Journal of Statistics Applications & Probability Letters
An International Journal

http://dx.doi.org/10.18576/jsapl/090203

Using Scramble Variable for the Estimation of Mean and the Sensitivity Level in Generalized Randomized Response Model

Chandra Mani Prasad

Department of Statistics, Central University of Rajasthan, Bandarsindri, Kishangarh, Ajmer, Rajasthan 305817, India

Received: 21 Apr. 2021, Revised: 27 Jan. 2022, Accepted: 5 Feb. 2022
Published online: 1 May 2022

Abstract: The present review built up a generalized optional randomized response (ORR) model for evaluating some unknown population parameters that are unbiased of a delicate variable when the appropriate responses are quantitative. In addition to evaluating unknown population parameters, it is demonstrated that the optional randomized response model is valuable in measuring the sensitivity levels of the inquiries in the individual meeting studies. This article built up two different type models and examined their properties. The proposed models are found to be more proficient than the usual existing estimator. The theoretical results are verified numerically and are illustrated as well.

Keywords: Randomized response models, parameter estimation, privacy, asymptotic distribution.

1 Introduction

Social surveys sometimes incorporate delicate things of enquiry including criticizing matters (i.e. the use of illegal drugs, homosexuality, tax evasion or abortion) which individuals jump at the chance to escape others. Coordinate question about them regularly yield untruthful reaction or non-reaction. In the pioneer work that was carried out by Warner [1], he developed a technique in order to estimate the population proportion of individuals having a sensitive characteristic from a survey data. After the spearheading work of Warner [1], a few randomized response models have been produced by analysts for gathering information on both the subjective and the quantitative factors for example see Greenberg et al. [2], Fox and Tracy [3], Chaudhuri and Mukherjee [4], Ryu et al. [5], Mangat and Singh [6], Tracy and Mangat [7], Tracy and Osahan [8] and Singh [9] and the references cited therein. In addition to hypothetical improvements, several scientists such as Kerkvliet [10] has demonstrated appropriateness of the randomized response models in the genuine circumstances.

It is notable that RR procedure is design for delicate qualities to accomplish legitimate sample data and decrease reaction bias. To reduce this kind of bias, Greenberg et al. [11] has estimation of mean of a delicate quantitative variable. However, Eichhorn and Hayre [12], Gupta et al. [13], Gupta et al. [14], Huang [15], Gupta et al. [16], [17], Mehta et al. [18] has also studied on quantitative response RRT models. In Eichhorn and Hayre [12] model, there is no way to reporting on delicate variable. In order to enhance the privacy of respondents, Gupta et al. [13] has consider a multiplicative option randomized response procedure where the respondent has the choice for a genuine or a scrambled response. Moreover, Gupta et al. [14] two-stage optional RR model and Mehta et al. [18] two-stage and three-stage optional RR models were both in view of added substance scrambling that disposed of the requirement for any guess by utilizing a split-sample approach.

In the present review, the delicate variable's mean as well its sensitivity level in the option RR model has been estimated. The sensitivity level is defined as the proportion of respondents in the population who consider the question delicate enough to not feel comfortable about giving a true response. To estimate the sensitivity level for certain materials of inquiry in optional RR model Huang [15] used a more general scrambling and the background of their model have been described in section 2 and denoted all notations. In the Section 3, we developed estimators by following the privacy

* Corresponding author e-mail: chandramanicuraj@gmail.com

of respondent. The empirical studies are carried out; which shows the efficiency of proposed estimators as shown in the upcoming sections.

2 Huang (2010) model

In Huang (2010) model, the sample of size n is divided into two independent sub-samples of sizes n_1 and n_1 i.e ($n = n_1 + n_2$) which drawn from the population by utilizing simple random sampling with replacement. Every respondent chosen in the i^{th} sample is associated with two randomization device $D_i(i = 1,2)$ that provide two independent scrambling variables S_i and R_i from some pre-assigned distribution with known $\mu_{R_i} = 1$, $\mu_{s_i} = \theta_i$, $\sigma^2_{R_i} = \gamma_i^2$ and $\sigma^2_{s_i} = \delta_i^2$. The RR procedures are following by the respondents and whole process is without revealed to the interviewer. Then the respondent selected by himself one of the following two statements:
(i) The respondent can report the correct response X, or
(ii) The respondent can report the scrambled response $R_iX + S_i$ with the following probabilities $1-W$ and W respectively. Thus, the reported response from a specific respondent in the i^{th} sample is given by

$$Z_{\text{i}} = (1 - W)X + W(R_iX{+}S_i)\,; i = 1,2 \quad (1)$$

It can be seen that,

$$E(Z_i) = (1 - W)\mu_X + W(\mu_{R_i}\mu_X + \mu_{S_i}) = \mu_X + W\theta_i \text{ ,Since } E(R_i) = \mu_{R_i} = 1. \quad (2)$$

Solving the equation Eq. (2) for μ_x and W, we have

$$\mu_X = \frac{\theta_2 E(Z_1) - \theta_1 E(Z_2)}{(\theta_2 - \theta_1)};\theta_1 \neq \theta_2 \quad (3)$$

and,

$$W = \frac{E(Z_2) - E(Z_1)}{(\theta_2 - \theta_1)};\theta_1 \neq \theta_2 \quad (4)$$

