Original
research
Article

Additive Optional Scrambled Randomized Response Model for Estimating Population Mean and Sensitivity Level of Sensitive Variable

Abstract: The paper proposed an additive optional randomized response technique model that improves upon Gjestvang and Singh (2009) model by effectively balancing respondents privacy protection and statistical estimation efficiency. Better efficiency and The proposed model establishes an unbiased estimator of the population mean under both simple random sampling and probability proportional to size sampling schemes. Findings show that the proposed model achieves a more balanced compromise between confidentiality protection and data quality than existing model.

Keywords: Scrambling Randomized Response, Privacy Protection, Statistical Estimation, Mean, Sensitive Variable

1 Introduction

Overtime, survey statisticians have been saddled with the task of collecting data on sensitive survey question such as use of illicit drug, sexual behaviour, tax invasion, examination malpractice etc due to social stigma attached to them. Traditional face-to-face interview has proved to be inefficient and ineffective for collecting data on such sensitive issue as those highlighted above.

Respondents wary of their privacy and confidentiality oftentimes either shy away from answering sensitive survey question or falsify their responses on a traditional direct interview, thus increasing the response bias. To minimize potential response bias and protect the respondents' privacy and confidentiality, Warner (1965) introduced randomized response technique for estimating population proportion π of those possessing sensitive attribute. This procedure is applicable to qualitative or binary data. Warner (1971) introduced additive randomized response technique for collecting data on quantitative sensitive variable. In furthering research on randomized response technique, authors like Gjestvang and Singh (2009), Narjis and Shabbir (2023), Azeem et al. (2024), Azzem (2023), Mojeed et al. (2025), Gupta et al. (2022), Ahmad et al. (2024), Eichhorn and Hayre (1983), Gupta et al. (2010), Parker et al. (2024), Zawar and Muhammad (2019), Neeraj and Prachi (2017) have have contributed to the literature of randomization response technique model.

Randomized response technique is classified into forced randomized response technique and optional randomized response technique. In forced randomized response model, each respondent is asked to report scrambled response irrespective of their perception of the sensitivity of the survey question. In contrast to forced randomized response technique, the concept of optional randomized response technique introduced by Gupta et al. (2002) gives respondents the option to report scrambled response if the survey question is considered sensitive by them or opt to report truthful response if the survey question is considered nonsensitive by them.

The interest of this paper is on optional randomized response technique model as research has shown that the trade-off between privacy protection and data accuracy is disrupted with the use of forced randomized response technique models when the false response rate is high. The paper proposed an additive optional randomized response technique for estimating the population mean and sensitivity level of the survey question.

The model will extend Gjestvang and Singh (2009) model to optional randomized response model which will give the respondents the option to choose either to report scrambled response if they considered the survey question sensitive or to report truthful response if they consider the survey question nonsensitive. It is expected that this will increase the privacy protection level of the respondents and their willingness to participate in the survey. The sample mean estimator and variance of the sample mean estimator under simple random sampling and probability proportional to size sample schemes as well as sensitivity level will be obtained.

2 Gjestvang and Singh (2009) Model

Briefly, Gjestvang and Singh (2009) model works as follows: Let Y be a sensitive quantitative variable with unknown mean μ_Y and variance σ_Y^2. Suppose the mean θ and variance σ_s^2 of scrambling variable S are known. S follows known probability distribution and is independent of Y. The respondent is asked to draw a card from a well shuffled deck of cards and to report the scrambled response according to statements contained in the card selected. Each of the cards contains either of the following statements:

1. Multiply scrambling variable S with α and add the result to the truthful value of sensitive variable Y with probability $p = \frac{\beta}{\alpha+\beta}$

2. Multiply scrambling variable S with β and subtract the result from the truthful value of sensitive variable Y with probability $1 - p = \frac{\alpha}{\alpha+\beta}$

Therefore, the i^{th} reported response h_i from the sampled respondent is given as

$$h_i = \begin{cases} y_i + \alpha s_i & with \quad probability \quad p = \frac{\beta}{\alpha+\beta} \\ y_i - \beta s_i & with \quad probability \quad (1-p) = \frac{\alpha}{\alpha+\beta} \end{cases} \tag{1}$$

The unbiased estimator of the population mean μ_Y is given as

$$\hat{\mu}_{gs} = \frac{1}{n}\sum_{i=1}^{n} h_i \tag{2}$$

with variance given as

$$V(\hat{\mu}_{gs}) = \frac{1}{n}\left[\alpha\beta\left(\sigma_s^2 + \theta^2\right) + S_y^2\right] \tag{3}$$

