An Effective Quantitative Randomized Response Survey Approach Employing Forced Answers

Habib Ullah Khan[1], Atta Ullah[2], Uzair Khan[3] and Wasif Khan[4]

https://doi.org/10.62345/jads.2025.14.2.71

Abstract

Over the past few decades, survey researchers have become increasingly interested in the randomized response technique. Numerous disciplines, including sociology, psychology, business, and education, have used the randomized response technique. Researchers can use it to address high non-response rates in sample surveys on delicate subjects such as monthly income, traffic infractions, income tax amount, and luxury spending. An optimal scrambling strategy that includes forced replies is presented in this research. The population mean estimator developed using the provided technique has been found to exhibit algebraic characteristics. The different model evaluation metrics were acquired and contrasted with existing models. The findings of this study indicate that the new quantitative method based on forced responses is more efficient than the current scrambling methods. Additionally, it combines model efficiency with a higher level of privacy, making it particularly valuable for practical applications. This paper also presents a helpful example of how this method can be utilized in sensitive sample surveys.

Keywords: Efficiency, Forced Responses, Privacy Protection, Sensitive Surveys.

Introduction

In almost every study, researchers encounter refusals and fake responses from survey participants. Non-response and false reporting are especially common in sample surveys with sensitive questions, such as those about illegal income, exam cheating, drug use, smoking habits, and luxury spending, among others. In an attempt to overcome non-response in sample surveys, Warner (1965) offered randomized responses. The Warner (1965) approach was intended to be used in sample surveys with binary data. Warner (1971) presented an adaptation of the original randomization method for quantitative variable-based sensitivity surveys. Gupta et al. (2002) suggested an optional version of the quantitative randomized response technique, allowing survey participants to report the proper response. Gjestvang and Singh (2009) presented a randomization method for data collection in sample surveys. Narjis and Shabbir (2023) developed an optimized form of Gjestvang and Singh's (2009) model that incorporates the valid response option. Diana and Perri (2011) proposed a quantitative scrambling algorithm that combines additive and multiplicative scrambling.

Mehta et al. (2012) suggested a three-stage scrambling technique and evaluated its properties. Gupta et al. (2013) proposed generalized scrambling algorithms for sensitive surveys. Gupta et al.

[1] *Lecturer, Department of Statistics, University of Malakand. Email: habibullah858@uom.edu.pk*
[2] *Department of Statistics, University of Malakand. Correspondence Author Email: attaullaharsalan34@gmail.com*
[3] *Department of Statistics, University of Malakand.*
[4] *Department of Statistics, University of Malakand.*

(2014) used a non-sensitive auxiliary variable to estimate the mean using various scrambling algorithms. Hussain et al. (2016) developed an optimal variation of quantitative scrambling models that employ additive noises. Khalil et al. (2021) investigated the effect of observational error on parameter estimating in sensitive surveys.

There are many different kinds of quantitative models for randomized responses available in the literature. In any particular data collection task including sensitive features, the technique chosen is governed by two major factors: survey participant privacy and model efficiency. These two aspects influence the efficacy of a quantitative technique. Gupta et al. (2018) proposed a new metric for evaluating survey methods. For other assessments of randomized response approaches in various contexts, readers can refer to Yan et al. (2008), Young et al. (2019), Murtaza et al. (2020), Zhang et al. (2021), Azeem and Ali (2023), and Azeem et al. (2024).

Gupta et al. (2022) developed an optional scrambling procedure that allowed survey respondents to report their results using either additive or multiplicative scrambling. Gupta et al. (2022) showed that their scrambling method outperformed Diana and Perri's (2011) model in terms of respondent privacy and model precision. Azeem (2023) recently introduced a single weighted metric for evaluating models.

