Variance Estimation Procedure Using Scrambled Responses and Multi-Auxiliary Variables In Multi-Phase Sampling

Nadia Mushtaq

Department of Statistics, Forman Christian College University, Lahore, Pakistan
E-mail: nadiamushtaq@fccollege.edu.pk

Abstract

Variations in the population can be estimated by variance estimation. In this study, we consider variance estimation procedure using scrambled randomized response for sensitive variable using multi-auxiliary variables in multi-phase sampling. Under Noor-ul-Amin et al. (2018) RRT model, generalized exponential regression type estimator for case-1and case-2 are derived. A simulation study is presented to illustrate the application and computational details. It is observed that proposed model showed better results under both cases.

Keywords: Variance estimation, multi-auxiliary variables, scrambled randomized response.

1 Introduction

The problems encountered in behavioral and social sciences surveys are expecting the refusal or biasness for sensitive questions such as related to drug addiction, HIV/AIDS disease, gambling, sexual behavior etc. In such

situations, participants of the surveys may refuse to respond or give false answers. So, refusal to answers or misleading answers comprise the bias and non-sampling errors. Warner (1965) introduced RRT (Randomized Response Technique) to reduce response error problems for sensitive questions. Eichhorn and Hayre (1983) proposed multiplicative RRT for quantitative data. Furthermore, scrambled response models suggested by different authors such as: additive scrambling model (Himmelfarb and Edgel, 1980), subtractive scrambling model (Hussain, 2012), or mixture of additive and multiplicative scrambling models (Huang, 2010).

Variations are occurring in many practical situations such as environmental, genetical, economic studies etc. Variations in the populations can be estimated by variance estimation of population about the existence of variability in it and used for future surveys, predictions, or sample size determinations etc. For estimating variance by using of auxiliary was considered by authors: Das and Tripathi (1978), Isaki (1983), Singh and Joarder (1998), Shabbir and Gupta (2010), Asghar et al. (2014), Yadav et al. (2015) Yasmeen et al. (2018). Neyman (1938) introduced the concept of two-phase, where sample selection can be done in two-phases, in the first phase information can be collected for the auxiliary variable for large sample and in second phase information can be observed for study variable for relatively small sample.

As in multipurpose surveys we need to estimate population parameters for several variables such as in socio-economic surveys multiple variables may be size of houses, monthly incomes, number of unemployed persons etc. The objective of this paper is to present variance estimation procedure using scrambled randomized response for sensitive study variable using multi-auxiliary variables in multi-phase sampling. From available literature, we have noticed that no contribution on variance estimation for sensitive variable using multi-auxiliary variables in multi-phase sampling. In computational procedure we discuss some special cases of the estimators.

2 Variance Estimation Procedure in RRT

Let Y be the sensitive study variable $X_i(i = 1, 2, 3, \ldots, n)$ be n non-sensitive auxiliary variables which are correlated with Y in population. Let S be the scrambling variables independent of Y and $X_i(i = 1, 2, 3, \ldots, n)$. We are interested to estimate population variance of sensitive variables using multi-auxiliary variables under following cases.

The sampling strategy of the cases given as below:

Case 1: The information on all auxiliary variables are known (we use single phase sampling).

Case 2: The information on all auxiliary variables are unknown. As on the first phase large sample $n_1(n_1 \subset U)$ is drawn to observe only auxiliary variables. The second phase sample $n_2(n_2 \subset n_1)$ is drawn to observe sensitive variable of interest Y and all non-sensitive auxiliary variables $(n_2 < n_1)$.

Let $S^2_{x_i}$ be the variance of the population and $s^2_{x_{i(l)}}$ be the sample variances of the auxiliary variable $x_i(i = 1, 2, \ldots, n)$, at the l-th phases $(l = 1, 2, \ldots, q)$. The reported optional scrambled response for Y is given by $Z = Y + kS$, following Noor-ul-Amin et al. (2018).

Let $s^2_z = \frac{1}{n_{(l)}-1} \sum_{j=1}^{n_{(l)}} (z_j - \bar{z}_{(l)})^2$, $s^2_{x_{i(l)}} = \frac{1}{n_{(l)}-1} \sum_{i=1}^{n_{(l)}} (x_i - \bar{x}_{(l)})^2$ are the sample variances of the reported response and the auxiliary variables $x_i(i = 1, 2, \ldots, n)$ respectively for the l-th phases $(l = 1, 2, \ldots, q)$ having sample means $\bar{z}_{(l)} = \frac{1}{n_{(l)}} \sum_{j=1}^{n_{(l)}} z_j$ and $\bar{x}_{i(l)} = \frac{1}{n_{(l)}} \sum_{j=1}^{n_{(l)}} x_{ij}$.

