STATISTICS IN TRANSITION new series, December 2023
Vol. 24, No. 5, pp. 45–61, https://doi.org/10.59170/stattrans-2023-063
Received – 25.06.2022; accepted – 04.05.2023

Ratio estimation of two population means in two-phase stratified random sampling under a scrambled response situation

Pitambar Das[1], **Garib Nath Singh**[2], **Arnab Bandyopadhyay**[3]

Abstract

In this paper, we have described the development of an effective two-phase stratified random sampling estimation procedure in a scrambled response situation. Two different exponential, regression-type estimators were formed separately for different structures of two-phase stratified sampling schemes. We have studied the properties of the suggested strategy. The performance of the proposed strategy has been demonstrated through numerical evidence based on a data set of a natural population and a population generated through simulation studies. Taking into consideration the encouraging findings, suitable recommendations for survey statisticians are prepared for the application of the proposed strategy in real-life conditions.

Key words: stratified random sampling, scrambled response, auxiliary variable, mean square error, simulation study.

2010 AMS Subject Classifications: 62D05

1. Introduction

In sample surveys, the population may be formed of heterogeneous units. For example, in socio-economic surveys, people may live in hospital, hostel, residential houses and jail, etc. The whole population is divided into certain internally homogeneous and externally heterogeneous groups, called strata, and then independent samples of different sizes are selected from each stratum. Stratified sampling is one of the most widely used sampling techniques as it increases the precision of the estimate of the survey variable when units of the population are from

[1] Department of Mathematics, Netaji Nagar Day College, Kolkata- 700092, India. E-mail: pitambardas.in@gmail.com.

[2] Department of Mathematics & Computing, Indian Institute of Technology (Indian School of Mines), Dhanbad-826004, India. E-mail: gnsingh@iitism.ac.in.

[3] Department of Mathematics, Asansol Engineering College, Asansol–713305, India. Email: arnabbandyopadhyay4@gmail.com. ORCID: https://orcid.org/0000-0002-0769-7491.

different portions of population. Many authors have discussed different types of estimators using the auxiliary information in stratified random sampling, e.g. Singh and Sukhatme (1973), Kadilar and Cingi (2000, 2003), Shabbir and Gupta (2005), Singh and Vishwakarma (2005), Koyuncu and Kadilar (2008, 2009), Singh *et al.* (2009), etc. It is noted that most of recently developed the estimation of ratio of population mean in simple random sampling only, limited attempts have been taken to estimate the ratio of population mean under stratified random sampling scheme.

In socio -economic surveys, an estimate of the population ratio of two characters for the stratified random sampling may be of considerable interest. For example, the ratio of per month total income and total expenditure of people of different classes in a locality.

In practice, it may also be observed that characteristics under study on sensitive data that someone wants to response but prefers to hide the true value for avoiding possible social stigma and harassment, etc. As many people prefers to hide their exact responses, available sample of returns is camouflaged. Basic idea behind scrambling the survey data is that the response given by the respondent against any query related to some sensitive character is camouflaged by adding or multiplying the data with any random number generated by the respondent by using any random device which will not be known by the surveyor although surveyor may know the mean and variance. This procedure is applied in such a way that the respondents feels that their privacy is protected. Commanding work was done by Greenberg *et al.* (1971), Pollok and Bek (1976), Eichhorn and Hayre (1983), Giancarlo and Pier (2010), Dianna and Perri (2010). In socio -economic surveys, people may live in a different economic zone. If we know the ratio of income and expenditure of people of different classes in a locality, then an investment may be planned suitably. In this case income and expenditure are highly sensitive variables in which scramble response situation may be found. We estimate the ratio of domestic violence reported in a society and the number of premarital abortions that took place in a society, then we reduce the rate of abortions.

It may be noted that no significant attempt has been taken to estimate the ratio of population parameters through two-phase stratified random sampling in the presence of scrambled response situation. Influenced and convinced with the points discussed above, we have recommended a general procedure to estimate population mean in stratified random sampling in presence of the above -mentioned situation. The findings are demonstrated through numerical illustrations carried over the data set of natural population and population generated through simulation studies using different types of correlations. Suitable recommendations are made to the survey statisticians for possible applications.

2. Sample structures and notations

Consider a finite population $U = (U_1, U_2, ..., U_N)$ of size N units and divided into L strata, each of size N_h units (h = 1, 2, ..., L) such that $\sum_{h=1}^{L} N_h = N$. Let y be the study variable and (x, z) be the auxiliary variables respectively taking values y_{hi} and (x_{hi}, z_{hi}) respectively, for the i^{th} unit (i = 1, 2, ..., N_h) of the h^{th} stratum (h = 1, 2, ..., L).

The first phase sample S_{nh} of size n_h is drawn at random without replacement from each stratum containing N_h units (h = 1, 2, …, L). Again the second phase sample S_{mh} of size m_h units is drawn using SRSWOR scheme from each first phase sample of size n_h (h = 1, 2, ..., L).

We consider the following notations for their further use:

$\bar{Y}_h$: Population mean of the study variable y_h (h = 1, 2, ... , L).

$\bar{X}_h, \bar{Z}_h$: Population means of the auxiliary variables x_h and z_h respectively (h = 1, 2, ... , L).

$R_h = \frac{\bar{Y}_h}{\bar{X}_h}$: Ratio of population means of the variables y_h and x_h of the h^{th} stratum (h = 1, 2, ... , L).

$R = \sum_{h=1}^{L} W_h R_h$: Total ratio of the population means.

$R_{nh} = \frac{\bar{y}_{nh}}{\bar{x}_{nh}}$: Ratio of sample means of the variables y_{nh} and x_{nh} based on the sample of size n_h (h = 1, 2, ... , L).

$R_{mh} = \frac{\bar{y}_{mh}}{\bar{x}_{mh}}$: Ratio of sample means of the variables y_{mh} and x_{mh} based on the sample of size m_h (h = 1, 2, ... , L).

$\bar{Y}_h = \sum_{i=1}^{N_h} \frac{y_{hi}}{N_h}$, $\bar{X}_h = \sum_{i=1}^{N_h} \frac{x_{hi}}{N_h}$, $\bar{Z}_h = \sum_{i=1}^{N_h} \frac{z_{hi}}{N_h}$: Population means of the respective variables on the stratum h (h = 1, 2, ..., L).