Available Models

Consider a population consisting of N units, and let a sample of n elements be selected using a sampling with replacement design. Let us denote the main sensitive variable by Y, and let S denote a random variable. Let us also assume that $E(Y_i) = \mu_Y, V(Y_i) = \sigma_Y^2$, $E(S) = \theta$, and $V(S) = \sigma_S^2$. In addition, let T be another random variable with $E(T) = 1$, and. Let F be a static response established by the interrogator based on previous knowledge or a pilot survey. Let $E(F) = F$, and $Var(F) = 0$. To protect privacy, we assume that the variables are independent of one another.

Bar-Lev et al. (2004) Technique

The response function is described as under:

$$Z_B = \begin{cases} Y & \text{with probablity } P = \dfrac{\alpha}{\alpha+\beta} \\ TY & \text{with probablity } (1-P) = \dfrac{\alpha}{\alpha+\beta} \end{cases} \tag{1}$$

The formula for the estimation of mean might be express as:

$$\hat{\mu}_B = \frac{1}{n}\sum_{i=1}^{n} Z_i \tag{2}$$

The formula of variance as follows:

$$Var(\hat{\mu}_B) = \frac{1}{n}\left[\sigma_Y^2 + \sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right)\right] \tag{3}$$

Gjestvang and Singh (2009) Technique

The response function is described as under:

$$Z_{GS} = \begin{cases} Y_i + \alpha S & \text{with probablity } P = \dfrac{\beta}{\alpha+\beta} \\ Y_i - \beta S & \text{with probablity } 1-P = \dfrac{\alpha}{\alpha+\beta} \end{cases} \tag{4}$$

The mean estimator may be expressed as:

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

The formula of variance is as follows:

$$Var(\hat{\mu}_{GS}) = \frac{1}{n}\left[\sigma_Y^2 + \alpha\beta\left(\sigma_S^2 + \theta^2\right)\right] \tag{6}$$

Narjis and Shabbir (2023) Technique

The model is described as follows:

$$Z_{NS} = \begin{cases} Y_i + \alpha S & \text{with probablity } P_1 = \dfrac{\beta}{\alpha+\beta+\lambda} \\ Y_i - \beta S & \text{with probablity } P_2 = \dfrac{\alpha}{\alpha+\beta+\lambda} \\ Y_i & \text{with probablity } P_3 = \dfrac{\lambda}{\alpha+\beta+\lambda} \end{cases} \tag{7}$$

The estimator of mean might be express as:

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

The formula of variance is as follows:

$$Var(\hat{\mu}_{NS}) = \frac{1}{n}\left[\frac{\alpha\beta\left(\sigma_S^2 + \theta^2\right)}{\alpha+\lambda+\beta} + \sigma_Y^2\right] \tag{9}$$

The Gupta et al. (2024) Technique

The model is described as follows:

$$Z_G = \begin{cases} Y & \text{with probablity } P_1 = \dfrac{\gamma}{\alpha+\beta+\lambda} \\ Y + S & \text{with probablity } P_2 = \dfrac{\beta}{\alpha+\beta+\lambda} \\ TY + S & \text{with probablity } P_3 = \dfrac{\alpha}{\alpha+\beta+\lambda} \end{cases} \tag{10}$$

The estimator of the mean is given by:

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

The sampling variance is given by:

$$Var\left(\hat{\mu}_G\right) = \frac{1}{n}\left[\sigma_Y^2 + \frac{\alpha}{\alpha+\beta+\lambda}\left(\sigma_Y^2+\mu_Y^2\right)\sigma_T^2 + \sigma_S^2\frac{\alpha+\beta}{\alpha+\beta+\lambda}\right] \tag{12}$$

Proposed Forced Randomized Response Model

By taking motivation from the study of Narjis and Shabbir (2023), we develop a forced quantitative technique where the option of a fixed response is added to the model. The response noted from the proposed new model is described as follows:

$$Z_i = \begin{cases} TY + \alpha S & \text{with probability} \quad P_1 = \dfrac{\beta}{(\alpha+\beta)\lambda} \\ TY - \beta S & \text{with probability} \quad P_2 = \dfrac{\alpha}{(\alpha+\beta)\lambda} \\ F & \text{with probability} \quad P_3 = 1 - \dfrac{1}{\lambda} \end{cases}. \tag{13}$$

Where α, β, and λ represent predetermined constants the interrogator assigns based on earlier information or by polite survey. According to the model presented in equation (13), an unbiased mean estimator is provided by:

$$\hat{\mu}_P = \frac{\frac{1}{n}\sum_{i=1}^{n} Z_i - FP_3}{P_1+P_2}. \tag{14}$$

We now analyze the properties of the mean estimator given in equation (14).

