REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000
https://doi.org/00.00000/revstat.v00i0.000

Incorporating Direct Responses into Optional Randomized Response Models Without Compromising Respondents' Privacy

Authors: Muhammad Azeem ✉
– Department of Statistics, University of Malakand, Chakdara, Pakistan
azeemstats@uom.edu.pk

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- Survey researchers use randomized response models for collecting data in sensitive surveys. The existing optional scrambling models lack the option of direct response. In practice, some of the respondents may be willing to provide the direct response. An efficient randomization technique is proposed where each respondent is free to provide either direct or protected response. The findings reveal that the option of direct response increases the efficiency of the models without compromising on data privacy. A real-life application of the proposed technique is also presented and the improvement in efficiency is shown for various choices of parameters.

Keywords:

- *Efficiency comparison; direct response; protected response; privacy protection; sensitive surveys.*

AMS Subject Classification:

- 62D05.

✉ Corresponding author

1. INTRODUCTION

In data collection on sensitive characteristics such as illegal income, the number of cigarettes consumed in a certain period, cheating in an examination, and the number of violation of rules by employees, refusals and incorrect responses by the respondents is a common practice. The large number of refusals result in high rates of non-response which may badly influence the estimates of population parameter. Attempting to get reliable information from the respondents on sensitive variables, Warner [19] devised a strategy commonly known as the randomized response technique. The Warner's [19] procedure was limited to sample surveys on binary sensitive variables. For estimation of the parameters of quantitative sensitive variables, Warner [20] presented another technique based on an additive scrambling variable. Duncan [7] studied the calculation methods for analyzing mutual information. Eichhorn and Hayre [8] proposed a quantitative scrambling strategy where multiplicative noise is utilized in place of additive noise.
Gupta et al. [10] presented a quantitative randomized response technique called optional randomized response model. In the Gupta et al. [10] technique, the respondents are given the choice to either provide the true response or provide a random response. A multiplicative-type optional scrambling technique was presented by Bar-Lev et al. [5]. Pluim et al. [18] reviewed the medical image processing literature using mutual-information registration. Kraskov et al. [15] suggested two new classes of estimators for mutual information. Yan et al. [21] proposed a method for measuring the respondents' privacy level in a given quantitative randomized response model. Zamanzade and Arghami [23] developed a new estimator of entropy for continuous variables and proved its consistency. In another study, Zamanzade and Mahdizadeh [24] analyzed entropy estimators under ranked set sampling design. Diana and Perri [6] presented an efficient randomized response procedure by utilizing both additive and multiplicative noises. An additive cum subtractive scrambling technique was introduced by Al-Sobhi et al. [1]. Zamanzade and Mahdizadeh [25] studied goodness-of-fit tests under Phi-divergence.

Attempting to measure the overall quality of a given model, Gupta et al. [11] presented a joint measure of privacy level and efficiency. Narjis and Shabbir [17] introduced an efficient quantitative scrambling version of the Gjestvang and Singh [9] technique and proved the improvement over the previous models. The research study of Khalil et al. [14] demonstrated the effect of measurement errors on the mean estimators. Gupta et al. [12] presented an efficient scrambling technique and proved its improvement over the Diana and Perri [6] model. Zhou et al. [27] proposed a new mutual information-driven Pan-sharpening framework. Other research studies related to various aspects of randomized response techniques were conducted by Kalucha et al. [13], Young et al. [22], Murtaza et al. [16], Zhang et al. [26], Azeem et al. [4], and Azeem [2] etc.

Recently, Azeem and Salam [3] introduced an efficient randomized response technique which provides the following two options to the respondents.

(i) Report a direct response,

(ii) Report a scrambled response.

While the Azeem and Salam [3] model achieved the improvement in efficiency of the previous models, one drawback of the Azeem and Salam [3] model was that it forced all of the survey respondents to disclose to the researcher about their choice. The Azeem and Salam [3] technique ignored a real-life situation where some of the respondents may be willing to provide the true protected response. That is, some of the respondents may not want to disclose to the researcher about their chosen option.

The proposed technique is more versatile in the sense that it gives the following two options to each respondent.

(i) Report a direct response (unprotected true response),

(ii) Report a protected response, which further provides two options:
(a) true response, (b) scrambled response.

In other words, the proposed procedure incorporates the option of true response for those who opt for protected response, which was ignored by the Azeem and Salam [3] model.

2. GUPTA et al. [12] OPTIONAL RANDOMIZED RESPONSE MODEL

Let the population under consideration consists of N units and a simple random sample of n units is obtained with replacement. Let the sensitive variable under study be denoted by Y, and let the additive scrambling variable be denoted by S. Moreover, it is assumed that $E(Y_i) = \mu_y$, $E(S) = 0$, $V(Y_i) = \sigma_Y^2$, $V(S) = \sigma_S^2$. Further, let T denote a multiplicative quantitative scrambling variable with $E(T) = 1$, and $V(T) = \sigma_S^2$, where σ_Y^2, σ_T^2 and σ_S^2 are the population variances of variables Y, T, and S, respectively, whereas μ_y denotes the mean of the sensitive quantitative variable Y. Further, it is also assumed that all three variables work independently of one another.
Gupta et al. [12] presented the following quantitative model:

$$Z = \begin{cases} Y & \text{with probability } 1 - W \\ Y + S & \text{with probability } WA \\ TY + S & \text{with probability } W(1 - A), \end{cases} \tag{2.1}$$

where W is the sensitivity level, and A is a constant, $0 < A < 1$. An unbiased estimator of the mean on using the Gupta et al. [12] scrambling model can be written as:

$$\hat{\mu}_G = \frac{1}{n}\sum_{i=1}^{n} Z_i \tag{2.2}$$

The variance of $\hat{\mu}_G$ can be derived as:

$$Var(\hat{\mu}_G) = \frac{1}{n}[W(1-A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_Y^2 + W\sigma_S^2] \tag{2.3}$$

The Gupta et al. [12] optional model gives the following three options to the respondents.