For l-th phases $(l = 1, 2, \ldots, q)$, we define the following notations:

Let $s^2_{z(l)} = S^2_z(1 + e_{z(l)})$, $s^2_{x_{i(l)}} = S^2_{x_i}(1 + e_{x_{i(l)}})$

Such that

$$\theta_{(l)} = \frac{1}{n_{(l)}}, \ (l = 1, 2, \ldots, q)$$

$$\Lambda_{z(1\times1)} = [e_{z(l)}], \quad \Lambda_{x(1\times n)} = [e_{x_{1(l)}}, e_{x_{2(l)}}, \ldots, e_{x_{n(l)}}]$$

$$E(\Lambda'_z \Lambda_z) = \theta_l \sum_{z(1\times1)}, \quad E(\Lambda'_x \Lambda_x) = \theta_l \sum_{x(n\times n)},$$

$$E(\Lambda'_x \Lambda_z) = \theta_l \sum_{xz(n\times1)}.$$

Where

$$\sum_{x(2\times2)} = \begin{bmatrix} var(x_1) & cov(x_1 x_2) \\ cov(x_2 x_1) & var(x_2) \end{bmatrix}$$

Let S^2_z be the population variance and its unbiased variance given as:

$$t_o = S^2_z \quad \text{and} \quad var(t_o) = \theta_l S' S \sum_{z(1\times1)}. \tag{1}$$

2.1 Proposed Generalized Exponential Regression Type Estimator Using Randomized Response Technique (RRT) in Case 1

For estimating the variance of the population for sensitive variable using multi-auxiliary non-sensitive variables under case 1 given as:

$$t_z = [t_z]_{(1\times 1)}; \tag{2}$$

where

$$t_z = s^2_{z(l)} + \sum_{i=1}^{n} \beta_i (S^2_{x_i} - s^2_{x_i(l)}) exp \sum_{i=1}^{n} \left(\frac{S^2_{x_i} - s^2_{x_i(l)}}{S^2_{x_i} + s^2_{x_i(l)}} \right); \tag{3}$$

By solving (2), we have following expressions given below:

$$t_z = \left[S^2_z(1 + e_{z(l)}) + \sum_{i=1}^{n} \beta_i (S^2_{x_i} - S^2_{x_i}(1 + e_{x_{i(l)}})) \right]$$

$$\times exp \sum_{i=1}^{n} \left(\frac{S^2_{x_i} - S^2_{x_i}(1 + e_{x_{i(l)}})}{S^2_{x_i} - S^2_{x_i}(1 + e_{x_{i(l)}})} \right)$$

$$t_z = \left[S^2_z + S^2_z e_{z(l)} - \sum_{i=1}^{n} \beta_i S^2_{x_i} e_{x_{i(l)}} \right]$$

$$\times exp \sum_{i=1}^{n} \frac{-e_{x_{i(l)}}}{2} \left(1 + \frac{-e_{x_{i(l)}}}{2} \right)^{-1},$$

$$t_z = \left[S^2_z + S^2_z e_{z(l)} - \sum_{i=1}^{n} \beta_i S^2_{x_i} e_{x_{i(l)}} \right]$$

$$\times \left[1 + \sum_{i=1}^{n} \left\{ \frac{-e_{x_{i(l)}}}{2} \left(1 - \frac{e_{x_{i(l)}}}{2} + \frac{e^2_{x_{i(l)}}}{4} \pm \cdots \right) \right\} \right]$$

$$t_z = \left(S^2_z + S^2_z e_{z(l)} - \sum_{i=1}^{n} \beta_i S^2_{x_i} e_{x_{i(l)}} \right) \left(1 - \sum_{i=1}^{n} \frac{e_{x_{i(l)}}}{2} \right)$$

$$t_z = \left(S^2_z + S^2_z e_{z(l)} - \sum_{i=1}^{n} \beta_i S^2_{x_i} e_{x_{i(l)}} - S^2_{z_j} \sum_{i=1}^{n} \frac{e_{x_{i(l)}}}{2} \right)$$

$$t_z = \left(S_z^2 + S_z^2 e_{z(l)} - \frac{1}{2} \sum_{i=1}^{n} \left(2\beta_i \frac{S_{x_i}^2}{S_z^2} + 1 \right) e_{x_{i(l)}} \right)$$

$$t_z - S_z^2 = S_z^2 \left(e_{z(l)} - \frac{1}{2} \sum_{i=1}^{n} \left(2\beta_i \Psi_i + 1 \right) e_{x_{i(l)}} \right)$$