Theorem 1: Using the new forced technique, the sample mean $\hat{\mu}_P$ is an unbiased estimator of μ_Y

$$\hat{\mu}_P = \frac{\frac{1}{n}\sum_{i=1}^{n} Z_i - P_3 F}{P_1+P_2}$$

Proof: Applying expectation on the estimator presented in equation (12), we get:

$$E\left(\hat{\mu}_P\right) = E\left(\frac{\frac{1}{n}\sum_{i=1}^{n} Z_i - FP_3}{P_1+P_2}\right),$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n} E\left(Z_i\right) - E(F)P_3}{P_1+P_2},$$

$$= \frac{E(Z_i) - FP_3}{P_1+P_2},$$

Now

$$E(Z_i) = E\left[(TY_i + \alpha S)P_1 + (TY_i - \beta S)P_2 + FP_3\right] \tag{15}$$

Putting the value of $E\left(Z_i\right)$, we get

$$= P_1[E(TY_i) + \alpha\, E(S)] + P_2[E(TY) - \beta\,(S)] + P_3 E(F).$$

Additional simplification yields:

$E(\hat{\mu}_P) = \mu_Y$.

Theorem 2: The mathematical formula for the variance of the estimator $\hat{\mu}_P$ can be written as:

$$Var(\hat{\mu}_Y) = \frac{1}{n}\begin{bmatrix} \left\{\left(\sigma_Y^2+\mu_Y^2\right)+\left(\sigma_Y^2+\mu_Y^2\right)\sigma_T^2+\left(\theta^2+\sigma_S^2\right)\alpha\beta\right\}\lambda-\lambda F^2 \\ -\left\{\mu_Y^2+F^2+\left(\mu_Y\lambda-\mu_Y-\lambda F\right)2F\right\} \end{bmatrix}.$$

Proof: We can obtain the variance of Z_i as:

$$Var(Z) = E\left(Z_i^2\right)-\left\{E\left(Z_i\right)\right\}^2. \tag{16}$$

In equation (14), $E\left(Z_i^2\right)$ can be solved as follows:

$$E\left(Z_i\right)^2 = E\left[(TY_i+\alpha S)^2 P_1 + (TY_i-\beta S)^2 P_2 + F^2 P_3\right]. \tag{17}$$

We first introduce a few assumptions given below:

$$\left.\begin{array}{l} E(Y)=\mu_Y,\ E(T)=1,\ E(F)=F,\ E(S)=\theta, \\ E\left(Y^2\right)=\sigma_Y^2+\mu_Y^2,\ E\left(S^2\right)=\sigma_S^2, \\ E\left(T^2\right)=\sigma_T^2+1,\ E(F^2)=F^2,\ \ Var(F)=0. \end{array}\right\} \tag{18}$$

Putting equation (15) in equation (16), we get:

$$= \frac{\left(\sigma_S^2+\theta^2\right)\alpha\beta+\left(\sigma_Y^2+\mu_Y^2\right)\sigma_T^2+(\lambda-1)F^2+\left(\sigma_Y^2+\mu_Y^2\right)}{\lambda}. \tag{19}$$

