STATISTICS IN TRANSITION new series, September 2023
Vol. 24, No. 4, pp. 123–138, https://doi.org/10.59170/stattrans-2023-055
Received – 09.01.2021; accepted – 17.05.2022

Advances in estimation by the item sum technique in two move successive sampling

Kumari Priyanka[1], **Pidugu Trisandhya**[2]

Abstract

The present article proposes an estimator using the Item Sum Technique (IST) for the estimation of dynamic sensitive population mean using non-sensitive auxiliary information in the two-move successive sampling. Properties of the proposed IST estimator have been analysed. Possible allocation designs for allocating long-list and short-list samples pertaining to the IST have been elaborated. The comparison between various allocation designs has been carried out. Theoretical considerations have been integrated with numerical as well as simulation studies to show the working version of the proposed IST estimators in the two-move successive sampling.

Key words: Sensitive variable, Successive moves, Population mean, Variance, Mean squared error, Optimum matching fraction.

1. Introduction

In many social surveys, data gathering on sensitive issues such as incidence of domestic violence, drug addiction, eve teasing, negligence of government rules, duration of suffering from AIDS, use of harmful pesticides in agriculture, sexual behaviour, etc., are a challenging task in the present scenario. Hence, in such circumstances, many respondents either refuse to participate or give false or evasive responses in social surveys. Therefore, to overcome mis-reporting on sensitive issues and to protect respondents confidentiality, the Randomized Response (RR) technique, the Scrambled Response (SR) technique, Item Count Technique (ICT), etc., may be used.

The RR technique was first initiated by Warner (1965) which was followed by Horvitz *et al.* (1967), Greenberg *et al.* (1971), Franklin (1989), Arcos *et al.* (2015), etc. However, SR technique was introduced by Pollock and Bek (1976) and was further explored by Eichhorn and Hayre (1983), Diana and Perri (2010, 2011), Perri and Diana (2013), etc. The ICT is used in surveys that require the study of qualitative sensitive variable and was introduced by Miller (1984). Subsequently the literature addressing ICT was enhanced by Droitcour *et al.* (1991), Wimbush and Dalton (1997), LaBrie and Earleywine (2000), Rayburn *et al.* (2003), and Tsuchiya *et al.* (2007), Holbrook and Krosnick (2010)etc. For estimating quantitative sensitive variable, the concept of ICT was generalized by Chaudhuri and

[1]Department of Mathematics, Shivaji College(University of Delhi), New Delhi-110027, India. E-mail: priyanka.ism@gmail.com. ORCID:https://orcid.org/0000-0002-6241-1708.
[2]Department of Applied Sciences, Bharati Vidyapeeth's College of Engineering, New Delhi-110063, India. E-mail: trisandhya.09@gmail.com. ORCID:https://orcid.org//0000-0002-6671-6541.

Christofides (2013). Trappmann *et al.* (2014) named this technique IST. Perri *et al.* (2018) discussed the possibility of optimal sample size allocation in IST.

As the issues are sensitive, single time survey will not be sufficient, one need to monitor continuously over a period of time. So, to observe the situation at different point of time, a statistical tool generally recommended in the literature is 'Successive Sampling'. In order to address the dynamic sensitive variable Arnab and Singh (2013), Yu *et al.* (2015), Priyanka *et al.* (2018, 2019), Priyanka and Trisandhya (2019), etc. added valuable literature. To handle sensitive issues all these researchers dealt with SR technique or RR technique on two move successive sampling. As the IST is now emerging as an alternative technique to deal with sensitive issues, in the present article an attempt has been made to apply IST in successive sampling framework to estimate a sensitive population mean. The concept of linear estimators has been used under IST set-up on successive move which is a methodological advancement to the theory. Hence, IST class of estimator has been proposed and studied under general allocation design advocated by Trappmann *et al.* (2014) as well as optimum allocation design suggested by Perri *et al.* (2018). Properties of the proposed class of estimator have been discussed in detail. Empirical as well as simulation studies have been incorporated to justify the requirement and application of the proposed estimator using a natural population. The proposed estimator has also been compared with respect to direct version of the estimator to show the amount of loss incurred due to sensitivity management of the variable under study by IST.

2. Outline of the Item Sum Technique(IST)

A promising indirect questioning technique, called Item Count Technique [Miller (1984)] is proved to be a very useful technique to estimate the qualitative sensitive variable. In this technique each respondent is provided with a list of items describing behaviors and asked to count and report in how many he or she is engaged in and not in which ones. For example, a random subsample (say subsample A) is provided with three-item list that includes the socially disproved behaviour item; the remaining respondents (say subsample B) are given an identical (two-item) list from which the disapproved item has been removed. By comparing responses from the two subsamples, an estimate of sensitive behavior has been obtained.

This method of estimating qualitative sensitive characteristics was generalized by Chaudhuri and Christofides (2013) to work for estimation of quantitative sensitive variates. Later Trappmann *et al.*(2014) explored it and named the technique as Item Sum Technique, which is described as follows.

Two random sub-samples (say s_1 and s_2) are drawn from a random sample (say s). The respondent belonging to sub-sample s_1 is given a long list (LL) of items containing sensitive question and a number of non-sensitive questions. However, the respondents in sub-sample s_2 are confronted with a short list (SL) of items containing only the same non-sensitive questions present in the long list (LL). In both the sub-samples, the respondents are asked to respond only the total score of all the items given to them, without revealing the individual scores for the items. Finally, the mean difference of answers between the samples s_1 and s_2 is used as an unbiased estimator of the population mean of sensitive variable. The pivotal point in IST is how to split a single sample(s) into two-samples (s_1 and s_2). Trappmann *et al.* (2014), allocated equal number of units in each sub-sample irrespective of the variation of items in the *LL* and *SL*. Let us name this allocation design 'General Allocation' for further use. However, Perri *et al.* (2018) advocated the concept of 'Optimum allocation design' for allocating units in two sub-samples (s_1 and s_2) instead of assigning equal number of units to both the sub-samples. In the next section the IST set-up is modified to be applied in successive sampling to estimate population mean of dynamic sensitive variable.