Where $\Psi_i = \frac{S_{x_i}^2}{S_z^2}$

$$t_z = [t_z]_{(1\times 1)} = \left[S_z^2 \left(e_{z(l)} - \frac{1}{2} \sum_{i=1}^{n} (2\beta_{ij}\Psi_i + 1) e_{x_{i(l)}} \right) \right]_{(1\times 1)} \tag{4}$$

For expectations, we proceed as,

$$\sum_{tz_{(1\times 1)}} = E(t_{z_{(1\times 1)}} - S_{(1\times 1)})'(t_{z_{(1\times 1)}} - S_{(1\times 1)})$$

$$\sum_{tz_{(1\times 1)}} = S_z^4 E \left(e_{z(l)(1\times 1)} - \frac{1}{2} e_{x(l)(1\times m)} \Phi_{(n\times 1)} \right)'$$

$$\left(e_{z(l)(1\times 1)} - \frac{1}{2} e_{x(l)(1\times n)} \Phi_{(n\times 1)} \right) \tag{5}$$

Where,
$$\Phi_{(n\times 1)} = (2\beta_i \Psi_i + 1)_{(n\times 1)}$$

By applying expectations, we have the following results

$$\sum_{tz_{(1\times 1)}} = S'S\theta_l \left(\sum_{z(1\times 1)} - \frac{1}{2} \sum_{zx(1\times n)} \Phi_{(n\times 1)} - \frac{1}{2} \Phi'_{(1\times n)} \sum_{zx(n\times 1)} \right.$$

$$\left. + \frac{1}{4} \Phi'_{(1\times n)} \sum_{x(n\times n)} \Phi_{(n\times 1)} \right) \tag{6}$$

We differentiate the (5) w.r.t Φ and get the optimum value of Φ as,

$$\Phi_{opt(n\times 1)} = 2 \frac{\sum_{zx_{(n\times 1)}}}{\sum_{x_{(n\times n)}}} \tag{7}$$

Using (6) in (5), we get the minimum value of MSE of as:

$$min \sum_{tz_{(1\times1)}} = S'S\theta_l \left(\sum_{z(1\times1)} - \sum_{zx(1\times n)} \sum_{x(n\times n)}^{-1} \sum_{zx(n\times1)} \right) \qquad (8)$$

Remarks 1:

1. If we replace $l = 1$ in (2), so it is single phase sampling using q auxiliary variables given as:

$$t_{z1} = s_z^2 + \sum_{i=1}^{n} \beta_i(S_{x_i}^2 - s_{x_i}^2)exp \sum_{i=1}^{n} \left(\frac{S_{x_i}^2 - s_{x_i}^2}{S_{x_i}^2 + s_{x_i}^2} \right) \qquad (9)$$

2. For single auxiliary variables we replace $i = 1$ and $l = 1$ in (2), we have single phase sampling as:

$$t_{z2} = s_z^2 + \beta(S_x^2 - s_x^2)exp \left(\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2} \right) \qquad (10)$$

3. If we replace $l = 1$ in (2), and for two auxiliary variables in single phase sampling given following expressions:

$$t_{z3} = s_z^2 + \sum_{i=1}^{2} \beta_i(S_{x_i}^2 - s_{x_i}^2)exp \sum_{i=1}^{2} \left(\frac{S_{x_i}^2 - s_{x_i}^2}{S_{x_i}^2 + s_{x_i}^2} \right) \qquad (11)$$

2.2 Generalized Exponential Regression Type Estimator Using RRT in Case 2

Let $s_{y_{(l+1)}}^2$ be the sample variance of the sensitive study variables is selected at the $(l + 1)$th phases. Also $s_{x_{i(l)}}^2$ and $s_{x_{i(l+1)}}^2$ be the sample variances of the auxiliary variable $x_i(i = 1, 2, \ldots, n)$, at the l-th and $(l + 1)$-th phases of size $n_{(l)}$ and $n_{(l+1)}$ respectively. The population variance $S_{x_i}^2$ of all multi-auxiliary variables is unknown.

