Heliyon 10 (2024) e31690

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

A novel randomized scrambling technique for mean estimation of a finite population

Muhammad Azeem [a,*], Asadullah [a], Musarrat Ijaz [b], Sundus Hussain [c], Najma Salahuddin [c], Abdul Salam [a]

[a] *Department of Statistics, University of Malakand, Khyber Pakhtunkhwa, Pakistan*
[b] *Department of Statistics, Rawalpindi Women University, Rawalpindi, Pakistan*
[c] *Department of Statistics, Shaheed Benazir Bhutto Women University, Peshawar, Pakistan*

ARTICLE INFO

Keywords:
Efficiency comparison
Mean estimator
Privacy level
Randomized response
Sensitive variable

ABSTRACT

In recent decades, the randomized response technique has attracted researchers due to its usefulness in sensitive surveys. The randomized response procedure is used for the collection of responses on sensitive issues such as cheating in examination, income earned through illegal sources, expenditure on luxury items, and amount of tax paid, etc. This study introduces a new variant of quantitative randomized response models for use with sample surveys where the variable of interest is quantitative. The properties of a mean estimator based on the new technique have been studied. Further, the combined and separate evaluation metrics for efficiency and privacy level have also been derived and compared with those of the existing methods. Further, a simulation study has been conducted to prove the improvement in the degree of privacy protection and efficiency. The findings reveal that the suggested randomized response technique is not only more efficient than the existing techniques, but also improves the joint measure of efficiency and respondents' privacy, making it preferable over the existing techniques. A real-world example of a sample survey through the suggested model is also presented which illustrates its usefulness in practical surveys on sensitive issues.

1. Introduction

In data collection process, a serious issue encountered by almost every researcher is the difficulty in getting truthful responses on sensitive issues such as use of drugs, illegally earned income, cheating in examination, the amount spent on cigarettes per day, and the amount of tax paid etc. In such scenarios, the use of randomized responses can be the only way of obtaining reliable data from the respondents. Warner [1] attempted to cope with sensitive questioning in sample surveys by introducing a privacy-protection technique, commonly called the randomized response technique. The study of Warner [2] introduced a scrambling variable - based randomized response procedure. Gupta et al. [3] suggested an efficient version of the Warner's [2] method, where the survey participants have the choice to either provide the true or a randomized answer, which was based on the fact that some of the respondents may have no problem in reporting the true response. Bar-Lev et al. [4] presented a multiplicative scrambling – based version of the Gupta et al. [3] technique. An efficient randomized response procedure was introduced by Diana and Perri [5], using additive scrambling and multiplicative scrambling in a single technique. By presenting an improved quantitative randomized response model,

* Corresponding author.
E-mail address: azeemstats@uom.edu.pk (M. Azeem).

https://doi.org/10.1016/j.heliyon.2024.e31690
Received 4 March 2024; Received in revised form 20 May 2024; Accepted 20 May 2024
Available online 21 May 2024

Hussain et al. [6] utilized different scrambling methods to prove the improvement. Khalil et al. [7] conducted a study in the analysis of the influence of observational errors on the mean estimators in sample surveys of sensitive-type variables. An enhanced form of optional quantitative randomized response models was also presented by Narjis and Shabbir [8].

The current literature on randomized response techniques offers many variants of the scrambling - based methods to choose from. The researcher's decision to select a particular randomized response technique out of many available options in a given problem of data collection, depends upon two major features: respondents' privacy protection, and efficiency. To quantify the degree of privacy, Yan et al. [9] suggested a quantification method for the evaluation of a given model. The Yan et al. [9] metric was based only on respondents' privacy, and it ignored model efficiency, which is an important consideration when selecting a randomized response technique for real-world surveys. In order to incorporate both privacy and efficiency into a single measure, Gupta et al. [10] presented a joint metric to quantify the degree of privacy and efficiency. Another weighted unified metric of model-evaluation has recently been developed by Azeem [11].

Murtaza et al. [12] suggested a randomization strategy based on correlated scrambling variables. Zhang et al. [13] studied optional randomized response techniques to estimate the population mean under non-response and measurement error. Gupta et al. [14] presented a versatile randomization strategy for efficient estimation of the finite population mean. Shuja et al. [15] studied estimation of the mean under measurement errors in two-phase sampling design. Zapata et al. [16] presented a modified version of the Warner's technique. Kumar and Kour [17] studied the combined effect of non-response and measurement error under optional randomized techniques. Sanaullah et al. [18] suggested a generalized randomized method for use with two-phase sampling designs. Azeem and Ali [19] compared six existing randomized models using various evaluation measures. Azeem et al. [20] developed an efficient modification of the Narjis and Shabbir [8] technique.

Besides estimation of population mean, some recent research studies have presented efficient estimators of population variance based on randomized techniques. In this regard, the study of Gupta et al. [21] motivated survey researchers to explore efficient estimators of variance by using randomized scrambling models. The studies of Aloraini et al. [22], Saleem et al. [23] and Kumar et al. [24] presented efficient estimators of population variance under randomized models. Azeem et al. [25] also utilized a linear scrambling technique to suggest a new estimator of population variance.

In sensitive surveys, there is always a need to develop new randomized response techniques which achieve improvement in efficiency and/or the degree of privacy over the existing techniques. The current study introduces a new optional quantitative model which achieves improved efficiency and offers a higher degree of privacy protection than the available models. We analyze the mathematical properties of the new model and show the improvement over the available models.

Section 2 presents a few available models for comparison. Section 3 presents the new suggested randomized response model. In Section 4, the metrics of privacy and efficiency have been presented for the suggested model and the available models. In Section 5, an example of a real-world survey using the suggested model has been provided. Section 6 presents a comparative analysis of different models for various values of parameters. In Section 7, a simulation study has been conducted and the findings of the simulation have been presented. Section 8 presents the results and discussion based on the current study. Finally, Section 9 presents the conclusion of the study. Suggestions for possible future research have also been given.

2. Selected existing models

Suppose the population under consideration consists of N units and suppose a random sample of size n is selected. We consider the main variable Y and suppose S denotes a random variable. Further, let us use some parametric assumptions $E(Y_i) = \mu_Y$, $E(S) = 0$, $V(Y_i) = \sigma_Y^2$, $V(S) = \sigma_S^2$. Likewise, T and X denote two random variables with $E(T) = 1$, $V(T) = \sigma_T^2$, $E(X) = \theta$, and $Var(X) = \sigma_X^2$. We also assume that all three variables are uncorrelated with each other. Following are some of the existing models in the current literature.