The unbiased estimator $\hat{\mu}_X$ and $\hat{W}$ for μ_x and W respectively, obtained by estimating $E(Z_i) = \bar{Z}_i\,(i = 1, 2)$ are given by

$$\hat{\mu}_X = \frac{\theta_2\bar{Z}_1 - \theta_1\bar{Z}_2}{(\theta_2 - \theta_1)} \; and \; \hat{W} = \frac{\bar{Z}_2 - \bar{Z}_1}{(\theta_2 - \theta_1)},\theta_1 \neq \theta_2 \quad (5)$$

$$E(\hat{\mu}_X) = E\left(\frac{\theta_2\bar{Z}_1 - \theta_1 E\bar{Z}_2}{(\theta_2 - \theta_1)}\right) = \mu_X \; and \; E(\hat{W}) = E\left(\frac{\bar{Z}_2 - \bar{Z}_1}{(\theta_2 - \theta_1)}\right) = W. \quad (6)$$

Where $Z_i(i = 1,2)$ denoted the sample mean of responses in the i^{th} sub-samples. The variance of these estimators are given by respectively

$$V(\hat{\mu}_X) = \frac{1}{(\theta_2 - \theta_1)^2}\left[\theta_2^2\frac{\sigma^2_{Z_1}}{n_1} + \theta_1^2\frac{\sigma^2_{Z_2}}{n_2}\right],\theta_1 \neq \theta_2 \quad (7)$$

and,

$$V(\hat{W}) = \frac{1}{(\theta_2 - \theta_1)^2}\left[\frac{\sigma^2_{Z_1}}{n_1} + \frac{\sigma^2_{Z_2}}{n_2}\right],\theta_1 \neq \theta_2 \quad (8)$$

where,

$$\sigma^2_{Z_i} = \sigma^2_X + (\sigma^2_X + \mu^2_X)W\gamma_i^2 + (\delta_i^2 + \theta_i^2)W - \theta_i^2 W^2, i = 1, 2. \quad (9)$$

3 Proposed procedure

In this context, we propose a two-stage linear combination model and three-stage linear combination model by following the study of Huang (2010) model and show that how it can be improved this model by using two-stage linear combination model and three-stage linear combination model.

3.1 Two-stage optional RR model

In this sub-section, a pre-determined proportion (F) of respondents are given the instruction to provide a scramble response to the sensitive questions and remaining proportion $(1-F)$ of respondents have a choice of scrambling response additively $Z = RX + S$, if the question deemed sensitive or else they provide true response $Z = X$, if the question non-sensitive. In two-stage optional RR model, a sample of size n is split into two sub-samples of sizes n_1 and n_2 i.e. $n_1 + n_2 = n$, the total sample size needed. Let R_i and $S_i (i = 1,2)$ be scrambling variable associated with randomization device $D_i (i = 1,2)$ with known $\mu_{R_i} = 1$, $\mu_{s_i} = \theta_i$, $\sigma^2_{R_i} = \gamma^2_i$ and $\sigma^2_{s_i} = \delta^2_i$. Let us consider, X be true response having mean μ_x and variance σ^2_x , both are unknown. Suppose W is the sensitivity level of sensitive question. It is considered that the variables X, S_1, S_2, R_1 and R_2 are mutually independent. Let $Z_i (i = 1,2)$ be reported answer in the i^{th} sub-samples then it may be written as

$$Z_i = \begin{cases} X & with\ probability\ \ (1-F)(1-W) \\ R_i X + S_i & with\ probability\ F + (1-F)W \end{cases} \tag{10}$$

Note that from Eq. (10), for $(i = 1,2)$

$$E(Z_i) = \mu_X + \theta_i [F + (1-F)W] \tag{11}$$

Eliminating μ_x and W from equation Eq. (11), we have

$$\mu_X = \frac{\theta_2 E(Z_1) - \theta_1 E(Z_2)}{(\theta_2 - \theta_1)},\ \theta_1 \neq \theta_2 \tag{12}$$

$$W = \frac{1}{(1-F)} \left[\frac{E(Z_2) - E(Z_1)}{(\theta_2 - \theta_1)} - F \right],\ \theta_1 \neq \theta_2,\ F \neq 1 \tag{13}$$

It is easy to show that $\hat{\mu}_x$ and $\hat{W}$ are unbiased estimator of μ_x and W respectively with variances

$$V(\hat{\mu}_X) = \frac{1}{(\theta_2 - \theta_1)^2} \left[\theta_2^2 \frac{\sigma^2_{Z_1}}{n_1} + \theta_1^2 \frac{\sigma^2_{Z_2}}{n_2} \right],\ \theta_1 \neq \theta_2 \tag{14}$$

$$V(\hat{W}) = \frac{1}{(\theta_2 - \theta_1)^2 (1-F)^2} \left[\frac{\sigma^2_{Z_1}}{n_1} + \frac{\sigma^2_{Z_2}}{n_2} \right],\ \theta_1 \neq \theta_2,\ F \neq 1 \tag{15}$$

where,

$$\sigma^2_{Z_i} = \sigma^2_X + \delta^2_i [F + (1-F)W] + \gamma^2_i \sigma^2_x [F + (1-F)W] + \gamma^2_i \mu^2_x [F + (1-F)W] \\ + \theta^2_i [F + (1-F)W] [1 - \{F + (1-F)W\}]$$