$\bar{Y} = \sum_{h=1}^{L} \bar{Y}_h W_h$, $\bar{X} = \sum_{h=1}^{L} \bar{X}_h W_h$, $\bar{Z} = \sum_{h=1}^{L} \bar{Z}_h W_h$: Total population means of the respective variables,

where $W_h = \frac{N_h}{N}$:Weight of the h^{th} stratum (h = 1, 2, ... , L).

$\bar{z}_{nh} = \frac{1}{n_h} \sum_{i=1}^{n_h} z_{hi}$: Mean of the variable z based on the sample S_{nh} of size n_h (h = 1, 2, ... , L).

$\bar{z}_{mh}$: Mean of the variable z based on the sample S_{mh} of size m_h (h = 1, 2, ... , L).

$\bar{x}_{nh} = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}$: Mean of the variable x based on the sample S_{nh} of size n_h (h = 1, 2, ... , L).

$\bar{x}_{mh}$: Mean of the variable x based on the sample S_{mh} of size m_h (h = 1, 2, ... , L).

$\bar{y}_{nh} = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$: Mean of the variable y based on the sample S_{nh} of size n_h (h = 1, 2, ... , L).

$\bar{y}_{mh}$: Mean of the variable y based on the sample S_{mh} of size m_h (h = 1, 2, ... , L).

$S_{yh}^2 = \frac{1}{N_h-1}\sum_{i=1}^{N_h}(y_{hi}-\bar{Y}_h)^2$: Population mean square of the variable y based on the stratum h (h = 1, 2, ..., L).

S_{xh}^2, S_{zh}^2: Population mean squares of the variables based on the stratum h (h = 1, 2, ... , L).

$C_{yh} = \frac{S_{yh}}{\bar{Y}_h}$, $C_{xh} = \frac{S_{xh}}{\bar{X}_h}$, $C_{zh} = \frac{S_{zh}}{\bar{Z}_h}$: Coefficient of variations for the variables y, x and z for the h^{th} stratum (h = 1, 2, ..., L).

ρ_{yx_h}, ρ_{yz_h}, ρ_{xz_h}: Correlation coefficients between (y, x), (y, z), and (x, z) respectively in the h^{th} stratum (h = 1, 2, ... , L).

3. Scrambling technique

Eichhorn and Hayre (1983) studied a multiplicative randomized response method for obtaining responses to sensitive questions when the answers are quantitative. The method involves the respondent multiplying his sensitive answer by a random number from a known distribution, and giving the product to the interviewer, who does not know the value of the random number and thus receives a scrambled response. Some particular distributions for the random scrambling number are proposed and studied, and ways of generating the scrambling numbers are discussed.

We have denoted the scrambled variables as $y^* = yT$, $x^* = xT$, where T is the random number multiplied by the study variables y and x to yield y^* and x^*. The device is so selected that the mean of T, i.e. E (T) = 1 to minimize the effect of scrambling. Also, it is important to assume that any two random numbers created are mutually independent as well as the random number T and sensitive variables y and x are also mutually independent.

We use the following notation on scramble variables (i.e. y^* and x^*)

$R_h^* = \frac{\bar{Y}_h^*}{\bar{X}_h^*}$: Ratio of population means of the scramble variables y_h^* and x_h^* of the h^{th} stratum (h = 1, 2, ... , L).

$R^* = \sum_{h=1}^{L} W_h R_h^*$: Total ratio of the population means of the scramble variables.

$R_{nh}^* = \frac{\bar{y}_{nh}^*}{\bar{x}_{nh}^*}$: Ratio of sample means of the scramble variables y_{nh}^* and x_{nh}^* based on the sample of size n_h (h = 1, 2, ... , L).

$R_{mh}^* = \frac{\bar{y}_{mh}^*}{\bar{x}_{mh}^*}$: Ratio of sample means of the scramble variables y_{mh}^* and x_{mh}^* based on the sample of size m_h (h = 1, 2, ... , L).

$\bar{Y}_h^* = \sum_{i=1}^{N_h} \frac{y_{hi}^*}{N_h}$, $\bar{X}_h^* = \sum_{i=1}^{N_h} \frac{x_{hi}^*}{N_h}$: Population means of the respective scramble variables on the stratum h (h = 1, 2, ..., L).

$\bar{Y}^* = \sum_{h=1}^{L} \bar{Y}_h^* W_h$, $\bar{X}^* = \sum_{h=1}^{L} \bar{X}_h^* W_h$: Total population means of the respective scramble variables,

where $W_h = \frac{N_h}{N}$: Weight of the h^{th} stratum (h = 1, 2, ... , L).

$\bar{x}^*_{nh} = \sum_{i \in S_{nh}} \frac{x^*_{hi}}{n_h}$: Mean of the scramble variable x^* based on the sample S_{nh} of size n_h (h = 1, 2, ... , L).

$\bar{x}^*_{mh}$: Mean of the scramble variable x^* based on the sample S_{mh} of size m_h (h = 1, 2, ... , L).

$\bar{y}^*_{nh} = \sum_{i \in S_{nh}} \frac{y^*_{hi}}{n_h}$: Mean of the scramble variable y^* based on the sample S_{nh} of size n_h (h = 1, 2, ... , L).

$\bar{y}^*_{mh}$: Mean of the scramble variabley^* based on the sample S_{mh} of size m_h (h = 1, 2, ... , L).

$S^{2^*}_{yh} = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(y^*_{hi} - \bar{Y}^*_h)^2$: Population mean square of the scramble variable y^* based on the stratum h (h = 1, 2, ..., L).

$S^{2^*}_{xh}$: Population mean square of the scramble variable based on the stratum h (h = 1, 2, ..., L).

$C^*_{yh} = \frac{S^*_{yh}}{\bar{Y}^*_h}$, $C^*_{xh} = \frac{S^*_{xh}}{\bar{X}^*_h}$: Coefficient of variations for the scramble variablesy^*, x^* for the h^{th} stratum (h = 1, 2, ..., L).

$\rho^*_{yx_h}$, $\rho^*_{yz_h}$, $\rho^*_{xz_h}$: Correlation coefficients between(y^*, x^*),(y^*, z), and (x^*, z)respectively in the h^{th} stratum (h = 1, 2, ..., L).

4. Formulations of the estimators

To estimate the ratio of population means $R^* \left(= \frac{\bar{Y}}{\bar{X}}\right)$on the two-phases stratified random sampling scheme. It is also noted that we are getting scrambled response only for the variables y and x in terms of the variables $y^*$and $x^*$from the respective samples.

