scientific reports

Check for updates

OPEN

Efficient estimation of population variance of a sensitive variable using a new scrambling response model

Iram Saleem[1], Aamir Sanaullah[2]✉, Laila A. Al-Essa[3], Shakila Bashir[1] & Aned Al Mutairi[3]

This study introduces a pioneering scrambling response model tailored for handling sensitive variables. Subsequently, a generalized estimator for variance estimation, relying on two auxiliary information sources, is developed following this novel model. Analytical expressions for bias, mean square error, and minimum mean square error are meticulously derived up to the first order of approximation, shedding light on the estimator's statistical performance. Comprehensive simulation experiments and empirical analysis unveil compelling results. The proposed generalized estimator, operating under both scrambling response models, consistently exhibits minimal mean square error, surpassing existing estimation techniques. Furthermore, this study evaluates the level of privacy protection afforded to respondents using this model, employing a robust framework of simulations and empirical studies.

Information regarding complex characteristics for example family income, induced abortions, criminal activities, etc. may cause refusal bias, response bias, or both. The randomized response technique (RRT) was introduced by Warner[1] to overcome such a difficult situation where the response is qualitative. In RRT, the information attained ensures the privacy of the respondent. This work was extended to quantitative response models with scrambling responses. Pollock and Bek[2] presented the theory of additive and multiplicative scrambling randomized response (SRR) technique. On the choice of scrambling mechanism, many researchers made an effort to develop models such as Himmelfarb and Edgell[3], Eichhron and Hayre[4], Gupta et al.[5], Diana and Perri[6], Hussain and Khan[7], Zaman et al.[8], and Azeem[9]. Besides, developments by different survey researchers are made to the estimation stage using additive models. Sousa et al.[10] first presented ratio-type estimators estimating the population mean of the complex study variable using the non-delicate auxiliary variable. Additionally, Koyuncu et al.[11], Gupta et al.[12], Saleem et al.[13], Shahzad et al.[14], Sanaullah et al.[15,16], Saleem and Sanaullah[17], Khalid et al.[18] and Juarez-Moreno et al.[19] presented various estimators for the population mean of the complex variable using different RRT models.

In human regular life, variation remains existent everywhere. Naturally, not the two individuals or things are identical. In all fields, we necessitate estimating the population variance, such as the climate factors from place to place, the degree of blood pressure, etc. The medical researcher needs a suitable understanding of the level of variation of a particular HIV treatment dose curing or affecting from person to person to be able to plan whether to reduce or change the treatment for a particular person. Practically, several situations can be seen where the estimation of population variance can be observed for complex issues. In survey sampling, the auxiliary variable is used to intensify the precision of population variance estimators at the stage of estimation. The work on the estimation of population variance for the non-sensitive variable of interest was done by countless statisticians such as Gupta and Shabbir[20], Asghar et al.[21], Sanaullah et al.[22], Niaz et al.[23] and Zaman and Bulut[24]. Singh et al.[25] first introduced a new estimator to estimate the population variance of the sensitive variable of interest centered on a multiplicative scrambled response model using auxiliary information. They presented different procedures for estimating variance using Das and Tripathi[26] and Isaki[27] estimators. Later, Gupta et al.[28] presented three variance estimators under Diana and Perri's[6] RRT model using two scrambling variables. Aloraini et al.[29] proposed some separate and combined variance estimators using stratified sampling following the strategy presented by Gupta et al.[28].

[1]Department of Statistics, Forman Christian College (Chartered University), Lahore, Pakistan. [2]Department of Statistics, COMSATS University Islamabad, Lahore Campus, Lahore, Pakistan. [3]Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O.Box 84428, Riyadh 11671, Saudi Arabia. ✉email: chaamirsanaullah@yahoo.com

The present study follows the methodology of Gupta et al.[28] and suggested a new generalized exponential estimator, to estimate the variance of the finite population which is complex in nature. The jargon of the bias and the mean square error of the proposed estimator originated up to the first order of approximation. The outline of the article is organized as follows: In "Sampling strategy for scrambled response model" section the sampling strategy for the scrambled response model presented by Diana and Perri[6] is discussed. "The proposed estimator and its class of estimators" section, displays the proposed generalized estimator for two auxiliary variables under the existing model along with the expressions of bias and MSE. In "The proposed RRT model and estimator-II" section, we also propose a generalized randomized response model. The unbiased variance estimator, ratio estimator, and proposed generalized estimator are modified under the proposed model in the same section. The privacy protection measure for the models are discussed in "Privacy levels" section. To support the proposed methodology a simulation study is presented in "An application of the proposed model" section and some concluding interpretations are given in "Simulation study" section.