Unbiased estimators of $\mathrm{V}(\hat{\mu}_\mathrm{X})$ and $\mathrm{V}(\hat{\mathrm{W}})$ are respectively given by

$$\hat{V}(\hat{\mu}_X) = \frac{1}{(\theta_2 - \theta_1)^2} \left[\theta_2^2 \frac{S^2_{Z_1}}{n_1} + \theta_1^2 \frac{S^2_{Z_2}}{n_2} \right], \theta_1 \neq \theta_2$$

$$\hat{V}(\hat{W}) = \frac{1}{(\theta_2 - \theta_1)^2 (1-F)^2} \left[\theta_2^2 \frac{S^2_{Z_1}}{n_1} + \theta_1^2 \frac{S^2_{Z_2}}{n_2} \right], \theta_1 \neq \theta_2, F \neq 1$$

where,

$$S^2_{Z_i} = \frac{1}{(n_i - 1)} \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_i)^2, i = 1, 2$$

are sample variance of the responses in the two sub-samples.
Asymptotic normality for $\hat{\mu}_X$ and $\hat{W}$ is evident from central limit theorem and independence of two sub-samples. Now, the comparison of variances $V(\hat{\mu}_X)$ of our proposed two-stage model denoted by $V(\hat{\mu}_X)_A$ with variance $V(\hat{\mu}_X)$ and Huang [15] one-stage model denoted by $V(\hat{\mu}_X)_H$ with the condition that the scrambling variables of both the models have same mean and variance as well.

Theorem 3.1: For all value of F and W, $V(\hat{\mu}_X)_A \leq V(\hat{\mu}_X)_H$.

$$\Rightarrow F \geq \frac{1}{(1-W)} \left[\frac{n_1 \theta_1^2 (\delta_2^2 + \gamma_2^2 \sigma_x^2 + \gamma_2^2 \mu_x^2) + n_2 \theta_2^2 (\delta_1^2 + \gamma_1^2 \sigma_x^2 + \gamma_1^2 \mu_x^2)}{n \theta_1^2 \theta_2^2} + 1 - 2W \right] = F^*$$

Proof:

$$V(\hat{\mu}_X)_A \le V(\hat{\mu}_X)_H$$
$$\Leftrightarrow \frac{\theta_2^2}{n_1}\left[\sigma_x^2+\delta_1^2\{F+(1-F)W\}+\gamma_1^2\left(\sigma_x^2+\mu_x^2\right)\{F+(1-F)W\}+\theta_1^2\{F+(1-F)W\}\{1-\{F+(1-F)W\}\}\right]$$
$$+\frac{\theta_1^2}{n_2}\left[\sigma_x^2+\delta_2^2\{F+(1-F)W\}+\gamma_2^2\left(\sigma_x^2+\mu_x^2\right)\{F+(1-F)W\}+\theta_2^2\{F+(1-F)W\}\{1-\{F+(1-F)W\}\}\right]$$
$$\le \frac{\theta_2^2}{n_1}\left[\sigma_x^2+W\delta_1^2+W\gamma_1^2\left(\sigma_x^2+\mu_x^2\right)+W\theta_1^2\{1-W\}\right]$$
$$+\frac{\theta_1^2}{n_2}\left[\sigma_x^2+W\delta_2^2+W\gamma_2^2\left(\sigma_x^2+\mu_x^2\right)+W\theta_2^2\{1-W\}\right]$$
$$\Leftrightarrow \frac{\theta_2^2}{n_1}\left[\left(\delta_1^2+\gamma_1^2\sigma_x^2+\gamma_1^2\mu_x^2+\theta_1^2\right)\{F+(1-F)W-W\}-\theta_1^2\left\{(F+(1-F)W)^2-W^2\right\}\right]$$
$$+\frac{\theta_1^2}{n_2}\left[\left(\delta_2^2+\gamma_2^2\sigma_x^2+\gamma_2^2\mu_x^2+\theta_2^2\right)\{F+(1-F)W-W\}-\theta_2^2\left\{(F+(1-F)W)^2-W^2\right\}\right]\le 0$$
$$\Leftrightarrow \left[n_2\theta_2^2\left(\delta_1^2+\gamma_1^2\sigma_x^2+\gamma_1^2\mu_x^2\right)+n_1\theta_1^2\left(\delta_2^2+\gamma_2^2\sigma_x^2+\gamma_2^2\mu_x^2\right)+n\theta_1^2\theta_2^2\{1-\{F(1-W)+2W\}\}\right]\le 0$$

$$\Leftrightarrow \left[\frac{n_2\theta_2^2\left(\delta_1^2+\gamma_1^2\sigma_x^2+\gamma_1^2\mu_x^2\right)+n_1\theta_1^2\left(\delta_2^2+\gamma_2^2\sigma_x^2+\gamma_2^2\mu_x^2\right)}{n\theta_1^2\theta_2^2}+1-F(1-W)-2W\right]\le 0$$
$$\Leftrightarrow F \ge \frac{1}{(1-W)}\left[\frac{n_1\theta_1^2\left(\delta_2^2+\gamma_2^2\sigma_x^2+\gamma_2^2\mu_x^2\right)+n_2\theta_2^2\left(\delta_1^2+\gamma_1^2\sigma_x^2+\gamma_1^2\mu_x^2\right)}{n\theta_1^2\theta_2^2}+1-2W\right]=F^*$$