3 Proposed Additive Optional Randomized Response Model

Let's consider a finite population $\Omega = \Omega_1, \Omega_2, \ldots, \Omega_N$ and a sample of size n drawn from Ω with simple random sampling without replacement (SRSWOR). Let Y be sensitive study variable and S be a scrambled variable independent of Y, and Z be the scrambled response variable. For the sensitive study variable Y, respondent is to report scrambled response according to instruction given.
Let Y_i be the true response from i^{th} respondent and S_i be scramble variable selected by the i^{th} respondent in the population, let Z_i be reported response. The respondent is given the survey question to decide if it is sensitive or non-sensitive.

i) if sensitive, report scramble response with probability W.
The respondents who considered the survey question sensitive and will scramble their response with probability W will choose one of the following three scramble rules with weights proportional to β, and α:

Rule 1:) Report $Z_i = Y_i + \alpha S_i$ with probability $p_1 = \frac{W\beta}{\alpha+\beta}$

Rule 2:) Report $Z_i = Y_i - \beta S_i$ with probability $p_2 = \frac{W\alpha}{\alpha+\beta}$

ii) if non-sensitive, report non-scramble response with probability $1 - W$.
The respondents who consider the survey question non-sensitive and will report non-scramble response with probability $1 - W$, will report truthful value of the sensitive variable according to following rule.

Rule 3:) Report $Z_i = Y_i$ with probability $p_3 = 1 - W$

Let's define,

$$T_i = \begin{cases} 1 & if \ Sensitive \ for \ i^{th} \ respondent \\ 0 & if \ Non-Sensitve \ for \ i^{th} \ respondent \end{cases} \tag{4}$$

then the reported response Z_i for the i^{th} respondent satisfies

$$Z_i = \begin{cases} T_i = 1 : \begin{cases} Y_i + \alpha S_i & with\ probability\ \ p_1 \\ Y_i - \beta S_i & with\ probability\ \ p_2 \end{cases} \\ \\ T_i = 0 : \quad Y_i \qquad\qquad with\ probability\ \ p_3 \end{cases} \tag{5}$$

From (5), the reported response distribution model is then obtained as

$$Z_i = [Z_i|T_i = 1] + [Z_i|T_i = 0]$$

$$= p_1\Big(Y_i + \alpha S_i\Big) + p_2\Big(Y_i - \beta S_i\Big) + p_3 Y_i$$

$$= \left(\frac{\beta W}{\alpha + \beta}\right)\Big(Y_i + \alpha S_i\Big) + \left(\frac{\alpha W}{\alpha + \beta}\right)\Big(Y_i - \beta S_i\Big) + \Big(1 - W\Big)Y_i \tag{6}$$

where α and, β are predetermined real valued positive constants such that $p_1 + p_2 + p_3 = 1$, and p_1 and p_2 are unconditional probabilities for the scramble rules. $W, (0 \leq W \leq 1)$ is the unknown sensitivity level of the underlying sensitive question defined as the proportion of respondents in a survey that find the question sensitive enough to feel uncomfortable to answer it in a face-to-face interview.
It should be noted that if $W = 1$ the new model reduced to Gjestvang and Singh (2009) model and if the $W = 0$ the new model reduced to traditional direct interview, no longer masked the respondent's response to sensitive question hence, respondent's privacy is not guaranteed.

4 Statistical Estimation Under Simple Random Sampling Scheme

To estimate the mean μ_Y of the sensitive variable Y and the sensitivity level W, $(0 \leq W \leq 1)$, let a sample of size n respondents be drawn from a finite population Ω_N with SRSWOR. Let y_i be the true response from i^{th} respondent and s_i be scramble variable selected by the i^{th} respondent. Suppose that the scrambled variable S independent of Y has the sample mean and variance defined as $E(s) = \theta$ and $Var(s) = \sigma_s^2$ respectively, both are assumed to be known. Suppose z_i is the reported scrambled response by i^{th} respondent, then from (6), the reported sample response distribution to the survey question for selected i^{th} respondent is given as

$$z_i = \left[\left(\frac{W\beta}{\alpha + \beta}\right)\Big(y_i + \alpha s_i\Big) + \left(\frac{W\alpha}{\alpha + \beta}\right)\Big(y_i - \beta s_i\Big) + \Big(1 - W\Big)y_i\right] \qquad for \quad i = 1, 2, \ldots, n \tag{7}$$

