Pak. J. Statist.
2016 Vol. 32(5), 393-404

ESTIMATION OF A POPULATION MEAN OF A SENSITIVE VARIABLE IN STRATIFIED TWO-PHASE SAMPLING

Nadia Mushtaq[1], **Muhammad Noor-ul-Amin**[2] and **Muhammad Hanif**[1]

[1] National College of Business Administration and Economics Lahore. Pakistan. Email: dia_qau@yahoo.com drmianhanif@gmail.com

[2] COMSATS Institute of Information and Technology Lahore, Pakistan. Email: nooramin.stats@gmail.com

ABSTRACT

In this study, we present the estimation of population mean of a sensitive variable in stratified sampling using two-phase sampling based on randomized response technique. We introduce a ratio, a regression and general class of estimators for the mean of sensitive variable using non-sensitive auxiliary variable based on randomized response technique in stratified two-phase sampling. Under stratified two phase sampling, the expression of bias and mean square error (MSE) up to the first-order approximations are derived. Simulation studies and real data are presented to demonstrate the performance of proposed estimators.

1. INTRODUCTION

In social science research, survey respondents hesitate to answer sensitive questions. This explains why traditional self-report surveys often suffer from high levels of non-response and dishonest answers. To overcome these problems, Warner (1965) introduced Randomized Response Technique (RRT), which helps interviewers extract reliable data corresponding to sensitive questions while maintaining respondent anonymity. RRT models allow respondents to mask their actual response by giving a scrambled response which the researcher is later able to unscramble at an aggregate level but not at an individual level.

The stratified random sampling is used to estimate the parameters if the population consist of heterogeneous units. Many authors have presented the ratio and regression estimators such as Kadilar and Cingi (2003), Kadilar and Cingi (2005), Kadilar et al. (2007), Shabbir and Gupta (2007, 2010), Koyuncu and Kadilar (2010), Sousa et al. (2010), Gupta et al. (2012) and Sousa et al. (2014).

Two-phase sampling is a procedure in which we obtain the information about auxiliary variable(s) from a larger sample at first phase and relatively small sample from the second phase. Many authors suggested some improved ratio, regression and product type estimators included as Sukhatme (1962), Singh and Vishwakarma (2007), Sahoo et al. (2010), Noor-ul-Amin and Hanif (2012) Sanaullah et al. (2014), Rashid et al. (2015) etc.

In this study we presents ratio, regression estimator of population mean of a sensitive variable using non-sensitive auxiliary variable using RRT methodology in stratified two phase sampling. We also propose a general class of estimators for estimating population mean of a sensitive variable using non-sensitive auxiliary variable using RRT methodology in stratified two-phase sampling.

2. TERMINOLOGY

Consider a finite population of size N which is stratified into L homogenous strata. Let N_h be the size of h^{th} stratum $(h=1,...,L)$ such that $\sum_{h=1}^{L} N_h = N$. Let Y be the study variable, a sensitive variable which cannot be observed directly due to respondent bias and X be non-sensitive auxiliary variable which is correlated with Y. Let S be scrambling variable independent of Y and X. So the reported response is given as $Z=Y+S$.

And (y_{hi}, x_{hi}) be the observations of the study variable (y) and the auxiliary variable (x) on the i^{th} unit of h^{th} stratum respectively. When information on $\bar{X}_h$ is unknown, a first large sample of size n'_h is selected from each h^{th} stratum to estimate $\bar{X}_h$.

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, where $W_h = \dfrac{N_h}{N}$.

Let $\bar{Z} = \sum_{i=1}^{N_h} W_h \bar{Z}_h$ be the population mean for the reported variable Z. To obtain the bias and mean square error (MSE) under stratified two-phase sampling, let us define

$$\bar{Z} = \sum_{i=1}^{N_h} W_h \bar{Z}_h \,,\ \bar{z}_h = \bar{Z}_h\left(1+e_{0h}\right),\ e_0 = \frac{\sum_{h=1}^{L} W_h \bar{Z}_h e_{0h}}{\bar{Z}}$$

$$\bar{X} = \sum_{i=1}^{N_h} W_h \bar{X}_h \,,\ \bar{x}'_h = \bar{X}_h\left(1+e'_{1h}\right),\ e'_1 = \frac{\sum_{h=1}^{L} W_h \bar{X}'_h e'_{1h}}{\bar{X}},$$

$$\bar{x}_h = \bar{X}_h\left(1+e_{1h}\right),\ e_1 = \frac{\sum_{h=1}^{L} W_h \bar{X}_h e_{1h}}{\bar{X}}.$$