(i) Report the true response (i.e., no scrambling),

(ii) Report the scrambled response using additive scrambling,

(iii) Report the scrambled response using additive and multiplicative scrambling.

3. PROPOSED MODEL

Motivated by Gupta et al. [12], a quantitative randomized response technique is introduced. In the proposed technique, before asking the question on sensitive variable, the interviewer first asks the respondent to select one of the two options - the direct response and the protected response. The researcher records the choice of each response along with the reported value of the quantitative sensitive variable. After the survey is finished, the researcher knows the number of respondents opting for direct responses and the number of those who chose to provide a protected response. Let n_1 out of n be the number of respondents opting for direct response, and let $n_2 = n - n_1$ be the number of respondents opting for protected response for privacy protection. Thus, each respondent belongs to one of the two categories of the respondents:

(i) The n_1 respondents who choose to provide the direct response Y,

(ii) The n_2 respondents who choose to provide a protected response.

The second group further consists of the three types of respondents of the Gupta et al. [12] technique. It only differs in that the group size is n_2 in place of n.

The proposed model offers two options - unprotected/direct response, and randomized response to each respondent before asking the sensitive question. The researcher notes the option chosen by the respondent along with his response to the question being asked. In this way, the values of n_1 and n_2 are random as they vary from sample to sample. That is, the number of respondents opting for the protected response varies from sample to sample. In one sample, 30 out of 100 respondents may choose the direct response option, however, in another sample from the same population, 40 out of 100 respondents may choose the direct response option. As opposed to the existing randomized response models where the researcher doesn't know the exact number of respondents opting for a direct response, the values of n_1 and n_1 in the proposed model are known to the researcher after the completion of the survey. Keeping values of n_1 and n_1 as random values are in line with the real-world situations.
The mean of the first group is:

$$\bar{Y} = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_i \tag{3.1}$$

The mean response based on the second group can be written as:

$$\bar{Z} = \frac{1}{n_2}\sum_{i=1}^{n_2} Z_i \tag{3.2}$$

where Z_i are the reported responses defined in equation 2.1.
The mean estimator can be expressed as the weighted mean of the two groups of respondents. That is,

$$\hat{\mu}_G = \frac{n_1\bar{Y} + n_2\bar{Z}}{n_1 + n_2} \tag{3.3}$$

where $n_1 + n_2 = n$.

4. MEAN AND SAMPLING VARIANCE

The mathematical proofs of the unbiasedness of the mean estimator and the derivation of variances using the proposed model, are given in the following theorems.

Theorem 4.1. *The mean estimator $\hat{\mu}_p$ is unbiased for the population mean .*

Proof: Taking expectation on equation (3.3) gives:

$$E(\hat{\mu}_p) = E(\frac{n_1\bar{Y} + n_2\bar{Z}}{n_1 + n_2}) = \frac{n_1E(\bar{Y}) + n_2E(\bar{Z})}{n_1 + n_2}. \tag{4.1}$$

Taking expectation of equation (3.1) yields:

$$E(\bar{Y}) = E(\frac{1}{n_1}\sum_{i=1}^{n_1} Y_i) = \mu_Y. \tag{4.2}$$

Taking expectation of equation (3.2) yields:

$$E(\bar{Z}) = E(\frac{1}{n_2}\sum_{i=1}^{n_2} Z_i). \tag{4.3}$$

Now,

$$E(Z_i) = (1-W)E(Y) + (WA)E(Y+S) + W(1-A)E(TY+S),$$

or,

$$E(Z_i) = (1-W)\mu_y + (WA)(\mu_y+0) + W(1-A)(\mu_y+0) = \mu_Y. \tag{4.4}$$

Using equation (4.4) in (4.3) yields:

$$E(\bar{Z}) = \frac{1}{n_2}\sum_{i=1}^{n_2} \mu_y = \mu_Y. \tag{4.5}$$

Using equation (4.2) and (4.5) in equation (4.1) yields:

$$E(\hat{\mu_p}) = \frac{n_1\mu_Y + n_1\mu_Y}{n_1+n_2} = \mu_Y. \tag{4.6}$$

□

Theorem 4.2. *The variance of the mean estimators $\hat{\mu_p}$ is given by:*

$$Var(\hat{\mu_p}) = \frac{\sigma_Y^2}{n} + \frac{n_2}{n^2}[W(1-A)\sigma_T^2(\sigma_Y^2+\sigma_T^2) + W\sigma_S^2] \tag{4.7}$$

Proof: Applying variance on equation (3.3) gives:

$$Var(\mu_G) = \frac{n_1^2 Var(\bar{Y}) + n_2^2 Var(\bar{Z})}{(n_1+n_2)^2} \tag{4.8}$$

$$Var(\bar{Y}) = \frac{1}{n_1^2}\sum_{i=1}^{n_1} Var(Y_i) = \frac{\sigma_Y^2}{n_1}. \tag{4.9}$$