3. Survey Design

Let us consider a finite population U of size N units for sampling over two successive moves. Let $y_1(y_2)$ denote quantitative sensitive variable at first (second) move respectively. Similarly, let x and t be non-sensitive auxiliary variables available at both the moves. Let $\bar{Y}_1, \bar{Y}_2$, $\bar{X}$, and $\bar{T}$ be the population mean of y_1, y_2, x, and t respectively. The aim is to estimate the population mean of sensitive variable y_2 at current move under IST set-up for two move successive sampling. The sampling design under IST frame work is as follows.

At the first move, a sample s_n of size n is drawn using simple random sampling without replacement(SRSWOR). Two independent samples are drawn at the second move by considering the partial overlap case, one is matched sample s_m of size $m = n(1-\mu) = n\lambda$ drawn as sub sample from the sample s_n and other is the sample s_u which is of size $u = n-m = n\mu$ drawn afresh at the second move. Let s_{u^*} denote the left out units from s_n after drawing the sub sample s_m. Moreover all the available samples s_{u^*}, s_m and s_u are split in to two sub samples called *LL* sample and *SL* sample respectively for embedding the IST set-up in two move successive sampling, which is given in Table 1:

Table 1: *LL* and *SL* Sample on two moves

Move	**Sample**	*LL*− **Sample**	*SL*− **Sample**
I	s_{u^*}	$s_{u_1^*}$	$s_{u_2^*}$
	s_m	s_{m_1}	s_{m_2}
II	s_m	s_{m_1}	s_{m_2}
	s_u	s_{u_1}	s_{u_2}

Note: s_m denotes matched sample and s_u denotes unmatched sample at current (second) move

The response received and the corresponding IST estimate on two moves under IST set-up are presented in Table 2.

Table 2: Response received and IST estimate

Move	Sample size	Response received	IST estimate
I	u^*	$z_{1i} = \begin{cases} y_{1i} + t_i & \text{if } i \varepsilon s_{u_1^*} \\ t_i & \text{if } i \varepsilon s_{u_2^*} \end{cases}$	$\hat{\bar{y}}_{1u^*} = \bar{z}_{1u_1^*} - \bar{t}_{u_2^*}$
	m	$z_{1i} = \begin{cases} y_{1i} + t_i & \text{if } i \varepsilon s_{m_1} \\ t_i & \text{if } i \varepsilon s_{m_2} \end{cases}$	$\hat{\bar{y}}_{1m} = \bar{z}_{1m_1} - \bar{t}_{m_2}$
II	m	$z_{2i} = \begin{cases} y_{2i} + t_i & \text{if } i \varepsilon s_{m_1} \\ t_i & \text{if } i \varepsilon s_{m_2} \end{cases}$	$\hat{\bar{y}}_{2m} = \bar{z}_{2m_1} - \bar{t}_{m_2}$
	u	$z_{2i} = \begin{cases} y_{2i} + t_i & \text{if } i \varepsilon s_{u_1} \\ t_i & \text{if } i \varepsilon s_{u_2} \end{cases}$	$\hat{\bar{y}}_{2u} = \bar{z}_{2u_1} - \bar{t}_{u_2}$

Note: z_{ji} ; $j = 1, 2$ denote the observed response at first and second move respectively and $\bar{z}_{1j}$; $j \in \{u_1^*, m_1\}, \bar{z}_{2j}$; $j \in \{m_1, u_1\}$ and $\bar{t}_k$; $k \in \{u_2^*, m_2, u_2\}$ are the sample means based on sample size j and k.

4. Proposed class of IST Estimators

Inspired by the classic work of Patterson (1950), who considered a general linear unbiased estimator of population mean at the current move, we intend to propose an estimator for estimation of sensitive population mean at the current move in IST set-up using all the information available at the current move.

In sampling theory, the role of additional auxiliary variable is well known and its availability and use in estimation procedures can do wonders and enhance the results to a great extent. Hence, in IST set-up, the availability of additional auxiliary variable has been embedded and class of IST estimator has been proposed to estimate sensitive population mean at current move as under:

$$\mathbb{T} = \zeta_1 \hat{\bar{y}}^*_{1u^*} + \zeta_2 \hat{\bar{y}}^*_{1m} + \zeta_3 \hat{\bar{y}}^*_{2m} + \zeta_4 \hat{\bar{y}}^*_{2u}, \tag{1}$$

where, the constants ζ_j ; $j = 1, 2, 3,$ and 4 are to be chosen suitably and $\hat{\bar{y}}^*_{2u} = \hat{\bar{z}}^*_{2u} - \hat{\bar{t}}^*_u$ with $\hat{\bar{z}}^*_{2u} = g_1(\bar{z}_{2u_1}, \bar{x}_{u_1})$, and $\hat{\bar{t}}^*_u = h_1(\bar{t}_{u_2}, \bar{x}_{u_2})$,
$\hat{\bar{y}}^*_{2m} = \hat{\bar{z}}^*_{2m} - \hat{\bar{t}}^*_m$ with $\hat{\bar{z}}^*_{2m} = g_2(\bar{z}_{2m_1}, \bar{x}_{m_1})$, and $\hat{\bar{t}}^*_m = h_2(\bar{t}_{m_2}, \bar{x}_{m_2})$,
$\hat{\bar{y}}^*_{1m} = \hat{\bar{z}}^*_{1m} - \hat{\bar{t}}^*_m$ with $\hat{\bar{z}}^*_{1m} = g_3(\bar{z}_{1m_1}, \bar{x}_{m_1})$,
$\hat{\bar{y}}^*_{1u^*} = \hat{\bar{z}}^*_{1u^*} - \hat{\bar{t}}^*_{u^*}$ with $\hat{\bar{z}}^*_{1u^*} = g_4(\bar{z}_{1u_1^*}, \bar{x}_{u_1^*})$, and $\hat{\bar{t}}^*_{u^*} = h_3(\bar{t}_{u_2^*}, \bar{x}_{u_2^*})$.