The expectations are defined given as:

$$E\left(e_0\right) = E\left(e_1\right) = E\left(e'_1\right) = 0,\ V_{r,s} = \sum_{h=1}^{L} W_h^{r+s} \frac{E\left(\bar{z}_h - \bar{Z}_h\right)^r E\left(\bar{x}_h - \bar{X}_h\right)^s}{\bar{Z}^r \bar{X}^s}$$

$$E\left(e_0\right)^2 = \frac{1}{\bar{Z}^2}\sum_{h=1}^{L} W_h^2 \lambda_h S_{zh}^2 = V_{20}\ ,\ E\left(e_1\right)^2 = \frac{1}{\bar{X}^2}\sum_{h=1}^{L} W_h^2 \lambda_h S_{xh}^2 = V_{02}$$

$$E\left(e'_1\right)^2 = \frac{1}{\bar{X}^2}\sum_{h=1}^{L} W_h^2 \lambda'_h S_{xh}^2 = V'_{02},\ E\left(e_0 e_1\right) = \frac{1}{\bar{Z}\bar{X}}\sum_{h=1}^{L} W_h^2 \lambda_h S_{zxh} = V_{11}$$

$$E\left(e_0 e_1'\right)=\frac{1}{\bar{Z}\bar{X}}\sum_{h=1}^{L}W_h^2\lambda_h' S_{zxh}=V_{11}',\ \vartheta_{02}=V_{02}-V_{02}',\ \vartheta_{11}=V_{11}-V_{11}'.$$

$$\lambda_h=\left(\frac{1}{n_h}-\frac{1}{N_h}\right),\ \lambda_h'=\left(\frac{1}{n_h'}-\frac{1}{N_h}\right),\ \Lambda_h=\lambda_h-\lambda_h',\ \rho_{zxh}=\frac{\rho_{yxh}}{\sqrt{1+\frac{s_{sh}^2}{s_{yh}^2}}},$$

$$S_{zxh}=\rho_{zxh}S_{zh}S_{xh}.$$

The procedure of stratified two-phase sampling is as follows:

i. Select a sample of size n_h' from the h^{th} stratum using simple random sampling without replacement $(SRSWOR)$ such that $\sum_{h=1}^{L}n_h'=n'$ and observe auxiliary characteristic for these units. This is called s stratified first- phase sample.

ii. Select another stratified random sample of size n_h from each $n_h'\ (n_h<n_h')$ using $SRSWOR$ such that $\sum_{h=1}^{L}n_h=n$ and collect information on sensitive variable of the interest. This is called a second-phase sample.

Under stratified two-phase sampling, usual unbiased estimator for population mean of the sensitive variable is given by

$$t_{Ystd}=\sum_{h=1}^{L}W_h\bar{z}_h \tag{2.1}$$

which is unbiased estimator of population mean $\bar{Y}$ and using $\bar{Z}=\bar{Y}$.

The Var of t_{Ystd} is given by

$$Var\left(t_{Ystd}\right)=\bar{Y}^2\left(\frac{\sum_{h=1}^{L}W_h^2\lambda_h\left(S_{yh}^2+S_{sh}^2\right)}{\bar{Y}^2}\right)=\bar{Y}^2\,\frac{\sum_{h=1}^{L}W_h^2\lambda_h S_{zh}^2}{\bar{Y}^2}=\bar{Y}^2V_{20} \tag{2.2}$$