Now using equation (17) in equation (14), we get:

$$= \begin{bmatrix} \dfrac{\left\{\left(\sigma_Y^2+\mu_Y^2\right)+\alpha\beta\left(\sigma_S^2+\theta^2\right)+\left(\sigma_Y^2+\mu_Y^2\right)\sigma_T^2\right\}}{\lambda}+\dfrac{(\lambda-1)F^2}{\lambda} \\ -\left\{F\left(1-\dfrac{1}{\lambda}\right)+\dfrac{\mu_Y}{\lambda}\right\}^2 \end{bmatrix}. \tag{20}$$

Taking the variance of equation (12) gives:

$$Var(\hat{\mu}_P) = \frac{1}{n^2\left(P_1+P_2\right)^2}\sum_{i=1}^{n} Var\left(Z_i\right). \tag{21}$$

Using equation (18) in equation (19) yields:

$$Var(\hat{\mu}_Y) = \frac{1}{n}\begin{bmatrix} \lambda\left\{\alpha\beta\left(\sigma_S^2+\theta^2\right)+\left(\sigma_Y^2+\mu_Y^2\right)+\left(\sigma_Y^2+\mu_Y^2\right)\sigma_T^2\right\}-\lambda F^2 \\ -\left\{\mu_Y^2+F^2+2F\left(\mu_Y\lambda-\mu_Y-\lambda F\right)\right\} \end{bmatrix}. \tag{22}$$

Evaluation of Models

Yan et al. (2008) quantified the degree of privacy in the following measure:

$$\nabla = E\left[Z-Y\right]^2. \tag{23}$$

A unified metric of performance evaluation offered by Gupta et al. (2018) may be expressed as:

$$\delta = \frac{MSE}{\nabla}. \tag{24}$$

Upon examining equations (21) and (22), it is clear that a preferable model has a greater value of ∇ and a smaller value of δ.

The metric privacy using of Bar-lev et al technique can be expressed as follows:

$$\nabla_B = \sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right)(1-P). \tag{25}$$

The unified metric using the Bar-Lev et al technique can be quantified as follows:

$$\delta_B = \frac{1}{n}\left[\frac{\left(\sigma_Y^2+\mu_Y^2\right)\sigma_T^2+\sigma_Y^2}{\left(\sigma_Y^2+\mu_Y^2\right)(1-P)\sigma_T^2}\right]. \tag{26}$$

The metric privacy using the Gjestvang and Singh can be expressed as follows:

$$\nabla_{GS} = \sigma_S^2\beta\alpha.$$

The unified metric according to the Gjestvang and Singh model can be formulated as follows:

$$\delta_{GS} = \frac{1}{n}\left[\frac{\left(\sigma_S^2+\theta^2\right)\alpha\beta+\sigma_Y^2}{\alpha\beta\sigma_S^2}\right]. \tag{27}$$

The metric of privacy using the Narjis and Shabbir technique is given by:

$$\nabla_{NS} = \frac{\alpha\beta\left(\alpha+\beta\right)\sigma_S^2}{\alpha+\beta+\lambda}. \tag{28}$$

The unified metric using the Narjis and Shabbir model is given by the formula:

$$\delta_{NS} = \frac{1}{n}\left[\frac{\dfrac{\alpha\beta\left(\alpha+\beta\right)\sigma_S^2}{\alpha+\beta+\gamma}}{\sigma_Y^2+\dfrac{\left(\sigma_S^2+\theta^2\right)\alpha\beta}{\alpha+\beta+\gamma}}\right]. \tag{29}$$

The measure of privacy of Gupta et al:

$$\nabla_G = \left(\frac{\alpha}{\alpha+\beta}\right)\left\{\left(\sigma_Y^2+\mu_Y^2\right)\sigma_T^2\right\}+\sigma_S^2. \tag{30}$$

The unified metric using the Gupta et al model can be expressed as:

$$\delta_G = \frac{1}{n}\left[\frac{\sigma_Y^2+\dfrac{\alpha}{\alpha+\lambda+\beta}\sigma_T^2\left(\sigma_Y^2+\mu_Y^2\right)+\dfrac{\alpha+\beta}{\alpha+\lambda+\beta}\sigma_S^2}{\left\{\left(\sigma_Y^2+\mu_Y^2\right)\sigma_T^2\right\}\left(\dfrac{\alpha}{\alpha+\beta}\right)+\sigma_S^2}\right]. \tag{31}$$