Motivated with the previous work, we have proposed the following chain type exponential and regression type estimators for ratio of population means R^*on the two-phase stratified random sampling scheme as

$$\xi_1 = \sum_{h=1}^{L} W_h R^*_{mh} \exp\left(\frac{R'^*_{nh} - R'^*_{mh}}{R'^*_{nh} + R'^*_{mh}}\right) \quad (1)$$

$$\xi_2 = \sum_{h=1}^{L} W_h \left[R^*_{mh} + b_{2h}\left(R'^*_{nh} - R'^*_{mh}\right)\right] \quad (2)$$

where

$$R'^*_{nh} = R^*_{nh} + b_{1h}(\bar{Z}_h - \bar{z}_{nh}) \text{ and } R'^*_{mh} = R^*_{mh} + b'_{1h}(\bar{Z}_h - \bar{z}_{mh}).$$

We first estimate the ratio of the population means R^*_h on the stratum h (h = 1, 2, ..., L) using the proposed estimators and then estimate the total ratio of the population means R^*.

5.1. Mean square errors of the proposed estimators

The mean square errors (MSEs) of the proposed estimators ξ_1 and ξ_2 up to the first order of approximation are derived under large sample approximation using the following transformations:

$\bar{y}^*_{mh} = \bar{Y}_h(1+e_0)$, $\bar{x}^*_{mh} = \bar{X}_h(1+e_1)$,

$\bar{y}^*_{\mathrm{nh}} = \bar{\mathrm{Y}}_h(1+\mathrm{e}_2)$, $\bar{\mathrm{x}}^*_{\mathrm{nh}} = \bar{\mathrm{X}}_h(1+\mathrm{e}_3)$,

$\bar{z}_{\mathrm{mh}} = \bar{\mathrm{Z}}_h(1+\mathrm{e}_4)$, $\bar{\mathrm{z}}_{\mathrm{nh}} = \bar{\mathrm{Z}}_h(1+\mathrm{e}_5)$,

such that $\mathrm{E}(e_i) = 0$ and $|e_i| < 1$, for all i = 0, 1, ..., 5.

Under the above transformations the estimators take the following forms:

$$\xi_1 = \sum_{\mathrm{h}=1}^{L} W_h R_h^* (1+\mathrm{e}_0)(1+\mathrm{e}_1)^{-1} \exp\left[\begin{array}{l}\left\{(1+\mathrm{e}_2)(1+\mathrm{e}_3)^{-1} - (1+\mathrm{e}_0)(1+\mathrm{e}_1)^{-1} + \frac{b_{1\mathrm{h}}\bar{Z}_h}{R_h^*}(e_4 - \mathrm{e}_5)\right\} \\ \times \frac{1}{2}\left\{1 + \frac{1}{2}\left\langle (e_2 - \mathrm{e}_3) + (e_0 - \mathrm{e}_1) - \frac{b_{1\mathrm{h}}\bar{Z}_h}{R_h^*}(e_4 + \mathrm{e}_5)\right\rangle\right\}^{-1}\end{array}\right] \tag{3}$$

$$\xi_2 = \sum_{\mathrm{h}=1}^{L} W_h R_h^* \left[(1+\mathrm{e}_0)(1+\mathrm{e}_1)^{-1} + \mathrm{b}_{2\mathrm{h}}\{(1+\mathrm{e}_2)(1+\mathrm{e}_3)^{-1} - (1+\mathrm{e}_0)(1+\mathrm{e}_1)^{-1}\} + \frac{b_{3\mathrm{h}}\bar{Z}_h}{R_h^*}(e_4 - \mathrm{e}_5)\right] \tag{4}$$

Again, we obtain the following expression for expectations:

$E(e_0^2) = \mathrm{f}_1 C_{y_h}^{*2}$, $\mathrm{E}(e_1^2) = \mathrm{f}_1 C_{x_h}^{*2}$, $\mathrm{E}(e_2^2) = \mathrm{f}_2 C_{y_h}^{*2}$, $\mathrm{E}(e_3^2) = \mathrm{f}_2 C_{x_h}^{*2}$,

$E(e_4^2) = \mathrm{f}_1 C_{z_h}^2$, $\mathrm{E}(e_5^2) = \mathrm{f}_2 C_{z_h}^2$,

$E(e_0 e_1) = \mathrm{f}_1 \rho^*_{\mathrm{yx}_h} C^*_{y_h} C^*_{x_h}$, $\mathrm{E}(e_1 e_2) = \mathrm{f}_2 \rho^*_{\mathrm{yx}_h} C^*_{y_h} C^*_{x_h}$, $\mathrm{E}(e_2 e_3) = \mathrm{f}_2 \rho^*_{\mathrm{yx}_h} C^*_{y_h} C^*_{x_h}$,

$E(e_3 e_4) = \mathrm{f}_2 \rho^*_{\mathrm{xz}_h} C^*_{x_h} C_{z_h}$, $\mathrm{E}(e_3 e_5) = \mathrm{f}_2 \rho^*_{\mathrm{xz}_h} C^*_{x_h} C_{z_h}$, $\mathrm{E}(e_0 e_4) = \mathrm{f}_1 \rho_{\mathrm{yz}_h} C^*_{y_h} C_{z_h}$,

$E(e_0 e_5) = \mathrm{f}_2 \rho^*_{\mathrm{yz}_h} C^*_{y_h} C_{z_h}$, $\mathrm{E}(e_4 e_5) = \mathrm{f}_2 C^2_{z_h}$, $\mathrm{E}(e_1 e_4) = \mathrm{f}_1 \rho^*_{\mathrm{xz}_h} C^*_{x_h} C_{z_h}$,

$E(e_0 e_4) = \mathrm{f}_1 \rho^*_{\mathrm{yz}_h} C^*_{y_h} C_{z_h}$, $E(e_0 e_2) = \mathrm{f}_2 C_{y_h}^{*2}$, $\mathrm{E}(e_0 e_3) = \mathrm{f}_2 \rho^*_{\mathrm{yx}_h} C^*_{y_h} C^*_{x_h}$,

$E(e_1 e_3) = \mathrm{f}_2 C_{x_h}^{*2}$, $\mathrm{E}(e_0 e_5) = \mathrm{f}_2 \rho^*_{\mathrm{yz}_h} C^*_{y_h} C^*_{z_h}$, $\mathrm{E}(e_1 e_5) = \mathrm{f}_2 \rho^*_{\mathrm{xz}_h} C^*_{x_h} C_{z_h}$,

$E(e_2 e_4) = \mathrm{f}_2 \rho^*_{\mathrm{yz}_h} C^*_{y_h} C_{z_h}$, $\mathrm{E}(e_2 e_5) = \mathrm{f}_2 \rho^*_{\mathrm{xz}_h} C^*_{x_h} C_{z_h}$,

where $\mathrm{f}_1 = \left(\frac{1}{m_h} - \frac{1}{N_h}\right)$ and $\mathrm{f}_2 = \left(\frac{1}{n_h} - \frac{1}{N_h}\right)$.