Following Srivastava and Jhajj (1980) and Tracy *et al.* (1996), $g_1(\bar{z}_{2u_1}, \bar{x}_{u_1})$ is assumed as a function of $\bar{z}_{2u_1}$ and $\bar{x}_{u_1}$ such that:

(i) The point $(\bar{z}_{2u_1}, \bar{x}_{u_1})$ assumes the value in a closed convex subset $\mathbb{R}^2$ of two dimensional real space containing the point $(\bar{Z}_2, \bar{X})$.

(ii) The function g_1 $(\bar{z}_{2u_1}, \bar{x}_{u_1})$ is continuous and bounded in $\mathbb{R}^2$.

(iii) $g_1(\bar{Z}_2, \bar{X}) = \bar{Z}_2$.

(iv) The first and second order partial derivatives of $g_1\ (\bar{z}_{2u_1},\ \bar{x}_{u_1})$ exist and are continuous and bounded in $\mathbb{R}^2$.

The similar regularity conditions holds for $g_2, g_3, g_4,\ h_1,\ h_2,$ and h_3 respectively as that of g_1.

5. Properties of proposed class of IST estimator

Since the proposed IST estimator $\mathbb{T}$ has to be linear Unbiased Estimator, therefore, following Garcia and Artes (2002), we consider the following assumptions:

$$E(\hat{\bar{y}}^*_{2u}) = E(\hat{\bar{y}}^*_{2m}) \cong \bar{Y}_2, \tag{2}$$

$$E(\hat{\bar{y}}^*_{1u^*}) = E(\hat{\bar{y}}^*_{1m}) \cong \bar{Y}_1. \tag{3}$$

Now, using the results in equations (2) and (3) into the expression for the proposed estimator in equation (1), we have

$$E(\mathbb{T}) = (\zeta_1 + \zeta_2)\bar{Y}_1 + (\zeta_3 + \zeta_4)\bar{Y}_2. \tag{4}$$

In order to satisfy the assumption in equation (2), we have the following conditions:

$$\zeta_1 + \zeta_2 = 0 \text{ and } \zeta_3 + \zeta_4 = 1. \tag{5}$$

Now, using the conditions in equation (5), the final structure of unbiased IST estimator for estimating the sensitive population mean at current move is given as:

$$\mathbb{T} = \zeta_1(\hat{\bar{y}}^*_{1u^*} - \hat{\bar{y}}^*_{1m}) + \zeta_3\hat{\bar{y}}^*_{2m} + (1 - \zeta_3)\hat{\bar{y}}^*_{2u}. \tag{6}$$

Following Mukhopadhyay (2014), as the estimators $\hat{\bar{y}}^*_{1u^*}$, and $\hat{\bar{y}}^*_{1m}$ are based on two independent samples u^* and m respectively, so $\text{Cov}(\hat{\bar{y}}^*_{1u^*},\ \hat{\bar{y}}^*_{1m}) = 0$. Similarly

$$\text{Cov}(\hat{\bar{y}}^*_{1u^*},\ \hat{\bar{y}}^*_{2m}) = \text{Cov}(\hat{\bar{y}}^*_{1u^*},\ \hat{\bar{y}}^*_{2u}) = \text{Cov}(\hat{\bar{y}}^*_{1m},\ \hat{\bar{y}}^*_{2u}) = \text{Cov}(\hat{\bar{y}}^*_{2m},\ \hat{\bar{y}}^*_{2u}) = 0. \tag{7}$$

Also, $\hat{\bar{z}}^*_{1u^*}$ is based on *LL* sample and $\hat{\bar{t}}^*_{u^*}$ is based on corresponding *SL* sample, therefore $\text{Cov}(\hat{\bar{z}}^*_{1u^*},\ \hat{\bar{t}}^*_{u^*}) = 0$. Similarly,

$$\text{Cov}(\hat{\bar{z}}^*_{2u},\ \hat{\bar{t}}^*_{u}) = \text{Cov}(\hat{\bar{z}}^*_{1m},\ \hat{\bar{t}}^*_{u}) = \text{Cov}(\hat{\bar{z}}^*_{2m},\ \hat{\bar{t}}^*_{u}) = 0.$$

Properties of the proposed IST estimator are discussed under above conditions, and the following assumptions:
$\bar{z}_{2u_1} = \bar{Z}_2\ (1+e_0)$, $\bar{x}_{u_1} = \bar{X}\ (1+e_1)$, $\bar{t}_{u_2} = \bar{T}\ (1+e_2)$, $\bar{x}_{u_2} = \bar{X}\ (1+e_3)$, $\bar{z}_{2m_1} = \bar{Z}_2\ (1+e_4)$, $\bar{x}_{m_1} = \bar{X}\ (1+e_5)$, $\bar{t}_{m_2} = \bar{T}\ (1+e_6)$, $\bar{x}_{m_2} = \bar{X}\ (1+e_7)$, $\bar{z}_{1m_1} = \bar{Z}_1\ (1+e_8)$, $\bar{z}_{1u_1^*} = \bar{Z}_1\ (1+e_9)$, $\bar{x}_{u_1^*} = \bar{X}\ (1+e_{10})$, $\bar{t}_{u_2^*} = \bar{T}\ (1+e_{11})$, $\bar{x}_{u_2^*} = \bar{X}\ (1+e_{12})$, such that, $E(e_i) = 0$; $|e_i| < 1$ where, $i = 0,\ 1,\ 2,\ 3,\ \ldots,\ 12$.

5.1. General and optimum allocations on two moves

As discussed in Section (2), under the general allocation design, the following allocation will be applicable to *LL* and *SL* samples on two moves:
$u_1^* = u_2^* = u_1 = u_2 = \frac{u}{2}$, $m_1 = m_2 = \frac{m}{2}$.