Applying variance on both sides of equation (3.2) yields:

$$Var(\bar{Z}) = \frac{1}{n_2^2}\sum_{i=1}^{n_2} Var(Z_i). \tag{4.10}$$

By definition,

$$Var(Z_i) = E(Z_i^2) - [E(Z_i]^2. \tag{4.11}$$

$E(Z_i^2)$ can be simplified as:

$$E(Z_i^2) = (1-W)E(Y^2) + (WA)(Y+S)^2 + W(1-A)(TY+S)^2,$$

or,

$$E(Z_i^2) = (1-W)E(Y^2) + (WA)(Y^2+S^2+2SY) + W(1-A)(T^2Y^2+S^2+2STY).$$

Using independence of variables, and the assumptions given in Section 2, the above equation simplifies to:

$$E(Z_i^2) = (1-W)(\sigma_Y^2+\mu_Y^2) + (WA)(\sigma_Y^2+\mu_Y^2+\sigma_S^2) + [W(1-A)(\sigma_T^2+1)(\sigma_Y^2+\mu_Y^2) + \sigma_T^2+1].$$

On further simplification, the above equation reduces to:

$$E(Z_i^2) = W(1-A)\sigma_T^2(\sigma_Y^2+\mu_Y^2) + \sigma_Y^2 + \mu_Y^2 + W\sigma_S^2. \tag{4.12}$$

Using equation (4.4) and (4.12) in equation (4.11) yields:

$$Var(Z_i) = W(1-A)\sigma_T^2(\sigma_Y^2+\mu_Y^2) + \sigma_Y^2 + W\sigma_S^2. \tag{4.13}$$

Using equation (4.13) in equation (4.10) yields:

$$Var(Z_i) = \frac{1}{n_2}[W(1-A)\sigma_T^2(\sigma_Y^2+\mu_Y^2) + \sigma_Y^2 + W\sigma_S^2]. \tag{4.14}$$

Using equation (4.9) and (4.14) in (4.8) yields:

$$Var(\hat{\mu}_p) = \frac{n_1\sigma_Y^2 + n_2[W(1-A)\sigma_T^2(\sigma_Y^2+\mu_Y^2) + \sigma_Y^2 + W\sigma_S^2]}{(n_1+n_2)^2},$$

or,

$$Var(\hat{\mu}_p) = \frac{(n_1+n_2)\sigma_Y^2 + n_2[W(1-A)\sigma_T^2(\sigma_Y^2+\mu_Y^2) + \sigma_Y^2 + W\sigma_S^2]}{(n_1+n_2)^2},$$

or,

$$Var(\hat{\mu}_p) = \frac{\sigma_Y^2}{n} + \frac{n_2}{n^2}[W(1-A)\sigma_T^2(\sigma_Y^2+\mu_Y^2) + \sigma_Y^2 + W\sigma_S^2].$$

□

Remark: Clearly, the $Var(\hat{\mu}_p)$ is a function of n_2, the number of respondents opting for protected responses. This implies that the less the number of respondents going for protected responses, the more efficient the model is.

5. PERFORMANCE EVALUATION

The measure of respondent-privacy introduced by Yan et al. [21] for evaluation of model-performance can be written as:

$$\Delta = E[Z - Y]^2. \tag{5.1}$$

A higher value of Δ translates to a higher the level of respondent-privacy contained in a given model.
The combined measure under the Gupta et al. [11] model as follows:

$$\delta = \frac{MSE}{\Delta}. \tag{5.2}$$

Equation (5.2) reveals that lower values of δ are preferable.
Based on the Gupta et al. [12] quantitative technique, the measure of privacy is given by:

$$\Delta_G = (1 - A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_S^2. \tag{5.3}$$

The combined measure of model-efficiency and respondent-privacy for the Gupta et al. [12] model can be obtained as:

$$\delta_G = \frac{Var(\hat{\mu_G})}{\Delta_G} = \frac{1}{n}\left[\frac{W(1 - A)\sigma_T^2(\sigma_Y^2 + \sigma_T^2) + W\sigma_S^2}{1 - A)\sigma_T^2(\sigma_Y^2 + \mu_y^2) + \sigma_S^2}\right]. \tag{5.4}$$

Since the proposed technique contains n_2 respondents in the second group in place of n, and since equation (5.3) is independent of the sample size n, so the measure of privacy for the proposed model produces the same quantity as in the Gupta et al. [12] method. That is,

$$\Delta_P = (1 - A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_S^2. \tag{5.5}$$

This means that although the proposed model gives the option of direct response, there is no loss in the privacy protection level.
The joint measure of model-efficiency and respondent-privacy for the proposed technique is given as:

(5.6)

$$\delta_P = \frac{Var(\hat{\mu_P})}{\Delta_P} = \frac{\sigma_S^2}{n[1 - A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_S^2]} + \frac{n_2}{n^2}\left[\frac{W(1 - A)\sigma_T^2(\sigma_Y^2 + \sigma_T^2) + W\sigma_S^2}{1 - A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_S^2}\right].$$

Equation (5.6) is a function of, the number of respondents opting for protected response. This implies that the less the number of respondents going for protected responses, the better the quality of the model is.