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

We Consider ratio and regression estimators under stratified two-phase sampling as,

$$t_{rstd}=\bar{z}_{st}\left(\frac{\bar{x}_{st}'}{\bar{x}_{st}}\right)=\sum_{h=1}^{L}W_h\bar{z}_h\left(\frac{\sum_{h=1}^{L}W_h\bar{x}_h'}{\sum_{h=1}^{L}W_h\bar{x}_h}\right)$$

and

$$t_{regstd}=\bar{z}_{st}+\beta\left(\bar{x}_{st}'-\bar{x}_{st}\right)=\sum_{h=1}^{L}W_h\bar{z}_h+\left(\sum_{h=1}^{L}W_h\bar{x}_h'-\sum_{h=1}^{L}W_h\bar{x}_h\right) \tag{2.3}$$

Using $\bar{Z}=\bar{Y}$ and the mean square errors of the above estimators, up to first order approximation are

$$MSE\left(t_{rstd}\right)\cong \bar{Y}^2\left\{V_{20}+\vartheta_{02}-2\vartheta_{11}\right\}$$

$$MSE\left(t_{regstd}\right)=\bar{Y}^2 V_{20}\left\{\lambda_h\left(1-\rho_{zx}^2\right)+\lambda'_h\rho_{zx}^2\right\}.$$

3.1 Proposed Estimators

Sousa et al. (2014) have introduced ratio and regression estimators for the mean of sensitive variables based on a Randomized Response Technique (RRT) in stratified sampling. Motivated by this, we propose a general family of estimators for the mean of sensitive variable based on a Randomized Response Technique (RRT) in stratified two-phase sampling, is given by the following expression

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

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

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

Some special cases of the general family of estimators t_{Sistd} given in (3.1)

Generalized Exponential Type Estimators $\alpha=0$	**Generalized Ratio Type Estimators** $\alpha=1$	b_{st}
$t_{S0std}=\left[k_1\bar{z}_{st}+k_2\left(\bar{x}'_{st}-\bar{x}_{st}\right)\right]\left[\exp\left(\frac{\left(\bar{x}'_{st}-\bar{x}_{st}\right)}{\left(\bar{x}'_{st}+\bar{x}_{st}\right)}\right)\right]$	$t_{S3std}=\left[k_1\bar{z}_{st}+k_2\left(\bar{x}'_{st}-\bar{x}_{st}\right)\right]\left[\left(\frac{\bar{x}'_{st}}{\bar{x}_{st}}\right)\right]$	0
$t_{S1std}=\left[k_1\bar{z}_{st}+k_2\left(\bar{x}'_{st}-\bar{x}_{st}\right)\right]\left[\exp\left(\frac{\left(\bar{x}'_{st}-\bar{x}_{st}\right)}{\left(\bar{x}'_{st}+\bar{x}_{st}\right)+2\psi_1}\right)\right]$	$t_{S4std}=\left[k_1\bar{z}_{st}+k_2\left(\bar{x}'_{st}-\bar{x}_{st}\right)\right]\left[\left(\frac{\bar{x}'_{st}+\psi_1}{\bar{x}_{st}+\psi_1}\right)\right]$	ψ_1
$t_{S2std}=\left[k_1\bar{z}_{st}+k_2\left(\bar{x}'_{st}-\bar{x}_{st}\right)\right]\left[\exp\left(\frac{\left(\bar{x}'_{st}-\bar{x}_{st}\right)}{\left(\bar{x}'_{st}+\bar{x}_{st}\right)+2\psi_2}\right)\right]$	$t_{S5std}=\left[k_1\bar{z}_{st}+k_2\left(\bar{x}'_{st}-\bar{x}_{st}\right)\right]\left[\left(\frac{\bar{x}'_{st}+\psi_2}{\bar{x}_{st}+\psi_2}\right)\right]$	ψ_2