$$t_s = [t_s]_{(1\times1)}; \qquad (12)$$

Where

$$t_s = s_{z(l+1)}^2 + \sum_{i=1}^{n} \beta_i(s_{x_{i(l)}}^2 - s_{x_i(l+1)}^2)exp \sum_{i=1}^{n} \left(\frac{s_{x_{i(l)}}^2 - s_{x_{i(l+1)}}^2}{s_{x_{i(l)}}^2 + s_{x_{i(l+1)}}^2} \right); \qquad (13)$$

By solving (12), we have the following expressions:

$$t_s = \left[S_z^2(1 + e_{z(l+1)}) + \sum_{i=1}^{n} \beta_i (S_{x_i}^2 (e_{x_{i(l)}} - e_{x_{i(l+1)}})) \right]$$

$$\times \exp \sum_{i=1}^{n} \left(\frac{S_{x_i}^2 (e_{x_{i(l)}} - e_{x_{i(l+1)}})}{2S_{x_i}^2 + S_{x_i}^2 (e_{x_{i(l)}} + e_{x_{i(l+1)}})} \right)$$

$$t_s = \left[S_z^2 + S_z^2 e_{z(l+1)} - \sum_{i=1}^{n} \beta_i S_{x_i}^2 (e_{x_{i(l)}} - e_{x_{i(l+1)}}) \right]$$

$$\times \exp \sum_{i=1}^{n} \frac{(e_{x_{i(l)}} - e_{x_{i(l+1)}})}{2} \left(1 + \frac{(e_{x_{i(l)}} - e_{x_{i(l+1)}})}{2} \right)^{-1},$$

$$t_s = \left[S_z^2 + S_z^2 e_{z(l+1)} - \sum_{i=1}^{n} \beta_i S_{x_i}^2 (e_{x_{i(l)}} - e_{x_{i(l+1)}}) \right]$$

$$\times \left[1 + \sum_{i=1}^{n} \left\{ \frac{(e_{x_{i(l)}} - e_{x_{i(l+1)}})}{2} \left(1 - \frac{(e_{x_{i(l)}} - e_{x_{i(l+1)}})}{2} \pm \cdots \right) \right\} \right]$$

$$t_s = \left(S_z^2 + S_z^2 e_{z(l+1)} - \sum_{i=1}^{n} \beta_i S_{x_i}^2 (e_{x_{i(l)}} - e_{x_{i(l+1)}}) \right)$$

$$\times \left(1 - \sum_{i=1}^{n} \frac{(e_{x_{i(l)}} - e_{x_{i(l+1)}})}{2} \right)$$

$$t_s = \left(S_z^2 + S_z^2 e_{z(l+1)} - \sum_{i=1}^{n} \beta_i S_{x_i}^2 (e_{x_{i(l)}} - e_{x_{i(l+1)}}) \right.$$

$$\left. -S_{z_j}^2 \sum_{i=1}^{n} \frac{(e_{x_{i(l)}} - e_{x_{i(l+1)}})}{2} \right)$$

$$t_s = \left(S_z^2 + S_z^2 e_{z(l+1)} - \frac{1}{2} \sum_{i=1}^{n} \left(2\beta_i \frac{S_{x_i}^2}{S_z^2} + 1 \right) (e_{x_{i(l)}} - e_{x_{i(l+1)}}) \right)$$

$$t_s - S_z^2 = S_z^2 \left(e_{z(l+1)} - \frac{1}{2} \sum_{i=1}^{n} (2\beta_i \Psi_{ij} + 1) (e_{x_{i(l)}} - e_{x_{i(l+1)}}) \right)$$

Where $\Psi_i = \frac{S_{x_i}^2}{S_z^2}$

$$t_s = [t_s]_{(1\times1)} = \left[S_z^2 \left(e_{z_{(l+1)}} - \frac{1}{2} \sum_{i=1}^{n} (2\beta_i \Psi_i + 1)(e_{x_{i(l)}} - e_{x_{i(l+1)}}) \right) \right]_{(1\times1)} \tag{14}$$