Following Perri *et al.* (2018) and applying the optimum allocation design to allocate *LL* and *SL* samples on two moves, the following assumptions will be applicable:
$u_1^* = u\frac{S_{z_1}}{S_{z_1}+S_t} = u\beta_1$ (say), $u_2^* = u(\frac{S_t}{S_{z_1}+S_t}) = u\beta_2$ (say), $m_1 = m(\frac{S_{z_2}}{S_{z_2}+S_t}) = m\beta_3$ (say), $m_2 = m(\frac{S_t}{S_{z_2}+S_t}) = m\beta_4$ (say), $u_1 = u\beta_3$ and $u_2 = u\beta_4$.

Hence, utilizing the two allocation designs, we further discuss the properties of the proposed IST estimator.

Theorem 5.1 *The variance of the estimator $\mathbb{T}$ under general allocation design as well as optimum allocation design is obtained and given as*

$$[V(\mathbb{T})]_i = \frac{1}{n}[(\zeta_1^i)^2(\{\frac{\lambda^i+\mu_f^i}{\lambda^i\mu_f^i}\}s_1^i) + (\zeta_3^i)^2(\{\frac{1}{\lambda^i}\}s_2^i) + (1-\zeta_3^i)^2(\{\frac{1}{\mu_f^i}\}s_2^i) - 2\zeta_1^i\zeta_3^i(\{\frac{1}{\lambda^i}\}s_3^i)],$$

where, $i = \begin{cases} g & \text{for general allocation design} \\ o & \text{for optimum allocation design} \end{cases}$,
$s_1^g = 2S_{z_1}^2 - 2\rho_{z_1x}^2S_{z_1}^2 + 2S_t^2 - 2\rho_{tx}^2S_t^2$, $s_2^g = 2S_{z_2}^2 - 2\rho_{z_2x}^2S_{z_2}^2 + 2S_t^2 - 2\rho_{tx}^2S_t^2$, $s_3^g = 2S_{z_1}S_{z_2}(\rho_{z_1z_2} - \rho_{z_1x}\rho_{z_2x}) + 2S_t^2(1-\rho_{tx}^2)$, $s_1^o = \frac{S_{z_1}^2-\rho_{z_1x}^2S_{z_1}^2}{\beta_1} + \frac{S_t^2-\rho_{tx}^2S_t^2}{\beta_2}$, $s_2^o = \frac{S_{z_2}^2-\rho_{z_2x}^2S_{z_2}^2}{\beta_3} + \frac{S_t^2-\rho_{tx}^2S_t^2}{\beta_4}$,
$s_3^o = \frac{S_{z_1}S_{z_2}(\rho_{z_1z_2}-\rho_{z_1x}\rho_{z_2x})}{\beta_3} + \frac{S_t^2(1-\rho_{tx}^2)}{\beta_4}$.

Proof 5.1 *The variance of $\mathbb{T}$ is given by*

$$\begin{aligned}[V(\mathbb{T})]_i =& (\zeta_1^i)^2[V(\hat{\bar{y}}_{1u^*}^*) + V(\hat{\bar{y}}_{1m}^*)]_i + (\zeta_3^i)^2V(\hat{\bar{y}}_{2m}^*)_i + (1-\zeta_3^i)^2V(\hat{\bar{y}}_{2u}^*)_i - \\ & 2\zeta_1^iCov(\hat{\bar{y}}_{1u^*}^*, \hat{\bar{y}}_{1m}^*) + 2\zeta_1^i\zeta_3^i[Cov(\hat{\bar{y}}_{1u^*}^*, \hat{\bar{y}}_{2m}^*) - Cov(\hat{\bar{y}}_{1m}^*, \hat{\bar{y}}_{2m}^*)]_i + \\ & 2\zeta_1^i(1-\zeta_3^i)[Cov(\hat{\bar{y}}_{1u^*}^*, \hat{\bar{y}}_{2u}^*) - Cov(\hat{\bar{y}}_{1m}^*, \hat{\bar{y}}_{2u}^*)] + 2\zeta_3^i(1-\zeta_3^i)Cov(\hat{\bar{y}}_{2m}^*, \hat{\bar{y}}_{2u}^*). \end{aligned} \quad (8)$$

Using equation (7), we have

$$\begin{aligned}[V(\mathbb{T})]_i =& (\zeta_1^i)^2[V(\hat{\bar{y}}_{1u^*}^*) + V(\hat{\bar{y}}_{1m}^*)]_i + (\zeta_3^i)^2V(\hat{\bar{y}}_{2m}^*)_i + (1-\zeta_3^i)^2V(\hat{\bar{y}}_{2u}^*)_i - \\ & 2\zeta_3^i\zeta_1^i[Cov(\hat{\bar{y}}_{1m}^*, \hat{\bar{y}}_{2m}^*)]_i, \end{aligned} \quad (9)$$

where for large population size, the variance of $\hat{\bar{y}}_{1u^*}^*$ is computed below

$$V(\hat{\bar{y}}_{1u^*}^*) = V(\hat{\bar{z}}_{1u^*}^*) + V(\hat{\bar{t}}_{u^*}^*) - Cov(\hat{\bar{z}}_{1u^*}^*, \hat{\bar{t}}_{u^*}^*). \quad (10)$$

For this expanding $\hat{\bar{z}}_{1u^*}^*$ about the point $G = (\bar{Z}_1, \bar{X})$ using Taylor's series expansion, *retaining terms up to the first order approximations, we have*

$$\begin{aligned}\hat{\bar{z}}^{*}_{1u^{*}} =& g_4[\bar{Z}_1 + (\bar{z}_{1u_1^*} - \bar{Z}_1),\ \bar{X} + (\bar{x}_{u_1^*} - \bar{X})] \\ =& \bar{Z}_1 + (\bar{z}_{1u_1^*} - \bar{Z}_1)G_1 + (\bar{x}_{u_1^*} - \bar{X})G_2 + [(\bar{z}_{1u_1^*} - \bar{Z}_1)^2 G_{11} + (\bar{x}_{u_1^*} - \bar{X})^2 G_{22} \\ &+ 2(\bar{z}_{1u_1^*} - \bar{Z}_1)(\bar{x}_{u_1^*} - \bar{X})G_{12} + \ldots].\end{aligned}$$