Taking expectations on both sides of equations (3) and (4), we obtained the mean square error of the estimators ξ_1 and ξ_2 up to the first order of approximations as

$$\begin{aligned} M(\xi_1) &= \mathrm{E}(\xi_1 - \mathrm{R}^*)^2 \\ &= \frac{1}{4}\sum_{\mathrm{h}=1}^{L} W_h^2 R_h^{*2} \left[2(f_1 + \mathrm{f}_2)\left(C_{x_h}^{*2} + \mathrm{C}_{y_h}^{*2}\right) - 2(f_1 + 2\mathrm{f}_2)\rho^*_{\mathrm{yx}_h} C^*_{y_h} C^*_{x_h} \right. \\ &\qquad \left. + \frac{b_{1\mathrm{h}}\bar{Z}_h}{R_h^*}(f_1 - \mathrm{f}_2)\rho^*_{\mathrm{yz}_h} C^*_{y_h} C_{z_h} + \frac{b_{1\mathrm{h}}^2 \bar{Z}_h^2}{R_h^{*2}}(f_1 - \mathrm{f}_2) C^2_{z_h}\right] \\ &= \frac{1}{4}\sum_{\mathrm{h}=1}^{L} W_h^2 R_h^{*2} \left[A_1 + \frac{b_{1\mathrm{h}}\bar{Z}_h}{R_h^*} B_1 + \frac{b_{1\mathrm{h}}^2 \bar{Z}_h^2}{R_h^{*2}} C_1\right] \end{aligned} \tag{5}$$

where $\mathrm{A}_1 = 2(f_1 + \mathrm{f}_2)\left(C_{x_h}^{*2} + \mathrm{C}_{y_h}^{*2}\right) - 2(f_1 + 2\mathrm{f}_2)\rho^*_{\mathrm{yx}_h} C^*_{y_h} C^*_{x_h}$, $\mathrm{B}_1 = (f_1 - \mathrm{f}_2)\rho^*_{\mathrm{yz}_h} C^*_{y_h} C_{z_h}$,

$C_1 = (f_1 - \mathrm{f}_2) C^2_{z_h}$.

$$M(\xi_2) = \mathrm{E}(\xi_2 - \mathrm{R}^*)^2$$

$$\mathrm{M}(\xi_2) = \sum_{h=1}^{L} \mathrm{W}_h^2 \mathrm{R}_h^{*2} \left[\begin{array}{l} \mathrm{f}_1 \left(\mathrm{C}_{y_h}^{*2} + \mathrm{C}_{x_h}^{*2} - 2\rho_{yx_h}^{*} \mathrm{C}_{y_h}^{*} \mathrm{C}_{x_h}^{*} \right) + \mathrm{b}_{2h}^2 \left\{ \left(\mathrm{f}_1 + \mathrm{f}_2 \right) \left(\mathrm{C}_{y_h}^{*2} + \mathrm{C}_{x_h}^{*2} \right) + 2 \left(\mathrm{f}_2 - \mathrm{f}_1 \right) \rho_{yx_h}^{*} \mathrm{C}_{y_h}^{*} \mathrm{C}_{x_h}^{*} \right\} + \\ \dfrac{\mathrm{b}_{3h}^2 \overline{\mathrm{Z}}_h}{\mathrm{R}_h^{*2}} \left(\mathrm{f}_1 - \mathrm{f}_2 \right) \mathrm{C}_{z_h}^2 + 2\mathrm{b}_{2h} \left\{ \left(\mathrm{f}_2 - \mathrm{f}_1 \right) \left(\mathrm{C}_{y_h}^{*2} + \mathrm{C}_{x_h}^{*2} - 2\rho_{yx_h}^{*} \mathrm{C}_{y_h}^{*} \mathrm{C}_{x_h}^{*} \right) \right\} + \dfrac{2\mathrm{b}_{3h} \overline{\mathrm{Z}}_h}{\mathrm{R}_h^{*}} \left(\mathrm{f}_1 - \mathrm{f}_2 \right) \\ \left(\rho_{yz_h}^{*} \mathrm{C}_{y_h}^{*} \mathrm{C}_{z_h} - \rho_{xz_h}^{*} \mathrm{C}_{x_h}^{*} \mathrm{C}_{z_h} \right) + \dfrac{2\mathrm{b}_{4h} \overline{\mathrm{Z}}_h}{\mathrm{R}_h^{*}} \left\{ \left(2\mathrm{f}_2 - \mathrm{f}_1 \right) \left(\rho_{yz_h}^{*} \mathrm{C}_{y_h}^{*} \mathrm{C}_{z_h} - \rho_{xz_h}^{*} \mathrm{C}_{x_h}^{*} \mathrm{C}_{z_h} \right) \right\} \end{array} \right]$$

$$= \sum_{h=1}^{L} \mathrm{W}_h^2 \mathrm{R}_h^{*2} \left[\mathrm{A}_2 + \mathrm{b}_{2h}^2 \mathrm{B}_2 + \frac{\mathrm{b}_{3h}^2 \overline{\mathrm{Z}}_h^2}{\mathrm{R}_h^{*2}} \mathrm{C}_2 + \mathrm{b}_{2h} \mathrm{D}_2 + \frac{\mathrm{b}_{3h} \overline{\mathrm{Z}}_h}{\mathrm{R}_h^{*}} \mathrm{E}_2 + \frac{\mathrm{b}_{4h} \overline{\mathrm{Z}}_h}{\mathrm{R}_h^{*}} \mathrm{F}_2 \right] \tag{6}$$

where $\mathrm{A}_2 = f_1 \left(C_{y_h}^{*2} + \mathrm{C}_{x_h}^{*2} - 2\rho_{\mathrm{yx}_h}^{*} C_{y_h}^{*} C_{x_h}^{*} \right)$, $\mathrm{B}_2 = \left\{ (f_1 + \mathrm{f}_2) \left(C_{y_h}^{*2} + \mathrm{C}_{x_h}^{*2} \right) + 2(f_2 - \mathrm{f}_1) \rho_{\mathrm{yx}_h}^{*} C_{y_h}^{*} C_{x_h}^{*} \right\}$,