Sampling strategy for scrambled response model

Let a simple random sample done without replacement (SRSWOR) of size n be drawn a finite population of $U=\{U_1, U_2,\ldots, U_N\}$. Let Y be a true response of sensitive quantitative variables and X be the non-sensitive auxiliary variable, positively correlated to Y. Let $s_y^2=\frac{\sum_{i=1}^{n}(y_i-\overline{y})^2}{(n-1)}$, $s_{x1}^2=\frac{\sum_{i=1}^{n}(x_{1i}-\overline{x}_1)^2}{(n-1)}$, $s_{x2}^2=\frac{\sum_{i=1}^{n}(x_{2i}-\overline{x})^2}{(n-1)}$, $s_z^2=\frac{\sum_{i=1}^{n}(z_i-\overline{z})^2}{(n-1)}$, $\sigma_y^2=\frac{\sum_{i=1}^{N}(Y_i-\overline{Y})^2}{(N-1)}$, $\sigma_{x1}^2=\frac{\sum_{i=1}^{N}(X_{1i}-\overline{X}_1)^2}{(N-1)}$, $\sigma_{x2}^2=\frac{\sum_{i=1}^{N}(X_{2i}-\overline{X}_2)^2}{(N-1)}$, $\sigma_z^2=\frac{\sum_{i=1}^{N}(Z_i-\overline{Z})^2}{(N-1)}$, $\overline{X}_1=\frac{\sum_{i=1}^{N}X_{1i}}{N}$, $\overline{X}_2=\frac{\sum_{i=1}^{N}X_{2i}}{N}$, $\overline{Y}=\frac{\sum_{i=1}^{N}Y_i}{N}$, $\overline{Z}=\frac{\sum_{i=1}^{N}Z_i}{N}$, $\overline{y}=\frac{\sum_{i=1}^{n}y_i}{n}$, $\overline{x}_1=\frac{\sum_{i=1}^{n}x_{1i}}{n}$, $\overline{x}_2=\frac{\sum_{i=1}^{n}x_{2i}}{n}$, $\overline{z}=\frac{\sum_{i=1}^{n}z_i}{n}$.

Let us define the following assumptions and expectations to get the bias and mean square error: $s_z^2=\sigma_z^2(1+\delta_z)$, $s_{x_1}^2=\sigma_{x_1}^2(1+\delta_{X_1})$, $s_{x_2}^2=\sigma_{x_2}^2(1+\delta_{X_2})$, and $\overline{z}=\overline{Z}(1+e_z)$, where $\delta_z=\frac{s_z^2-\sigma_z^2}{\sigma_z^2}$, $\delta_{X_1}=\frac{s_{x_1}^2-\sigma_{x_1}^2}{\sigma_{x_1}^2}$, $\delta_{X_2}=\frac{s_{x_2}^2-\sigma_{x_2}^2}{\sigma_{x_2}^2}$ and $e_z=\frac{\overline{z}-\overline{z}}{\overline{z}}$, such that $E(\delta_z)=E(\delta_x)=E(e_z)=0$, $E\left(\delta_z^2\right)=\theta(\lambda_{400}-1)$, $E\left(\delta_{x1}^2\right)=\theta(\lambda_{040}-1)$, $\mathrm{E}\left(\delta_{x2}^2\right)=\theta(\lambda_{004}-1)$, $E(e_z^2)=\theta C_z^2$, $E(\delta_z e_z)=\theta\lambda_{300}C_z$, $E(\delta_{x1}e_z)=\theta\lambda_{120}C_z$, $E(\delta_{x2}e_z)=\theta\lambda_{102}C_z$, $E(\delta_{x1}\delta_{x2})=\theta(\lambda_{022}-1)$, $E(\delta_z\delta_{x1})=\theta(\lambda_{220}-1)$, and $E(\delta_z\delta_{x2})=\theta(\lambda_{202}-1)$.