3.2 The Bias and Mean Square Error of the General Family of Estimators

Using notations from section 2, the general family of estimators given in (3.1) may be expressed as given below:

$$t_{Sistd} \cong \left[k_1 \bar{Z}\left(1+e_0\right)+k_2 \bar{X}\left(e_1'-e_1\right)\right]$$
$$\left[\alpha\left(1+ge_1'\right)\left(1+ge_1\right)^{-1}+\left(1-\alpha\right)\exp\left\{\frac{1}{2}g\left(e_1'-e_1\right)\left(1+\frac{1}{2}g\left(e_1'+e_1\right)\right)^{-1}\right\}\right]$$

where

$$g=\frac{\bar{X}}{\bar{X}+b} \tag{3.2}$$

$$t_{Sistd}-\bar{Y} \cong \left(k_1-1\right)\bar{Y}+k_1\bar{Y}\left[e_0-\frac{1}{2}g\left(1+\alpha\right)\left(e_1-e_1'\right)+\frac{1}{8}g^2\left(3+5\alpha\right)\left(e_1^2-e_1'^2\right)\right.$$
$$\left.-\frac{1}{2}g\left(1+\gamma\right)\left(e_0e_1-e_0e_1'\right)\right]$$
$$-k_2\bar{X}\left[\left(e_1-e_1'\right)-\frac{1}{2}g\left(1+\alpha\right)\left(e_1^2-e_1e_1'\right)\right]. \tag{3.3}$$

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

$$Bias\left(t_{sistd}\right) \cong \left(k_1-1\right)\bar{Y}+k_1\bar{Y}\left(\Lambda_h\frac{1}{8}g^2\left(3+5\alpha\right)\vartheta_{02}-\frac{1}{2}g\left(1+\alpha\right)\Lambda_h\vartheta_{11}\right)$$
$$+k_2\bar{X}\left[\frac{1}{2}g\left(1+\alpha\right)\Lambda_h\vartheta_{02}\right] \tag{3.4}$$

$$MSE\left(t_{Sistd}\right) \cong \bar{Y}^2\left[\left(k_1-1\right)^2\right.$$
$$+k_1^2\left\{\lambda_h V_{20}+\Lambda_h\left(\frac{1}{4}g^2\vartheta_{02}\left(\alpha^2+7\alpha+4\right)-2g\vartheta_{11}\left(1+\alpha\right)\right)\right\}$$
$$-2k_1\Lambda_h\left\{\frac{1}{8}g^2\left(5\alpha+3\right)\vartheta_{02}-\frac{1}{2}g\left(1+\alpha\right)\vartheta_{11}\right\}+k_2^2\frac{\bar{X}^2}{\bar{Y}^2}\Lambda_h\vartheta_{02}$$
$$\left.-2k_2\frac{\bar{X}}{\bar{Y}}\frac{1}{2}g\Lambda_h\left(1+\alpha\right)\vartheta_{02}-2k_1k_2\frac{\bar{X}}{\bar{Y}}\Lambda_h\left(\vartheta_{11}-g\left(1+\alpha\right)\vartheta_{02}\right)\right] \tag{3.5}$$

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

$$k_{1(opt)} = \frac{1-\frac{1}{8}\Lambda_h g^2\left(4\alpha^2+3\alpha+1\right)\vartheta_{02}}{1+\left\{\lambda_h V_{20}\left(1-\rho_{zxh}^2\right)-g^2\frac{1}{4}\left(\alpha+3\alpha^2\right)\Lambda_h\vartheta_{02}\right\}}$$

$$k_{2(opt)} = \frac{\bar{Y}}{\bar{X}}\left\{\frac{1}{2}g\left(1+\alpha\right)+k_{1(opt)}\left(\frac{V_{11}}{V_{02}}-g\left(1+\alpha\right)\right)\right\}$$

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

$$MSE\left(t_{Sistd}\right)_{\min} \cong \bar{Y}^2\left[1-\frac{1}{4}g^2\left(1+\alpha\right)^2\Lambda_h\vartheta_{02} - \frac{\left\{1-\frac{1}{8}g^2\left(4\alpha^2+3\alpha+1\right)\Lambda_h\vartheta_{02}\right\}^2}{\left\{1+\left\{\lambda_h V_{20}-\Lambda_h V_{20}\rho_{zxh}^2-g^2\frac{1}{4}\left(\alpha+3\alpha^2\right)\Lambda_h\vartheta_{02}\right\}\right\}}\right] \tag{3.6}$$

By using (3.6), for different values of b_{st} and $\alpha=0$ or $\alpha=1$, we can get the minimum MSE_s of t_{Sistd} $\left(i=0,1,2,3,4,5\right)$.