2.1. *The Warner [2] additive technique*

Using the Warner's [2] scrambling technique, the observed response may be expressed as:

$$Z = Y + S. \tag{1}$$

The mean of Y may be unbiasedly estimated by:

$$\widehat{\mu}_W = \frac{1}{n}\sum_{i=1}^{n} Z_i. \tag{2}$$

The sampling variance of the estimator defined in equation (2) may be derived as:

$$Var(\widehat{\mu}_W) = \frac{1}{n}\left(\sigma_Y^2 + \sigma_S^2\right). \tag{3}$$

The result given in equation (3) can be used to evaluate the scrambling technique of Warner [2].

2.2. *Diana and Perri [5] technique*

The observed response using the Diana and Perri [5] randomized response technique may be expressed as:

$$Z = YT + S. \tag{4}$$

In equation (1) and equation (4), Z denotes the response reported by the survey participant. Under the Diana and Perri [5] technique, the population mean may be unbiasedly estimated by:

$$\hat{\mu}_{DP} = \frac{1}{n}\sum_{i=1}^{n} Z_i. \tag{5}$$

The sampling variance of the estimator presented in equation (5) may be derived as:

$$Var(\hat{\mu}_{DP}) = \frac{1}{n}\left[\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_Y^2 + \sigma_S^2\right]. \tag{6}$$

Equation (6) can be used to study the efficiency of the mean in the Diana and Perri [5] technique.

2.3. Gupta et al. [14] Optional technique

Gupta et al. [14] suggested an optional quantitative model, presented as follows:

$$Z = \begin{cases} Y, \text{with probability } 1 - W, \\ S + Y, \text{with probability } WA, \\ YT + S, \text{with probability } W(1 - A). \end{cases} \tag{7}$$

In equation (7), the constant W denotes the level of sensitivity, with A being a constant for which $0 < A < 1$. Based on the Gupta et al. [14] method, the finite population mean of the variable under study can be unbiasedly estimated by:

$$\hat{\mu}_G = \frac{1}{n}\sum_{i=1}^{n} Z_i. \tag{8}$$

The sampling variance of the estimator defined in equation (8) may be obtained as:

$$Var(\hat{\mu}_G) = \frac{1}{n}\left[W(1 - A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_Y^2 + W\sigma_S^2\right]. \tag{9}$$

It may be noted that the sampling variance presented in equation (9) depends upon the sensitivity level W.

2.4. Azeem et al. [25] Model

Azeem et al. [25] suggested a novel quantitative model, presented as follows:

$$Z = \gamma(Y + S) + (1 - \gamma)(Y + YS). \tag{10}$$

In equation (10), γ is a constant such that $0 < \gamma < 1$. The finite population mean can be unbiasedly estimated by:

$$\hat{\mu}_{Az} = \frac{1}{n}\sum_{i=1}^{n} Z_i. \tag{11}$$

The sampling variance of the estimator presented in equation (11) may be obtained as:

$$Var(\hat{\mu}_{Az}) = \frac{1}{n}\left[J\sigma_Y^2 + \left\{\gamma^2 + (1 - \gamma)^2\mu_Y^2\right\}\sigma_S^2\right]. \tag{12}$$

In equation (12), the symbol J is defined as:

$$J = \gamma^2 + (1 - \gamma)^2(\sigma_S^2 + 1).$$

3. Proposed model

Motivated by Gupta et al. [14], we propose a new optional randomized response technique. The reported responses based on the new suggested technique can be expressed as:

$$Z_i = \begin{cases} Y, \text{with probability } 1 - W, \\ Y + S - X, \text{with probability } WA, \\ TY + SX, \text{with probability } W(1 - A). \end{cases} \tag{13}$$

In equation (13), W denotes the sensitivity level, and A is a constant pre-determined by the interviewer on the basis of his/her prior knowledge, such that $0 < A < 1$. Using the proposed technique, the population mean can be unbiasedly estimated by:

$$\hat{\mu}_P = \frac{1}{n}\sum_{i=1}^{n} Z_i + WA\theta. \tag{14}$$

In equation (14), W denotes the sensitivity level, θ is the mean of variable X, and A is a predefined constant.

Theorem 3.1. *The estimator $\hat{\mu}_P$ is unbiased for population mean μ_Y.*

Proof: Appling expectation on equation (11) *gives:*

$$E(\hat{\mu}_P) = E\left(\frac{1}{n}\sum_{i=1}^{n} Z_i + WA\theta\right),$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(Z_i) + WA\theta,$$

$$= E(Z_i) + WA\theta,$$

$$= \mu_Y - WA\theta + WA\theta = \mu_Y.$$

This completes the result.

Theorem 3.2. *The sampling variance of $\hat{\mu}_P$ may be obtained as:*

$$Var(\hat{\mu}_P) = \frac{1}{n}\left[\sigma_Y^2 + WA\left(\sigma_S^2 + \sigma_X^2 + \theta^2 - 2\theta\mu_Y\right) + W(1-A)\left\{\sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right) + \sigma_S^2\left(\sigma_X^2 + \theta^2\right)\right\}\right].$$

Proof: The variance of Z_i is obtained as:

$$Var(Z) = \sigma_z^2 = E\left(Z^2\right) - [E(Z)]^2. \tag{15}$$

We can simplify $E\left(Z^2\right)$ as:

$$E\left(Z^2\right) = (1-W)E\left(Y^2\right) + WAE(Y+S-X)^2 + W(1-A)E(TY+SX)^2, \tag{16}$$

Using the following assumptions:

$$\left.\begin{array}{c} E(Y) = \mu_Y, E(T) = 1, E(S) = 0, E\left(Y^2\right) = \sigma_Y^2 + \mu_Y^2, E\left(T^2\right) = \sigma_T^2 + 1, \\ E\left(S^2\right) = \sigma_S^2, E(X) = \theta, E\left(X^2\right) = \sigma_X^2 + \theta^2, Var(X) = \sigma_X^2. \end{array}\right\} \tag{17}$$

Using equation (17) *in* equation (16) *yields:*

$$E\left(Z^2\right) = \sigma_Y^2 + \mu_Y^2 + WA\left(\sigma_S^2 + \sigma_X^2 + \theta^2 - 2\theta\mu_Y\right) + W(1-A)\left\{\sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right) + \sigma_S^2\left(\sigma_X^2 + \theta^2\right)\right\}, \tag{18}$$

Using equation (18) *in* equation (15) *leads to:*

$$Var(Z) = \sigma_Y^2 + WA\left(\sigma_S^2 + \sigma_X^2 + \theta^2 - 2\theta\mu_Y\right) + W(1-A)\left\{\sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right) + \sigma_S^2\left(\sigma_X^2 + \theta^2\right)\right\}. \tag{19}$$