Based on the Diana and Perri[6] RRT model $Z=TY+S$, Gupta et al.[28] introduced basic variance and some ratio-type estimators. The basic variance estimator is as

$$t_0=\sigma_y^2=\frac{\sigma_z^2-\sigma_S^2-\left(\sigma_T^2\overline{Z}^2\right)}{\sigma_T^2+1}. \tag{1}$$

The MSE of t_0 is as,

$$MSE(t_0)=\theta\left(\frac{1}{\left(\sigma_T^2+1\right)^2}\right)\left(\sigma_z^4(\mu_{400}-1)+4\sigma_T^2\overline{Z}^2C_z^2-4\sigma_z^2\sigma_T^2\overline{Z}^2\mu_{300}C_z\right). \tag{2}$$

The ratio estimators is given by,

$$t_{ratio}=\frac{s_z^2-\sigma_s^2-\sigma_T^2\times\overline{z}^2}{\sigma_T^2+1}\times\left(\frac{\sigma_x^2}{s_x^2}\right). \tag{3}$$

The MSE of t_{ratio} estimator is as,

$$\begin{aligned}MSE(t_{ratio})=\theta\frac{1}{\left(\sigma_T^2+1\right)^2}\Big[\sigma_z^4(\mu_{400}-1)-2\sigma_z^2\sigma_{x1}^2(\mu_{220}-1)\left(\sigma_T^2+1\right)+\sigma_s^2(\mu_{040}-1)\left(\sigma_T^2+1\right)^2\Big]\\+\theta\frac{1}{\left(\sigma_T^2+1\right)^2}4C_z\left(\sigma_T^2\overline{Z}^4C_z-\sigma_z^2\sigma_T^2\overline{Z}^2\lambda_{30}+\sigma_T^2\overline{Z}^2\lambda_{12}\left(\sigma_T^2+1\right)\right).\end{aligned} \tag{4}$$

The generalized ratio estimator is as,

$$t_{gratio}=\left(\left(\frac{s_z^2-\sigma_s^2-\sigma_T^2*\overline{z}^2}{\sigma_T^2+1}\right)+\left(\sigma_x^2-s_x^2\right)\right)\times\left(\frac{(\alpha\sigma_x^2+\beta)}{\omega\left(\alpha s_x^2+\beta\right)+(1-\omega)(\alpha\sigma_x^2+\beta)}\right)^g. \tag{5}$$

The MSE of t_0 is as,

$$\begin{aligned}\min MSE\left(t_{gratio}\right)=\theta\left(\frac{\sigma_z^4(\mu_{400}-1)+4\sigma_T^2\overline{Z}^2C_z^2-4\sigma_z^2\sigma_T^2\overline{Z}^2\mu_{300}C_z}{\left(\sigma_T^2+1\right)^2}\right)\\-\theta\left(\frac{1}{(\mu_{040}-1)}\right)\left(\sigma_z^2(\mu_{220}-1)-2\sigma_T^2\overline{Z}^2\mu_{120}C_z\right)^2,\end{aligned} \tag{6}$$

where $\mu_{rsa}=\frac{\mu_{rsa}}{\eta_{200}^{\frac{r}{2}}\eta_{020}^{\frac{s}{2}}\eta_{020}^{\frac{a}{2}}}$, $\eta_{rsa}=\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-\overline{Y})^r(X_{1i}-\overline{X})^s(X_{2i}-\overline{X})^a$, $\theta=\frac{1}{n}$.

The proposed estimator and its class of estimators

The proposed estimator

In this section, a generalized exponential estimator is presented following Koyuncu et al.[11]. The form of the proposed estimator is given by,

$$t_{1D} = \left[k_1\left(\frac{s_z^2-\sigma_s^2-\sigma_T^2\overline{z}^2}{\sigma_T^2+1}\right)+k_2\left(\sigma_{x1}^2-s_{x1}^2\right)+k_3\left(\sigma_{x2}^2-s_{x2}^2\right)\right]\left[\left\{\exp\left(\frac{\sigma_{x1}^2-s_{x1}^2}{\sigma_{x1}^2+s_{x1}^2}\right)\right\}^{\lambda_1}\left(\frac{\sigma_{x2}^2}{s_{x2}^2}\right)^{\lambda_2}\right], \quad (7)$$

where k_1, k_2, k_3, are the three optimizing and unrestricted constants which need to be estimated such that the MSE of the estimator is minimum, and λ_1 and λ_2 are the generalization constants which need to be placed with some suitable values, known parameters, or function of known parameters to get different efficient and or existing estimators. A few examples are shown in Table 1 by setting different values to the constants.