The sample mean of the response distribution is obtained as

$$\begin{aligned} \bar{z} &= \frac{1}{n}\sum_{i=1}^{n} z_i \\ &= \left[\frac{1}{n}\sum_{i=1}^{n}\left\{\left(\frac{W\beta}{\alpha + \beta}\right)\Big(y_i + s_i\alpha\Big) + \left(\frac{W\alpha}{\alpha + \beta}\right)\Big(y_i - s_i\beta\Big) + \Big(1 - W\Big)y_i\right\}\right] \\ &= \left(\frac{1}{\alpha + \beta}\right)\left[\frac{1}{n}\sum_{i=1}^{n}(W(\alpha + \beta) + (\alpha + \beta)(1 - W))y\right] \\ &= \mu_y \end{aligned} \tag{8}$$

Hence the following theorem
Theorem 1: *An unbiased estimator of* μ_Y is given as

$$\hat{\mu}_y = \frac{1}{n}\sum_{i=1}^{n} z_i \tag{9}$$

Proof: Let the expectation over the sampling design d be E_d and expectation over the randomization device R be E_R, then

$$
\begin{aligned}
E(\hat{\mu}_y) &= E_d E_R\left(\frac{1}{n}\sum_{i=1}^{n} z_i\right) \\
&= E_d E_R\left[\frac{1}{n}\sum_{i=1}^{n}\left\{\left(\frac{W\beta}{\alpha+\beta}\right)\left(y_i + s_i\alpha\right) + \left(\frac{W\alpha}{\alpha+\beta}\right)\left(y_i - s_i\beta\right) + \left(1-W\right)y_i\right\}\right] \\
&= E_d\left[\frac{1}{n}\sum_{i=1}^{n} E_R\left\{\left(\frac{W\beta}{\alpha+\beta}\right)\left(y_i + s_i\alpha\right) + \left(\frac{W\alpha}{\alpha+\beta}\right)\left(y_i - s_i\beta\right) + \left(1-W\right)y_i\right\}\right] \\
&= \left(\frac{1}{\alpha+\beta}\right) E_d\left[\frac{1}{n}\sum_{i=1}^{n}(W(\alpha+\beta) + (\alpha+\beta)(1-W))y\right] \\
&= \mu_Y
\end{aligned} \tag{10}
$$

Hence the proof.
Theorem 2: *The variance of $\hat{\mu}_y$* is given as

$$
\frac{\sigma_s^2}{n}\left[W\alpha\beta\left(1 + C_s^{-2}\right) + R_y\right] \tag{11}
$$

Proof: By definition,

$$
\begin{aligned}
V(\hat{\mu}_y) = V(z) &= \frac{1}{n}\left(E(z^2) - \mu_y^2\right) \\
&= \frac{1}{n}\left(E_S(z^2) + E_T(z^2) - \mu_y^2\right)
\end{aligned} \tag{12}
$$

where, E_S and E_T are expectations with respect to scrambled response and truthfully response respectively. Thus,

$$
\begin{aligned}
V(\hat{\mu}_y) &= \frac{1}{n}\left[E_S\left\{\left(\frac{W\beta}{\alpha+\beta}\right)(y+\alpha s)^2 + \left(\frac{W\alpha}{\alpha+\beta}\right)(y-\beta s)^2\right\} + E_T\left\{(1-W)y^2\right\} - \mu_y^2\right] \\
&= \frac{1}{n}\left\{W(E_S(y^2) + \alpha\beta(E(s^2))\} + (1-W)E_T(y^2) - \mu_y^2\right] \\
&= \frac{1}{n}\left[W\alpha\beta(\sigma_s^2 + \theta^2) + \sigma_y^2\right] \\
&= \frac{\sigma_s^2}{n}\left[W\alpha\beta(1 + C_s^{-2}) + R_y\right]
\end{aligned} \tag{13}
$$

Hence the proof

The first term in (13) is the penalty for using optional randomized response model.

5 Statistical Estimation Under Probability Proportional to Size Sampling Scheme

Probability proportional to size is a sampling scheme where auxiliary information is utilized at sample selection stage. Here the probability of selection is proportional to size of the auxiliary information associated with particular unit of the study variable.
Let $Y_i, (i = 1, 2, ..., N)$ be the units of the sensitive study variable Y in the finite population Ω. Let $X_i, (i = 1, 2, ..., N)$ be the nonsensitive auxiliary variable X associated with the Y_i units. Let a sample of size n be selected with probability proportional to size sample scheme with replacement ($PPSWR$) and probability of selecting i^{th} unit in the population be $p_i = \frac{X_i}{X}$ such that $\sum_{i}^{N} p_i = 1$
The response design follows as describe in simple random sampling scheme above. Thus, the reported response