Note that,

$$F^* \le 1 \Leftrightarrow \frac{n_1\theta_1^2\left(\delta_2^2+\gamma_2^2\sigma_x^2+\gamma_2^2\mu_x^2\right)+n_2\theta_2^2\left(\delta_1^2+\gamma_1^2\sigma_x^2+\gamma_1^2\mu_x^2\right)}{n\theta_1^2\theta_2^2}+1-2W \le 1-W$$

$$\Leftrightarrow \frac{n_1\theta_1^2\left(\delta_2^2+\gamma_2^2\sigma_x^2+\gamma_2^2\mu_x^2\right)+n_2\theta_2^2\left(\delta_1^2+\gamma_1^2\sigma_x^2+\gamma_1^2\mu_x^2\right)}{n\theta_1^2\theta_2^2} \le W$$

Thus, the above relation obviously holds true when W is large. In the regarding of this theorem, we can say that the variance $\mathrm{V}(\hat{\mu})_{\mathrm{A}}$ may be increased by using two-stage linear combination model. Now, to quantify the merits of the proposed model over the modelof Huang [15] computed numerical comparison of variance of the mean estimators for the sample size $n = 1000$ and the random variables X, R_1, R_2, S_1 and S_2 have Poisson distribution with means 4, 1, 1, 2 and 5 respectively. Table 1 provide comparisons of the proposed two-stage linear combination model and Huang [15] model. It is clear from Table 1 that for larger values of , the values of are continuously increasing with an increasing in F. Hence, for more sensitive questions, involving F in

Table 1. $V(\hat{\mu})_A$ of mean estimator (**in bold**) and $V(\hat{\mu})_H$ of mean estimator for different values of F and W, $n = 1000$, $n_1 = n_2 = 500$, $X \sim P(4)$, $R_1 \sim P(1)$, $R_2 \sim P(1)$, $S_1 \sim P(2)$ and $S_2 \sim P(5)$

F	$W=0.1$	$W=0.3$	$W=0.5$	$W=0.7$	$W=0.9$
0	**0.0442** 0.0442	**0.0784** 0.0784	**0.1091** 0.1091	**0.1362** 0.1362	**0.1598** 0.1598
0.1	**0.0601** 0.0442	**0.0896** 0.0784	**0.1162** 0.1091	**0.1400** 0.1362	**0.1609** 0.1598
0.3	**0.0896** 0.0442	**0.1106** 0.0784	**0.1298** 0.1091	**0.1473** 0.1362	**0.1630** 0.1598
0.5	**0.1162** 0.0442	**0.1298** 0.0784	**0.1424** 0.1091	**0.1542** 0.1362	**0.1651** 0.1598
0.7	**0.1400** 0.0442	**0.1473** 0.0784	**0.1542** 0.1091	**0.1609** 0.1362	**0.1672** 0.1598
0.9	**0.1609** 0.0442	**0.1630** 0.0784	**0.1651** 0.1091	**0.1672** 0.1362	**0.1692** 0.1598

linear combination model provides a little higher level of anonymity as compared to Huang [15] model. It may be noted that the estimated variance initially doing well with the competitor variance for all values when F is included. Thus, this is clear that by using a two-stage linear combination model shows higher level of sensitivity.

3.2 Three-stage optional RR model

In the last sub-section, it was considered that the two-stage model perform better than existing model under some reasonably mild condition. In this sub-section, we proposed a model that uses both (T) and (F) simultaneously. Again, the sample of size n is split into two sub-samples of sizes n_1 and n_2 i.e. $n = n_1 + n_2$. Suppose (F) is a fix pre-determined proportion of respondents is instructed to scramble their response and (T) is a fix pre-determine proportion of respondents is instructed to tell the truth if the question sensitive. The remaining $(1-T-F)$ proportions of respondents have a choice of scrambling their response additively $Z = RX + S$, if the question deemed sensitive or else they provide true response $Z = X$,if the question non-sensitive. Let X be true response with unknown mean μ_x and unknown variance σ_x^2. Let R_i and $S_i (i = 1,2)$ be scrambling variable associated with randomization device $D_i (i = 1,2)$ with known $\mu_{R_i} = 1$, $\mu_{S_i} = \theta_i$, $\sigma_{R_i}^2 = \gamma_i^2$ and $\sigma_{S_i}^2 = \delta_i^2$. Assumed that the variables X, S_1, S_2, R_1 and R_2 are mutually independent. Thus, the reported answer $Z_i (i = 1,2)$ in the $i^{th} (i = 1,2)$ sub-samples may be written as

$$Z_i = \begin{cases} X & with\ probability\ T + (1-T-F)(1-W) \\ R_i X + S_i & with\ probability\ F + (1-T-F)W \end{cases} \tag{16}$$

From Eq. (16) for $i = 1,2$, we have

$$E(Z_i) = \mu_X + \theta_i [F + (1-F)W - WT] \tag{17}$$

From Eq. (17), one may easily solve for μ_x and W, and we get the following estimators

$$\hat{\mu}_X = \frac{\theta_2 \bar{Z}_1 - \theta_1 \bar{Z}_2}{(\theta_2 - \theta_1)},\ \theta_1 \neq \theta_2 \tag{18}$$

$$\hat{W} = \frac{1}{(1-T-F)} \left[\frac{\bar{Z}_2 - \bar{Z}_1}{(\theta_2 - \theta_1)} - F \right],\ \theta_1 \neq \theta_2,\ T + F \neq 1 \tag{19}$$

Thus, we have the following hypothesis.