Taking variance on both sides of equation (14) *yields:*

$$Var(\hat{\mu}_P) = \frac{1}{n^2}\sum_{i=1}^{n} Var(Z). \tag{20}$$

Putting equation (19) *in* equation (20) *gives the required result as:*

$$Var(\hat{\mu}_P) = \frac{1}{n}\left[\sigma_Y^2 + WA\left(\sigma_S^2 + \sigma_X^2 + \theta^2 - 2\theta\mu_Y\right) + W(1-A)\left\{\sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right) + \sigma_S^2\left(\sigma_X^2 + \theta^2\right)\right\}\right]. \tag{21}$$

We can observe that the sampling variance in equation (21) *depends on the sensitivity level W.*

Remark 1. *If the sensitivity level W and the constant A are known, an estimator of the sampling variance may be obtained as:*

$$var(\hat{\mu}_P) = \frac{s_z^2}{n},$$

or

$$var(\hat{\mu}_P) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(z_i - \overline{Z}\right)^2. \tag{22}$$

In equation (22), *$\overline{Z}$ and s_z^2 denote the mean and variance for the sample data, respectively.*

Remark 2. *An algebraic solution of* equation (19) *for* σ_Y^2 *gives*:

$$\sigma_Y^2 = \frac{\sigma_z^2 - WA\left(\sigma_S^2 + \sigma_X^2 + \theta^2 - 2\theta\mu_Y\right) - W(1-A)\left\{\sigma_T^2\mu_Y^2 + \sigma_S^2\left(\sigma_X^2 + \theta^2\right)\right\}}{1 + W(1-A)\sigma_T^2}, \tag{23}$$

where σ_z^2 *denotes the variance of variable Z.*

Using the sample statistics in place of population parameters of Z in equation (23), *we get an estimator of the population variance as*:

$$s_Y^2 = \frac{s_z^2 - WA\left(\sigma_S^2 + \sigma_X^2 + \theta^2 - 2\theta\overline{Z}\right) - W(1-A)\left\{\sigma_T^2\overline{Z}^2 + \sigma_S^2\left(\sigma_X^2 + \theta^2\right)\right\}}{1 + W(1-A)\sigma_T^2}. \tag{24}$$

It may be noted that the estimator defined in equation (24) *depends upon the sensitivity level W.*

4. Model-quality metrics

An important feature of randomized response techniques is to secure the privacy of the survey respondents. At the same time, a randomized response technique should also provide efficient estimates of population parameters. There are various quantitative metrics available to measure the quality of randomized response techniques. In this section, we derive some model evaluation metrics for existing models.

The respondents' privacy level can be quantified as follows:

$$\Delta = E[Z - Y]^2. \tag{25}$$

Another model-quality metric for simultaneous consideration of the variance and privacy is given as follows:

$$\delta = \frac{Var}{\Delta}. \tag{26}$$

Azeem [11] developed a weighted metric of efficiency and privacy as:

$$\log\varphi = \log\left[\frac{w_1(Eff) + w_2(P)}{w_1 + w_2}\right], \tag{27}$$

where w_1 and w_2 are the relative weights of efficiency and privacy, respectively. It may be noted that the measure defined in equation (27) is a unitless measure and a model having $\log\varphi > 0$ is desirable for practical use. For further details, one may refer to the work of Azeem [11]. Moreover, the efficiency and privacy metrics are defined in equation (28) and equation (29).

$$Eff = \frac{Var(\widehat{\mu}_P)}{Var(\widehat{\mu}_i)}, \tag{28}$$

and

$$P = \frac{\Delta_P}{\Delta_i}, \tag{29}$$

for $i = \widehat{\mu}_G, \widehat{\mu}_{DP}, \widehat{\mu}_{Az}$.

Now we derive the evaluation metrics for different models.

Using the Warner's [2] additive technique, the degree of privacy defined in equation (25) may be derived in the form:

$$\Delta_W = E[Y + S - Y]^2 = \sigma_S^2. \tag{30}$$

The δ value defined in equation (26) using the Warner's [2] technique can be written in the form:

$$\delta_W = \frac{Var(\widehat{\mu}_W)}{\Delta_W} = \frac{1}{n}\left[\frac{\sigma_Y^2 + \sigma_S^2}{\sigma_S^2}\right]. \tag{31}$$

The metrics presented in equation (30) and equation (31) can be used in the analysis of model-evaluation.

The level of privacy may be obtained as in the form:

$$\Delta_{DP} = E[TY + S - Y]^2 = \sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right) + \sigma_S^2. \tag{32}$$

The δ value using the Diana and Perri [5] linear technique may be obtained in the form:

$$\delta_{DP} = \frac{Var(\widehat{\mu}_{DP})}{\Delta_{DP}} = \frac{1}{n}\left[\frac{\sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right) + \sigma_S^2 + \sigma_Y^2}{\sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right) + \sigma_S^2}\right]. \tag{33}$$

The metrics presented in equation (32) and equation (33) can be used in the analysis of model-evaluation.

The degree of privacy protection under the Gupta et al. [14] procedure may be obtained in the form:

$$\Delta_G = (1 - A)\left[\sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right)\right] + \sigma_S^2. \tag{34}$$

An enhanced version of the metric defined in equation (34) is a unified metric which can be obtained as follows:

$$\delta_G = \frac{Var(\widehat{\mu}_G)}{\Delta_G} = \frac{1}{n}\left[\frac{W(1 - A)\sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right) + \sigma_Y^2 + W\sigma_S^2}{(1 - A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_S^2}\right]. \tag{35}$$

The metric presented in equation (35) is a joint consideration of privacy and efficiency.

The privacy level metric for the proposed model can be obtained as:

$$\Delta_P = WA\left(\sigma_S^2 + \sigma_X^2 + \theta^2\right) + W(1 - A)\left\{\sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right) + \sigma_S^2\left(\sigma_X^2 + \theta^2\right)\right\}. \tag{36}$$

An enhanced version of the metric defined in equation (36) is a unified metric which can be obtained as follows:

$$\delta_P = \frac{Var(\widehat{\mu}_P)}{\Delta_P},$$

or

$$\delta_P = \frac{1}{n}\left[\frac{\sigma_Y^2 + WA\left(\sigma_S^2 + \sigma_X^2 + \theta^2 - 2\theta\mu_Y\right) + W(1 - A)\left\{\sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right) + \sigma_S^2\left(\sigma_X^2 + \theta^2\right)\right\}}{WA\left(\sigma_S^2 + \sigma_X^2 + \theta^2\right) + W(1 - A)\left\{\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_S^2\left(\sigma_X^2 + \theta^2\right)\right\}}\right]. \tag{37}$$

The metric presented in equation (37) can be used in model-evaluation analysis.