The privacy metric using the proposed model might be obtained as:

$$\nabla_P = \frac{1}{\lambda}\left[(\sigma_Y^2+\mu_Y^2)\sigma_T^2+(\sigma_S^2+\theta^2)\alpha\beta+\{(\sigma_Y^2+\mu_Y^2)+F^2-2\mu_Y F\}(\lambda-1)\right] \tag{32}$$

The unified metric using the proposed model might be obtained as:

$$\delta_P = \frac{\dfrac{1}{n}\left[\lambda\left\{\left(\sigma_Y^2+\mu_Y^2\right)+\left(\sigma_Y^2+\mu_Y^2\right)\sigma_T^2+\left(\sigma_S^2+\theta^2\right)\alpha\beta\right\}-\lambda F^2-\left\{\mu_Y^2+F^2+2F\left(\mu_Y\lambda-\mu_Y-\lambda F\right)\right\}\right]}{\dfrac{1}{\lambda}\left[(\sigma_Y^2+\mu_Y^2)\sigma_T^2+(\sigma_S^2+\theta^2)\alpha\beta+\{(\sigma_Y^2+\mu_Y^2)+F^2-2\mu_Y F\}(\lambda-1)\right]} \tag{33}$$

The measures provided in equations (31) to (32) may be used for the comparison of models.

Model Comparison

Table 1 presents the sampling variances for our proposed forced model and the competing models, with various parameter values. Table 2 displays the δ values for different models with varied parameter values.

Table 1: Sampling variance under various models for various values of constants and parameters

$n=500$, $\mu = 2$ $\sigma^2 = 10$, $\theta = 0.7$, $\sigma^2 = 0.5$, $\sigma^2 = 0.5$, $F = 2$

λ	α	β	$Var(\hat{u}_G)$	$Var(\hat{u}_{NS})$	$Var(\hat{u}_B)$	$Var(\hat{u}_{GS})$	$Var(\hat{u}_P)$
0.5		1.5	0.02550	0.02128	0.02840	0.02153	0.00177
	1	2	0.02486	0.02175	0.02933	0.02204	0.00202
		2.5	0.02438	0.02223	0.03000	0.02255	0.00228
		1.5	0.02788	0.02268	0.02600	0.02306	0.00253
	2	2	0.02711	0.02363	0.02700	0.02408	0.00304
		2.5	0.02650	0.02459	0.02778	0.02510	0.00355
0.6		1.5	0.02532	0.02123	0.02840	0.02153	0.00852
	1	2	0.02472	0.02170	0.02933	0.02204	0.00882
		2.5	0.02427	0.02218	0.03000	0.02255	0.00913
		1.5	0.02768	0.02261	0.02600	0.02306	0.00944
	2	2	0.02696	0.02355	0.02700	0.02408	0.01005
		2.5	0.02637	0.02450	0.02778	0.02510	0.01066
0.7		1.5	0.02516	0.02120	0.02840	0.02153	0.01527
	1	2	0.02460	0.02165	0.02933	0.02204	0.01563
		2.5	0.02417	0.02213	0.03000	0.02255	0.01599
		1.5	0.02750	0.02255	0.02600	0.02306	0.01634
	2	2	0.02681	0.02347	0.02700	0.02408	0.01706
		2.5	0.02625	0.02441	0.02778	0.02510	0.01777

Table 2: δ values under various models for various values of constants and parameters

$n=500$, $\mu_y = 2$, $\sigma_y^2 = 10$, $\theta = 0.7$, $\sigma_S^2 = 0.5$, $\sigma_T^2 = 0.5$, $F = 2$