4. SIMULATION STUDY

We use the simulation studies for efficiency comparison by empirically and theoretically. Two populations for simulation studies of size 1000 each from bivariate normal populations for (Y, X), with different covariance matrices are used. The Scrambling variable $S \sim N\left(0, 0.1\sigma_x\right)$ and $Z=Y+S$ is the reported response.

Mean of $\left[Y, X\right]$ given as $\mu=\left[5,5\right]$

Population 1: $\Sigma=\begin{bmatrix}9 & 3.2\\ 3.2 & 4\end{bmatrix}$ $\rho_{XY1}=0.5333; \rho_{XY2}=0.5672$

Population 2: $\Sigma=\begin{bmatrix}5 & 3\\ 3 & 2\end{bmatrix}, \rho_{XY1}=0.9522; \rho_{XY2}=0.9487$

For each population we considered three sample sizes for first phase: n' = 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 and for second phase given as: n = 25, 55, 115 respectively. Table 1 and 2 gives the empirical and theoretical MSE's for the various estimators based on 1st order approximation. The empirical MSE's are computed by using t_{β} and theoretical MSE's are computed by using $MSE\left(t_{\beta}\right)$.

We estimate the empirical MSE using 5000 samples of size n' for first phase and n for second phase. We use the following expression to find the percent relative efficiency (PRE) of study estimators as compared to the ordinary sample mean:

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

where $\beta = R_{std}, \text{Re}\,g_{std}, S0_{std}, S1_{std}, S2_{std}, S3_{std}, S4_{std}, S5_{std}$.

5. NUMERICAL EXAMPLE

For this analysis, we consider the real population used by Sousa et al. (2014). The data come from a sample from the survey on Information and Communication Technologies (ICT) usage in enterprises in 2010 with seat in Portugal (Smilhily and Storm, 2010). Let Y be the purchase orders in 2010, X is the enterprises of turnover. And $S \sim N\left(0, 0.1\sigma_x\right)$ so 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 for first phase n' =100, 250, 500 and for second phase n = 45, 100, 205. The proportional allocation has been used for allocating sample size of each stratum.

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 with random sample of sizes for first phase n' =100, 250, 500 and for second phase n = 45, 100, 205.

According to the MSE and PRE results in Table 3, the proposed a general family of estimators for estimating sensitive mean estimator based on randomized response technique in stratified two phase sampling is considerably better than the existing estimators i.e., usual estimator, ratio estimator and regression estimator in stratified two phase sampling.

6. CONCLUSION

In this study, we consider a ratio, a regression and propose a general class of estimators for mean of sensitive variable based on randomized response technique in stratified two-phase sampling. The expression for bias and MSE are derived. From Tables 1-3, it is observed that the theoretical and empirical MSE and PRE of the family of estimators are performed better than the usual estimator, ratio and regression estimator in stratified two phase sampling based on randomized response technique. These results are computed with a simulation studies and using a real data set.

REFERENCES

1. Gupta, S., Shabbir, J. and Sehra, S. (2010). Mean and sensitivity estimation in optional randomized response models. *Journal of Statistical Planning and Inference*, 140(10), 2870-2874.
2. Kadilar, C. and Cingi, H. (2005). A new estimator in stratified random sampling. *Communication in Statistics-Theory and Methods*, 34, 597-602.
3. Kadilar, C., Candan, M. and Cingi, H. (2007). Ratio estimators using robust regression. *Hacettepe Journal of Mathematics and Statistics*, 36(2), 181-188.
4. Kadilar, C. and Cingi, H. (2003). Ratio estimator in stratified sampling. *Biometrical Journal*, 45, 218-225.
5. Koyuncu, N. and Kadilar, C. (2010). On the family of estimators of population mean in stratified random sampling. *Pakistan Journal of Statistics*, 26(2), 427-443.
6. Noor-ul-Amin, M. and Hanif, M. (2012). Some exponential estimators in survey sampling. *Pakistan Journal of Statistics*, 28(3), 367-374.
7. Rashid, R., Noor-ul-Amin, M. and Hanif, M. (2015). Exponential estimators for population mean using the transformed auxiliary variables. *Applied Mathematics Information and* Science, 9(4), 2107-2112.
8. Sahoo, L., Mishra, G., and Nayak, S. (2010). On two different classes of estimators in two-phase sampling using multi-auxiliary variables. *Model Assisted Statistics and Applications*, 5(1), 61-68.
9. Sanaullah, A., Ali, H.A., Noor-ul-Amin, M. and Hanif, M. (2014). Generalized Exponential chain ratio estimators under stratified two-phase random sampling. *Applied Mathematics and Computation*, 226, 541-547.
10. Shabbir, J. and Gupta, S. (2007). On improvement in variance estimation using auxiliary information. *Commun. in Statist.-Theory and Methods*, 36(12), 2177-2185.
11. Shabbir, J. and Gupta, S. (2010). Estimation of the finite population mean in two-phase sampling when auxiliary variables are attribute. *Hacettepe Journal of Mathematics and Statistics*, 39(1), 121-129.
12. 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.
13. Smilhily, M. and Storm, H. (2010). ICT usage in enterprises - 2009. *Eurostat Publications*, Issue 1.