$$C_2 = (f_1 - \mathrm{f}_2) C_{z_h}^2, \ \mathrm{D}_2 = 2 \left\{ (f_2 - \mathrm{f}_1) \left(C_{y_h}^{*2} + \mathrm{C}_{x_h}^{*2} - 2\rho_{\mathrm{yx}_h}^{*} C_{y_h}^{*} C_{x_h}^{*} \right) \right\},$$

$$E_2 = \frac{2\bar{\mathrm{Z}}_h}{R_h^{*}} (f_1 - \mathrm{f}_2) \left(\rho_{\mathrm{yz}_h}^{*} C_{y_h}^{*} C_{z_h} - \rho_{\mathrm{xz}_h}^{*} C_{x_h}^{*} C_{z_h} \right), \ \mathrm{F}_2 = \frac{2\bar{\mathrm{Z}}_h}{R_h^{*}} \left\{ (2\mathrm{f}_2 - \mathrm{f}_1) \left(\rho_{\mathrm{yz}_h}^{*} C_{y_h}^{*} C_{z_h} - \rho_{\mathrm{xz}_h}^{*} C_{x_h}^{*} C_{z_h} \right) \right\}.$$

5.2. Minimum mean square errors of the proposed estimators

It may be observed from equation (5) and (6) that the expressions for $M(\xi_1)$ and $M(\xi_2)$ depend on the values of b_{1h}, b_{2h}, b_{3h} and b_{4h}(h =1, 2, . . ., L), which are real constants. Therefore, we need to find the optimum values of b_{1h}, b_{2h}, b_{3h} and b_{4h} which can minimize the MSE of the estimators ξ_1 and ξ_2 respectively.The optimum values of b_{1h}, b_{2h}, b_{3h} and b_{4h}(h = 1, 2, . . ., L) are found as

$$b_{1h} = -\frac{R_h^{*} B_1}{2\mathrm{C}_1 \bar{Z}_h} \tag{7}$$

$$b_{2h} = -\frac{D_2}{2\mathrm{B}_2} \tag{8}$$

$$b_{3h} = -\frac{E_2 R_h^{*}}{2\mathrm{C}_2 \bar{Z}_h} \tag{9}$$

$$b_{4h} = \frac{D_2 E_2 R_h^{*}}{4\mathrm{B}_2 C_2 \bar{Z}_h} \tag{10}$$

Thus, substituting the values of b_{1h}, b_{2h}, b_{3h} and b_{4h} from equations (7), (8), (9) and (10) to the equations (5) and (6),we have derived the optimum mean square errors of the proposed estimators as

$$\mathrm{M}(\xi 1)\mathrm{opt} = \frac{1}{4}\sum_{\mathrm{h}=1}^{L} W_h^2 R_h^{*2}\left[A_1 - \frac{B_1^2}{4\mathrm{C}_1}\right] \tag{11}$$

$$M(\xi_2)_{\mathrm{opt}} = \sum_{\mathrm{h}=1}^{L} W_h^2 R_h^{*2}\left[A_2 - \frac{D_2^2}{4\mathrm{B}_2} - \frac{E_2^2}{4\mathrm{C}_2} + \frac{D_2 E_2 F_2}{4\mathrm{B}_2 C_2}\right] \tag{12}$$

Remark: 5.2.1. It is to be noted that the optimality condition of the estimators in equation (11) and (12) the optimum values b_{1h}, $b_{2h,}$ b_{3h} and b_{4h} depend on unknown population parameters such as R_h^{*2}, $C_{x_h}^{2*}$, $C_{y_h}^{2*}$, $\rho_{xy_h}^{*}$ and $\rho_{yz_h}^{*}$ (h = 1, 2, ..., L).

Thus, to make the classes of estimators practicable, these unknown population parameters may be estimated with their respective sample estimates or from past data or guessed from experience gathered over time. Such problems are also considered by Reddy (1978), Tracy *et al.* (1996) and Singh *et al.* (2007).

6. Efficiency comparison

It is important to investigate the situation under which our proposed estimators are more efficient than the conventional ones. Since no estimator has been improved yet for the ratio of population means in stratified random sampling under scrambled response. Therefore, we have consider the natural ratio of sample means estimator r^* under scramble response

situation, where $r^* = \sum_{\mathrm{h}=1}^{L} W_h \frac{\bar{y}_{\mathrm{mh}}^*}{\bar{x}_{\mathrm{mh}}^*}$.

Its variance may be derived under scrambled response situation as

$$V(r^*) = \sum_{\mathrm{h}=1}^{L} W_h f_1 R_h^{*2} C_{y_h}^{*2}.$$

The percent relative efficiency of the estimators with respect to ratio of sample means estimator r* (in the presence of scrambled response situation) is defined as

$$\mathrm{PRE}(E_1) = \frac{V(r^*)}{M(\xi_1)_{\mathrm{opt.}}} \times 100 \tag{13}$$

$$\mathrm{PRE}(E_2) = \frac{V(r^*)}{M(\xi_2)_{\mathrm{opt.}}} \times 100 \tag{14}$$

6.1. Numerical illustration

It is important to investigate the situation where our suggested methodologies are superior to the conventional ones. The performances of the suggested technique are demonstrated through empirical investigations carried over the data set of natural population and artificially generated population. For the sake of minimum impact of scrambling on the actual data we have considered random variable $T^* = 1$and $S_T^{*2} = 0.16$. It is to be noted that the parametric values of the scrambled variables x* and y* for different populations may be obtained by using the statistical parameters of the variables x and y shown in sections.

6.2. Numerical illustration using known natural population

6.2.1. Natural population data set 1: We have selected natural population data sets on abortion rates form Statistical Abstract of the United States: 2011 to elucidate the efficacious performance of our proposed estimator. The nature of the variables y, x and z and the values of the various parameters are given below.

y, x, z: the number of abortions reported in the state of US during the years 2008, 2007 and 2005 respectively. The detailed calculations on various parameters are given in Table 1.