Theorem 3.2. $\hat{\mu}_X$ is asymptotically normal with a mean of μ_x and a variance of

$$V(\hat{\mu}_X) = \frac{1}{(\theta_2 - \theta_1)^2} \left[\theta_2^2 \frac{\sigma_{Z_1}^2}{n_1} + \theta_1^2 \frac{\sigma_{Z_2}^2}{n_2} \right], \theta_1 \neq \theta_2 \tag{20}$$

Theorem 3.3. $\hat{W}$ is asymptotically normal with a mean of W and a variance of

$$V(\hat{W}) = \frac{1}{(\theta_2 - \theta_1)^2 (1-T-F)^2} \left[\theta_2^2 \frac{\sigma_{Z_1}^2}{n_1} + \theta_1^2 \frac{\sigma_{Z_2}^2}{n_2} \right],\ \theta_1 \neq \theta_2,\ T + F \neq 1 \tag{21}$$

where,

$$\sigma_{Z_i}^2 = \sigma_X^2 + \delta_i^2 [F + (1-T-F)W] + \gamma_i^2 \sigma_x^2 [F + (1-T-F)W] + \gamma_i^2 \mu_x^2 [F + (1-T-F)W] \\ + \theta_i^2 [F + (1-T-F)W] [1 - \{F + (1-T-F)W\}]$$

The unbiased estimators of the variance of $\hat{\mu}_X$ and $\hat{W}$ are respectively given by

$$\hat{V}(\hat{\mu}_X) = \frac{1}{(\theta_2 - \theta_1)^2} \left[\theta_2^2 \frac{S_{Z_1}^2}{n_1} + \theta_1^2 \frac{S_{Z_2}^2}{n_2} \right], \theta_1 \neq \theta_2$$

And,

$$\hat{V}(\hat{W}) = \frac{1}{(\theta_2 - \theta_1)^2 (1-T-F)^2} \left[\theta_2^2 \frac{S_{Z_1}^2}{n_1} + \theta_1^2 \frac{S_{Z_2}^2}{n_2} \right], \theta_1 \neq \theta_2, T + F \neq 1$$

where, $S_{Z_i}^2 = \frac{1}{(n_i - 1)} \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_i)^2, i = 1,\ 2$ are sample variance of the responses in the two sub-samples. Asymptotic normality for $\hat{\mu}_X$ and $\hat{W}$ is obvious using central limit theorem and independence of $\bar{Z}_1$ and $\bar{Z}_2$.

Now, we compare the variances in the following theorems under the assumption that the values of scrambling variables have the same mean and variance in all models. Let $V(\hat{\mu}_X)_{A1}$ denoted the variance of three-stage linear combination model and $V(\hat{\mu}_X)_H$ denoted the variance of Huang [15] model, $V(\hat{\mu}_X)_A$ denoted the variance of proposed two-stage model respectively.

Theorem 3.4. For all value of T and F, $V(\hat{\mu}_X)_{A1} \leq V(\hat{\mu}_X)_H$

$$\Rightarrow F \geq \frac{1}{(1-W)}\left[1+W(T-2)+\frac{n_1\theta_1^2\left\{\left(\sigma_x^2+\mu_x^2\right)\gamma_2^2+\delta_2^2\right\}+n_2\theta_2^2\left\{\left(\sigma_x^2+\mu_x^2\right)\gamma_1^2+\delta_1^2\right\}}{n\theta_1^2\theta_2^2}\right]=F^* \tag{22}$$

Proof:

$$\begin{aligned}
&V(\hat{\mu}_X)_{A1} \leq V(\hat{\mu}_X)_H \\
&\Leftrightarrow \tfrac{\theta_2^2}{n_1}\left[\sigma_x^2+\delta_1^2\{F+(1-T-F)W\}+\gamma_1^2\left(\sigma_x^2+\mu_x^2\right)\{F+(1-T-F)W\}\right. \\
&+\theta_1^2\{F+(1-T-F)W\}\{1-\{F+(1-T-F)W\}\}\big]+\tfrac{\theta_1^2}{n_2}\left[\sigma_x^2+\delta_2^2\{F+(1-T-F)W\}\right. \\
&+\gamma_2^2\left(\sigma_x^2+\mu_x^2\right)\{F+(1-T-F)W\}+\theta_2^2\{F+(1-T-F)W\}\{1-\{F+(1-T-F)W\}\}\big] \\
&\leq \tfrac{\theta_2^2}{n_1}\left[\sigma_x^2+W\delta_1^2+W\gamma_1^2\left(\sigma_x^2+\mu_x^2\right)+W\theta_1^2\{1-W\}\right]+\tfrac{\theta_1^2}{n_2}\left[\sigma_x^2+W\delta_2^2+W\gamma_2^2\left(\sigma_x^2+\mu_x^2\right)+W\theta_2^2\{1-W\}\right] \\
&\Leftrightarrow \{F+(1-T-F)W-W\}\left[\left(n_2\theta_2^2\delta_1^2+n_1\theta_1^2\delta_2^2+n\theta_1^2\theta_2^2\right)+\left(n_2\theta_2^2\gamma_1^2+n_1\theta_1^2\gamma_2^2\right)\left(\sigma_x^2+\mu_x^2\right)\right. \\
&-n\theta_1^2\theta_2^2\{F+(1-T-F)W+W\}\leq 0
\end{aligned}$$