6. A PRACTICAL DATA COLLECTION EXAMPLE

The suggested technique was used to estimate the true mean μ_Y of the Grade Point Average (GPA) of students. The population under study consisted

of the 175 students enrolled in the Department of Statistics of the University of Malakand, Pakistan. For selection of the sample, the simple random sampling technique was applied to choose 40 students out of the total 175 students of undergraduate level. Each of the 40 selected students was given the choice to either provide the direct response or go for a protected response. In the case of opting for a protected response, the student had the choice to either provide the true GPA or report a scrambled GPA. If a respondent did not want to report the direct response, he/she was provided with a deck consisting of 100 cards and a calculator. Each card carried two random values – one each for variable T and S. Keeping in view the situation at hand, the random numbers for both of the scrambling variable S and T were generated by the researcher by utilizing a normal distribution. For generation of random numbers for the additive scrambling variable S, a normal distribution was used having mean 0 and variance 0.5. For generation of random numbers for the multiplicative scrambling variable T, a normal distribution was used having mean 1 and variance 0.5. Those respondents who chose to provide a protected response were instructed not to show the card selected by him/her to the interviewer, and hence it was ensured that the respondent's privacy was protected. Out of 40 selected students, 14 students chose to provide the direct response, and the remaining 26 students chose to go for a protected response.
The values of W and A are decided by the researcher based on his/her prior knowledge about proportion of people who feel that the question is of sensitive nature. In the absence of prior information, a pilot survey may be carried out to get an estimate of W and A. In this case, the researcher decided to choose , and , so that , , and . Converting these proportions into percentages, one of the following three statements were recorded on each card.

(i) 60 of 100 cards carried the statement: "Report your true GPA in last exam."

(ii) 20 of 100 cards carried the statement: "Add the value of S to your true GPA and report the result of the addition."

(iii) 20 of 100 cards carried the statement: "Multiply the value of T with your true GPA and then add the value of S and report the result." The students who opted for protected responses were asked to draw one card at random from the 100 cards and add or multiply the numbers as per the instruction on the selected card.

The responses reported by the 40 sampled students are presented in Table 1.

In Table 1, it may be observed that some of the protected responses exceeded 4.0 despite the fact that the students' actual GPA was based on the scale of 4.0. Any observed response greater than 4.0 clearly indicates that the respondent used some sort of scrambling, although the privacy of the respondent

Direct Responses	**Protected Responses**
3.76, 2.43, 2.73, 3.16, 3.15, 2.26, 2.69, 3.30, 2.92, 3.68, 2.88, 1.76, 2.51, 3.28	1.9667, 4.3816, 3.7816, 3.3618, 3.5346, 2.4980, 1.9408, 3.7091, 2.7870, 3.6079, 4.7927, 3.8478, 1.3655, 2.9668, 4.5787, 1.5271, 2.1962, 1.4274, 1.8135, 3.9360, 2.1973, 3.3528, 3.3941, 2.8133, 4.4831, 2.6674

Table 1: Observed Responses.

is protected. To estimate the true mean GPA, one may calculate the weighted mean of the two types of responses given in Table 1. To see the accuracy of the estimator, one may collect the data of the results of all 175 students from department office, calculate the value of population mean, and observe the difference between the estimator and parameter.

7. EFFICIENCY COMPARISON

The variance of the mean estimator on the basis of the Gupta et al. [12] model given in equation (2.3) may be re-written in the form:

$$Var(\hat{\mu}_G) = \frac{\sigma_Y^2}{n} + \frac{n}{n^2}[W(1-A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + W\sigma_S^2]. \tag{7.1}$$

The proposed model will be more efficient than the model suggested by Gupta et al. [12] if:

$$Var(\hat{\mu_p}) \leq Var(\hat{\mu_G}),$$

or,

$$\frac{\sigma_Y^2}{n} + \frac{n_2}{n^2}[W(1-A)\sigma_T^2(\sigma_Y^2+\mu_Y^2)+W\sigma_S^2] \leq \frac{\sigma_Y^2}{n} + \frac{n}{n^2}[W(1-A)\sigma_T^2(\sigma_Y^2+\mu_Y^2)+W\sigma_S^2],$$

or,

$$n_2 \leq n. \tag{7.2}$$

Condition (7.2) always holds. Thus, it is clear that the suggested model is more efficient than the Gupta et al. [12] model. An extreme case where the two models are equally efficient is the situation in which none of the respondents opts for direct response. Otherwise, if at least one respondent opts for direct response, then the suggested technique becomes more efficient than the Gupta et al. [12] model. Since the measure of privacy protection produces the same value for both models, so the proposed model is better when the overall quality of both models is taken into account.

8. GENERALIZATION OF THE FINDINGS TO OTHER MODELS

In addition to the Gupta et al. [12] optional randomized response model, the direct response option can be incorporated into any of the available optional models, and efficiency condition will always hold with no loss in the level of respondent privacy. To see this, let us consider two optional randomized response models: the Gupta et al. [10] and the Bar-Lev et al. [5] technique.
The observed response based on the Gupta et al. [10] model is:

$$Z = \begin{cases} Y & \text{with probability } p \\ Y + S & \text{with probability } 1 - p. \end{cases} \tag{8.1}$$

The mean estimator using the Gupta et al. [10] model can be written as:

$$\hat{\mu}_{G1} = \frac{1}{n} \sum_{i=1}^{n} Z_i, \tag{8.2}$$

where Z is defined in equation (8.1). The sampling variance of the mean can be derived as:

$$Var(\hat{\mu}_{G1}) = \frac{\sigma_Y^2}{n} + \frac{n}{n^2}(1-p)\sigma_S^2. \tag{8.3}$$