14. Sousa, R., Shabbir, J., Real, P.C. and Gupta, S. (2010). Ratio Estimation of the Mean of a Sensitive Variable in the Presence of Auxiliary Information. *Journal of Statistical Theory and Practice*, 4(3), 495-507.
15. Sousa, R., Gupta, S. Shabbir, J. and Real, P.C. (2014). Improved Mean Estimation of a Sensitive Variable Using Auxiliary Information in Stratified Sampling. *Journal of Statistics & Management System*, 17(5&6), 503-518.
16. Sukhatme, R.V. and Sukhatme, B.V. (1970). *Sampling Theory of Surveys with Applications*, Iowa State University Press.
17. Warner, S.L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60, 63-69.
18. Yadav, S.K., Kadilar, C., Shabbir, J. and Gupta, S. (2015). Improved family of estimators of population variance in simple random sampling. *Journal of Statistical Theory and Practice*, 9(2), 9219-226.

Table 1
Empirical and Theoretical MSE, PRE for the Estimators Relative to RRT Mean Estimator in Stratified Two Phase Random Sampling for Population 1

N	N_h	ρ_{XYh}	n'	n	MSE Estimation			
					Estimation	Empirical	Theoretical	PRE
1000	$N_1 = 550$	$\rho_{XY1} = 0.5397$	60	25	t_{Ystd}	0.3825	0.3798	100
	$N_2 = 450$	$\rho_{XY2} = 0.5410$			t_{rstd}	0.3269	0.3277	115.89
					t_{regstd}	0.3172	0.3197	118.79
					t_{S0std}	0.3015	0.3019	125.80
					t_{S1std}	0.2933	0.2906	130.69
					t_{S2std}	0.2948	0.2967	128.00
					t_{S3std}	0.2752	0.2733	138.96
					t_{S4std}	0.2844	0.2828	134.29
					t_{S5std}	0.2800	0.2877	132.01
			150	55	t_{Ystd}	0.1661	0.1656	100
					t_{rstd}	0.1390	0.1378	120.17
					t_{regstd}	0.1319	0.1313	126.12
					t_{S0std}	0.1284	0.1309	126.51
					t_{S1std}	0.1241	0.1211	136.74
					t_{S2std}	0.1282	0.1256	131.85
					t_{S3std}	0.1204	0.1183	139.98
					t_{S4std}	0.1215	0.1278	129.57
					t_{S5std}	0.1290	0.1255	131.95
			300	115	t_{Ystd}	0.0728	0.0733	100
					t_{rstd}	0.0601	0.0611	119.96
					t_{regstd}	0.0582	0.0581	126.16
					t_{S0std}	0.0576	0.0579	126.59
					t_{S1std}	0.0542	0.0530	138.30
					t_{S2std}	0.0547	0.0552	132.78
					t_{S3std}	0.0532	0.0530	138.30
					t_{S4std}	0.0579	0.0590	124.23
					t_{S5std}	0.0566	0.0585	125.29