Table 1: Formation of 4 different strata (zone wise) out of 51 states of United States and corresponding parametric values

Strata	Constituent States	Statistical Parameters
Strata 1	Wyoming, Missouri, Mississippi, Kentucky, Oklahoma , Arkansas, Indiana, Nebraska, South Carolina, Wisconsin, Utah, South Dakota, Idaho, West Virginia.	N_h=14, n_h=13, m_h= 7, $\bar{X}_h$= 6.551, $\bar{Y}_h$= 6.59, $\bar{Z}_h$= 6.720, h = 1.
Strata 2	Alaska, Montana, New Hampshire, Minnesota, Vermont, Ohio, Arizona, New Mexico, North Dakota, Maine, Michigan, Massachusetts, Washington, Kansas, Virginia, North Carolina, Oregon, Pennsylvania, Texas, Louisiana, Colorado, Tennessee, Iowa, Alabama , Georgia.	N_h= 25, n_h= 22, m_h=15, $\bar{X}_h$=15.031, $\bar{Y}_h$=15.11, $\bar{Z}_h$=14.851, h = 2.
Strata 3	Hawaii, Rhode, Island, Connecticut, Nevada, Florida, California, Illinois.	N_h= 7, n_h= 6, m_h= 3, $\bar{X}_h$= 24.562, $\bar{Y}_h$= 24.48, $\bar{Z}_h$= 24.095, h = 3.
Strata 4	Maryland, District of Columbia, New Jersey, New York, Delaware.	N_h= 5, n_h= 4, m_h= 2, $\bar{X}_h$= 33.528, $\bar{Y}_h$= 33.55, $\bar{Z}_h$= 36.533, h = 4.

The computed values of population mean squares and correlation coefficients of the respective variables based on the strata h (h = 1, 2, . . ., 4) are shown in Table 2.

Table 2: Population mean squares and correlation coefficients of the respective variables

Strata	S_{yh}^2	S_{xh}^2	S_{zh}^2	ρ_{xy_h}	ρ_{yz_h}	ρ_{xz_h}
Strata 1	4.56	4.51	5.21	0.9784	0.9725	0.9484
Strata 2	6.91	6.93	8.87	0.9413	0.8780	0.8988
Strata 3	31.42	37.22	65.29	0.9885	0.9442	0.9545
Strata 4	19.31	29.03	53.75	0.9751	-0.3926	-0.5396

We have obtained the PRE of the proposed estimator with respect to natural ratio sample mean estimator for different values of sample sizes and outcomes are given in Table 3.

Table 3: PRE of the proposed strategy based on natural population data set 1

Sample Sizes		PRE (E_1)	PRE (E_2)	Sample Sizes		PRE (E_1)	PRE (E_2)
$n_h = 4$ (h = 1, 2, 3, 4)	$m_1 = 3$	546.7479	3356.626	$n_1 = 8$	$m_h = 3$ (h = 1, 2, 3, 4)	629.1048	4414.462
	$m_2 = 2$			$n_2 = 13$			
	$m_3 = 3$			$n_3 = 6$			
	$m_4 = 2$			$n_4 = 4$			
$n_h = 4$ (h = 1, 2, 3, 4)	$m_1 = 2$	585.2504	3862.067	$n_1 = 9$	$m_h = 3$ (h = 1, 2, 3, 4)	631.2996	4526.913
	$m_2 = 2$			$n_2 = 14$			
	$m_3 = 3$			$n_3 = 4$			
	$m_4 = 3$			$n_4 = 4$			
$n_h = 4$ (h = 1, 2, 3, 4)	$m_1 = 3$	606.8738	4256.116	$n_1 = 10$	$m_h = 3$ (h = 1, 2, 3, 4)	635.5419	4688.513
	$m_2 = 3$			$n_2 = 15$			
	$m_3 = 2$			$n_3 = 5$			
	$m_4 = 2$			$n_4 = 4$			

6.2.2. Natural population data set 2: We have taken another natural population data set on literacy rates in India based on the Census: 2011. The nature of the variables y, x and z and the values of the various parameters are given in Table 4.

y, x, z: the number of literates (persons) during the years 2001, 2011 and the female literacy rate (2011) respectively.

Table 4: Formation of 4 different strata out of 34 states of India (zone wise) and corresponding

Strata	Constituent States	Statistical Parameters
Strata 1	Andhra Pradesh (S), Karnataka (S), Kerala (S), Tamil Nadu (S), Chhattisgarh (C), Madhya Pradesh (C).	N_h= 6, n_h= 5, m_h= 3, $\bar{X}_h$= 68.06, $\bar{Y}_h$= 75.13, $\bar{Z}_h$= 65.18, h =1.
Strata 2	West Bengal(E), Jharkhand (E), Odisha (E), Bihar (E), Manipur (NE), Meghalaya (NE), Nagaland (NE), Arunachal Pradesh (NE), Sikkim (NE), Assam(NE), Tripura(NE).	N_h=11, n_h= 9, m_h= 6, $\bar{X}_h$= 62.2, $\bar{Y}_h$= 62.80, $\bar{Z}_h$= 59.7, h = 2.
Strata 3	Haryana(N), Himachal Pradesh (N), Jammu & Kashmir (N), Uttar Pradesh (N), Uttarakhand (N), Goa (W), Gujarat(W), Maharashtra (W), Punjab (W), Rajasthan (W).	N_h=10, n_h= 8, m_h= 5, $\bar{X}_h$= 68.62, $\bar{Y}_h$= 78.18, $\bar{Z}_h$= 68.12, h = 3.
Strata 4	A.&N. Islands, Chandigarh, D.&N. Haveli, Daman & Diu, Delhi, Lakshadweep, Pondicherry.	N_h= 7, n_h= 5, m_h= 3, $\bar{X}_h$= 78.07, $\bar{Y}_h$= 69.35, $\bar{Z}_h$= 79.54, h = 4.

The computed values of population mean squares and correlation coefficients of the respective variables based on the strata h (h = 1, 2, 3, 4) are shown in Table 5.

Table 5: Population mean squares and correlation coefficients of the respective variables

Strata	S^2_{yh}	S^2_{xh}	S^2_{zh}	ρ_{xy_h}	ρ_{yz_h}	ρ_{xz_h}
Strata 1	127.70	172.18	194.02	0.9929	0.9980	0.9970
Strata 2	56.01	80.53	107.77	0.9631	0.9631	0.9631
Strata 3	24.75	17.68	214.00	0.9633	0.9851	0.9988
Strata 4	21.76	89.89	52.67	0.9532	0.9822	0.9862

We have obtained the PRE of the proposed estimator with respect to natural ratio sample mean estimator for different values of sample sizes and outcomes are given in Table 6.