5. Example of data collection through the proposed model

To apply the proposed method to a practical sample survey, a sample of 50 undergraduate students was collected from the presently enrolled students at the University of Malakand. The objective of the study was to analyze the fairness of the evaluation system of the university, in terms of the number of times the students cheated. In order to gather data on cheating in examination, each of the students selected in the sample was offered a set of 100 cards and. Each of the cards presented values of the three random numbers for variable S, T, and X. We generated random numbers for T, X, and S, using a normal distribution. For variable S, the random values were obtained by using a normal population having parameters zero and 16. For variable X, the random numbers were generated from a normal population having mean 3 and variance 10. For variable T, the random values were generated from a normal population having parameters one and 0.5. The researcher utilized his prior knowledge about population to choose the values of the constants W and A. It was decided by the researcher to choose $W = 0.6$, and $A = 0.5$, so that $1 - W = 0.4$, $WA = 0.6 \times 0.5 = 0.3$, and $W(1 - A) = 0.6 \times 0.5 = 0.3$. Based on these values, the reported response using our suggested model maybe expressed as follows:

$$Z_i = \begin{cases} Y & \text{with probability } 0.4 \\ Y + S - X & \text{with probability } 0.3\,. \\ TY + SX & \text{with probability } 0.3, \end{cases} \tag{38}$$

Corresponding to equation (38), each card presented anyone of the following three types of statements.

(i) 40 of 100 cards had the instruction: "Report the true number of times you cheated in examination."
(ii) 30 of 100 cards presented the statement: "Add the value of S with the number of times you cheated and then subtract the value of X and report the number you get."
(iii) 30 of 100 cards presented the statement: "Multiply the value of T with the number of times you cheated and then add the product of S and X and report the number you get."

The students were advised not to disclose the card chosen to the interviewer, thus ensuring the privacy protection of the students. The observed responses by the survey participants are presented in Table 1. We observed that the sample mean of these reported responses is $\overline{Z} = 4.18$ and the variance is $s_z^2 = 28.13$. An estimate of the sampling variance of the mean is calculated as $\text{var}(\widehat{\mu}_P) = \frac{s_z^2}{n} = 0.5626$.

We applied our proposed model to collect the data on cheating in examination. We have observed that in the institution where the data was collected, the cheating behavior of students is approximately normal, as the number of average cheaters is higher than extreme cheaters. This makes it feasible to use a normal distribution to generate the values of the variables S, X, and T. However, in

Table 1
Reported responses.

5.13	4.46	9.76	−3.69	1.83	7.12	−6.23	9.67	−0.55	3.98
11.18	−2.91	5.14	2.81	−2.16	13.07	0.91	−2.86	10.01	−3.14
−0.59	6.39	1.33	8.63	−4.81	6.62	11.69	8.35	9.27	7.98
8.66	−6.36	7.59	−3.90	4.47	−3.12	1.78	5.08	10.96	4.13
8.89	3.98	4.84	9.13	11.86	5.91	8.16	4.86	5.77	−2.27

real-world sample surveys, if it is known that the characteristic of interest follows a non-normal distribution, the researcher can use non-normal distributions to generate the data. Further, we used the proposed model to obtain data on cheating in examination as cheating is generally regarded as a sensitive variable. Survey researchers can, however, apply the proposed model on other sensitive variables.

We suggest survey researchers to be careful when choosing the values of parameters while generating random numbers. In the case of data collection on students' grades on the scale 0 to 4, small values of the parameters will be suitable. Likewise, in data collection on monthly income, the values of the parameters should ideally be large as monthly income is generally measured in thousands of units of local currency. Wrong choice of the values of parameters may lead to misleading results.

6. Comparison of models

In this section, the proposed randomized response model is compared with the available models in terms of separate and joint metrics of the degree of privacy and efficiency. Table 2 displays the sampling variance using the proposed and available models for different values of W, and A, and for select values of the parameters. As far as respondents' privacy protection level is concerned, the values of Δ for various models under study have been presented in Table 3. Likewise, the values of δ for the proposed and already available models have been presented in Table 4. The improvement in terms of efficiency, Δ, and δ can be easily observed, making the proposed model a better choice for practical surveys than the available models.

The values of the weighted metric presented in equation (22) have been computed and presented in Table 7.

7. Comparison using simulation

A simulation study was conducted to compare the performance measures of the new suggested model with the available randomized models. An artificial population of 1000 units was generated by using a normal distribution. The location and scale parameters were set at 30 and 10, respectively. As far as the parameters of the random variables X, S, and T are concerned, we considered different values for the parameters, in line with the assumptions given in Section 2. The sample size was set at $n = 100$ and the variances of the sample mean were simulated over 1000 iterations using the proposed and the available models. In Table 5, the improvement in efficiency can be clearly observed for different sensitivity levels. Table 6 presents the values for Δ and δ using various values of the parameters. One may notice that the values of Δ for the suggested model are much higher than the existing techniques, which indicates that the suggested model offers the highest degree of privacy protection. Likewise, Table 6 clearly indicates that the values of δ are the smallest for the proposed model, which makes the proposed model the best choice for application in real-world surveys related to sensitive quantitative variables. The improvement may also be observed by examining Fig. 1. Likewise, Fig. 2 shows the results of privacy with Fig. 3 presenting the values of the unified metric δ.

8. Results and discussion

The current study presented a new randomized response method which was observed to perform better than the existing models. Our proposed optional model compensates for those respondents who may perceive the question being asked as non-sensitive, and hence they may be willing to report their true response. As opposed to one-variable models, the new proposed model uses two scrambling variables and thus offers a higher level of privacy protection. This is because with two-variable models, the scrambling

Table 2
Variance of the mean using the proposed and existing models for $\theta = 0.5$, $n = 400$, $\sigma_X^2 = 0.5$.