To obtain the Bias, and the MSE, we define the following error terms, rewriting Eq. (7), we have

$$t_{1D} = \left[k_1\left(\frac{\sigma_z^2(1+\delta_z)-\sigma_S^2-\sigma_T^2\overline{Z}^2(1+e_z)^2}{\sigma_T^2+1}\right)+k_2\left(\sigma_{x1}^2-\sigma_{x1}^2(1+\delta_{x1})\right)+k_3\left(\sigma_{x2}^2-\sigma_{x2}^2(1+\delta_{x2})\right)\right]$$
$$\left[\left\{\exp\left(\frac{\sigma_{x1}^2-\sigma_{x1}^2(1+\delta_{x1})}{\sigma_{x1}^2+\sigma_{x1}^2(1+\delta_{x1})}\right)\right\}^{\lambda_1}\left(\frac{\sigma_{x2}^2}{\sigma_{x2}^2(1+\delta_{x2})}\right)^{\lambda_2}\right]. \quad (8)$$

$$\begin{aligned} Bias(t_{1D}) = {} & \sigma_y^2(k_1-1)-k_1\sigma_y^2\theta\frac{1}{2}\left[\lambda_1\left\{\lambda_2(\mu_{022}-1)-\frac{1}{4}\lambda_1(\mu_{040}-1)\right\}-\lambda_2\frac{1}{2}(\lambda_2-1)(\mu_{004}-1)\right] \\ & +k_1\sigma_z^2\theta\frac{1}{2(\sigma_T^2+1)}[\lambda_1(\mu_{220}-1)-2\lambda_2(\mu_{202}-1)]-\frac{\overline{Z}^2\sigma_T^2}{(\sigma_T^2+1)}k_1\theta[C_z+\lambda_1\mu_{120}-2\mu_{120}\lambda_2]C_z \\ & -k_2\sigma_{x1}^2\theta\left[\frac{\lambda_1}{4}((\mu_{040}-1)-\lambda_2(\mu_{022}-1)\right]-k_3\sigma_{x2}^2\theta\left[\frac{\lambda_1}{2}((\mu_{022}-1)-\lambda_2(\mu_{004}-1)\right]. \end{aligned} \quad (9)$$

The mean square error of the generalized estimator t_{1D} is given by

$$\begin{aligned} MSE(t_{1D}) = {} & \sigma_y^4(k_1-1)^2+k_1^2\theta A_1+k_2^2\sigma_{x1}^4\theta(\mu_{040}-1)+k_3^2\sigma_{x2}^4\theta(\mu_{004}-1) \\ & +2k_1k_2\sigma_{x1}^2\theta A_2-2k_1k_3\sigma_{x2}^2\theta A_3+2k_{21}k_3\sigma_{x1}^2\sigma_{x2}^2\theta(\mu_{022}-1). \end{aligned} \quad (10)$$

Differentiate Eq. (10) with respect to k_1, k_2, and k_3, and after the simplification optimum values of the constants are given by,

$$k_{11} = \frac{\sigma_y^4}{A_7}, k_{21} = \frac{-k_1A_4}{\sigma_{x1}^2A_6}, \text{ and } k_{31} = \frac{k_1A_5}{\sigma_{x2}^2A_6},$$

and utilizing the optimum values of the constants into Eq. (10), the simplified form of the minim MSE of the estimator is given by

