				t_{S4st}	0.6314	0.6299	315.54
				t_{S5st}	0.6401	0.6379	311.58
				t_{S6st}	0.6334	0.6310	314.99
				t_{S7st}	0.6296	0.6300	315.49
				t_{S8st}	0.6354	0.6325	314.25
				t_{S9st}	0.6310	0.6307	315.14
			250	$t_{Y_{st}}$	0.7290	0.7152	100
				t_{Rst}	0.2700	0.2709	264.00
				$t_{\mathrm{Re}gst}$	0.2678	0.2626	272.35
				t_{S0st}	0.2518	0.2511	284.83
				t_{S1st}	0.2226	0.2228	321.00
				t_{S2st}	0.2126	0.2156	331.72
				t_{S3st}	0.2026	0.2018	354.41
				t_{S4st}	0.2038	0.2028	352.67
				t_{S5st}	0.2408	0.2410	296.76
				t_{S6st}	0.2137	0.2132	335.46
				t_{S7st}	0.2069	0.2070	345.51
				t_{S8st}	0.2167	0.2178	328.37
				t_{S9st}	0.2208	0.2220	322.16
			500	$t_{Y_{st}}$	0.2844	0.2859	100
				t_{Rst}	0.1040	0.1034	276.49
				$t_{\mathrm{Re}gst}$	0.0985	0.0977	292.63
				t_{S0st}	0.0802	0.0805	355.15
				t_{S1st}	0.0786	0.0770	371.29

Table 3: Continued

N	N_h	ρ_{XYh}	n	*Estimation*	*MSE Estimation* *Empirical*	*Theoretical*	*PRE*
				t_{S2st}	0.0724	0.0744	384.27
				t_{S3st}	0.0708	0.0710	402.67
				t_{S4st}	0.0716	0.0720	397.08
				t_{S5st}	0.0738	0.0739	386.87
				t_{S6st}	0.0740	0.0741	385.83
				t_{S7st}	0.0766	0.0764	374.21
				t_{S8st}	0.0772	0.0770	371.29
				t_{S9st}	0.0798	0.0795	359.62