distribution is obtained as

$$Z_i = [Z_i|T_i = 1] + [Z_i|T_i = 0]$$

$$= p_1\Big(Y_i + \alpha S_i\Big) + p_2\Big(Y_i - \beta S_i\Big) + p_3 Y_i$$

$$= \Big(\frac{\beta W}{\alpha+\beta}\Big)\Big(Y_i + \alpha S_i\Big) + \Big(\frac{\alpha W}{\alpha+\beta}\Big)\Big(Y_i - \beta S_i\Big) + \Big(1 - W\Big)Y_i \tag{14}$$

Let a sample of size n respondents be drawn from a finite population Ω_N with PPSWR. Let y_i be the true response from i^{th} respondent and s_i be scramble variable selected by the i^{th} respondent. Suppose that the scrambled variable S independent of Y has the sample mean and variance defined as $E(s) = \theta$ and $Var(s) = \sigma_s^2$ respectively, both are assumed to be known. Suppose z_i is the reported scrambled response by i^{th} respondent, then from (14), the reported sample response distribution to the survey question for selected i^{th} respondent is given as

$$z_i = \Bigg[\Big(\frac{W\beta}{\alpha+\beta}\Big)\Big(y_i + \alpha s_i\Big) + \Big(\frac{W\alpha}{\alpha+\beta}\Big)\Big(y_i - \beta s_i\Big) + \Big(1 - W\Big)y_i\Bigg] \quad for \quad i = 1, 2, \ldots, n \tag{15}$$

Theorem 3: *An unbiased estimator of* μ_Y under $PPSWR$ is given as

$$\hat{\mu}_y = \frac{1}{nN}\sum_{i=1}^{n} z_i \tag{16}$$

Proof: Let the expectation over the sampling design d be E_d and expectation over the randomization device R be E_R, then

$$E(\hat{\mu}_y) = E_d E_R(\hat{\mu}_y) = E_d E_R(\bar{z}) = E_R E_d(\bar{z})$$

$$= E_R\Bigg[\frac{1}{nN}E_d\sum_{i=1}^{n}\frac{1}{p_i}\Bigg\{\frac{\beta W}{\alpha+\beta}\Big(y_i + \alpha s_i\Big) + \frac{\gamma W}{\alpha+\beta}\Big(y_i - \beta s_i\Big) + (1 - W)y_i\Bigg\}\Bigg]$$

$$= E_R\Bigg[\frac{1}{nN}\sum_{i=1}^{n}E_d\Bigg(\frac{1}{p_i}\Bigg\{\frac{\beta W}{\alpha+\beta}\Big(y_i + \alpha s_i\Big) + \frac{\gamma W}{\alpha+\beta}\Big(y_i - \beta s_i\Big) + (1 - W)y_i\Bigg\}\Bigg)\Bigg]$$

$$= E_R\Bigg[\frac{1}{nN}\sum_{i=1}^{n}\sum_{i=1}^{N}p_i\Bigg(\frac{1}{p_i}\Bigg\{\frac{\beta W}{\alpha+\beta}\Big(y_i + \alpha s_i\Big) + \frac{\gamma W}{\alpha+\beta}\Big(y_i - \beta s_i\Big) + (1 - W)y_i\Bigg\}\Bigg)\Bigg]$$

$$= E_R\Bigg[\frac{1}{n}\sum_{i=1}^{n}\Bigg\{\frac{\beta W}{\alpha+\beta}\Big(\mu_Y + \alpha\theta\Big) + \frac{\gamma W}{\alpha+\beta}\Big(\mu_Y - \beta\theta\Big) + (1 - W)\mu_Y\Bigg\}\Bigg]$$

$$= E_R\Bigg[\frac{1}{n}\frac{1}{\alpha+\beta}\sum_{i=1}^{n}\Bigg\{W(\alpha+\beta)\mu_Y + (\alpha+\beta)(1 - W)\mu_Y\Bigg\}\Bigg] = \mu_Y \tag{17}$$

Theorem 4: *The variance of* $\hat{\mu}_y$ under ppswr is given as

$$\frac{\sigma_s^2}{nN^2}\Big[W\alpha\beta\left(1 + C_s^{-2}\right) + R_{y(pps)}\Big] \tag{18}$$