μ_Y	σ_Y^2	W	A	$\sigma_S^2 = 300,\ \gamma = 0.4,\ \sigma_T^2 = 0.3$				$\sigma_S^2 = 600\gamma = 0.8,\ \sigma_T^2 = 0.5$			
				$\widehat{\mu}_{DP}$	$\widehat{\mu}_G$	$\widehat{\mu}_{Az}$	$\widehat{\mu}_P$	$\widehat{\mu}_{DP}$	$\widehat{\mu}_G$	$\widehat{\mu}_{Az}$	$\widehat{\mu}_P$
30	4	0.2	0.3	1.4	0.3	244.2	0.2	2.6	0.5	55.2	0.4
	8	0.5	0.5	1.5	0.6	245.3	0.5	2.7	1.1	55.5	0.9
	12	0.8	0.7	1.5	0.8	246.4	0.7	2.7	1.5	55.7	1.4
50	8	0.2	0.3	2.7	0.4	677.3	0.4	4.7	0.8	151.5	0.7
	12	0.5	0.5	2.7	0.9	678.4	0.8	4.7	1.6	151.7	1.4
	20	0.8	0.7	2.7	1.1	680.5	1.0	4.7	2.0	152.2	1.8
100	10	0.2	0.3	8.3	1.2	2702.8	1.2	14.0	2.1	601.6	2.0
	20	0.5	0.5	8.3	2.3	2705.5	2.2	14.1	3.9	602.2	3.8
	30	0.8	0.7	8.3	2.5	2708.3	2.3	14.1	4.3	602.8	4.1
200	20	0.2	0.3	30.8	4.4	10805.5	4.3	51.6	7.4	2402.2	7.3
	35	0.5	0.5	30.9	8.0	10809.6	7.8	51.6	13.3	2403.1	13.1
	50	0.8	0.7	30.9	7.9	10813.7	7.6	51.7	13.3	2404.0	13.0
500	80	0.2	0.3	188.5	26.6	67521.8	26.5	314.3	44.3	15005.9	44.1
	120	0.5	0.5	188.6	47.6	67532.7	47.2	314.5	79.2	15008.4	78.8
	150	0.8	0.7	188.7	46.0	67540.8	45.3	314.6	76.6	15010.2	75.8
1000	80	0.2	0.3	751.0	105.4	270021.8	105.2	1251.8	175.5	60005.9	175.3
	150	0.5	0.5	751.2	188.3	270040.8	187.6	1252.1	313.7	60010.2	313.0
	200	0.8	0.7	751.6	181.3	270067.9	179.8	1252.4	301.9	60016.4	300.4

Table 3
Values of Δ under the proposed and existing models for $n = 400$.

σ_X^2	W	μ_Y	σ_Y^2	A	$\sigma_T^2 = 0.1,\ \sigma_S^2 = 2,\ \theta = 250$			$\sigma_T^2 = 0.5,\ \sigma_S^2 = 10,\ \theta = 350$		
					Δ_{DP}	Δ_G	Δ_P	Δ_{DP}	Δ_G	Δ_P
10	0.9	30	10	0.9	93.0	10.0	61894.7	465.0	50.0	209541.2
			20	0.8	94.0	18.4	67528.8	470.0	91.8	308815.2
	0.8	60	40	0.9	366.0	30.7	55039.4	1830.0	153.6	186368.0
			50	0.8	367.0	60.0	60069.3	1835.0	300.0	274720.8
60	0.7	90	70	0.9	819.0	58.6	48229.7	4095.0	293.0	163297.1
			80	0.8	820.0	115.9	52666.0	4100.0	579.6	240795.8
	0.6	120	100	0.9	1452.0	88.2	41377.7	7260.0	441.0	140158.8
			110	0.8	1453.0	175.3	45218.3	7265.0	876.6	206776.2
110	0.5	150	130	0.9	2265.0	114.2	34549.6	11325.0	570.8	117049.8
			140	0.8	2266.0	227.4	37793.2	11330.0	1137.0	172790.0
	0.4	180	160	0.9	3258.0	131.0	27679.4	16290.0	655.2	93938.4
			170	0.8	3259.0	261.4	30314.0	16295.0	1306.8	138629.2
160	0.3	210	190	0.9	4431.0	133.5	20811.2	22155.0	667.4	70583.3
			200	0.8	4432.0	266.4	22823.9	22160.0	1332.0	104365.8

Table 4
δ values under the proposed and existing models.

σ_X^2	W	μ_Y	σ_Y^2	A	$\sigma_T^2 = 0.1,\ \sigma_S^2 = 2,\ \theta = 250,\ n = 400$			$\sigma_T^2 = 0.5,\ \sigma_S^2 = 10,\ \theta = 350,\ n = 800$		
					δ_{DP}	δ_G	δ_P	δ_{DP}	δ_G	δ_P
10	0.9	30	10	0.9	0.0028	0.0050	0.0008	0.0013	0.0015	0.0008
			20	0.8	0.0030	0.0052	0.0013	0.0013	0.0015	0.0010
	0.8	60	40	0.9	0.0028	0.0058	0.0010	0.0013	0.0016	0.0008
			50	0.8	0.0028	0.0046	0.0014	0.0013	0.0015	0.0010
60	0.7	90	70	0.9	0.0027	0.0055	0.0012	0.0013	0.0015	0.0009
			80	0.8	0.0027	0.0042	0.0016	0.0013	0.0014	0.0011
	0.6	120	100	0.9	0.0027	0.0053	0.0014	0.0013	0.0015	0.0009
			110	0.8	0.0027	0.0041	0.0017	0.0013	0.0014	0.0011
110	0.5	150	130	0.9	0.0026	0.0053	0.0016	0.0013	0.0015	0.0010
			140	0.8	0.0027	0.0040	0.0018	0.0013	0.0014	0.0011
	0.4	180	160	0.9	0.0026	0.0056	0.0018	0.0013	0.0016	0.0010
			170	0.8	0.0026	0.0041	0.0020	0.0013	0.0014	0.0011
160	0.3	210	190	0.9	0.0026	0.0061	0.0020	0.0013	0.0016	0.0011
			200	0.8	0.0026	0.0044	0.0021	0.0013	0.0014	0.0012
	0.2	240	220	0.9	0.0026	0.0072	0.0022	0.0013	0.0017	0.0012
			230	0.8	0.0026	0.0050	0.0023	0.0013	0.0015	0.0012

process randomizes the responses twice. This makes it difficult for the interviewer to guess the true status of the respondent, thus enhancing the degree of privacy protection. The option of true response in the proposed model also makes it superior to the recently developed model of Azeem et al. [20] which lacks the true response option to the respondents.

The real-world survey example in Section 5 illustrates the practical implementation of the proposed model. We observed that all of the sampled participants reported their response with no issue of non-response whatsoever. We didn't face any hurdle in data collection as the respondents easily calculated their scrambled responses with the help of a calculator or mobile application. The advent of smartphones has made it much easier for the respondents to scramble their responses, compared to the earlier simple models.

In Tables 2 and 5, we observed the superiority of the proposed model over the previous models. Compared to the recently developed model, the proposed model produces more efficient estimators of the population mean. In Table 5, we can see that as the level of sensitivity, W, enhances, the variance of the mean based on our proposed scrambling model also increases. Further, Table 6 reveals that our proposed scrambling model achieves highest simulated values of Δ and smallest values of δ. This makes our proposed model the best of all competitor models presented in Table 6.

Table 7 presents the computed values of the weighted measure $\log \varphi$ using different weights for efficiency by taking different values of parameters. As opposed to the absolute measure presented in Table 2, the $\log \varphi$ is a relative measure of model performance, comparing the proposed model with the Gupta et al. [14] model. We can see that almost all values of $\log \varphi$ are positive which indicates that the proposed model is better than the model of Gupta et al. [14].