$$\begin{aligned}
&\Leftrightarrow \tfrac{n_1\theta_1^2\left\{\left(\sigma_x^2+\mu_x^2\right)\gamma_2^2+\delta_2^2\right\}+n_2\theta_2^2\left\{\left(\sigma_x^2+\mu_x^2\right)\gamma_1^2+\delta_1^2\right\}}{n\theta_1^2\theta_2^2}+1-\{F(1-W)+W(2-T)\}\leq 0 \\
&\Rightarrow F \geq \tfrac{1}{(1-W)}\left[1+W(T-2)+\tfrac{n_1\theta_1^2\left\{\left(\sigma_x^2+\mu_x^2\right)\gamma_2^2+\delta_2^2\right\}+n_2\theta_2^2\left\{\left(\sigma_x^2+\mu_x^2\right)\gamma_1^2+\delta_1^2\right\}}{n\theta_1^2\theta_2^2}\right]=F^*
\end{aligned}$$

Note that,

$$F^* \leq 1 \Leftrightarrow \tfrac{n_1\theta_1^2\left\{\left(\sigma_x^2+\mu_x^2\right)\gamma_2^2+\delta_2^2\right\}+n_2\theta_2^2\left\{\left(\sigma_x^2+\mu_x^2\right)\gamma_1^2+\delta_1^2\right\}}{n\theta_1^2\theta_2^2}+1+W(T-2)\leq 1-W$$

$$F^* \leq 1 \Leftrightarrow T \leq 1-\tfrac{n_1\theta_1^2\left\{\left(\sigma_x^2+\mu_x^2\right)\gamma_2^2+\delta_2^2\right\}+n_2\theta_2^2\left\{\left(\sigma_x^2+\mu_x^2\right)\gamma_1^2+\delta_1^2\right\}}{nW\theta_1^2\theta_2^2}$$

Theorem 3.5. For all value of T and F, $V(\hat{\mu}_X)_{A1} \leq V(\hat{\mu}_X)_A$

$$\Rightarrow F \geq \frac{1}{2(1-W)}\left[1+W(T-2)+\frac{n_1\theta_1^2\left\{\left(\sigma_x^2+\mu_x^2\right)\gamma_2^2+\delta_2^2\right\}+n_2\theta_2^2\left\{\left(\sigma_x^2+\mu_x^2\right)\gamma_1^2+\delta_1^2\right\}}{n\theta_1^2\theta_2^2}\right]=F^{**}$$

Proof:

$$\begin{aligned}
&V(\hat{\mu}_X)_{A1} \leq V(\hat{\mu}_X)_A \\
&\Leftrightarrow \tfrac{\theta_2^2}{n_1}\left[\sigma_x^2+\delta_1^2\{F+(1-T-F)W\}+\gamma_1^2\left(\sigma_x^2+\mu_x^2\right)\{F+(1-T-F)W\}\right. \\
&+\theta_1^2\{F+(1-T-F)W\}\{1-\{F+(1-T-F)W\}\}\big]+\tfrac{\theta_1^2}{n_2}\left[\sigma_x^2+\delta_2^2\{F+(1-T-F)W\}\right. \\
&+\gamma_2^2\left(\sigma_x^2+\mu_x^2\right)\{F+(1-T-F)W\}+\theta_2^2\{F+(1-T-F)W\}\{1-\{F+(1-T-F)W\}\}\big] \\
&\leq \tfrac{\theta_2^2}{n_1}\left[\sigma_x^2+\delta_1^2\{F+(1-F)W\}+\gamma_1^2\left(\sigma_x^2+\mu_x^2\right)\{F+(1-F)W\}+\theta_1^2\{F+(1-F)W\}\{1-\{F+(1-F)W\}\}\right] \\
&+\tfrac{\theta_1^2}{n_2}\left[\sigma_x^2+\delta_2^2\{F+(1-F)W\}+\gamma_2^2\left(\sigma_x^2+\mu_x^2\right)\{F+(1-F)W\}+\theta_2^2\{F+(1-F)W\}\{1-\{F+(1-F)W\}\}\right] \\
&\Leftrightarrow \left(n_2\theta_2^2\delta_1^2+n_1\theta_1^2\delta_2^2+n\theta_1^2\theta_2^2\right)\{F+(1-T-F)W-\{F+(1-F)W\}\} \\
&+\left(n_2\theta_2^2\gamma_1^2+n_1\theta_1^2\gamma_2^2\right)\left(\sigma_x^2+\mu_x^2\right)\{F+(1-T-F)W-\{F+(1-F)W\}\} \\
&-\left(n_1\theta_1^2\theta_2^2+n_2\theta_1^2\theta_2^2\right)\left\{(F+(1-T-F)W)^2-\{F+(1-F)W\}^2\right\}\leq 0
\end{aligned}$$