Table 6: PRE of the proposed strategy based on natural population data set 2

Sample Sizes		PRE (E1)	PRE (E2)	Sample Sizes		PRE (E1)	PRE (E2)
nh = 5 (h = 1, 2, 3, 4)	m1 = 3	560.256	3777.337	n1 = 5	mh = 4 (h = 1, 2, 3, 4)	587.743	4026.767
	m2 = 2			n2 = 8			
	m3 = 4			n3 = 9			
	m4 = 3			n4 = 6			
nh = 5 (h = 1, 2, 3, 4)	m1 = 4	603.106	4032.503	n1 = 5	mh = 4 (h = 1, 2, 3, 4)	592.549	4163.234
	m2 = 3			n2 = 9			
	m3 = 2			n3 = 8			
	m4 = 4			n4 = 6			
nh = 5 (h = 1, 2, 3, 4)	m1 = 2	669.103	5241.656	n1 = 5	mh = 4 (h = 1, 2, 3, 4)	595.730	4463.945
	m2 = 4			n2 = 10			
	m3 = 3			n3 = 7			
	m4 = 4			n4 = 6			

6.3. Numerical illustration using artificially generated population data set:

Efficiency comparison through artificial population generation technique helps in concluding whether a newly developed technique is better than the existing ones (see for instance the work of Singh *et al.* (2017)). Motivated by the artificial population generation techniques adopted by Singh and Deo (2003) and Singh *et al.* (2017).

We have generated the population artificially and taken 5 strata sequentially each of size 20. The PREs of the proposed estimators with respect to sample mean estimator are computed for different values of sample sizes (i.e. n and m) and correlation coefficients (i.e. rx1z1) as displayed in Table 7 and Table 8.

Table 7: PRE of the proposed strategy based on simulated population

Artificial data set -1 ryx = 0.7 ; rx1z1 = 0.5				**Artificial data set -2 ryx = 0.7; rx1z1 = 0.75**			
Sample Sizes		PRE (E_1)	PRE (E_2)	Sample Sizes		PRE (E_1)	PRE (E_2)
$n_h = 16$	$m_1 = 15$	573.1565	1193.307	$n_h = 16$	$m_1 = 15$	347.2386	515.4361
	$m_2 = 14$				$m_2 = 14$		
	$m_3 = 10$				$m_3 = 10$		
	$m_4 = 7$				$m_4 = 7$		
	$m_5 = 11$				$m_5 = 11$		
$n_h = 16$	$m_1 = 14$	585.493	1259.59	$n_h = 16$	$m_1 = 14$	359.9875	576.5938
	$m_2 = 13$				$m_2 = 13$		
	$m_3 = 9$				$m_3 = 9$		
	$m_4 = 6$				$m_4 = 6$		
	$m_5 = 10$				$m_5 = 10$		
$n_h = 16$	$m_1 = 13$	595.8184	1327.56	$n_h = 16$	$m_1 = 13$	371.6504	639.7454
	$m_2 = 12$				$m_2 = 12$		
	$m_3 = 8$				$m_3 = 8$		
	$m_4 = 5$				$m_4 = 5$		
	$m_5 = 9$				$m_5 = 9$		
$n_1 = 14$	$m_h = 10$	527.7561	1148.109	$n_1 = 14$	$m_h = 10$	295.8098	428.7791
$n_2 = 13$				$n_2 = 13$			
$n_3 = 12$				$n_3 = 12$			
$n_4 = 15$				$n_4 = 15$			
$n_5 = 16$				$n_5 = 16$			
$n_1 = 15$	$m_h = 10$	544.4047	1221.123	$n_1 = 15$	$m_h = 10$	312.4275	467.9867
$n_2 = 14$				$n_2 = 14$			
$n_3 = 13$				$n_3 = 13$			
$n_4 = 16$				$n_4 = 16$			
$n_5 = 17$				$n_5 = 17$			
$n_1 = 16$	$m_h = 10$	559.7598	1310.919	$n_1 = 16$	$m_h = 10$	328.579	504.79
$n_2 = 15$				$n_2 = 15$			
$n_3 = 14$				$n_3 = 14$			
$n_4 = 17$				$n_4 = 17$			
$n_5 = 18$				$n_5 = 18$			

Table 8: PRE of the proposed strategy based on simulated population

Artificial data set -3 ryx = 0.8 ; rx1z1 = 0.4				**Artificial data set -2 ryx = 0.7; rx1z1 = 0.75**			
Sample Sizes		PRE (E_1)	PRE (E_2)	Sample Sizes		PRE (E_1)	PRE (E_2)
$n_h = 15$	$m_1 = 8$	315.0862	484.5729	$n_h = 15$	$m_1 = 10$	331.263	523.5138
	$m_2 = 9$				$m_2 = 9$		
	$m_3 = 10$				$m_3 = 8$		
	$m_4 = 11$				$m_4 = 7$		
	$m_5 = 12$				$m_5 = 6$		
$n_h = 15$	$m_1 = 9$	305.04	447.2177	$n_h = 15$	$m_1 = 12$	313.6524	422.9525
	$m_2 = 10$				$m_2 = 11$		
	$m_3 = 11$				$m_3 = 10$		
	$m_4 = 12$				$m_4 = 13$		
	$m_5 = 13$				$m_5 = 14$		
$n_h = 15$	$m_1 = 10$	292.7033	422.6365	$n_h = 15$	$m_1 = 13$	291.9995	389.6614
	$m_2 = 11$				$m_2 = 14$		
	$m_3 = 12$				$m_3 = 11$		
	$m_4 = 13$				$m_4 = 13$		
	$m_5 = 14$				$m_5 = 12$		
$n_1 = 10$	$m_h = 9$	309.3693	427.3363	$n_1 = 11$	$m_h = 10$	274.0509	361.8313
$n_2 = 11$				$n_2 = 12$			
$n_3 = 13$				$n_3 = 13$			
$n_4 = 14$				$n_4 = 14$			
$n_5 = 15$				$n_5 = 15$			
$n_1 = 12$	$m_h = 9$	346.0258	528.5342	$n_1 = 12$	$m_h = 10$	307.0895	445.9498
$n_2 = 14$				$n_2 = 14$			
$n_3 = 15$				$n_3 = 15$			
$n_4 = 16$				$n_4 = 16$			
$n_5 = 18$				$n_5 = 18$			
$n_1 = 13$	$m_h = 9$	361.2879	610.2509	$n_1 = 15$	$m_h = 10$	340.2874	572.5906
$n_2 = 15$				$n_2 = 16$			
$n_3 = 17$				$n_3 = 17$			
$n_4 = 18$				$n_4 = 18$			
$n_5 = 19$				$n_5 = 19$			

7. Conclusion

We have noted the following observations from Tables 1 - 8.