Proof: By definition,

$$V(\hat{\mu}_y) = V(z) = \frac{1}{n} E\left(\frac{z_i}{Np_i} - \bar{z}\right)^2 = \frac{1}{n}\left(E\left(\frac{z_i}{Np_i}\right)^2 - \mu_Y^2\right)$$
$$= \frac{1}{nN^2}\left(E_S E_d\left(\frac{z_i}{p_i}\right)^2 + E_T E_d\left(\frac{z_i}{p_i}\right)^2 - N^2\mu_y^2\right)$$
$$= \frac{1}{nN^2}\left(E_S \sum_{i=1}^{n} p_i \frac{1}{p_i^2} z_i^2 + E_T \sum_{i=1}^{n} p_i \frac{1}{p_i^2} z_i^2 - N^2\mu_y^2\right)$$
$$= \frac{1}{nN^2}\left(\sum_{i=1}^{n} \frac{1}{p_i} E_S(z_i^2) + \sum_{i=1}^{n} \frac{1}{p_i} E_T(z_i^2) - N^2\mu_y^2\right)$$
$$= \frac{1}{nN^2}\left(\sum_{i=1}^{n} \frac{1}{p_i} E_S\left(\frac{W\beta}{\alpha+\beta}\left(y_i + \alpha s_i\right)^2 + \frac{W\alpha}{\alpha+\beta}\left(y_i - \beta s_i\right)^2\right) + \sum_{i=1}^{n} \frac{1}{p_i} E_T\left((1-W)y_i^2\right) - N^2\mu_y^2\right)$$
$$= \frac{1}{nN^2}\left(W\sum_{i=1}^{n} \frac{Y_i}{p_i} + W\alpha\beta(\sigma_s^2 + \theta^2) + \sum_{i=1}^{n} \frac{1}{p_i}(1-W)Y_i - N^2\mu_Y^2\right)$$
$$= \frac{1}{nN^2}\left(W\alpha\beta(\sigma_s^2 + \theta^2) + \sum_{i=1}^{n} \frac{Y_i}{p_i} - N^2\mu_Y^2\right)$$
$$= \frac{\sigma_s^2}{nN^2}\left(W\alpha\beta(1 + C_s^{-2}) + R_{y(pps)}\right) \tag{19}$$

where,

$$R_{y(pps)} = \frac{1}{\sigma_s^2}\left(\sum_{i=1}^{n} \frac{Y_i}{p_i} - N^2\mu_Y^2\right) \tag{20}$$

Hence the proof.

6 Estimation of Sensitivity Level

To estimate the sensitivity level W, let n_0 be the number of respondents that opted for scramble response out of n selected respondents. Then, the sensitivity level of the survey question W is estimated thus,

$$\hat{W} = \frac{n_0}{n} \tag{21}$$

7 Efficiency Comparison

The proposed model will be more efficient than Gjestvang and Singh (2009) model under simple random sampling scheme if the relative efficiency given as

$$\theta_{re} = \frac{V_{gs}(\hat{\mu}_y)}{V_{pro}^{(srs)}(\hat{\mu}_y)}$$

$$= \frac{\alpha\beta(1 + C_s^{-2}) + R_y}{W\alpha\beta(1 + C_s^{-2}) + R_y} > 1 \tag{22}$$

The proposed model will be more efficient than Gjestvang and Singh (2009) model under probability proportional to size sampling scheme if the relative efficiency given as

$$\theta_{re} = \frac{V_{gs}(\hat{\mu}_y)}{V_{pro}^{(pps)}(\hat{\mu}_y)}$$

$$= \frac{N^2(\alpha\beta(1 + C_s^{-2}) + R_y)}{W\alpha\beta(1 + C_s^{-2}) + R_{y(pps)}} > 1 \tag{23}$$

The relative gain in efficiency of the proposed model over Gjestvang and Singh (2009) model is use to compare performance of the proposed model and Gjestvang and Singh (2009) model under both sampling scheme.
The relative gain in efficiency under simple random sampling scheme is defined as

$$\theta_{rg} = 1 - \frac{V_{pro}^{(srs)}(\hat{\mu}_y)}{V_{gs}(\hat{\mu}_y)}$$

$$= 1 - \frac{W\alpha\beta(1 + C_s^{-2}) + R_y}{\alpha\beta(1 + C_s^{-2}) + R_y} \tag{24}$$

and under probability proportional to size with replacement sampling scheme defined as

$$\theta_{rg} = 1 - \frac{V_{pro}^{(pps)}(\hat{\mu}_y)}{V_{gs}(\hat{\mu}_y)}$$

$$= 1 - \frac{W\alpha\beta(1 + C_s^{-2}) + R_{y(pps)}}{N^2(\alpha\beta(1 + C_s^{-2}) + R_y)} \tag{25}$$

Thus, the proposed model will have

1. more gain in efficiency than any existing model if $0 < \theta_{rg} \leq 1$.
2. equal gain in efficiency with any existing model if $\theta_{rg} = 0$.
3. less gain in efficiency than any existing model if $-\infty < \theta_{rg} < 0$.