$$\begin{aligned}
&\Leftrightarrow \{F+(1-T-F)W-\{F+(1-F)W\}\}\left[\left(n_2\theta_2^2\delta_1^2+n_1\theta_1^2\delta_2^2+n\theta_1^2\theta_2^2\right)\right. \\
&+\left(n_2\theta_2^2\gamma_1^2+n_1\theta_1^2\gamma_2^2\right)\left(\sigma_x^2+\mu_x^2\right)-n\theta_1^2\theta_2^2\{F+(1-T-F)W+\{F+(1-F)W\}\}\big]\leq 0
\end{aligned}$$

$$\begin{aligned}
&\Leftrightarrow \tfrac{n_1\theta_1^2\left\{\left(\sigma_x^2+\mu_x^2\right)\gamma_2^2+\delta_2^2\right\}+n_2\theta_2^2\left\{\left(\sigma_x^2+\mu_x^2\right)\delta_1^2+\gamma_1^2\right\}}{n\theta_1^2\theta_2^2}+1-\{2F(1-W)+W(2-T)\}\leq 0 \\
&\Rightarrow F \geq \tfrac{1}{2(1-W)}\left[1+W(T-2)+\tfrac{n_1\theta_1^2\left\{\left(\sigma_x^2+\mu_x^2\right)\gamma_2^2+\delta_2^2\right\}+n_2\theta_2^2\left\{\left(\sigma_x^2+\mu_x^2\right)\gamma_1^2+\delta_1^2\right\}}{n\theta_1^2\theta_2^2}\right]=F^{**}
\end{aligned}$$

Note that,

$$F^{**} \le 1 \Leftrightarrow \frac{n_1\theta_1^2\{(\sigma_x^2+\mu_x^2)\gamma_2^2+\delta_2^2\}+n_2\theta_2^2\{(\sigma_x^2+\mu_x^2)\gamma_1^2+\delta_1^2\}}{n\theta_1^2\theta_2^2}+1+W(T-2) \le 2(1-W)$$
$$F^{**} \le 1 \Leftrightarrow T \le \frac{1}{W} - \frac{n_1\theta_1^2\{(\sigma_x^2+\mu_x^2)\gamma_2^2+\delta_2^2\}+n_2\theta_2^2\{(\sigma_x^2+\mu_x^2)\gamma_1^2+\delta_1^2\}}{nW\theta_1^2\theta_2^2}$$

Thus, if the criteria of theorems 3.4 and 3.5 fulfil, we can say that the variance $V(\hat{\mu}_X)$ may be decreased further by using three-stage linear combination model.

Next, we do numerical comparison of variances for the sample size $n = 1000$ and the random variables X, R_1, R_2, S_1 and S_2 have Poisson distribution with means 4, 1, 1, 2 and 5 respectively. Tables 2 and 3 provide variances values of the proposed three-stage linear combination model, Huang [15] model and proposed two-stage model respectively.

From Tables 2 and 3 it observed that, the values of estimated variance are no more than Huang [15] and proposed two-stage linear combination model for various combination values of F and T. It is further noticed that the values of estimated variance decreases as the values of T increases which was expected from theorem 3.4 and 3.5. Thus, the proposed three-stage linear combination model performs better than Huang [15] and proposed two-stage linear combination model.

It is clear from above proceeding after introducing F and T in Huang [15] model, the advantage of using a three-stage linear combination model for questions that seems highly sensitive.

Table 2.$V(\hat{\mu})_{A1}$ of mean estimator (**in bold**) and $V(\hat{\mu})_H$ of mean estimator for different values of F and $W = 0.9$, $n = 1000$, $n_1 = n_2 = 500$, $X \sim P(4)$, $R_1 \sim P(1)$, $R_2 \sim P(1)$, $S_1 \sim P(2)$ and $S_2 \sim P(5)$

F	$T = 0$	$T = 0.1$	$T = 0.3$	$T = 0.5$	$T = 0.7$	$T = 0.9$
0	**0.1598**	**0.1496**	**0.1271**	**0.1018**	**0.0735**	**0.0424**
	0.1598	0.1598	0.1598	0.1598	0.1598	0.1598
0.1	**0.1609**	**0.1508**	**0.1285**	**0.1033**	**0.0752**	**0.0442**
	0.1598	0.1598	0.1598	0.1598	0.1598	0.1598
0.3	**0.1630**	**0.1531**	**0.1311**	**0.1062**	**0.0784**	**0.0478**
	0.1598	0.1598	0.1598	0.1598	0.1598	0.1598
0.5	**0.1651**	**0.1554**	**0.1337**	**0.1091**	**0.0817**	**0.0514**
	0.1598	0.1598	0.1598	0.1598	0.1598	0.1598
0.7	**0.1672**	**0.1576**	**0.1362**	**0.1120**	**0.0849**	**0.0549**
	0.1598	0.1598	0.1598	0.1598	0.1598	0.1598
0.9	**0.1692**	**0.1598**	**0.1387**	**0.1148**	**0.0880**	**0.0583**
	0.1598	0.1598	0.1598	0.1598	0.1598	0.1598