For expectations, we proceed as:

$$\sum_{ts_{(1\times1)}} = E(t_{s_{(1\times1)}} - S_{(1\times1)})'(t_{s_{(1\times1)}} - S_{(1\times1)})$$

$$\sum_{ts_{(1\times1)}} = S_z^4 E\left(e_{z_{(l+1)(1\times1)}} - \frac{1}{2}(e_{x_{(l)(1\times n)}} - e_{x_{(l+1)(1\times n)}})\Phi_{(n\times1)} \right)'$$

$$\times \left(e_{z_{(l+1)(1\times1)}} - \frac{1}{2}(e_{x_{(l)(1\times n)}} - e_{x_{(l+1)(1\times n)}})\Phi_{(n\times1)} \right) \tag{15}$$

Where,

$$\Phi_{(n\times1)} = (2\beta_i \Psi_i + 1)_{(n\times1)}$$

By applying expectations on (14), we have the following results

$$\sum_{ts_{(1\times1)}} = S'S \left(\begin{array}{c} \theta_{(l+1)} \displaystyle\sum_{z(1\times1)} -(\theta_{(l)} - \theta_{(l+1)})\frac{1}{2} \displaystyle\sum_{zx(1\times n)} \Phi_{(n\times1)} \\[2ex] -(\theta_{(l)} - \theta_{(l+1)})\frac{1}{2}\Phi'_{(1\times n)} \displaystyle\sum_{zx(1\times n)} \\[2ex] +(\theta_{(l)} - \theta_{(l+1)})\frac{1}{4}\Phi'_{(1\times n)} \displaystyle\sum_{x(n\times n)} \Phi_{(n\times1)} \end{array} \right) \tag{16}$$

We differentiate (15) w.r.t Φ and get the optimum value of Φ as,

$$\Phi_{opt(n\times1)} = 2\frac{\sum_{zx_{(n\times1)}}}{\sum_{x_{(n\times n)}}} \tag{17}$$

Using (16) in (15), we get the minimum value of MSE ofas:

$$min \sum_{ts_{(1\times1)}} = S'S \left(\theta_{(l+1)} \sum_{z(1\times1)} -(\theta_{(l)} - \theta_{(l+1)}) \sum_{zx(1\times n)} \sum_{x(n\times n)}^{-1} \sum_{zx(n\times1)} \right) \tag{18}$$

Remarks 2:

1. If we replace $l = 1$ in (12), so it is two-phase sampling using q multi-auxiliary variables:

$$t_{s1} = s^2_{z(2)} + \sum_{i=1}^{n} \beta_i (s^2_{x_{i(1)}} - s^2_{x_{i(2)}}) exp \sum_{i=1}^{n} \left(\frac{s^2_{x_{i(1)}} - s^2_{x_{i(2)}}}{s^2_{x_{i(1)}} + s^2_{x_{i(2)}}} \right) \quad (19)$$

2. We replace $i = 1$ and $l = 1$ in (12) so it is two-phase sampling for single auxiliary variable:

$$t_{s2} = s^2_{z(2)} + \beta (s^2_{x_{(1)}} - s^2_{x_{(2)}}) exp \left(\frac{s^2_{x_{(1)}} - s^2_{x_{(2)}}}{s^2_{x_{(1)}} + s^2_{x_{(2)}}} \right) \quad (20)$$

3. If we replace $l = 1$ in (12), it is two auxiliary variables two-phase sampling given following expressions:

$$t_{s3} = s^2_{z(2)} + \sum_{i=1}^{2} \beta_i (s^2_{x_{i(1)}} - s^2_{x_{i(2)}}) exp \sum_{i=1}^{2} \left(\frac{s^2_{x_{i(1)}} - s^2_{x_{i(2)}}}{s^2_{x_{i(1)}} + s^2_{x_{i(2)}}} \right) \quad (21)$$

3 Computational Procedure

We consider computational and application of proposed estimators by using two type of simulation studies given below:

3.1 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 $S \sim N(0, 0.1\sigma_x)$ and $Z = Y + kS, k = -1, -0.5, 0.5, 1$.

$$\text{Mean of } [Y, X] \text{ given as } \mu = [2, 2]$$

Population 1:

$$\sum = \begin{bmatrix} 9 & 1.9 \\ 1.9 & 4 \end{bmatrix}, \quad \rho_{XY} = 0.3209$$

Population 2:

$$\sum = \begin{bmatrix} 10 & 3 \\ 3 & 2 \end{bmatrix}, \quad \rho_{XY} = 0.6746;$$

Population 3:

$$\sum = \begin{bmatrix} 6 & 3 \\ 3 & 2 \end{bmatrix}, \quad \rho_{XY} = 0.8684;$$

For each population we considered three sample sizes for first phase: $n = 250, 500$ and for second phase given as: $n1 = 100, 200$ respectively.