Expressing above equation in terms of $e_i^{'}s$ and retaining terms up to the first order approximations we have

$$\hat{\bar{z}}^{*}_{1u_1^*} - \bar{Z}_1 = (\bar{Z}_1 e_9 G_1 + \bar{X} e_{10} G_2 + [\bar{Z}_1^2 e_9^2 G_{11} + \bar{X}^2 e_{10}^2 G_{22} + \bar{Z}_1 \bar{X} e_9 e_{10} G_{12}]), \tag{11}$$

where,

$$G_1 = \frac{\partial g_4}{\partial \bar{z}_{1u_1^*}}|_G = 1,\ G_2 = \frac{\partial g_4}{\partial \bar{x}_{u_1^*}}|_G,\ G_{11} = \frac{1}{2}\frac{\partial^2 g_4}{\partial \bar{z}^2_{1u_1^*}}|_G = 0,\ G_{22} = \frac{1}{2}\frac{\partial^2 g_4}{\partial \bar{x}^2_{u_1^*}}|_G,$$

$$\text{and } G_{12} = \frac{1}{2}\frac{\partial^2 g_4}{\partial \bar{z}_{1u_1^*}\partial \bar{x}_{u_1^*}}|_G.$$

Squaring both sides of equation (11) and further retaining terms up to the first order approximation, we have

$$\left(\hat{\bar{z}}^{*}_{1u_1^*} - \bar{Z}_1\right)^2 = (\bar{Z}_1^2 e_9^2 + \bar{X}^2 e_{10}^2 G_2^2 + 2\bar{Z}_1 \bar{X} e_9 e_{10} G_2). \tag{12}$$

Taking expectations on both sides of the above equation, the variance of $\hat{\bar{z}}^{*}_{1u_1^*}$ is obtained as

$$V(\hat{\bar{z}}^{*}_{1u_1^*}) = \frac{1}{u_1}\left[S^2_{z_1} - S^2_{z_1}\rho^2_{z_1 x}\right],$$

similarly,

$$V(\hat{\bar{t}}^{*}_{u_2^*}) = \frac{1}{u_2}\left[S^2_t - S^2_t \rho^2_{tx}\right].$$

Following similar procedure for $V(\hat{\bar{y}}^{*}_{1m})_i$, $V(\hat{\bar{y}}^{*}_{2m})_i$, $V(\hat{\bar{y}}^{*}_{2u})_i$ and $Cov(\hat{\bar{y}}^{*}_{1m}, \hat{\bar{y}}^{*}_{2m})_i$ and substituting in equation (9), we have the expression of the variance of $[\mathbb{T}]_i$ under general allocation design and optimum allocation design as described in Theorem 5.1.

6. Constants under IST Allocation Designs

It is observed that $[V(\mathbb{T})]_i$ is a function of unknown constants ζ_1^i and ζ_3^i. Hence, they are minimized with respect to ζ_1^i and ζ_3^i respectively to obtain the optimum value of constants. The optimum values obtained are given in Table 3.

Table 3: Optimum Value of Constants

General Allocation Design	**Optimum Allocation Design**
$\zeta_1^g = \frac{s_2^g s_3^g \lambda^g \mu_f^g}{s_1^g s_2^g - (s_3^g)^2 (\mu_f^g)^2}$,	$\zeta_1^o = \frac{s_2^o s_3^o \lambda^o \mu_f^o}{s_1^o s_2^o - (s_3^o)^2 (\mu_f^o)^2}$,
$\zeta_3^g = \frac{s_1^g s_2^g \lambda^g}{s_1^g s_2^g - (s_3^g)^2 (\mu_f^g)^2}$	$\zeta_3^o = \frac{s_1^o s_2^o \lambda^o}{s_1^o s_2^o - (s_3^o)^2 (\mu_f^o)^2}$

Substituting the above optimum values of ζ_1^i and ζ_3^i in the expression of $[V(\mathbb{T})]_i$ respectively, we get the minimum variance of the proposed IST estimator as presented in Table 4.

Table 4: Optimum Variance

General Allocation Design	**Optimum Allocation Design**
$[V(\mathbb{T})_{opt.}]_g = \left(\frac{1}{n}\right)\left[\frac{s_2^g (s_1^g s_2^g - (s_3^g)^2 \mu_f^g)}{s_1^g s_2^g - (s_3^g)^2 (\mu_f^g)^2}\right]$	$[V(\mathbb{T})_{opt.}]_o = \left(\frac{1}{n}\right)\left[\frac{s_2^o (s_1^o s_2^o - (s_3^o)^2 \mu_f^o)}{s_1^o s_2^o - (s_3^o)^2 (\mu_f^o)^2}\right]$

6.1. Optimum Replacement policy and Minimum Variance

In surveys repeated over time, the objective is to obtain efficient estimates with minimum cost of the survey. This is technically achieved by maintaining a high overlap between two successive moves. However, the best strategy would be to minimize the variance of the estimator in order to determine the optimum value of μ or λ. Hence, $[\mathbb{V}(\mathbb{T})_{opt.}]_i$ is further minimized with respect to μ_f^i respectively, and the obtained optimum values of μ_f^i say $\hat{\mu}_f^i$ are as:

$$\hat{\mu}_f^i = min\left\{\frac{I_2^i + \sqrt{(I_2^i)^2 - I_1^i I_3^i}}{I_1^i}, \frac{I_2^i - \sqrt{(I_2^i)^2 - I_1^i I_3^i}}{I_1^i}\right\} \; \varepsilon \; [0, \; 1], \tag{13}$$

where,

$$i = \begin{cases} g & \text{for general allocation design} \\ o & \text{for optimum allocation design} \end{cases},$$

$I_1^g = (s_3^g)^4 s_2^g$, $I_2^g = (s_3^g)^2 s_1^g (s_2^g)^2$, $I_3^g = s_1^g (s_2^g)^2 (s_3^g)^2$, $I_1^o = (s_3^o)^4 s_2^o$, $I_2^o = (s_3^o)^2 s_1^o (s_2^o)^2$ and $I_3^o = s_1^o (s_2^o)^2 (s_3^o)^2$.