$$\min MSE(t_{1D}) = \sigma_y^6\left(1-\frac{1}{A_7}\right), \quad (11)$$

Class of estimators	k_1	k_2	k_3	λ_1	λ_2
$t_{1D(1)} = \left[k_1\left(\frac{s_z^2-\sigma_s^2-\sigma_T^2\overline{z}^2}{\sigma_T^2+1}\right)+k_2\left(\sigma_{x1}^2-s_{x1}^2\right)+k_3\left(\sigma_{x2}^2-s_{x2}^2\right)\right]\exp\left(\frac{\sigma_{x1}^2-s_{x1}^2}{\sigma_{x1}^2+s_{x1}^2}\right)$	k_1	k_2	0	1	0
$t_{1D(2)} = \left[k_1\left(\frac{s_z^2-\sigma_s^2-\sigma_T^2\overline{z}^2}{\sigma_T^2+1}\right)+k_2\left(\sigma_{x1}^2-s_{x1}^2\right)+k_3\left(\sigma_{x2}^2-s_{x2}^2\right)\right]\left(\frac{\sigma_{x2}^2}{s_{x2}^2}\right)$	k_1	k_2	0	0	1
$t_{1D(3)} = \left(\frac{s_z^2-\sigma_s^2-\sigma_T^2\overline{z}^2}{\sigma_T^2+1}\right)\left(\frac{\sigma_{x2}^2}{s_{x2}^2}\right)$ Gupta et al.[28]	1	0	0	0	1
$t_{1D(4)} = \left(\frac{s_z^2-\sigma_s^2-\sigma_T^2\overline{z}^2}{\sigma_T^2+1}\right)+k_2\left(\sigma_{x1}^2-s_{x1}^2\right)+k_3\left(\sigma_{x2}^2-s_{x2}^2\right)$	1	k_2	k_3	0	0
$t_{1D(5)} = \left(\left(\frac{s_z^2-\sigma_s^2-\sigma_T^2\overline{z}^2}{\sigma_T^2+1}\right)+k_2\left(\sigma_{x1}^2-s_{x1}^2\right)\right)\exp\left(\frac{\sigma_{x1}^2-s_{x1}^2}{\sigma_{x1}^2+s_{x1}^2}\right)$	1	k_2	0	1	0
$t_{1D(6)} = \left[\left(\frac{s_z^2-\sigma_s^2-\sigma_T^2\overline{z}^2}{\sigma_T^2+1}\right)+k_2\left(\sigma_{x1}^2-s_{x1}^2\right)\right]\left(\frac{\sigma_{x2}^2}{s_{x2}^2}\right)$	1	k_2	0	0	1
$t_{1D(7)} = \left(\frac{s_z^2-\sigma_s^2-\sigma_T^2\overline{z}^2}{\sigma_T^2+1}\right)\left\{\exp\left(\frac{\sigma_{x1}^2-s_{x1}^2}{\sigma_{x1}^2+s_{x1}^2}\right)\right\}^{\lambda_1}$	1	0	0	λ_1	0
$t_{1D(8)} = \left(\frac{s_z^2-\sigma_s^2-\sigma_T^2\overline{z}^2}{\sigma_T^2+1}\right)\left(\frac{\sigma_{x2}^2}{s_{x2}^2}\right)^{\lambda_2}$	1	0	0	0	λ_2

Table 1. The class of estimators for different choices of constant's values.

where $A_1=\sigma_y^4\left[\frac{\lambda_1^2}{4}(\mu_{040}-1)+\lambda_2^2(\mu_{004}-1)-\lambda_1\lambda_2(\mu_{022}-1)\right]+\frac{\sigma_z^4(\mu_{400}-1)}{(\sigma_T^2+1)^2}+4\frac{\overline{Z}^4\sigma_T^4C_z^2}{(\sigma_T^2+1)^2}+\frac{\overline{Z}^2\sigma_T^2\sigma_z^2\mu_{300}C_z}{(\sigma_T^2+1)}$ $-2\frac{\sigma_y^2}{\sigma_T^2+1}\left[\sigma_z^2\left\{\frac{\lambda_1}{2}((\mu_{220}-1)-\lambda_2(\mu_{202}-1)\right\}+2\overline{Z}^2\sigma_T^2\left\{\frac{\lambda_1}{2}\mu_{120}-\lambda_2\mu_{102}\right\}C_z\right]$, $A_2=\sigma_y^2\left[\frac{\lambda_1}{2}(\mu_{040}-1)-\lambda_2(\mu_{022}-1)\right]-\frac{\sigma_z^2(\mu_{220}-1)}{(\sigma_T^2+1)}+2\frac{\overline{Z}^2\sigma_T^2\mu_{120}}{(\sigma_T^2+1)}C_z$, $A_3=\sigma_y^2\left[\frac{\lambda_1}{2}(\mu_{022}-1)-\lambda_2(\mu_{004}-1)\right]-\frac{\sigma_z^2(\mu_{202}-1)}{(\sigma_T^2+1)}+2\frac{\overline{Z}^2\sigma_T^2\mu_{120}}{(\sigma_T^2+1)}C_z$, $A_4=A_2(\mu_{004}-1)+A_3(\mu_{022}-1)$, $A_5=A_2(\mu_{022}-1)+A_3(\mu_{040}-1)$, $A_6=(\mu_{040}-1)(\mu_{004}-1)-(\mu_{022}-1)^2$, and $A_7=\sigma_y^4+\theta A_1+\theta(\mu_{040}-1)\left(\frac{A_4}{A_6}\right)^2+\theta(\mu_{004}-1)\left(\frac{A_5}{A_6}\right)^2+2\theta\frac{A_2A_4}{A_6}-2\theta\frac{A_3A_5}{A_6}+2\theta(\mu_{022}-1)\frac{A_4A_5}{A_6^2}$.