3.2 In this computational procedure, we consider single and two auxiliary variables for both cases given as:

$Y_i = RX_i + e_i$, where $e_i \sim N(0, 1)$ and   (Sensitive Study Variable)
$R = 1.5$.
$X_i \sim G(a, b)$, where $a = 2$ and $b = 3$.   (Auxiliary Variable)
$w_i \sim G(a, b)$, where $a = 2$ and $b = 10$.   (Another auxiliary variable)
$S \sim B(\alpha, \beta)$, where $\alpha = 6.5$ and   (Scrambling Variable)
$\beta = 0.5$.
$Z_i = Y_i + kS_i$, $i = 1, 2, 3, \ldots, n$,   (Randomized reported response)
$k = -1, -0.5, 0.5, 1$.
$N = 3000$,

We computed percent relative efficiencies of the proposed variance estimators regarding to t_0 for case-I and case-2 given as:

$$PRE = \frac{\sum_{t=1}^{5000} (t_0 - S_z^2)^2}{\sum_{t=1}^{5000} (t_m - S_z^2)^2} \times 100, \ m = z2, z3, s2, s3.$$

The results of the simulation study are as in Tables 1 and 2.

4 Main Findings

The results of the estimators for case-1 and case-2 are presented in Tables 1–2 for PRE. Major findings of the estimators are:

  i. PREs are higher than usual variance estimator, which are shown in Tables 1 and2 respectively, which show the efficiency of the proposed estimator estimators.
 ii. From Table 1, it is observed that proposed estimators have the highest efficiency when the value of correlation coefficient is highest.
iii. It is observed that the value of k varies from -1 to $+1$, the value of PRE is increase of the proposed estimators in both cases.

Table 1   Percent relative efficiencies of the proposed estimators for Section 3.1

N =1000		k		Case-I, PRE			Case-2, PRE		
			n	t_o	t_{z2}	$n1$	t_o	t_{s2}	
Pop. 1	$\rho_{XY} = 0.3209$	-1	500	100	100.88	200	100	100.35	
			250	100	101.35	100	100	100.96	
		-0.5	500	100	100.64	200	100	100.67	
			250	100	101.94	100	100	100.49	
		0.5	500	100	100.98	200	100	100.32	
			250	100	100.25	100	100	100.29	
		1	500	100	100.51	200	100	100.11	
			250	100	100.58	100	100	100.01	
Pop. 2	$\rho_{XY} = 0.6746$	-1	500	100	118.59	200	100	111.86	
			250	100	119.72	100	100	109.94	
		-0.5	500	100	121.78	200	100	110.25	
			250	100	128.28	100	100	115.20	
		0.5	500	100	122.51	200	100	126.78	
			250	100	124.03	100	100	116.63	
		1	500	100	127.95	200	100	112.84	
			250	100	123.67	100	100	128.88	
Pop. 3	$\rho_{XY} = 0.8684$	-1	500	100	228.51	200	100	150.35	
			250	100	211.63	100	100	145.95	
		-0.5	500	100	212.78	200	100	147.56	
			250	100	214.83	100	100	148.37	
		0.5	500	100	236.52	200	100	151.06	
			250	100	216.48	100	100	144.06	
		1	500	100	213.12	200	100	146.64	
			250	100	202.28	100	100	144.47	

iv. In Table 1, we consider single and two-phase sampling using single auxiliary variable. So, we also noticed that the value of sample size increases efficiency of the estimators also increases.

v. In Table 2, we consider Section 3.2 and compute results for single and two auxiliary variables for single and two-phase sampling. So, by increasing sample size efficiency of the estimators also increases.