In almost every practical survey, the researchers need improvement not only in respondent privacy but also in model efficiency, and thus it is desirable to compare randomized response models using the unified metric δ. Table 4 shows that the suggested randomized response model achieves smaller values of δ than the previous models. On the basis of the findings of this study, we recommend the suggested quantitative model for use in data collection on sensitive characteristics.

Table 5
Simulated variances for $N = 1000$, $n = 100$, $\mu_Y = 30$, $\sigma_Y^2 = 10$.

Parameters	γ	W	A	$Var(\hat{\mu}_G)$	$Var(\hat{\mu}_{DP})$	$Var(\hat{\mu}_{Az})$	$Var(\hat{\mu}_P)$
$\sigma_S^2 = 100$, $\sigma_X^2 = 0.1$, $\sigma_T^2 = 0.1$, $\theta = 0.2$	0.3	0.1	0.4	0.3229	2.0710	458.8490	0.2642
			0.8	0.2927	2.2739	473.6136	0.2719
	0.4	0.3	0.4	0.6659	2.2154	343.8397	0.5243
			0.8	0.5266	2.0710	342.4781	0.4845
	0.5	0.5	0.4	1.0055	2.2421	259.7957	0.7180
			0.8	0.7611	2.1127	247.6804	0.6617
	0.6	0.7	0.4	1.3217	2.1740	162.7971	0.9201
			0.8	1.0247	2.1604	166.196	0.9335
	0.7	0.9	0.4	1.4895	2.1706	90.5899	1.1047
			0.8	1.2823	2.1958	104.9238	1.1149
$\sigma_S^2 = 200$, $\sigma_X^2 = 0.3$, $\sigma_T^2 = 0.3$, $\theta = 0.4$	0.3	0.1	0.4	0.5258	4.8281	917.0671	0.4563
			0.8	0.4312	5.4026	947.0367	0.4047
	0.4	0.3	0.4	1.3384	5.1696	687.3942	1.1687
			0.8	0.9346	4.8281	684.3876	0.8957
	0.5	0.5	0.4	2.1739	5.3218	519.6751	1.7489
			0.8	1.4628	5.0528	495.5827	1.3363
	0.6	0.7	0.4	2.8395	5.0774	325.3101	2.2557
			0.8	2.0223	5.0855	332.2354	1.9316
	0.7	0.9	0.4	3.3021	5.1866	181.1832	2.7680
			0.8	2.5852	5.2193	209.7817	2.3429
$\sigma_S^2 = 300$, $\sigma_X^2 = 0.5$, $\sigma_T^2 = 0.5$, $\theta = 0.5$	0.3	0.1	0.4	0.7308	7.5857	1375.21	0.6782
			0.8	0.5688	8.5261	1420.457	0.5477
	0.4	0.3	0.4	2.0117	8.1091	1030.932	1.9100
			0.8	1.3435	8.1091	1026.232	1.3313
	0.5	0.5	0.4	3.3454	8.3988	779.5965	2.9428
			0.8	2.1685	7.9971	743.5494	2.0644
	0.6	0.7	0.4	4.3500	7.9716	487.8061	3.7849
			0.8	3.0204	8.0092	498.2778	3.0088
	0.7	0.9	0.4	5.1175	8.2020	271.8079	4.6712
			0.8	3.8913	8.2529	314.658	3.6557

Table 6
Simulated Δ and δ values under various models for $N = 1000$, $n = 100$, $\mu_Y = 30$, $\sigma_Y^2 = 10$.

Parameters	γ	W	A	Δ_G	Δ_{DP}	Δ_{Az}	Δ_P	δ_G	δ_{DP}	δ_{Az}	δ_P
$\sigma_S^2 = 2$, $\sigma_X^2 = 10$, $\sigma_T^2 = 0.1$, $\theta = 100$	0.3	0.1	0.4	5.693	94.699	934.743	1595.724	0.021	0.003	0.002	0.001
			0.8	2.021	94.923	932.602	1199.841	0.051	0.003	0.002	0.001
	0.4	0.3	0.4	17.234	94.893	699.111	4821.444	0.008	0.003	0.002	0.002
			0.8	6.068	94.699	697.323	3604.466	0.020	0.003	0.002	0.001
	0.5	0.5	0.4	28.735	94.673	494.830	8002.526	0.006	0.003	0.002	0.001
			0.8	10.261	94.611	494.565	6020.684	0.013	0.003	0.002	0.001
	0.6	0.7	0.4	40.281	94.805	326.416	11272.140	0.005	0.003	0.003	0.002
			0.8	14.661	94.654	327.469	8426.028	0.009	0.003	0.003	0.001
	0.7	0.9	0.4	52.228	94.983	193.278	14532.760	0.004	0.003	0.003	0.002
			0.8	18.648	94.837	193.821	10902.620	0.008	0.003	0.003	0.001
$\sigma_S^2 = 6$, $\sigma_X^2 = 30$, $\sigma_T^2 = 0.3$, $\theta = 150$	0.3	0.1	0.4	17.080	284.097	2804.230	8949.912	0.009	0.003	0.003	0.002
			0.8	6.062	284.770	2797.807	4490.515	0.019	0.003	0.002	0.001
	0.4	0.3	0.4	51.701	284.681	2097.336	27087.460	0.004	0.003	0.003	0.002
			0.8	18.204	284.097	2091.970	13503.520	0.008	0.003	0.002	0.001
	0.5	0.5	0.4	86.205	284.020	1484.490	44924.780	0.004	0.003	0.003	0.002
			0.8	30.783	283.835	1483.696	22587.820	0.006	0.003	0.002	0.001
	0.6	0.7	0.4	120.843	284.416	979.250	63458.850	0.003	0.003	0.003	0.002
			0.8	43.982	283.963	982.407	31719.540	0.005	0.003	0.003	0.001
	0.7	0.9	0.4	156.685	284.949	579.834	81713.280	0.003	0.003	0.003	0.002
			0.8	55.944	284.512	581.464	41100.120	0.004	0.003	0.002	0.001
$\sigma_S^2 = 10$, $\sigma_X^2 = 50$, $\sigma_T^2 = 0.5$, $\theta = 200$	0.3	0.1	0.4	28.467	473.495	4673.717	25427.310	0.006	0.003	0.002	0.002
			0.8	10.104	474.617	4663.012	11162.610	0.012	0.003	0.002	0.001
	0.4	0.3	0.4	86.168	474.468	3495.560	76986.840	0.004	0.003	0.003	0.002
			0.8	30.340	473.495	3486.616	33578.050	0.006	0.003	0.002	0.001
	0.5	0.5	0.4	143.675	473.368	2474.15	127665.100	0.003	0.003	0.003	0.002
			0.8	51.305	473.059	2472.826	56199.480	0.005	0.003	0.002	0.001
	0.6	0.7	0.4	201.405	474.028	1632.085	180436.600	0.003	0.003	0.003	0.002
			0.8	73.303	473.272	1637.346	79030.780	0.004	0.003	0.003	0.002
	0.7	0.9	0.4	261.141	474.916	966.3912	232282.000	0.003	0.003	0.003	0.002
			0.8	93.240	474.187	969.107	102467.000	0.003	0.003	0.002	0.001

Table 7
Values of the weighted measure log φ for the proposed model relative to the Gupta et al. [14] model for $\theta = 0.2$, $A = 0.8$, $n = 400$, $\sigma_X^2 = 5$, $\sigma_S^2 = 300$, $\gamma = 0.2$, $\sigma_T^2 = 0.1$.