λ	α	β	δ_G	δ_{NS}	δ_{NS}	δ_{GS}	δ_P
0.5		1.5	0.007727	0.034040	0.006760	0.028710	0.000320
		2	0.008773	0.025370	0.006290	0.022040	0.000330
	1	2.5	0.009750	0.020330	0.006000	0.018040	0.000350
		1.5	0.006194	0.017280	0.008670	0.015370	0.000360
		2	0.006778	0.013290	0.007710	0.012040	0.000380
	2	2.5	0.007338	0.010930	0.007140	0.010040	0.000390
0.6		1.5	0.007674	0.035110	0.006760	0.028710	0.001360
		2	0.008725	0.026040	0.006290	0.022040	0.001320
	1	2.5	0.009707	0.020780	0.006000	0.018040	0.001280
		1.5	0.006152	0.017660	0.008670	0.015370	0.001250
		2	0.006739	0.013540	0.007710	0.012040	0.001200
	2	2.5	0.007303	0.011110	0.007140	0.010040	0.001150

0.7		1.5	0.007623	0.036170	0.006760	0.028710	0.002240
		2	0.008680	0.026710	0.006290	0.022040	0.002180
	1	2.5	0.009667	0.021240	0.006000	0.018040	0.002120
		1.5	0.006111	0.018040	0.008670	0.015370	0.002070
		2	0.006702	0.013790	0.007710	0.012040	0.001980
	2	2.5	0.007269	0.011280	0.007140	0.010040	0.001900

A real-world survey

A survey was conducted using the proposed randomized response model with a sample of 100 undergraduate students from the Statistics Department at the University of Malakand, Khyber Pakhtunkhwa, Pakistan. The purpose was to determine the average examination grades on a 4.0-point scale. To generate random variables, 100 integers for the additive-type scrambling variable S were created using a normal distribution with parameters 2 and 3, and another 100 integers for the multiplicative-type scrambling variable T with parameters 1 and 2. A deck of 100 cards was prepared, each with values for S and T, and was utilized as a randomization technique. Students were instructed to randomly draw a card and follow its instructions while keeping their chosen card hidden from the interviewer to ensure response confidentiality.

Furthermore, the values of parameters and constants were selected as $\alpha = 3$, $\beta = 2$, and $\gamma = 2$. The likelihood of choosing different cards was determined as:

$$P_1 = \frac{\alpha}{(\alpha+\beta)\lambda} = 0.30, \quad P_2 = \frac{\beta}{(\alpha+\beta)\lambda} = 0.20, \text{ and } P_3 = 1 - \frac{1}{\lambda} = 0.5$$

By applying these values in Equation (13), the response function we observe can be interestingly expressed as:

$$Z_i = \begin{cases} TY + 2S & \text{with probability } P_1 = 0.3 \\ TY - 3S & \text{with probability } P_2 = 0.2 \\ 3.4 & \text{with probability } P_3 = 0.5 \end{cases} \tag{33}$$

Every card in the deck showcased a powerful statement, highlighting one of four distinct ideas.

i. Thirty percent of the cards featured the directive: "Multiply T by your SGPA, then add αS and provide the resulting number".
ii. Twenty percent of the cards featured the directive: "Multiply T by your SGPA, then subtract βS and provide the resulting number".
iii. Fifty percent of the cards instructed: "Provide the given number".

The responses gathered from the 100 participants were analyzed for clarity and accuracy.

Table 3: Survey Observati n

-3.30	3.40	12.80	-2.49	10.20	3.40	3.40	-11.44	4.52	3.40
-11.93	3.40	3.40	3.40	3.40	3.40	3.40	-0.90	14.78	3.40
3.40	12.20	3.40	16.30	3.40	-2.74	8.24	10.09	-4.81	9.12
3.40	3.40	8.20	3.40	-4.58	10.09	3.40	-3.32	13.20	3.40
-3.69	3.40	5.65	9.80	3.01	8.99	3.40	3.40	11.40	3.40
3.40	3.40	3.40	9.50	-12.26	3.40	-5.60	3.40	-4.68	-10.40
3.40	17.20	12.45	3.40	3.40	3.40	11.20	-0.74	-2.90	3.40
3.40	-0.52	3.40	10.10	12.50	-10.69	-3.20	3.40	16.58	10.12
3.40	5.85	3.40	3.40	-3.02	-8.42	1.50	4.89	1.30	3.40
12.20	3.40	6.54	3.40	3.40	3.40	10.30	3.40	10.26	-10.15