Findings from natural population data sets (Tables 1 - 6)

a) It is clear that the selected natural data set 1 and data set 2 are heterogeneous as they have significantly different values of parameters. The values of the PREs, i.e. E_1 and E_2, are very high for different choices of the sample sizes, which indicates that

suggested estimators perform profoundly for the data sets belong to heterogeneous population. This phenomenon is a desirable one because most of the data sets we come across in practice belong to heterogeneous population. Thus, these performances of our estimators enhance their recommendation in practice.

b) For fixed values of the first phase sample sizes (i.e. n_h), the percent relative efficiencies of the proposed estimators in scrambled response situations, i.e. E_1 and E_2, are increasing when the values of the second phase sample sizes (i.e. m_h) are decreasing. This phenomenon indicates only a smaller fraction of the sample is to be drawn at the second phase but it may produce precise estimate. This reduces the cost of the survey.

Findings from artificially generated population data sets (Tables 7 - 8)

a) It may be observed that the data sets obtained from the artificially generated population are almost homogeneous as parametric values are close enough. It may be noted that our proposed strategies produce precise estimates as the values of PREs are high for different values of the sample sizes.
b) For the fixed values of n_h (first phase sample size), rx1z1 (correlation coefficient) and ryx (correlation coefficient), the percent relative efficiencies E_1 and E_2 are increasing when the values of the sample size at the second phase m_h is decreasing. This reduces the cost of the survey.
c) For the fixed values of n_h (first phase sample size), m_h (second phase sample size) and ryx (correlation coefficient) the percent relative efficiencies E_1 and E_2 are increasing when the values of rx1z1 (correlation coefficient) are decreasing. Thus, it is clear that for high values of the correlation coefficient between study variable and auxiliary variable our proposed strategy produces more precise estimates. This behaviour helps us in choosing a population for application of our strategy in real life.

Therefore, it is established that our proposed strategies may produce efficient estimators in comparison with the conventional ones and they are also applicable for homogeneous and heterogeneous population. Looking at the encouraging findings, our proposed methodologies are recommended to the survey practitioners for their applications in real life.

Acknowledgements

Authors are thankful to the editor and reviewers for their valuable and constructive suggestions, which motivated us to generate an improved version of the present manuscript.

References

Diana, G., Perri, P. F., (2010). New scrambled response models for estimating the mean of a sensitive quantitative character. *Journal of Applied Statistics*, Vol. 37, pp. 1875–1890.

Eichhorn, B., Hayre, L. S., (1983). Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and Inference*, Vol. 7, pp. 307–316.

Giancarlo, D., Pier, P. F., (2010). New scrambled response models for estimating the mean of a sensitive quantitative character. *Journal of Applied Statistics*, Vol. 37, pp.1875–1890, DOI: 10.1080/02664760903186031.

Greenberg, B. G., Kuebler, R. R., Abernathy, J. R., Horvitz, D. G., (1971). Application of the randomized response technique in obtaining quantitative data. *Journal of American Statistical Association*, Vol. 66, pp. 243–250.

Kadilar, C., Cingi, H., (2000). Ratio Estimator in stratified sampling. *Biometrical Journal*, Vol. 45, pp. 218–225.

Kadilar, C., Cingi, H., (2003). A new ratio Estimator in stratified sampling. *Communication in Statistics-Theory and Methods*, Vol. 34, pp. 597–602.

Pollock, K. H., Bek, Y., (1976). A comparison of three randomized response models for quantitative data. *Journal of American Statistical Association*, Vol. 71, pp. 884–886.

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

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

Reddy, V.N., (1978). A study on the use of prior knowledge on certain population parameters in estimation. *Sankhya*, Series C, 40, pp. 29–37.

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

Singh, S., Deo, B., (2003). Imputation by power transformation. *Statistical Papers*, Vol. 4, pp. 555–579.

Singh, H.P., Vishwakarma, G. K., (2005). Combined Ratio - product Estimator of Finite Population Mean in Stratified Sampling. *Metodologia de Encuestas*, Vol. 8, pp. 35–44.

Singh, R., Sukhatme, B. V., (1973). Optimum stratification with ratio and regression method of estimation. *Annals of the Institute of Statistical Mathematics*, Vol. 25, pp. 627–633.

Singh, R., Kumar, M., Chaudhary, M. K., Kadilar, C., (2009). Improved Exponential estimator in Stratified Random Sampling. *Pakistan Journal of Statistics and Operation Research*, Vol. 5, pp. 67–82.

Singh, H. P., Chandra, P., Joarder, A. H., Singh, S., (2007). Family of estimators of mean, ratio and product of a finite population using random non-response. *Test*, Vol. 16, pp. 565–597.

Singh, G. N., Sharma, A. K., Bandyopadhyay, A., (2017). Effectual Variance Estimation Strategy in Two Occasions Successive Sampling in Presence of Random Non-Response. *Communications in Statistics-Theory & Methods*, Vol. 46, pp. 7201–7224.

Tracy, D. S., Singh, H. P., Singh, R., (1996). An alternative to the ratio-cum-product estimator in sample surveys. *Journal of Statistical Planning and Inference*, Vol. 53, pp. 375–387.

Appendix

We have generated three sets of independent random numbers of size N (N = 100), namely $x[k]$, y$[k]$and z$[k]$(k $= 1, 2, 3, \ldots,$ N) from a standard normal distribution as presented below.

The following algorithm is used to generate the population artificially:

1. Generate there random variables x1 and z1 and a which are normally distributed with mean 0, S.D. = 1 and which are of size 100.
2. Define N = 100.
3. Define rx1z1 = 0.75, Sx1 $= \sqrt{50}$(s.d. of x1), Sz1 $= \sqrt{40}$ (s.d. of z1), mx1$= 20$ (i.e. mean of x1), mz1 $= 25$ (i.e. mean of z1) . [Note: x1, z1 are temporary variables]
4. a = sz1*sz1*(1-(rx1z1^2))
5. for (j in 1:N)
   {
   x[j] = 20.0+(sx1*x1[j])
   z[j] = 25+(√a)*z1[j])+(rx1z1*sz1*x1[j])
   }
6. Take output of the variables x and z
7. Generate the variable y with ryx = 0.7 from the variable
8. Take output of the variable y
9. Repeat the steps 1 to 8 with different values of rx1z1 (step 3), which will generate different population for different values of the correlation coefficients.