Mathematical comparison of the proposed class of estimators with t_0

i. $MSE(t_0)-MSE\left(t_{1D(1)}\right)=\frac{B_1}{B_2}>1$ is true iff $B_2>B_1$, where $B_1=\left(\frac{1}{(\sigma_T^2+1)^2}\right)\left(\sigma_z^4(\mu_{400}-1)+4\sigma_T^2\overline{Z}^2C_z^2-4\sigma_z^2\sigma_T^2\overline{Z}^2\mu_{300}C_z\right)$, $B_2=\frac{\sigma_y^4}{\theta}(k_1-1)^2+k_1^2B_{21}+k_2^2\sigma_{x1}^4(\mu_{040}-1)+2k_1k_2\sigma_{x1}^2B_{22}$, $B_{21}=\frac{\sigma_y^4}{4}(\mu_{040}-1)+\frac{\sigma_z^4(\mu_{400}-1)}{(\sigma_T^2+1)^2}+4\frac{\overline{Z}^4\sigma_T^4C_z^2}{(\sigma_T^2+1)^2}+\frac{\overline{Z}^2\sigma_T^2\sigma_z^2\mu_{300}C_z}{(\sigma_T^2+1)}-2\frac{\sigma_y^2}{\sigma_T^2+1}\left[\frac{\sigma_z^2}{2}((\mu_{220}-1)+\overline{Z}^2\sigma_T^2\mu_{120}C_z\right]$, $B_{22}=\frac{\sigma_y^2}{2}(\mu_{040}-1)-\frac{\sigma_z^2(\mu_{220}-1)}{(\sigma_T^2+1)}+2\frac{\overline{Z}^2\sigma_T^2\mu_{120}}{(\sigma_T^2+1)}C_z$.

ii. $MSE(t_0)-MSE\left(t_{1D(2)}\right)=\frac{B_1}{B_3}>1$ is true iff $B_3>B_1$, where $B_3=\frac{\sigma_y^4}{\theta}(k_1-1)^2+k_1^2B_{31}+k_2^2\sigma_{x1}^4(\mu_{040}-1)+2k_1k_2\sigma_{x1}^2B_{32}B_{31}=\sigma_y^4\lambda_2^2(\mu_{004}-1)+\frac{\sigma_z^4(\mu_{400}-1)}{(\sigma_T^2+1)^2}+4\frac{\overline{Z}^4\sigma_T^4C_z^2}{(\sigma_T^2+1)^2}+\frac{\overline{Z}^2\sigma_T^2\sigma_z^2\mu_{300}C_z}{(\sigma_T^2+1)}+2\frac{\sigma_y^2}{\sigma_T^2+1}\left[\sigma_z^2\lambda_2(\mu_{202}-1)+2\overline{Z}^2\sigma_T^2\lambda_2\mu_{102}C_z\right]$, $B_{32}=2\frac{\overline{Z}^2\sigma_T^2\mu_{120}}{(\sigma_T^2+1)}C_z-\sigma_y^2\lambda_2(\mu_{022}-1)-\frac{\sigma_z^2(\mu_{220}-1)}{(\sigma_T^2+1)}$.

iii. $MSE(t_0)-\left[MSE\left(t_{1D(3)}\right)=MSE(t_{ratio})\right]=\frac{B_1}{B_4}>1$ is true iff $B_4>B_1$, where $B_4=\sigma_y^4\lambda_2^2(\mu_{004}-1)+\frac{\sigma_z^4(\mu_{400}-1)}{(\sigma_T^2+1)^2}+4\frac{\overline{Z}^4\sigma_T^4C_z^2}{(\sigma_T^2+1)^2}+\frac{\overline{Z}^2\sigma_T^2\sigma_z^2\mu_{300}C_z}{(\sigma_T^2+1)}+2\frac{\sigma_y^2}{\sigma_T^2+1}\left[\sigma_z^2\lambda_2(\mu_{202}-1)+2\overline{Z}^2\sigma_T^2\lambda_2\mu_{102}C_z\right]$.

iv. $MSE(t_0)-MSE\left(t_{1D(4)}\right)=\frac{B_1}{B_5}>1$ is true iff $B_5>B_1$, where $B_5=\left[\frac{\sigma_z^4(\mu_{400}-1)}{(\sigma_T^2+1)^2}+4\frac{\overline{Z}^4\sigma_T^4C_z^2}{(\sigma_T^2+1)^2}+\frac{\overline{Z}^2\sigma_T^2\sigma_z^2\mu_{300}C_z}{(\sigma_T^2+1)}\right]+k_2^2\sigma_{x1}^4(\mu_{040}-1)+k_3^2\sigma_{x2}^4(\mu_{004}-1)+2k_2\sigma_{x1}^2\left[2\frac{\overline{Z}^2\sigma_T^2\mu_{120}}{(\sigma_T^2+1)}C_z-\frac{\sigma_z^2(\mu_{220}-1)}{(\sigma_T^2+1)}\right]-2k_1k_3\sigma_{x2}^2\left[2\frac{\overline{Z}^2\sigma_T^2\mu_{120}}{(\sigma_T^2+1)}C_z-\frac{\sigma_z^2(\mu_{202}-1)}{(\sigma_T^2+1)}\right]+2k_3\sigma_{x1}^2\sigma_{x2}^2(\mu_{022}-1)$.