If the option of direct responses is used, then the variance of the mean becomes:

$$Var(\hat{\mu}_{p1}) = \frac{\sigma_Y^2}{n} + \frac{n_2}{n^2}(1-p)\sigma_S^2. \tag{8.4}$$

Comparing equation (8.3) and (8.4), since $n_2 \leq n$, therefore:

$$\hat{\mu}_{p1} \leq \hat{\mu}_{G1}. \tag{8.5}$$

The measure of privacy for the Gupta et al. [10] model is:

$$\Delta_{G1} = (1-p)\sigma_S^2. \tag{8.6}$$

Since equation 8.6 is independent of the sample size n, so even if the option of direct responses is given to the respondents, there will be no effect on the measure of privacy.
Now consider the Bar-Lev et al. [5] model, the responses reported by the respondents are as follows:

$$Z = \begin{cases} Y & \text{with probability } p \\ TY & \text{with probability } 1 - p. \end{cases} \tag{8.7}$$

The mean estimator using the Bar-Lev et al. [5] model can be expressed as:

$$\hat{\mu}_B = \frac{1}{n} \sum_{i=1}^{n} Z_i,$$

where Z is defined in equation (8.7). The sampling variance can be derived as:

$$Var(\hat{\mu}_B) = \frac{\sigma_Y^2}{n} + \frac{n}{n^2}(1-p)\sigma_T^2 + (\sigma_Y^2 + \mu_Y^2). \tag{8.8}$$

If the option of direct responses is used, then the variance of the mean becomes:

$$Var(\hat{\mu}_{P2}) = \frac{\sigma_Y^2}{n} + \frac{n_2}{n^2}(1-p)\sigma_T^2 + (\sigma_Y^2 + \mu_Y^2). \tag{8.9}$$

Comparing equation (8.8) and (8.9), since $n_2 \leq n$, therefore:

$$Var(\hat{\mu_{p1}}) \leq Var(\hat{\mu_B}), \tag{8.10}$$

The measure of privacy based on the Bar-Lev et al. [5] model is:

$$\Delta_B = (1-p)\sigma_T^2 + (\sigma_Y^2 + \mu_Y^2). \tag{8.11}$$

Since equation (8.11) doesn't depend on the sample size n, so even if the option of direct responses is given to the respondents, there will be no effect on the measure of privacy.

9. COMPARISON OF MODELS

Table 2 displays the variances of the mean for the proposed model with respect to the Gupta et al. [12] model for different values of n_1 and n_2 . The improvement in efficiency can be clearly observed over the Gupta et al. [12] model. The improvement in efficiency can also be observed graphically in Figure 1, Figure 2, and Figure 3.

10. DISCUSSION AND CONCLUSION

A modified optional randomized response model using direct responses was presented in Section 3. The current study found that the proposed model is more efficient than the Gupta et al. [12] quantitative technique, with the same level of respondent privacy as the Gupta et al. [12] model. Table 2 the improvement in

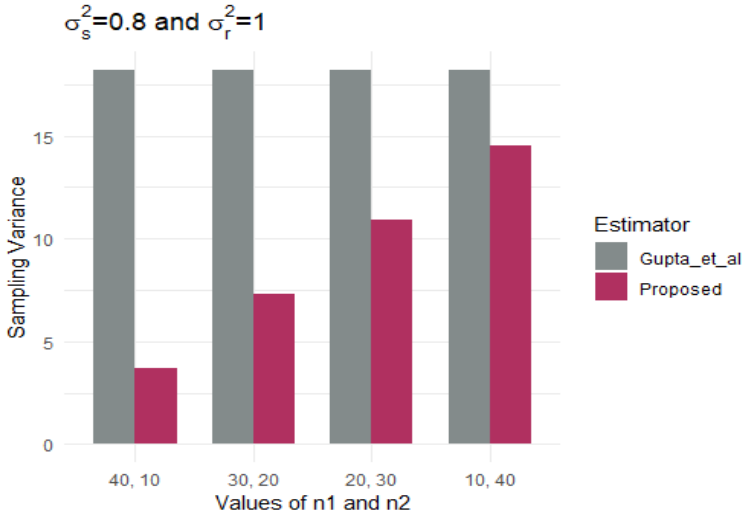

Figure 1: Variance of the Mean under the Gupta et al. [12] and the Proposed Model for $W = 0.2$.

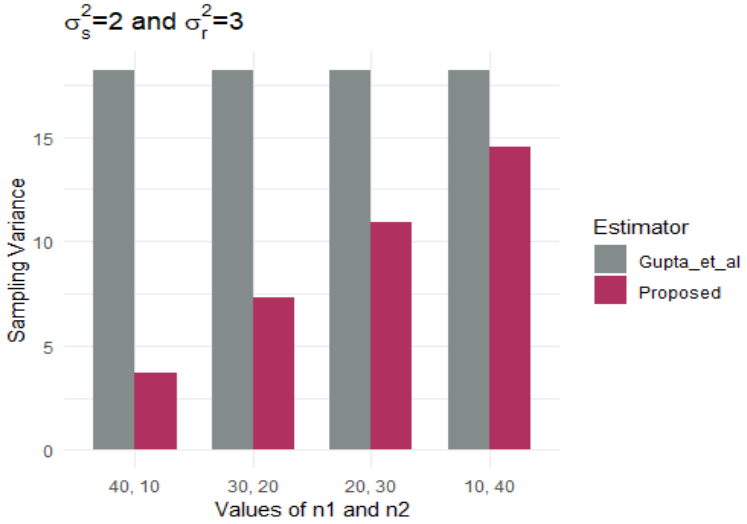

Figure 2: Variance of the Mean under the Gupta et al. [12] and the Proposed Model for $W = 0.5$.