Substituting the optimum values of $\hat{\mu}^i$ in $[V(\mathbb{T})_{opt.}]_i$, we have the minimum variance of the proposed IST estimator as presented in Table 5.

Table 5: Optimum Variance in terms of Optimum μ

Estimator	**General Allocation Design**	**Optimum Allocation Design**
$\mathbb{T}$	$[V(\mathbb{T})_{opt.^*}]_g = (\frac{1}{n}) \left[\frac{s_2^g(s_1^g s_2^g - (s_3^g)^2 \hat{\mu}_f^g)}{s_1^g s_2^g - (s_3^g)^2 (\hat{\mu}_f^g)^2}\right]$	$[V(\mathbb{T})_{opt.^*}]_o = (\frac{1}{n}) \left[\frac{s_2^o(s_1^o s_2^o - (s_3^o)^2 \hat{\mu}_f^o)}{s_1^o s_2^o - (s_3^o)^2 (\hat{\mu}_f^o)^2}\right]$

7. Comparison

To judge the efficiency of the proposed class of IST estimators $\mathbb{T}$, the IST estimator τ has been considered where no additional auxiliary is used at any move, which is given as

$$\tau = \kappa_1 \hat{\hat{y}}_{1u^*} + \kappa_2 \hat{\hat{y}}_{1m} + \kappa_3 \hat{\hat{y}}_{2m} + \kappa_4 \hat{\hat{y}}_{2u}, \tag{14}$$

where, κ_j ; $j = 1,\ 2,\ 3$ and 4 are suitably chosen constants.

The minimum variance of the IST estimator τ has been computed and is given as

$$[V(\tau)_{opt.}]_i = \left(\frac{1}{n}\right) \left[\frac{v_2^i(v_1^i v_2^i - (v_3^i)^2 \mu_1^i)}{v_1^i v_2^i - (v_3^i)^2 (\mu_1^i)^2}\right] \tag{15}$$

with

$$\hat{\mu}_1^i = min \left\{ \frac{I_{12}^i + \sqrt{(I_{12}^i)^2 - I_{11}^i I_{13}^i}}{I_{11}^i}, \frac{I_{12}^i - \sqrt{(I_{12}^i)^2 - I_{11}^i I_{13}^i}}{I_{11}^i} \right\} \ \varepsilon\ [0,\ 1], \tag{16}$$

where, $I_{11}^i = (v_3^i)^4 v_2^i$, $I_{12}^i = (v_3^i)^2 v_1^i (v_2^i)^2$, $I_{13}^i = v_1^i (v_2^i)^2 (v_3^i)^2$, $v_1^g = 2S_{z_1}^2 + 2S_t^2$, $v_2^g = 2S_{z_2}^2 + 2S_t^2$, $v_3^g = 2\rho_{z_1 z_2} S_{z_1} S_{z_2} + 2S_t^2$, $v_1^o = \frac{S_{z_1}^2}{\beta_1} + \frac{S_t^2}{\beta_2}$, $v_2^o = \frac{S_{z_2}^2}{\beta_3} + \frac{S_t^2}{\beta_4}$, $v_3^o = \frac{\rho_{z_1 z_2} S_{z_1} S_{z_2}}{\beta_3} + \frac{S_t^2}{\beta_4}$.

8. Performance of IST estimator

In this section, we check the percent relative efficiency of the IST class of estimators $\mathbb{T}$ with respect to the IST estimator τ which are the linear combination of the estimators based on all available samples considering the availability and non-availability of additional non-sensitive auxiliary variable respectively. The percent relative efficiency has been computed under both the general allocation design as well as optimum allocation design as under:

$$\mathbb{E}^i = \frac{[V(\tau)_{opt.^*}]_i}{[V(\mathbb{T})_{opt.^*}]_i} \times 100, \tag{17}$$

where, $i = \begin{cases} g & for\ general\ allocation\ design \\ o & for\ optimum\ allocation\ design \end{cases}$.

8.1. Simulation Study

To validate the theoretical results, simulation studies have been carried out using Monte Carlo Simulation by MATLAB. The simulation is performed by examining $5,000$ different samples at two moves and the process is repeated for varying sample sizes.

Population Source:[Free access to data by Statistical Abstracts of United States]
A real population consisting of $N = 51$ states has been considered for evaluation of the performance of proposed estimators. The variables considered under IST set-up for two moves are assumed as:

y_1 :Rate of abortions in the year 2005
y_2 :Rate of abortions in the year 2008
t :Rate of residents in the year 2004
x :Rate of residents in the year 2000.

From the above considered variables, it is obvious that the rate of abortions is sensitive in nature however the rate of residents is non-sensitive in nature. Therefore, the data are suitable to be used to test the performance of the proposed IST estimators.

The simulated percent relative efficiencies of τ with respect to $\mathbb{T}$ have been computed under general as well as optimum allocation design denoted by $\mathbb{E}^{si}$; $i \in \{g, o\}$.

The simulation results are presented in Table 6 and Figure 1.

Table 6: Simulation Results

	u/n	$\mathbb{E}^{sg}$	$\mathbb{E}^{so}$
$n = 24, u = 3$	0.125	688.4432	716.2598
$n = 24, u = 5$	0.208	643.9376	696.7527
$n = 24, u = 7$	0.291	605.1396	681.0255
$n = 24, u = 10$	0.416	556.1936	664.5010
$n = 24, u = 12$	0.5	527.6227	657.7056
$n = 24, u = 15$	0.625	493.9802	656.8039
$n = 24, u = 17$	0.708	476.4308	662.1325
$n = 24, u = 20$	0.833	461.5208	683.8706
$n = 24, u = 22$	0.916	475.8523	710.6313

8.2. Numerical Illustration

To judge the performance of the proposed estimator, numerical illustration has been done for the real data considered in Section (8.1). The percent relative efficiency of the proposed estimator has been computed under general as well as optimum allocation designs denoted by E^i; $i \in \{g, o\}$.