v. $MSE(t_0)-MSE\left(t_{1D(5)}\right)=\frac{B_1}{B_6}>1$ is true iff $B_6>B_1$, where $B_6=B_{21}+k_2^2\sigma_{x1}^4(\mu_{040}-1)+2k_2\sigma_{x1}^2B_{22}$.

vi. $MSE(t_0)-MSE\left(t_{1D(7)}\right)=\frac{B_1}{B_7}>1$ is true iff $B_7>B_1$, where $B_7=B_{31}+k_2^2\sigma_{x1}^4(\mu_{040}-1)+2k_2\sigma_{x1}^2B_{32}$.

vii. $MSE(t_0)-MSE\left(t_{1D(8)}\right)=\frac{B_1}{B_8}>1$ is true iff $B_8>B_1$, where $B_8=\sigma_y^4\left[\frac{\lambda_1^2}{4}(\mu_{040}-1)+(\mu_{004}-1)-\lambda_1(\mu_{022}-1)\right]+\frac{\sigma_z^4(\mu_{400}-1)}{(\sigma_T^2+1)^2}+4\frac{\overline{Z}^4\sigma_T^4C_z^2}{(\sigma_T^2+1)^2}+\frac{\overline{Z}^2\sigma_T^2\sigma_z^2\mu_{300}C_z}{(\sigma_T^2+1)}-2\frac{\sigma_y^2}{\sigma_T^2+1}\left[\sigma_z^2\left\{\frac{\lambda_1}{2}((\mu_{220}-1)-(\mu_{202}-1)\right\}+2\overline{Z}^2\sigma_T^2\left\{\frac{\lambda_1}{2}\mu_{120}-\mu_{102}\right\}C_z\right]$.

viii. $MSE(t_0)-MSE\left(t_{1D(9)}\right)=\frac{B_1}{B_9}>1$ is true iff $B_9>B_1$, where $B_9=\sigma_y^4\lambda_2^2(\mu_{004}-1)+\frac{\sigma_z^4(\mu_{400}-1)}{(\sigma_T^2+1)^2}+4\frac{\overline{Z}^4\sigma_T^4C_z^2}{(\sigma_T^2+1)^2}+\frac{\overline{Z}^2\sigma_T^2\sigma_z^2\mu_{300}C_z}{(\sigma_T^2+1)}+2\frac{\sigma_y^2}{\sigma_T^2+1}\left[\sigma_z^2\lambda_2(\mu_{202}-1)+2\overline{Z}^2\sigma_T^2\lambda_2\mu_{102}C_z\right]$.

The above (i)–(vii) expressions the conditions under which the proposed class of estimators performs better as compared to the estimators t_0.

The proposed RRT model and estimator-II
The proposed RRT model
Our scrambled randomized response model provides a combination of multiplicative, additive, and subtractive models. Since Y is the sensitive variable of interest and hence subject to social desirability bias. S and R are the two independent scrambling variables and are mutually uncorrelated with Y. We assume

$$Z_{NP}=g(Y+aS)+\left(1-g\right)R(Y+aS), \tag{12}$$

$$\sigma_{Z_{NP}}^2=g^2\left(\sigma_Y^2+a^2\sigma_S^2\right)+(1-g)^2\sigma_R^2\left(\sigma_Y^2+a^2\sigma_S^2\right), \tag{13}$$

$$\begin{aligned}\sigma^2_{Z_{NP}} =& g^2\left(\sigma^2_Y + a^2\sigma^2_S\right) + (1-g)^2\sigma^2_{RY} + a^2(1-g)^2\sigma^2_{RS}\\ =& g^2\left(\sigma^2_Y + a^2\sigma^2_S\right) + (1-g)^2\left(\left(\sigma^2_R\sigma^2_Y + \sigma^2_R\right)(E[Y])^2 + \sigma^2_Y(E[R])^2\right)\\ &+ a^2(1-g)^2\left(\sigma^2_R\sigma^2_S + \sigma^2_R(E[S])^2 + \sigma^2_S(E[R])^2\right)\\ =& \sigma^2_Y\left[g^2 + (1-g)^2\left(\sigma^2_R + 1\right)\right] + a^2\sigma^2_S\left[g^2 + (1-g)^2\left(\sigma^2_R + 1\right)\right] + (1-g)^2\sigma^2_R\mu^2_Y.\end{aligned} \tag{14}$$