The scrambling procedure employed for the responses resulted in some bad results or even exceeding the highest bound of the grading scale. In the proposed model, the responses observed had a mean value of 3.52. To test the estimator's accuracy, the researcher acquired marks for all enroll students in the Department of Statistics from the office of examination of the university and estimated the population mean, which was 3.39.

Discussion and Conclusion

We developed a novel forced scrambling strategy using forced answers, which has been shown to outperform existing techniques across several model assessment measures. As shown in Table 1, our proposed forced technique is more effective than the competing methods identified in earlier studies. Additionally, table 2 demonstrates that this innovative forced strategy outperforms existing methods across various constant and parameter values when applying Gupta et al. (2018) unified measure. As observed in table 1, decreasing the value of λ tends to decrease the variance of our model, which increases its efficiency and enhances the overall reliability compared to competitors' models.

The suggested model, such as Narjis and Shabir's (2023), alleviates some of the strain on survey respondents by not requiring a thorough understanding of mathematics. To participate in the survey, respondents must be able to perform basic addition and multiplication using an electronic calculator or a smartphone app. This improves the proposed model's applicability to real-world problems. Table 2 indicates that increasing the value of λ reduces the unified measure of our model and increases its efficiency, improving its overall dependability compared to competing models.

From a practical standpoint, the forced response model offers significant benefits for sensitive surveys in fields like sociology, psychology, and economics, where respondents might be reluctant to disclose truthful information due to privacy concerns or social desirability bias. By integrating fixed responses, respondents may feel more comfortable participating, thus reducing non-response rates, a pervasive challenge highlighted in the literature. The comparative analysis, as shown in Tables 1 and 2, demonstrates that the proposed model can achieve lower sampling variances and higher privacy metrics under various parameter settings, indicating superior performance in terms of efficiency and confidentiality.

The randomization process used for the responses has led to some results exceeding the upper limits of the grading scale or appearing as negative values. The mean value of the responses produced by the proposed model is 3.52, reflecting a moderate assessment within the grading framework. To assess the estimator's accuracy and reliability, the researchergathered the complete grade records of all current students from the Statistics Department through the office of the university examination. Calculations revealed that the average value of the population is 3.39. The notable discrepancy between the model's mean and the actual population mean indicates that the model is effective and reliable in capturing the general trends of student performance, while also revealing areas where response accuracy could be improved.

The use of unified metrics (equations 31 and 32) enables practitioners to evaluate the trade-offs between privacy and precision systematically. Notably, the model’s higher privacy protection, quantified through the metrics derived from Yan et al. (2008) and Gupta et al. (2018), underscores its suitability for real-world applications involving highly sensitive data, such as income, illegal activities, or illicit behaviors.

For future researchers, we advocate investigating the variance estimation of sensitive variables using the proposed approach. We can expect variance estimators to outperform the majority of existing estimators using the suggested model.