Table 2   PRE of the proposed estimators in case-1 and case-2 with respect to t_o for Section 3.2

		PRE for case-1				PRE for case-2		
k	n	t_o	t_{z2}	t_{z3}	$n1$	t_o	t_{s2}	t_{s3}
-1	250	100	127.43	114.77	100	100	115.59	120.09
	500	100	124.51	152.49	200	100	115.20	103.37
-0.5	250	100	128.69	117.27	100	100	114.97	111.31
	500	100	128.01	117.56	200	100	116.19	110.84
0.5	250	100	139.96	146.93	100	100	121.81	135.71
	500	100	123.28	133.87	200	100	112.52	104.30
1	250	100	113.89	131.99	100	100	113.66	115.00
	500	100	137.66	148.05	200	100	119.49	116.98

vi. From Remark 1, it is shown that t_{z2} and t_{z3} are the special cases of t_z that are for single and two-auxiliary variables respectively. From Table 2, case-1 the efficiency for proposed two-auxiliary variables estimator increased.

vii. From Table 2, we noticed that when the value of k increases the efficiency of the proposed estimators is increased.

From above discussions, we can conclude that the proposed estimators performed better as compared to population variance. So computational procedure supports the theoretical findings in both cases in randomized response technique.

5 Concluding Remarks

The objective of the survey sampling techniques is to estimate population characteristics with precision, and it can be increased by using proper methodology. In this study, we have suggested variance estimation procedure using scrambled randomized response and multi-auxiliary variables for multi-phase sampling. The proposed estimators are more efficient than usual variance estimator which is shown in Tables 1 and 2 for case-1 and case-II. We use different sample sizes and value of k lie between $(-1,1)$ for different scramble responses for estimating MSEs of the proposed estimators. From computational procedure, it can be shown that proposed estimators are more efficient and helpful in estimation of variance for sensitive variable using RRT.

References

[1] Asghar, A., Sanaullah, A. and Hanif, M. (2014). Generalized exponential type estimator for population variance in survey sampling. Revista Colombiana de Estadística, 37(1), 213–224.

[2] Das, A.K. and Tripathi, T.P. (1978): Use of auxiliary information in estimating the finite population variance. Sankhya, 40, C, 139–148.

[3] Eichhorn, B.H. and Hayre, L.S. (1983). Scrambled randomized response methods for obtaining sensitive quantitative data. Journal of Statistical Planning and Inference, 7(4), 307–316.

[4] Himmelfarb, S. and Edgell, S.E. (1980). Additive constants model: A randomized response technique for eliminating evasiveness to quantitative response questions. Psychological Bulletin, 87(3), 525–530.

[5] Huang, K.C. (2010). Unbiased estimators of mean, variance and sensitivity level for quantitative characteristics in finite population sampling. Metrika, 71(3), 341–352.

[6] Hussain, Z. (2012). Improvement of the Gupta and Thornton scrambling model through double use of randomization device. International Journal of Academic Research in Business and Social Sciences, 2(6), 91–97.

[7] Isaki, C.T. (1983). Variance estimation using auxiliary information. Journal of the American Statistical Association, 78(381), 117–123.

[8] Shabbir, J., Gupta, S. (2010) Some estimators of finite population variance of stratifed sample mean, Communication in Statistics – Theory and Methods 39, 3001–3008.

[9] Singh, S., Joarder, A.H. (1998). Estimation of finite population variance using random nonresponse in survey sampling, Metrika 47, 241–249, 1998.

[10] Neyman, J. (1938) Contributions to the Theory of Sampling Human Populations. Journal of the American Statistical Association, 33, 101–116.

[11] Noor-ul-Amin, M., Mushtaq, N., and Hanif, M. (2018). Estimation of mean using generalized optional scrambled responses in the presence of nonsensitive auxiliary variable, Journal of Statistics and Management Systems. 21(2), 287–304.

[12] Warner, S.L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 60(309), 63–69.

[13] 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), 9219–9226.

[14] Yasmeen, U., Noor-ul-Amin, M., Hanif, M. (2018). Exponential Estimators of Finite Population Variance Using Transformed Auxiliary Variables. Proceedings of the National Academy of Sciences, India Section A: Physical. https://doi.org/10.1007/s4001

Biographies

Nadia Mushtaq received her MSc and MPhil degrees in Statistics from Quaid-i-Azam University Islamabad, Pakistan and PhD degree in Statistics from National College of Business administration & Economics Lahore, Pakistan. Dr. Mushtaq is currently working as an Assistant Professor at Forman Christian College Lahore, Pakistan. She has more than fifteen years of teaching/research experience at university. Her research interests include sampling techniques, Time series analysis and statistical data analysis using different statistical software such as: SPSS, SAS, Minitab, and R-Language. She published ten research papers in national and international Journals.