8 Weighted Privacy - Efficiency Measure

In this section, comparative performance of proposed model and Gjestvang and Singh (2009) model is carried out using Azzem (2023) log weighted privacy - efficiency measure defined as

$$\log \phi = \log\left(\frac{\omega_1\theta_{re} + \omega_2 P}{\omega_1 + \omega_2}\right), \qquad (-\infty < \log \phi < \infty) \tag{26}$$

where,

$$P = \frac{\nabla_{pro}}{\nabla_{gs}} \tag{27}$$

The Yan et al. (2009) privacy protection level ∇ is defined as

$$\nabla = E(Z - Y)^2 \tag{28}$$

Thus, the privacy protection level of the proposed model is obtained as

$$\nabla_{pro} = W\alpha\beta\sigma_s^2(1 + C_s^{-2}) \tag{29}$$

and the privacy protection level of Gjestvang and Singh (2009) model is given as

$$\nabla_{gs} = \alpha\beta\sigma_s^2(1 + C_s^{-2}) \tag{30}$$

Thus (26) becomes

$$\log \phi = \log\left(\frac{1}{\omega_1 + \omega_2}\left\{\omega_1 \frac{\alpha\beta(1 + C_s^{-2}) + R_y}{W\alpha\beta(1 + C_s^{-2}) + R_y} + \omega_2 W\right\}\right) \tag{31}$$

$\log \phi > 0$ indicates that the proposed model performs better than Gjestvang and Singh (2009) model while $\log \phi < 0$ indicates that the proposed model performs poorly than Gjestvang and Singh (2009). For $\log \phi = 0$ indicates that both models are of equal performance.

9 Simulation Study

It should be observed that from (22) and (23) under both sampling schemes, the relative efficiency depends on scrambling parameter α and β, coefficient of variation of scrambling variable C_s, and sensitivity level W. Gjestvang and Singh (2009) has observed that the value of α and β should be such that their product should be closed to zero as much as possible for better efficiency. Cochran (1977) on the other hand suggested that the value of the coefficient of variation should be around the neighbourhood of 0.1

Thus, fixing $C_s = 0.1$, a simulation study was carried out for different values α, β and W to the relative efficiency, relative gain in efficiency of the proposed model. The results were shown in Table 1. Also, simulation study was carried out to observe the weighted privacy - efficiency measure of the proposed model against Gjestvang and Singh (2009) model. The result was shown in Table 2.

Table 1: Relative Efficiency and Gain in Efficiency for the Proposed Model

W	α	β	Relative Efficiency (θ_{re})		Relative Gain in Efficiency (θ_{rg})	
			Under SRS	Under PPS	Under SRS	Under PPS
0.1	0.09	0.10	2.3845	3815.1970	0.5806	0.9997
	0.12	0.08	2.4618	3938.8900	0.5938	0.9997
	0.03	0.09	1.4655	2344.7570	0.3176	0.9996
	0.08	0.08	2.0303	3248.5120	0.5075	0.9997
	0.09	0.05	1.7499	2799.8900	0.4285	0.9996
0.3	0.09	0.10	1.8235	2917.5620	0.4516	0.9997
	0.12	0.08	1.8582	2973.0930	0.4618	0.9997
	0.03	0.09	1.3281	2124.9950	0.2470	0.9995
	0.08	0.08	1.6521	2643.3020	0.3947	0.9996
	0.09	0.05	1.4999	2399.9370	0.3333	0.9996
0.5	0.09	0.10	1.4762	2361.8650	0.3226	0.9996
	0.12	0.08	1.4923	2387.6520	0.3299	0.9996
	0.03	0.09	1.2143	1942.8300	0.1765	0.9995
	0.08	0.08	1.3926	2228.1830	0.2819	0.9996
	0.09	0.05	1.3125	2099.9660	0.2381	0.9996
0.7	0.09	0.10	1.2400	1983.9830	0.1935	0.9995
	0.12	0.08	1.2468	1994.8420	0.1979	0.9995
	0.03	0.09	1.1184	1789.4600	0.1059	0.9994
	0.08	0.08	1.2036	1925.7520	0.1692	0.9995
	0.09	0.05	1.1667	1866.6500	0.1429	0.9995
0.9	0.09	0.10	1.0690	1710.3410	0.0645	0.9994
	0.12	0.08	1.0706	1713.0200	0.0660	0.9994
	0.03	0.09	1.0366	1658.5330	0.0353	0.9994
	0.08	0.08	1.0598	1695.6070	0.0564	0.9994
	0.09	0.05	1.0500	1679.9960	0.0476	0.9994