Table 3.$V(\hat{\mu})_{A1}$ of mean estimator (**in bold**) and $V(\hat{\mu})_A$ of mean estimator for different values of F and $W = 0.9$, $n = 1000$, $n_1 = n_2 = 500$, $X \sim P(4)$, $R_1 \sim P(1)$, $R_2 \sim P(1)$, $S_1 \sim P(2)$ and $S_2 \sim P(5)$

F	$T = 0$	$T = 0.1$	$T = 0.3$	$T = 0.5$	$T = 0.7$	$T = 0.9$
0	**0.1598**	**0.1496**	**0.1271**	**0.1018**	**0.0735**	**0.0424**
	0.1598	0.1598	0.1598	0.1598	0.1598	0.1598
0.1	**0.1609**	**0.1508**	**0.1285**	**0.1033**	**0.0752**	**0.0442**
	0.1609	0.1609	0.1609	0.1609	0.1609	0.1609
0.3	**0.1630**	**0.1531**	**0.1311**	**0.1062**	**0.0784**	**0.0478**
	0.1630	0.1630	0.1630	0.1630	0.1630	0.1630
0.5	**0.1651**	**0.1554**	**0.1337**	**0.1091**	**0.0817**	**0.0514**
	0.1651	0.1651	0.1651	0.1651	0.1651	0.1651
0.7	**0.1672**	**0.1576**	**0.1362**	**0.1120**	**0.0849**	**0.0549**
	0.1672	0.1672	0.1672	0.1672	0.1672	0.1672
0.9	**0.1692**	**0.1598**	**0.1387**	**0.1148**	**0.0880**	**0.0583**
	0.1692	0.1692	0.1692	0.1692	0.1692	0.1692

4 Perspective

It is clear from theorems 3.4 and 3.5 that for every values of there exists many values of in the range $[0, 1]$, for which three-stage linear combination model performs better for estimating than corresponding existing estimator. It is important to remark here that for highly sensitive question any large value of by introducing make the three-stage linear combination model more efficient than Huang [15] and proposed two-stage linear combination model. Moreover, three-stage linear scrambling combination protects respondent's anonymity however two-stage does not.

Acknowledgement

The author is grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

[1] Warner, S. L. Randomized response: a survey technique for eliminating evasive answer bias, Journal of the American Statistical Association, **60**, 63-69 (1965).
[2] B. G. Greenberg, A. L. A. Abul-Ela, W. R. Simmons and D. G. Horvitz, The unrelated question randomized response model theoretical framework, Journal of the American Statistical Association **64**, 520-539 (1969).
[3] J. A. Fox and P. E. Tracy, Randomized Response: A method of Sensitive Surveys, Newbury Park, CA: SEGE Publications. (1986).
[4] A. Chaudhuri and R. Mukerjee, Randomized Response: Theory and Techniques, Marcel-Dekker, New York USA. (1988).
[5] J. B. Ryu, K. H. Hong and G. S. Lee, Randomized response model, Freedom Academy, Seoul, Korea (1993).
[6] N. S. Mangat and R. Singh, An alternative randomized procedure, Biometrika, **77**, 439-442 (1990).
[7] D. S. Tracy and N. S. Mangat, Some developments in randomized response sampling during the last decade ? A follow up of review by Chaudhuri and Mukherjee, Journal of Applied Statistical Science, **4**, 147-158 (1996).
[8] D. S. Tracy and S. S. Osahan, An improved randomized response technique,Pakistan Journal of statistics, **15**, 1-6 (1999).
[9] S. Singh, Advanced Sampling Theory with Applications, Kluwer Academic Publishers Dordrecht, the Netherlands (2003).
[10] J. Kerkvliet, Estimating a log it model with randomized data: the case of cocaine use, Australian Journal of Statistics, **36**, 9-20 (1994).
[11] B. G. Greenberg, R. R. Kubler, J. R. Abernathy and D. G. Horvitz, Applications of the RR technique in obtaining quantitative data, Journal of the American Statistical Association **66**, 243-250 (1971).
[12] B. H. Eichhorn and L. S Hayre, Scrambled randomized response methods for obtaining sensitive quantitative data, Journal of Statistical Planning and Inference **1**, 307-316 (1983).
[13] S. N. Gupta, B. C. Gupta and S. Singh, Estimation of sensitivity level of personal interview survey questions, Journal of Statistical Planning and Inference **100**, 239-247 (2002).
[14] S. N. Gupta, J. Shabbir and S. Sehra, Mean and sensitivity estimation in optional randomized response models, Journal of Statistical Planning and Inference, **140**, 2870-2874 (2010).
[15] K. C. Huang, Unbiased estimators of mean, variance and sensitivity level for quantitative characteristics in finite population sampling, Metrika, **71**, 341-352 (2010).
[16] S. N. Gupta, S. Mehta, J. Shabbir and B. K. Dass, Some optimality issues in estimating two-stage optional randomized response models, American Journal of Mathematics and Management Sciences, **31**, 1-12 (2011).
[17] S. N. Gupta, S. Mehta, J. Shabbir and B. K. Dass, Generalized scrambling in quantitative optional randomized response models, Communications in Statistics-Theory and Methods, **42**42, 4034-4042 (2013).
[18] S. Mehta, B. K. Dass, J. Shabbir and S. Gupta, A three-stage optional randomized response model, Journal of Statistical Theory and Practice, **6**, 417-427 (2012).