Table 2
Empirical and Theoretical MSE, PRE for the Estimators Relative to RRT Mean Estimator in Stratified Two Phase Random Sampling for Population 2

N	N_h	ρ_{XYh}	n'	n	MSE Estimation			
					Estimation	Empirical	Theoretical	PRE
1000	$N_1 = 550$	$\rho_{XY1} = 0.9522$	60	25	t_{Ystd}	0.1971	0.2000	100
	$N_2 = 450$	$\rho_{XY2} = 0.9478$			t_{rstd}	0.1023	0.1056	189.39
					t_{regstd}	0.1001	0.0989	202.22
					t_{S0std}	0.0897	0.0909	220.02
					t_{S1std}	0.0963	0.0962	207.90
					t_{S2std}	0.0917	0.0931	214.82
					t_{S3std}	0.0823	0.0856	233.64
					t_{S4std}	0.0926	0.0967	206.82
					t_{S5std}	0.0874	0.0853	234.46
			150	55	t_{Ystd}	0.0969	0.1012	100
					t_{rstd}	0.0442	0.0464	218.10
					t_{regstd}	0.0410	0.0421	240.38
					t_{S0std}	0.0371	0.0388	260.82
					t_{S1std}	0.0346	0.0320	316.25
					t_{S2std}	0.0341	0.0340	297.64
					t_{S3std}	0.0353	0.0357	283.47
					t_{S4std}	0.0359	0.0371	272.77
					t_{S5std}	0.0338	0.0322	314.28
			300	115	t_{Ystd}	0.0394	0.0389	100
					t_{rstd}	0.0175	0.0171	227.48
					t_{regstd}	0.0154	0.0158	246.20
					t_{S0std}	0.0145	0.0147	264.63
					t_{S1std}	0.0134	0.0132	294.69
					t_{S2std}	0.0139	0.0135	288.15
					t_{S3std}	0.0138	0.0137	283.94
					t_{S4std}	0.0143	0.0149	261.07
					t_{S5std}	0.0136	0.0141	275.88

Table 3
Empirical and Theoretical MSE, PRE for the Estimators Relative to RRT Mean Estimator in Stratified Two Phase Random Sampling for Real Data Set

N	N_h	ρ_{XYh}	n'	n	MSE Estimation			
					Estimation	**Empirical**	**Theoretical**	**PRE**
1698	$N_1 = 979$	$\rho_{XY1} = 0.7802$	100	45	t_{Ystd}	4.0450	4.1340	100
	$N_2 = 362$	$\rho_{XY2} = 0.7952$			t_{rstd}	3.0010	2.9878	138.36
	$N_3 = 357$	$\rho_{XY3} = 0.8408$			t_{regstd}	2.7087	2.6812	154.18
					t_{S0std}	2.5310	2.4533	168.51
					t_{S1std}	2.588	2.4715	167.26
					t_{S2std}	2.6922	2.6649	155.13
					t_{S3std}	2.4288	2.5018	165.24
					t_{S4std}	2.4458	2.5109	164.64
					t_{S5std}	2.6487	2.7180	152.09
			250	100	t_{Ystd}	1.7922	1.8109	100
					t_{rstd}	1.1488	1.1899	152.18
					t_{regstd}	1.0561	1.0587	171.04
					t_{S0std}	1.0326	1.0389	174.31
					t_{S1std}	0.9299	0.9398	192.68
					t_{S2std}	0.9460	0.9606	188.52
					t_{S3std}	0.9933	0.9925	182.45
					t_{S4std}	0.9920	0.9985	181.36
					t_{S5std}	1.0413	1.0657	169.93
			500	205	t_{Ystd}	0.6955	0.6868	100
					t_{rstd}	0.4382	0.4377	156.91
					t_{regstd}	0.4269	0.4243	161.86
					t_{S0std}	0.4127	0.4156	165.25
					t_{S1std}	0.3814	0.3826	179.50
					t_{S2std}	0.4042	0.4055	169.37
					t_{S3std}	0.3824	0.3888	176.64
					t_{S4std}	0.3983	0.3992	172.04
					t_{S5std}	0.4013	0.4022	170.76