10 Discussion of Results

By adopting optional scrambled randomized response approach as against forced scrambled randomized response approach, the proposed model effectively balances respondent privacy protection with statistical efficiency — a key trade-off in survey designs involving sensitive questions and also offers improved alternative to Gjestvang and Singh (2009) model.

The relative efficiency of the proposed model under both sampling schemes is consistently greater than one ($\theta_{re} > 1$) for different values of scrambling parameters α, β and sensitivity level W. The implication of this would be, the proposed model will have lower variance and consequently will be more precise than Gjestvang and Singh (2009) model. The relative efficiency values of the proposed model are substantially higher under probability proportional to size sampling scheme than under simple random sampling scheme. This can be attributed to the use of auxiliary information at the sample selection stage through probability proportional to size sampling scheme which enhances the estimator's performance.

In the same vein, relative gain in efficiency is consistently greater than zero ($\theta_{rg} > 0$) for all values of scrambling parameters and sensitivity level, implying that the proposed model consistently improves estimation accuracy. However, as sensitivity level increases, the relative gain in efficiency slightly decreases. This is in line with the theoretical

Table 2: Log Weighted Privacy-Efficiency Measure
Proposed Model

W	α	β	ω_1	ω_2	$\log\phi$	
					Under SRS	Under PPS
0.1	0.09	0.10	0.2	0.8	-0.2542	2.8826
	0.12	0.08	0.5	0.5	0.1075	3.2944
	0.03	0.09	0.8	0.2	0.0764	3.2732
	0.08	0.08	0.2	0.8	-0.3133	2.8128
	0.09	0.05	0.5	0.5	-0.0339	3.1461
0.3	0.09	0.10	0.2	0.8	-0.2185	2.7662
	0.12	0.08	0.5	0.5	0.0331	3.1722
	0.03	0.09	0.8	0.2	0.0502	3.2305
	0.08	0.08	0.2	0.8	-0.2438	2.7234
	0.09	0.05	0.5	0.5	-0.0458	3.0792
0.5	0.09	0.10	0.2	0.8	-0.1579	2.6747
	0.12	0.08	0.5	0.5	-0.0017	3.0770
	0.03	0.09	0.8	0.2	0.0300	3.1916
	0.08	0.08	0.2	0.8	-0.1684	2.6494
	0.09	0.05	0.5	0.5	-0.0428	3.0213
0.7	0.09	0.10	0.2	0.8	-0.0926	2.5992
	0.12	0.08	0.5	0.5	-0.0117	2.9990
	0.03	0.09	0.8	0.2	0.0148	3.1559
	0.08	0.08	0.2	0.8	-0.0965	2.5863
	0.09	0.05	0.5	0.5	-0.0300	2.9702
0.9	0.09	0.10	0.2	0.8	-0.0298	2.5350
	0.12	0.08	0.5	0.5	-0.0064	2.9330
	0.03	0.09	0.8	0.2	0.0040	3.1229
	0.08	0.08	0.2	0.8	-0.0306	2.5313
	0.09	0.05	0.5	0.5	-0.0110	2.9245

expectations: as more respondents see the survey question as sensitive and opt for scrambled response, the amount of truthful response decreases, slightly increasing the estimator's variance. Nevertheless, even at high sensitivity level ($W = 0.9$), the proposed estimator maintains acceptable efficiency and unbiasedness.

Table 2 presents the log weighted privacy-efficiency measure ($\log\phi$) values for both sampling schemes. The results reveal that under simple random sampling scheme, the values of $\log\phi$ are negative in most cases when $\omega_1 = 0.2$ and $\omega_2 = 0.8$ suggesting that if privacy protection is a priority over efficiency, Gjestvang and Singh (2009) model would be preferable to proposed model under simple random sampling. However, under probability proportional to size sampling scheme, the values of $\log\phi$ are positive in all cases suggesting a better joint performance of the proposed model in protecting respondents privacy while maintaining estimation accuracy.