efficiency over the Gupta et al. [12] technique. Table 2 also reveals that as n_2, the number of respondents choosing to provide a protected response, decreases, the efficiency of the proposed model increases. Thus, it is recommended to the researchers to motivate the respondents to report direct responses as far as possible. A larger number of direct responses in a given survey will result in a smaller number of protected responses, which will result in getting efficient estimates of

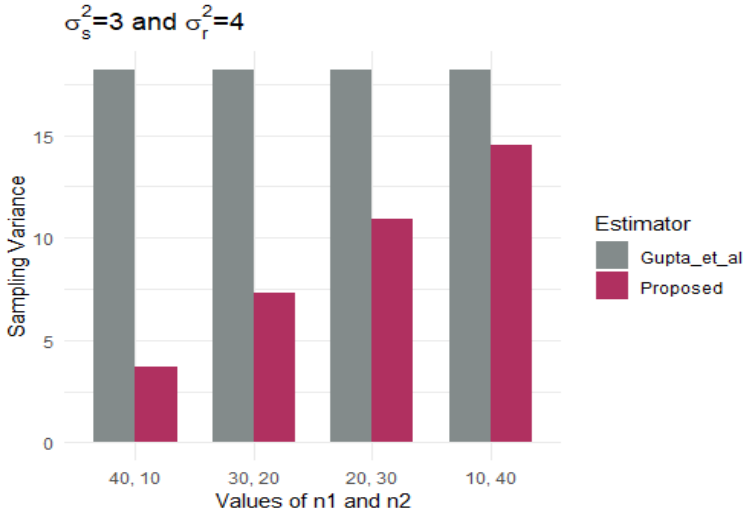

Figure 3: Variance of the Mean under the Gupta et al. [12] and the Proposed Model for $W = 0.8$.

the mean.

In order to compare the overall quality of the proposed and the Gupta et al. [12] model, δ values for both models were calculated and presented in Table 3 for various choices of n_1 and n_2. The δ values for the proposed model are smaller than the Gupta et al. [12] model which indicate that the suggested model is better than the Gupta et al. [12] model. Observing Table 3, it may also be observed that δ value decreases as n_2 decreases which indicate that a survey with lesser number of protected responses, or equivalently, larger number of direct responses, is desirable. In Section 9, it was also observed that the option of direct responses can be incorporated in any of the existing optional randomized response models and will always result in a more efficient model. It was also observed that giving the respondents the option of direct responses doesn't result in the value of Δ, the measure of respondent privacy.

11. FUTURE RESEARCH

The present study analyzed the efficiency of the mean estimator when the option of direct responses is given to the respondents. Further research studies on the estimation of other parameters like population median, and variance can also be conducted under the suggested technique. Moreover, in the current study, the properties of the proposed model have been analyzed under simple random sampling design. Future researchers may use other sampling designs to evaluate

W	A	$n1$	$n2$	$\sigma_s^s = 0.8$ $\sigma_T^s = 1$ $V(\hat{\mu_G})$	$\sigma_s^s = 0.8$ $\sigma_T^s = 1$ $V(\hat{\mu_P})$	$\sigma_s^s = 2$ $\sigma_T^s = 3$ $V(\hat{\mu_G})$	$\sigma_s^s = 2$ $\sigma_T^s = 3$ $V(\hat{\mu_P})$	$\sigma_s^s = 3$ $\sigma_T^s = 4$ $V(\hat{\mu_G})$	$\sigma_s^s = 3$ $\sigma_T^s = 4$ $V(\hat{\mu_P})$
0.2	**0.3**	**10**	**40**	**1.1916**	**0.96528**	**3.4532**	**2.77456**	**4.5856**	**3.68048**
		20	**30**	**1.1916**	**0.73896**	**3.4532**	**2.09592**	**4.5856**	**2.77536**
		30	**20**	**1.1916**	**0.51264**	**3.4532**	**1.41728**	**4.5856**	**1.87024**
		40	**10**	**1.1916**	**0.28632**	**3.4532**	**0.73864**	**4.5856**	**0.96512**
	0.7	**10**	**40**	**0.5468**	**0.44944**	**1.5188**	**1.22704**	**2.0064**	**1.61712**
		20	**30**	**0.5468**	**0.35208**	**1.5188**	**0.93528**	**2.0064**	**1.22784**
		30	**20**	**0.5468**	**0.25472**	**1.5188**	**0.64352**	**2.0064**	**0.83856**
		40	**10**	**0.5468**	**0.15736**	**1.5188**	**0.35176**	**2.0064**	**0.44928**
0.5	**0.3**	**10**	**40**	**2.889**	**2.3232**	**8.543**	**6.8464**	**11.374**	**9.1112**
		20	**30**	**2.889**	**1.7574**	**8.543**	**5.1498**	**11.374**	**6.8484**
		30	**20**	**2.889**	**1.1916**	**8.543**	**3.4532**	**11.374**	**4.5856**
		40	**10**	**2.889**	**0.6258**	**8.543**	**1.7566**	**11.374**	**2.3228**
	0.7	**10**	**40**	**1.277**	**1.0336**	**3.707**	**2.9776**	**4.926**	**3.9528**
		20	**30**	**1.277**	**0.7902**	**3.707**	**2.2482**	**4.926**	**2.9796**
		30	**20**	**1.277**	**0.5468**	**3.707**	**1.5188**	**4.926**	**2.0064**
		40	**10**	**1.277**	**0.3034**	**3.707**	**0.7894**	**4.926**	**1.0332**
0.8	**0.3**	**10**	**40**	**4.5864**	**3.68112**	**13.6328**	**10.9182**	**18.1624**	**14.54192**
		20	**30**	**4.5864**	**2.77584**	**13.6328**	**8.20368**	**18.1624**	**10.92144**
		30	**20**	**4.5864**	**1.87056**	**13.6328**	**5.48912**	**18.1624**	**7.30096**
		40	**10**	**4.5864**	**0.96528**	**13.6328**	**2.77456**	**18.1624**	**3.68048**
	0.7	**10**	**40**	**2.0072**	**1.61776**	**5.8952**	**4.72816**	**7.8456**	**6.28848**
		20	**30**	**2.0072**	**1.22832**	**5.8952**	**3.56112**	**7.8456**	**4.73136**
		30	**20**	**2.0072**	**0.83888**	**5.8952**	**2.39408**	**7.8456**	**3.17424**
		40	**10**	**2.0072**	**0.44944**	**5.8952**	**1.22704**	**7.8456**	**1.61712**