References

- Azeem, M., & Ali, S. (2023). A neutral comparative analysis of additive, multiplicative, and mixed quantitative randomized response models. *PLOS ONE*, *18*(4), e0284995.
- Azeem, M., & Salam, A. (2023). Introducing an efficient alternative technique to optional quantitative randomized response models. *Methodology*, *19*(1), 24-42.
- Azeem, M., & Salam, A. (2025). On the use of randomized response technique for precise estimation of population variance. *Quality & Quantity*, 1-13.
- Azeem, M. (2023). Introducing a weighted measure of privacy and efficiency for comparison of quantitative randomized response models. *Pakistan Journal of Statistics, 39*(3), 377-85.
- Azeem, M., Salahuddin, N. Hussain, S., Ijaz, M., & Salam, A., (2024). An efficient estimator of the population variance of a sensitive variable with a new randomized response technique. *Heliyon, 10*(5).
- Azeem, M., & Salam, A. (2023). Introducing an efficient alternative technique to optional quantitative randomized response models. *Methodology*, *19*(1), 24-42.
- Bar-Lev, S.K., Bobovitch, E., & Boukai, B. (2004). A note on randomized response models for quantitative data. *Metrika, 60*(3), 255–60.
- Diana, G., & Perri, P.F. (2011). A class of estimators of quantitative sensitive data. *Statistical Papers, 52*(3), 633-50.
- Gjestvang, C.R., Singh, S. (2009). An improved randomized response model: Estimation of mean. *Journal of Applied Statistics, 36*(12), 1361-7.
- Gupta, S., Gupta, B., & Singh, S. (2002). Estimation of the sensitivity level of personal interview survey questions. *Journal of Statistical Planning and Inference*, *100*(2), 239-47.
- Gupta, S., Kalucha, G., Shabbir, J., & Dass, B.K. (2014). Estimation of finite population mean using optional RRT models in the presence of non-sensitive auxiliary information. *American Journal of Mathematical and Management Sciences, 33*(2), 147-59.
- Gupta, S., Mehta, S., Shabbir, J., & Dass, B.K. (2013). Generalized scrambling in quantitative optional randomized response models. *Communications in Statistics – Theory and Methods*, *42*(22), 4034-42.
- Gupta, S., Mehta, S., Shabbir, & J., Khalil, S. (2018). A unified measure of respondent privacy and model efficiency in quantitative RRT models. *Journal of Statistical Theory and Practice, 12*(3), 506-11.
- Gupta, S., Zhang, J., Khalil, S., & Sapra, P. (2024). Mitigating lack of trust in quantitative randomized response technique models. *Communications in Statistics – Simulation and Computation, 53*(6), 2624-2632.
- Hussain, Z., Al-Sobhi, M.M., Al-Zahrani, B. Singh, H.P., & Tarray, T.A. (2016). Improved randomized response in additive scrambling models. *Mathematical Population Studies, 23*(4), 205-21.
- Khalil, S., Zhang, Q., & Gupta, S. 2021. Mean estimation of sensitive variables under measurement errors using optional rrt models. *Communications in Statistics – Simulation and Computation*, *50(*5), 1417-26.
- Lovig, M., Khalil, S., Rahman, S., Sapra, P. & Gupta, S. (2023). A mixture binary RRT model with a unified measure of privacy and efficiency. *Communications in Statistics – Simulation and Computation, 52*(6), 2727-37.
- Mehta, S., Dass, B.K., Shabbir, J., & Gupta, S. (2012). A three-stage optional randomized response model. *Journal of Statistical Theory and Practice*, *6*(3), 417-27.
- Murtaza, M., Singh, S., & Hussain, Z. (2021). Use of correlated scrambling variables in quantitative randomized response technique. *Biometrical Journal*, *63*(1), 134-47.
- Narjis, G., Shabbir, & J. (2023). An efficient new scrambled response model for estimating sensitive population mean in successive sampling. *Communications in Statistics – Simulation and Computation, 52*(11), 5327-44.
- Warner, S.L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, *60*(309), 63-9.

- Warner, S.L. (1971). The linear randomized response model. *Journal of the American Statistical Association, 66*(336), 884-8.
- Yan, Z., Wang, J., & Lai, J. (2008). An efficiency and protection degree-based comparison among the quantitative randomized response strategies. *Communications in Statistics – Theory and Methods, 38*(3), 400-8.
- Young, A., Gupta, S., & Parks, R. (2019). A binary unrelated-question rrt model accounting for untruthful responding. *Involve, A Journal of Mathematics, 12*(7), 1163-73.
- Zhang, Q., Khalil, S., & 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), 3492-504.