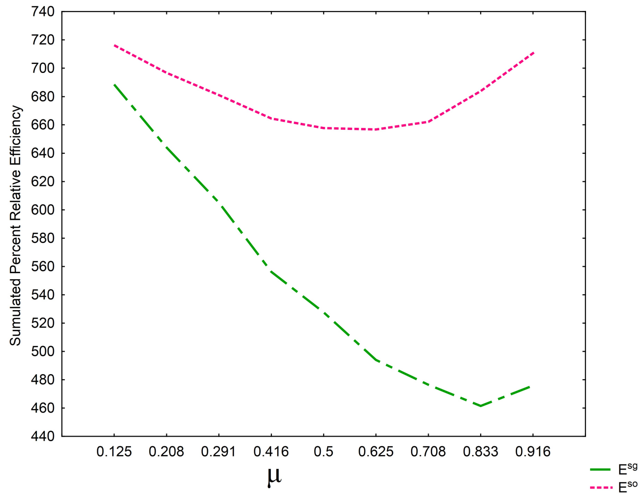

Figure 1: Graphical Representation of Simulation Results

Therefore, Table 7 represents the results obtained on performing the empirical calculation on the considered data in Section (8.1).

Table 7: Optimum value of $\mu' s$ and Percent relative efficiencies

$\hat{\mu}_1^g$	$\hat{\mu}_1^o$	$\hat{\mu}_f^g$	$\hat{\mu}_f^o$	E^g	E^o
0.8461	0.8584	0.6585	0.6712	656.6323	726.9219

9. Direct Method

The direct method of estimation is compared with the IST embedded method in order to observe the amount of loss in the precision of estimators that result due to application of IST. Some loss in precision is expected but application of direct method may not represent the true facts as the variable under consideration is sensitive in nature. As a result, privacy protection becomes an important issue for which the respondents need to be convinced. The direct version of the class of estimators $\mathbb{T}$ denoted by $\mathbb{T}_d$ is discussed as:

$$\mathbb{T}_d = \zeta_{d1}\hat{\bar{y}}^*_{1du^*} + \zeta_{d2}\hat{\bar{y}}^*_{1dm} + \zeta_{d3}\hat{\bar{y}}^*_{2dm} + \zeta_{d4}\hat{\bar{y}}^*_{2du}, \tag{18}$$

where, the constants ζ_{dj} ; $dj = 1, 2, 3$ and 4 are to be suitably chosen. Now, for computing the variance, we have the following steps as

$$E(\mathbb{T}_d) = (\zeta_{d1} + \zeta_{d2})\bar{Y}_1 + (\zeta_{d3} + \zeta_{d4})\bar{Y}_2 \tag{19}$$

with

$$\zeta_{d1} + \zeta_{d2} = 0 \text{ and } \zeta_{d3} + \zeta_{d4} = 1.$$

Following similar conditions as in Section (4), the optimum variance of the direct method is obtained and is given as

$$V(\mathbb{T}_d)_{opt.^*} = \left(\frac{1}{n}\right)\left[\frac{C_2(C_1C_2 - C_3^2\hat{\mu}_d)}{C_1C_2 - C_3^2(\hat{\mu}_d)^2}\right] \tag{20}$$

with

$$\hat{\mu}_d = min\left\{\frac{F_2 + \sqrt{F_2^2 - F_1F_3}}{F_1}, \frac{F_2 - \sqrt{F_2^2 - F_1F_3}}{F_1}\right\} \varepsilon\ [0,\ 1] \tag{21}$$

where, $F_1 = C_3^4C_2,\ F_2 = C_3^2C_1C_2^2,\ F_3 = C_1C_2^2C_3^2,\ C_1 = S_{y_1}^2 - \rho_{y_1x}^2S_{y_1}^2,$
$C_2 = S_{y_2}^2 - \rho_{y_2x}^2S_{y_2}^2,\ C_3 = S_{y_1}S_{y_2}(\rho_{y_1y_2} - \rho_{y_1x}\rho_{y_2x}).$

To examine the performance of the direct method we compare the estimator $\mathbb{T}_d$ with respect to the proposed IST class of estimator $\mathbb{T}$ under both general allocation design as well as optimum allocation design as

$$E_d^i = \frac{V(\mathbb{T}_d)_{opt.^*}}{[V(\mathbb{T})_{opt.^*}]_i} \times\ 100. \tag{22}$$

The numerical comparison has been done on the data considered in Section (8.1) and the results are presented in Table 8.

Table 8: Direct Method comparison with IST under general allocation design and optimum allocation design

$\hat{\mu}_d$	E_d^g	E_d^o
0.6000	33.4446	41.9547

10. Discussion of Results

The following interpretations can be drawn from empirical and simulation results:

1. The minimum variance unbiased estimation is feasible under IST set-up to estimate sensitive population mean on successive moves.

2. The simulation results in Table 6 and Figure 1 show that for all $\mu \in [0, 1]$, the percent relative efficiencies exist for both the allocation designs. As $E^{so} > E^{sg}\ \forall\ \mu$, this indicate that the IST estimator under optimum allocation design is more efficient than the general allocation design. The percent relative efficiencies exit for all considered variations in sample sizes. Also, both E^{sg} and $E^{so} > 0$, this indicates that the IST class of estimators $\mathbb{T}$ is better than that of IST estimator τ.

3. From Table 7, it is observed that the optimum fraction of fresh sample to be drawn afresh at current occasion exists for both the IST class of estimators under both allocation designs. Further, it is observed that IST estimator $\mathbb{T}$ is coming out to be more efficient than IST estimator τ under both the allocation designs. However, the estimators under the optimum allocation design are proved to be more efficient than that of the general allocation design.