Table 2: Variances of the mean under different models for $\mu_Y = 20, \sigma_Y^2 = 3, n = 50$.

the efficiency of the estimators under the proposed model.

ACKNOWLEDGMENTS

I acknowledge the valuable suggestions from the unknown referees which helped me improve the quality of the paper. I am also thankful to Ms. Irsa Sajjad, PhD Scholar at Central South University, China, for help in preparing the revised version of the paper.

W	A	$n1$	$n2$	$\sigma_s^s = 0.8$ $\sigma_T^s = 1$ δ_G	$\sigma_s^s = 0.8$ $\sigma_T^s = 1$ δ_P	$\sigma_s^s = 2$ $\sigma_T^s = 3$ δ_G	$\sigma_s^s = 2$ $\sigma_T^s = 3$ δ_P	$\sigma_s^s = 3$ $\sigma_T^s = 4$ δ_G	$\sigma_s^s = 3$ $\sigma_T^s = 4$ δ_P
0.2	**0.3**	**10**	**40**	**0.004212**	**0.003412**	**0.004071**	**0.003271**	**0.004053**	**0.003253**
		20	**30**	**0.004212**	**0.002612**	**0.004071**	**0.002471**	**0.004053**	**0.002453**
		30	**20**	**0.003412**	**0.001812**	**0.004071**	**0.001671**	**0.004053**	**0.001653**
		40	**10**	**0.004212**	**0.001012**	**0.004071**	**0.000871**	**0.004053**	**0.000853**
	0.7	**10**	**40**	**0.004493**	**0.003693**	**0.004165**	**0.003365**	**0.004123**	**0.003323**
		20	**30**	**0.004493**	**0.002893**	**0.004165**	**0.002565**	**0.004123**	**0.002523**
		30	**20**	**0.004493**	**0.002093**	**0.004165**	**0.001765**	**0.004123**	**0.001723**
		40	**10**	**0.004493**	**0.001293**	**0.004165**	**0.000965**	**0.004123**	**0.000923**
0.5	**0.3**	**10**	**40**	**0.010212**	**0.008212**	**0.010071**	**0.008071**	**0.010053**	**0.008053**
		20	**30**	**0.010212**	**0.006212**	**0.010071**	**0.006071**	**0.010053**	**0.006053**
		30	**20**	**0.010212**	**0.004212**	**0.010071**	**0.004071**	**0.010053**	**0.004053**
		40	**10**	**0.010212**	**0.002212**	**0.010071**	**0.002071**	**0.010053**	**0.002053**
	0.7	**10**	**40**	**0.010493**	**0.008493**	**0.010165**	**0.008165**	**0.010123**	**0.008123**
		20	**30**	**0.010493**	**0.006493**	**0.010165**	**0.006165**	**0.010123**	**0.006123**
		30	**20**	**0.010493**	**0.004493**	**0.010165**	**0.004165**	**0.010123**	**0.004123**
		40	**10**	**0.010493**	**0.002493**	**0.010165**	**0.002165**	**0.010123**	**0.002123**
0.8	**0.3**	**10**	**40**	**0.016212**	**0.013012**	**0.016071**	**0.012871**	**0.016053**	**0.012853**
		20	**30**	**0.016212**	**0.009812**	**0.016071**	**0.009671**	**0.016053**	**0.009653**
		30	**20**	**0.016212**	**0.006612**	**0.016071**	**0.006471**	**0.016053**	**0.006453**
		40	**10**	**0.016212**	**0.003412**	**0.016071**	**0.003271**	**0.016053**	**0.003253**
	0.7	**10**	**40**	**0.016493**	**0.013293**	**0.016165**	**0.012965**	**0.016123**	**0.012923**
		20	**30**	**0.016493**	**0.010093**	**0.016165**	**0.009765**	**0.016123**	**0.009723**
		30	**20**	**0.016493**	**0.006893**	**0.016165**	**0.006565**	**0.016123**	**0.006523**
		40	**10**	**0.016493**	**0.003693**	**0.016165**	**0.003365**	**0.016123**	**0.003323**

Table 3: Variances of the mean under different models for $\mu_Y = 15, \sigma_Y^2 = 5, n = 50$.