μ_Y	σ_Y^2	w_1	w_2	$W = 0.5$		$W = 0.8$	
				$\theta = 20$	$\theta = 50$	$\theta = 20$	$\theta = 50$
50	5	0.2	0.8	1.454	2.241	1.658	2.446
		0.4	0.6	1.329	2.116	1.534	2.321
		0.6	0.4	1.154	1.940	1.358	2.145
		0.8	0.2	0.853	1.639	1.057	1.843
100	10	0.2	0.8	1.302	2.087	1.506	2.291
		0.4	0.6	1.177	1.962	1.381	2.166
		0.6	0.4	1.002	1.786	1.205	1.990
		0.8	0.2	0.702	1.485	0.905	1.689
200	15	0.2	0.8	0.970	1.746	1.174	1.950
		0.4	0.6	0.846	1.621	1.050	1.825
		0.6	0.4	0.672	1.445	0.875	1.649
		0.8	0.2	0.376	1.144	0.577	1.348
500	25	0.2	0.8	0.367	1.075	0.566	1.279
		0.4	0.6	0.266	0.951	0.456	1.155
		0.6	0.4	0.133	0.776	0.308	0.979
		0.8	0.2	−0.060	0.478	0.082	0.681
1000	40	0.2	0.8	0.079	0.537	0.238	0.739
		0.4	0.6	0.112	0.421	0.228	0.620
		0.6	0.4	0.142	0.263	0.218	0.455
		0.8	0.2	0.170	0.011	0.208	0.186

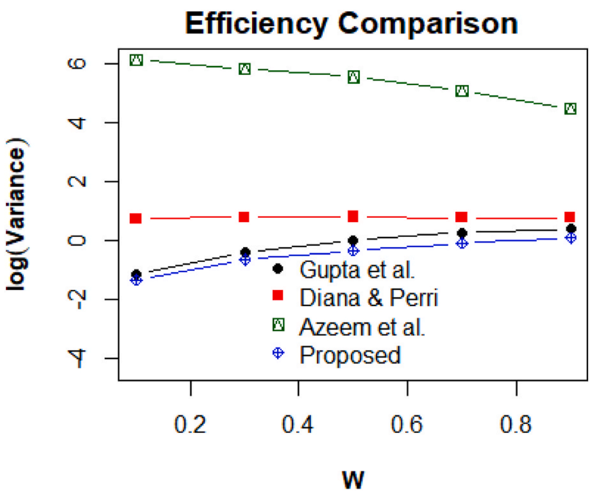

Fig. 1. Efficiency comparison.

9. Conclusion

Observing Table 2, one may clearly notice that the suggested quantitative randomized response model performs more precisely than both the Gupta et al. [14] and the Diana and Perri [5] models. It is also clearly noticeable from Table 2 that, as the respondents' level of sensitivity, W, decreases, the variance of the mean under the suggested randomized response model increases. Moreover, it is also clear that as the value of A increases, the variance of the mean under the suggested model also increases. As far as respondent-privacy is concerned, one may clearly observe the improvement over the available models from Table 3. It is also observed from Table 3 that as the value of A increases, the value of Δ decreases, which indicates that smaller values of A are preferable for better privacy protection. The improvement over the Gupta et al. [14] model is also clear from the simulation study results presented in Tables 5 and 6. It is also observed from Table 4 that as the value of A increases, the value of δ also increases, which indicates that smaller values of A are preferable.

For future research, we recommend survey statisticians to analyze the combined effect of various types of non-sampling errors on the estimators under the suggested model. Further, auxiliary variables can be used under the proposed technique to achieve further improvement in efficiency.

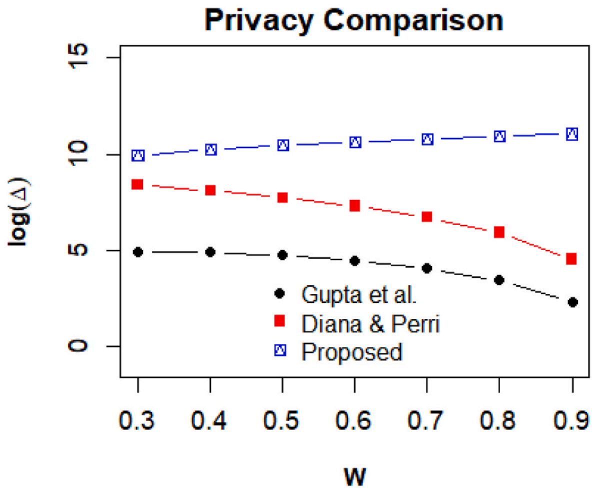

Fig. 2. Privacy comparison.

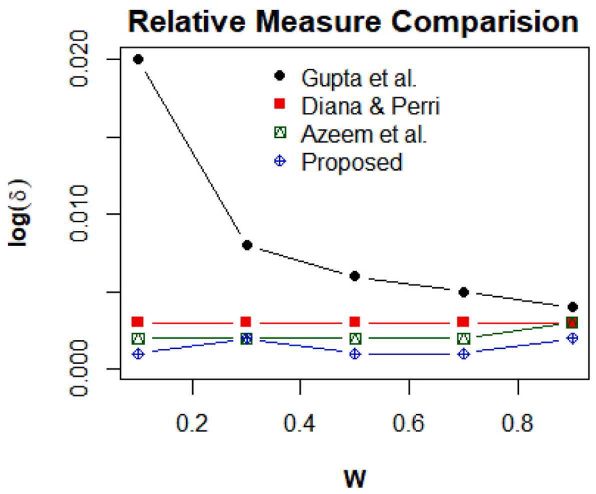

Fig. 3. Comparison of δ values.

Funding source

The authors received no funding for this study.

Ethics and consent

The institution where the data was collected from students did not require ethics approval.