4. Table 8 indicates that $E_d^i < 100\ \forall\ i \in \{g,\ o\}$, which means there is a loss in precision when the IST estimator is compared with the direct method under both the allocation designs. The loss is incurred due to the usage of IST in the estimator $\mathbb{T}$. However, dealing with sensitive issues when the direct method is used, may result in false response or even no response.

11. Concluding Remarks

From the interpretation of results, it is concluded that IST is an alternative technique to deal with sensitive issues in successive sampling. In IST setup, the estimator utilizing additional auxiliary variable is proved to be more efficient than the estimator in which no additional auxiliary variable is used. Out of the two allocation designs for allocating *LL* and *SL* samples, the IST class of estimators using optimum allocation design is coming out to be more efficient than the estimator using general allocation deign. Therefore, the IST estimators with optimum allocation designs may be recommended for their practical use by survey practitioners.

Acknowledgements

Authors are thankful to the honourable reviewer for deeply reading the manuscript and giving valuable suggestions leading to improvement in the quality of content and presentation over the original version of the manuscript. The first authors is thankful to Science and Engineering Research Board, New Delhi, India for providing the financial assistance to carry out the present work. Authors sincerely acknowledged the free access to data by statistical abstracts of United states.

References

Arcos, A., Rueda, M. and Singh, S. A., (2015). Generalized approach to randomized response for quantitative variables. *Quality and Quantity*, 49, pp. 1239 –1256.

Arnab R., Singh S., (2013). Estimation of mean of sensitive characteristics for successive sampling. *Communication in Statistics-Theory and Methods*, 42, 2499–2524.

Chaudhuri,A., Christofides,T. C., (2013). Indirect questioning in sample surveys. *Springer-Verlag, Berlin, Heidelberg, De.*

Diana, G., Perri, P. F., (2010). New scrambled response models for estimating the mean of a sensitive quantitative character. *Journal of Applied Statistics*, 37(11), pp. 1875–1890.

Diana, G., Perri, P. F., (2011). A class of estimators for quantitative sensitive data. *Statistical Papers,* 52(3), pp. 633–650.

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

Franklin, L. A., (1989). A comparison of estimators for randomized response sampling with continuous distribution from dichotomous populations. *Communications in Statistics-Theory and Methods*, 18, pp. 489–505.

Garcia, A. V., Artes, E. M., (2002). Improvement on estimating of current population ratio in successive sampling. *Brazilian Journal of Probability and Statistics*, 16(2), pp. 107–122.

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

Holbrook, A. L., Krosnick, J. A., (2010). Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique. *Public Opinion Quarterly*, 74, pp. 37–67.

Horvitz, D. G., Shah, B. V., Simmons, W. R., (1967). The unrelated question randomized response model. *Journal of the American Statistical Association*, pp. 65–72.

LaBrie, J. W., Earleywine, M., (2000). Sexual Risk Behaviors and Alcohol: Higher Base Rates Revealed Using the Unmatched-count Technique. *Journal of Sex Research*, 37, pp. 321–326.

Miller, J. D., (1984). A New Survey Technique For Studying Deviant Behavior.The George Washington University Ph.D. thesis.

Mukhopadhyay, P., (2014). Theory and Methods of Survey Sampling. *PHI Learning Private Limited*, Delhi.

Patterson, H. D., (1950). Sampling on successive occasions with partial replacement units. *Journal of royal statistical society, Wiley, series B(Mthodoligical)* 12(2), pp. 241–255.

Perri, P. F., Diana, G., (2013). Scrambled response models Based on auxiliary variables. In: Torelli, N., Pesarin, F., Bar-Hen, A. (eds.). *Advances in Theoretical and Applied Statistics, Springer, Berlin*, pp. 281–291.

Perri, P. F., Rueda, M. M. G., Cobo, B. R., (2018). Multiple sensitive estimation and optimal sample size allocation in the item sum technique. *Biometrical Journal*, 60, pp. 155–173.

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

Priyanka, K., Trisandhya, P. and Mittal, R., (2018). Dealing sensitive characters on successive occasions through a general class of estimators using scrambled response techniques. *Metron*, 76, pp. 203–230.

Priyanka, K., Trisandhya, P., (2019). A Composite Class of Estimators using Scrambled Response Mechanism for Sensitive Population mean in Successive Sampling. *Communications in statistics- Theory and Methods*, 48(4), pp. 1009–1032.

Priyanka, K., Trisandhya, P. and Mittal, R., (2019). Scrambled Response Techniques in Two Wave Rotation Sampling for Estimating Population Mean of Sensitive Characteristics with Case Study. *Journal of Indian Society of Agricultural Statistics*, 73(1), pp. 41–52.

Rayburn, N. R., M. Earleywine, and Davison, G. C., (2003). Base Rates of Hate Crime Victimization among College Students. *Journal of Interpersonal Violence*, 18, pp. 1209–1221.

Srivastava, S. K., Jhajj H. S., (1980). A class of estimators using auxiliary information for estimating finite population variance. *Sankya series C,* 42, pp. 87–96.

Tracy, D. S., Singh, H. P. and Singh, R., (1996). An alternative to the ratio-cum-product estimator in sample surveys. *Journal of Statistical Planning and Inference*, 53, pp. 375–397.

Trappmann, M., Krumpal, I., Kirchner, A., and Jann, B., (2014). Item sum: A new technique for asking quantitative sensitive questions. *Journal of Survey Statistics and Methodology*, 2, pp. 58–77.

Tsuchiya, T., Hirai, Y. and Ono, S., (2007). A Study of the Properties of the Item Count Technique. *Public Opinion Quarterly*, 71, pp. 253–272.

Warner, S. L., (1965). Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60, pp. 63–69.

Wimbush, J. C., Dalton, D. R., (1997). Base Rate for Employee Theft: Convergence of Multiple Methods. *Journal of Applied Psychology*, 82, pp. 756–763.

Yu, B., Jin, Z., Tian, J. and Gao, G., (2015). Estimation of sensitive proportion by randomized response data in successive sampling. *Computational and Mathematical Methods in Medicine*, pp 6, DOI: 10.1155/2015/172918.