REFERENCES

[1] Al-Sobhi, M.M., HUSSAIN, Z., AL-ZAHRANI, B., SINGH, H.P., and TARRAY, T.A. (2016). Improved randomized response approaches for additive scrambling models, *Mathematical Population Studies*, **23**, 4, 205–221.

[2] Azeem, M. (2023). Using the exponential function of scrambling variable in quantitative randomized response models, *Mathematical Methods in the Applied Sciences*, **46**, 13, 13882–13893.

[3] Azeem, M., and Salam, A. (2023). Introducing an efficient alternative technique to optional quantitative randomized response models , *Methodology*, **19**, 1, 24–42.

[4] Azeem, M., Hussain, S., Ijaz, M., and Salahuddin, N. (2024). An improved quantitative randomized response technique for data collection in sensitive surveys, *Quality and Quantity*, **58**, 1, 329–341.

[5] Bar-Lev, S.K., Bobovitch, E., and Boukai, B. (2004). A note on randomized response models for quantitative data, *Metrika*, **60**, 3, 255–260.

[6] Diana, G., and Perri, P.F. (2011). A class of estimators of quantitative sensitive data. Statistical Papers, *Metrika*, **52**, 3, 633–650.

[7] Duncan, T. E. (1970). On the calculation of mutual information, *Metrika*, **19**, 1, 215-220.

[8] 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.

[9] Gjestvang, C.R., and Sing, S. (2009). An improved randomized response model: Estimation of mean, *Journal of Applied Statistics*, **36**, 12, 1361-1367.

[10] Gupta, S., Gupta, B., and Sing, S. (2002). Estimation of sensitivity level of personal interview survey questions, *Journal of Statistical Planning and Inference*, **100**, 2, 239-247.

[11] Gupta, S., Mehta, S., Shabbir, J., and Khalil, S. (2018). A unified measure of respondent privacy and model efficiency in quantitative rrt models, *Journal of Statistical Theory and Practice*, **12**, 3, 506-511.

[12] Gupta, S., Zhang, J., Khalil, S., and Sapra, P. (2022). Mitigating lack of trust in quantitative randomized response technique models, *Communications in Statistics – Simulation and Computation*, 1-9.

[13] Kalucha, G., Gupta, S., and Shabbir, J. (2016). A two-step approach to ratio and regression estimation of finite population mean using optional randomized response models, *Hacettepe Journal of Mathematics and Statistics*, **45**, 6, 1819-1830.

[14] Khalil, S., Zhang, Q., and Gupta, S (2021). Mean estimation of sensitive variables under measurement errors using optional rrt models, *Communications in Statistics – Simulation and Computation*, **50**, 5, 1417-1426.

[15] Kraskpv, A., Stogbauer, H., and Grassberger, P. (2004). Estimating mutual information, *Physical Review E*, **69**, 6, 066138.

[16] Murtaza, M., Sing, S., and Hussain, Z. (2021). Use of correlated scrambling variables in quantitative randomized response technique, *Biometrical Journal*, **63**, 1, 134-147.

[17] Narjis, G., and Shabbir, J. (2021). An efficient new scrambled response model for estimating sensitive population mean in successive sampling, *Communications in Statistics – Simulation and Computation*, **52**, 11, 5327-5344.

[18] Pluim, J.P., Maintz, J.A., and Viergever, M.A. (2003). Mutual-information-based registration of medical images: a survey, . *IEEE Transactions on Medical Imaging*, **22**, 8, 986-1004.

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

[20] Warner, S.L. (1971). The linear randomized response model, *Journal of the American Statistical Association*, **66**, 336, 884-888.

[21] Yan, Z., Wang, J., and Lai, J (2008). An efficiency and protection degree-based comparison among the quantitative randomized response strategies, *Communications in Statistics – Theory and Methods*, **38**, 3, 400-408.

[22] Young, A., Gupta, S., and Parks, R. (2019). A binary unrelated-question rrt model accounting for untruthful responding , *Involve, A Journal of Mathematics*, **12**, 7, 1163-1173.

[23] Zamanzade, E., and Arghami, N. R (2011). Goodness-of-fit test based on correcting moments of modified entropy estimator, *Journal of Statistical Computation and Simulation*, **81**, 12, 2077-2093.

[24] Zamanzade, E., and Mahdizadeh, M. (2017). Entropy estimation from ranked set samples with application to test of fit, *Revista Colombiana de Estadística*, **40**, 2, 223-241.

[25] Zamanzade, E., and Mahdizadeh, M. (2017). Goodness of fit tests for Rayleigh distribution based on Phi-divergence, *Revista Colombiana de Estadística*, **40**, 2, 279-290.

[26] Zhang, Q., Khalil, S., and Gupta, S (2021). Mean estimation in the simultaneous presence of measurement errors and non-response using optional RRT models under stratified sampling, *Journal of Statistical Computation and Simulation*, **91**, 17, 3492-3504.

[27] Zhou, M., Yan, K., Huang, J., Yang, Z., Fu, X., and Zhao, F (2022). Mutual information-driven pan-sharpening, *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, , 1798-1808.