By and large, the model establishes robustness across a wide range of scrambling parameters (α, β) and sensitivity levels (W). It performs particularly well when α and β are close in magnitude and when the coefficient of variation $(C_s = 0.1)$ remains small, validating theoretical assertions by Gjestvang and Singh (2009) and Cochran (1977).

11 Conclusion

The research suggested an additive optional scrambled randomized response technique model which is an improvement on Gjestvang and Singh (2009) model. The model extends Gjestvang and Singh (2009) model to optional randomized response technique by allowing respondents the option to either report a scrambled or truthful response based on their perception of the sensitive nature of the survey question. By this way, both respondents cooperation and data reliability are enhanced while maintaining their confidentiality and privacy. The model is designed to improve the estimation of the population mean and sensitivity level.

Based on the theoretical results, the proposed model establishes an unbiased estimator of the population mean under both simple random sampling and probability proportional to size sampling schemes. Simulation study established that the proposed model consistently outperforms Gjestvang and Singh (2009) model in terms of efficiency and privacy protection. This high performance is more noticeable in probability proportional to size sampling, indicating the importance of using auxiliary information at sample selection stage.

Furthermore, the use of a weighted privacy-efficiency measure revealed that the proposed model achieves higher performance than Gjestvang and Singh (2009) model, indication better overall balance between respondents privacy and estimator precision. consequently, the proposed model is recommended for use in surveys involving sensitive variables.

References

Ahmad, M. A., Elshehawey, A. M., and Mohammad, A. Z. (2024). A New Improved Randomized Response Model With Application to Compulsory Motor Insurance. *Heliyon*, 10(e27252):1–10.

Azeem, M., Najma, S., Sundus, H., and Musarrat, I. (2024). An Efficient Estimator of population Variance of a Sensitive Variable With a New Randomized Response Technique. *Heliyon, 10(224) e27488*, pages 1–11.

Azzem, M. (2023). Introducing a Weighted Measure of Privacy and Efficiency for Comparison of Quantitative Randomized Response Models. *Pak. J. Statist*, 39(3):377–383.

Cochran, W. G. (1977). *Sampling Technique*. Wiley New York, third edition.

Eichhorn, B. H. and Hayre, S. L. (1983). Scrambled Randomized Response Methods for Obtaining Sensitive Quantitative Data. *Journal of Statistical Planning and Inference*, 7:307–316.

Gjestvang, C. R. and Singh, S. (2009). An Improved Randomized Response Model: Estimation of Mean. *Journal of Applied Statistics*, 36(12):1361–1367.

Gupta, S., Shabbir, J., and Sehra, S. (2010). Mean and Sensitivity Estimation in Optional Randomized Response Models. *Journal of Statistical Planning and Inference*, 140(10):2870–2874.

Gupta, S., Zhang, J., Khalil, S., and Sapra, P. (2022). Mitigating Lack of Trust in Quantitative Randomized Response Technique Models. *Communication in Statistics Part B - Simulation Computation*, 51:1–9.

Gupta, S. N., Gupta, B. C., and Singh, S. (2002). Estimation of Sensitivity Level of Personal Interview Survey Questions. *Journal of Statistical Planning and Inference*, 100:239–247.

Mojeed, A. Y., Ahmed, A., and Awwal, A. (2025). Robust Optional Randomized Response Technique Models For Seeking Information About Sensitive Issues. *International Journal of Scientific Research In Mathematical and Statistical Sciences.*, 12(3):01–07.

Narjis, G. and Shabbir, J. (2023). An Efficient New Scrambled Response Model for Estimating Sensitive Population Mean in Successive Sampling. *Communication in Statistics - Simulation and Computation*, 52(11):5317–5344.

Neeraj, T. and Prachi, M. (2017). Additive Randomized Response Model With Known Sensitivity Level. *International Journal of Computational and Theoretical Statistics*, 4(2):83–93.

Parker, M., Gupta, S., and Sadia, K. (2024). A Mixture Quantitative Randomized Response Model That Improves Trust in RRT Methodology. *Axioms*, 13(11):1–13.

Warner, S. L. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of The American Statistical Association*, 60:63–69.

Warner, S. L. (1971). The Linear Randomized Response Model. *Journal of The American Statistical Association*, 66(336):884–888.

Yan, Z., Wang, J., and Lai, J. (2009). An Efficiency and Protection Degree Based Comparison Among the Quantitative Randomized Response Strategies. *Communication in Statistics: Theory and Methods*, 38:400–408.

Zawar, H. and Muhammad, I. S. (2019). Improved Randomized Response in Optional Scrambling Models. *Journal of Statistical Theory and Applications*, 18(4):351–360.