Data availability statement

The data associated with our study has not been deposited into a publicly available repository. All relevant data is available within the article and its references.

CRediT authorship contribution statement

Muhammad Azeem: Writing – original draft, Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Asadullah:** Writing – review & editing, Software, Methodology, Investigation, Formal analysis, Data curation. **Musarrat Ijaz:** Writing – review & editing, Validation, Resources, Investigation, Data curation. **Sundus Hussain:** Writing – review & editing, Validation, Software, Investigation. **Najma Salahuddin:** Writing – review & editing, Visualization, Software, Methodology. **Abdul**

Salam: Validation, Software, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S.L. Warner, Randomized response: a survey technique for eliminating evasive answer bias, J. Am. Stat. Assoc. 60 (309) (1965) 63–69, https://doi.org/10.1080/01621459.1965.10480775.
[2] S.L. Warner, The linear randomized response model, J. Am. Stat. Assoc. 66 (336) (1971) 884–888, https://doi.org/10.1080/01621459.1971.10482364.
[3] S. Gupta, B. Gupta, S. Singh, Estimation of sensitivity level of personal interview survey questions, J. Stat. Plann. Inference 100 (2) (2002) 239–247, https://doi.org/10.1016/S0378-3758(01)00137-9.
[4] S.K. Bar-Lev, E. Bobovitch, B. Boukai, A note on randomized response models for quantitative data, Metrika 60 (3) (2004) 255–260.
[5] G. Diana, P.F. Perri, A class of estimators of quantitative sensitive data, Stat. Pap. 52 (3) (2011) 633–650, https://doi.org/10.1007/s00362-009-0273-1.
[6] Z. Hussain, M.M. Al-Sobhi, B. Al-Zahrani, H.P. Singh, T.A. Tarray, Improved randomized response approaches for additive scrambling models, Math. Popul. Stud. 23 (4) (2016) 205–221, https://doi.org/10.1080/08898480.2015.1087773.
[7] S. Khalil, Q. Zhang, S. Gupta, Mean estimation of sensitive variables under measurement errors using optional rrt models, Commun. Stat. Simulat. Comput. 50 (5) (2021) 1417–1426, https://doi.org/10.1080/03610918.2019.1584298.
[8] G. Narjis, J. Shabbir, An efficient new scrambled response model for estimating sensitive population mean in successive sampling, Commun. Stat. Simulat. Comput. 52 (11) (2023) 5327–5344, https://doi.org/10.1080/03610918.2021.1986528.
[9] Z. Yan, J. Wang, J. Lai, An efficiency and protection degree-based comparison among the quantitative randomized response strategies, Commun. Stat. Theor. Methods 38 (3) (2008) 400–408, https://doi.org/10.1080/03610920802220785.
[10] S. Gupta, S. Mehta, J. Shabbir, S. Khalil, A unified measure of respondent privacy and model efficiency in quantitative rrt models, Journal of Statistical Theory and Practice 12 (3) (2018) 506–511, https://doi.org/10.1080/15598608.2017.1415175.
[11] M. Azeem, Introducing a weighted measure of privacy and efficiency for comparison of quantitative randomized response models, Pakistan Journal of Statistics 39 (3) (2023) 377–385.
[12] M. Murtaza, S. Singh, Z. Hussain, Use of correlated scrambling variables in quantitative randomized response technique, Biom. J. 63 (1) (2021) 134–147.
[13] Q. Zhang, S. Khalil, S. Gupta, Mean estimation in the simultaneous presence of measurement errors and non-response using optional RRT models under stratified sampling, J. Stat. Comput. Simulat. 91 (17) (2021) 3492–3504, https://doi.org/10.1080/00949655.2021.1941018.
[14] S. Gupta, J. Zhang, S. Khalil, P. Sapra, Mitigating lack of trust in quantitative randomized response technique models, Commun. Stat. Simulat. Comput. (2022) 1–9, https://doi.org/10.1080/03610918.2022.2082477.
[15] U. Shuja, S. Khalil, S. Gupta, Mean estimation using scrambled responses with the impact of measurement errors in two-phase sampling, J. Interdiscipl. Math. 25 (6) (2022) 1675–1695, https://doi.org/10.1080/09720502.2021.1968581.
[16] Z. Zapata, S.A. Sedory, S. Singh, An innovative improvement in Warner's randomized response device for evasive answer bias, J. Stat. Comput. Simulat. 93 (2) (2023) 298–311, https://doi.org/10.1080/00949655.2022.2101649.
[17] S. Kumar, S.P. Kour, The joint influence of estimation of sensitive variable under measurement error and non-response using ORRT models, J. Stat. Comput. Simulat. 92 (17) (2022) 3583–3604.
[18] A. Sanaullah, I. Saleem, S. Gupta, M. Hanif, Mean estimation with generalized scrambling using two-phase sampling, Commun. Stat. Simulat. Comput. 51 (10) (2022) 5643–5657, https://doi.org/10.1080/03610918.2020.1778032.
[19] M. Azeem, S. Ali, A neutral comparative analysis of additive, multiplicative, and mixed quantitative randomized response models, PLoS One 18 (4) (2023) e0284995, https://doi.org/10.1371/journal.pone.0284995.
[20] M. Azeem, S. Hussain, M. Ijaz, N. Salahuddin, An improved quantitative randomized response technique for data collection in sensitive surveys, Qual. Quantity 58 (2024) 329–341, https://doi.org/10.1007/s11135-023-01652-5.
[21] S. Gupta, B. Aloraini, M.N. Qureshi, S. Khalil, Variance estimation using randomized response technique, REVSTAT – Statistical Journal 18 (2) (2020) 165–176.
[22] B. Aloraini, S. Khalil, M.N. Qureshi, S. Gupta, Estimation of population variance for a sensitive variable in stratified sampling using randomized response technique, REVSTAT-Statistical Journal (2022) 1–18. https://revstat.ine.pt/index.php/REVSTAT/article/view/508.
[23] I. Saleem, A. Sanaullah, L.A. Al-Essa, S. Bashir, Efficient estimation of population variance of a sensitive variable using a new scrambling response model, Sci. Rep. 13 (2023), https://doi.org/10.1038/s41598-023-45427-2. Article Number 19913.
[24] S. Kumar, S.P. Kour, H.P. Singh, Applying ORRT for the estimation of population variance of sensitive variable, Commun. Stat. Simulat. Comput. (2023) 1–11, https://doi.org/10.1080/03610918.2023.2292966.
[25] M. Azeem, N. Salahuddin, S. Hussain, M. Ijaz, A. Salam, An efficient estimator of population variance of a sensitive variable with a new randomized response technique, Heliyon 10 (5) (2024) e27488, https://doi.org/10.1016/j.heliyon.2024.e27488.