Rewriting, we get

$$\sigma^2_Y = \frac{\sigma^2_{Z_{NP}} - (1-g)^2\sigma^2_R\overline{Z}^2}{\left[g^2 + (1-g)^2\left(\sigma^2_R + 1\right)\right]} - a^2\sigma^2_S. \tag{15}$$

(i) Estimating σ^2_{ZNP} by its unbiased estimator s^2_Z,

$$t_{NP1} = \frac{s^2_Z - (1-g)^2\sigma^2_R * \overline{z}^2}{\left[g^2 + (1-g)^2\left(\sigma^2_R + 1\right)\right]} - a^2\sigma^2_S. \tag{16}$$

Rewrite (16) we get

$$t_{NP1} = \frac{\sigma^2_Z(1-\delta_Z) - (1-g)^2\sigma^2_R\overline{Z}^2(1+e_0)^2}{\left[g^2 + (1-g)^2\left(\sigma^2_R + 1\right)\right]} - a^2\sigma^2_S, \tag{17}$$

$$t_{NP1} = \frac{\sigma^2_Z - (1-g)^2\sigma^2_R\overline{Z}^2}{\left[g^2 + (1-g)^2\left(\sigma^2_R + 1\right)\right]} - a^2\sigma^2_S + \frac{\sigma^2_Z\delta_Z - (1-g)^2\sigma^2_R\overline{Z}^2(e^2_0 + 2e_0)}{\left[g^2 + (1-g)^2\left(\sigma^2_R + 1\right)\right]}, \tag{18}$$

$$t_{NP1} - \sigma^2_Y = \frac{\sigma^2_Z\delta_Z - (1-g)^2\sigma^2_R\overline{Z}^2(e^2_0 + 2e_0)}{\left[g^2 + (1-g)^2\left(\sigma^2_R + 1\right)\right]}. \tag{19}$$

By pertaining expectations together on (19), the bias we obtain is as

$$Bias\,(t_{NP1}) = -\theta\frac{(1-g)^2\sigma^2_R\overline{Z}^2}{\left[g^2 + (1-g)^2\left(\sigma^2_R + 1\right)\right]}C^2_Z, \tag{20}$$

$$MSE(t_{NP1}) = \theta\frac{1}{\left[g^2 + (1-g)^2\left(\sigma^2_R + 1\right)\right]^2}\left[\sigma^4_Z(\lambda_{400} - 1) + 4(1-g)^4\sigma^4_R\overline{Z}^2C^2_Z - 4(1-g)^2\sigma^2_R\overline{Z}^2\lambda_{300}C_Z\right]. \tag{21}$$

(ii) Ratio estimator under the proposed randomized model:

$$t_{NP2} = \left[\frac{s^2_Z - (1-g)^2\sigma^2_R \times \overline{z}^2}{\left[g^2 + (1-g)^2\left(\sigma^2_R + 1\right)\right]} - a^2\sigma^2_S\right]\left[\frac{\sigma^2_{X1}}{s^2_{X1}}\right]. \tag{22}$$

Rewrite (22) we get

$$t_{NP2} = \left[\frac{\sigma^2_Z(1-\delta_Z) - (1-g)^2\sigma^2_R\overline{Z}^2(1+e_0)^2}{\left[g^2 + (1-g)^2\left(\sigma^2_R + 1\right)\right]} - a^2\sigma^2_S\right]\left[\frac{\sigma^2_{X1}}{\sigma^2_{X1}(1-\delta_{X1})}\right], \tag{23}$$

$$t_{NP2} = \left[\sigma^2_{Y+}\frac{\sigma^2_Z\delta_Z - (1-g)^2\sigma^2_R\overline{Z}^2(e^2_0 + 2e_0)}{\left[g^2 + (1-g)^2\left(\sigma^2_R + 1\right)\right]}\right](1-\delta_{X1})^{-1}. \tag{24}$$

By pertaining expectations together on (24), the Bias and mean square error we obtain are as

$$Bias\,(t_{NP2}) = \frac{\theta}{G}\left[(\lambda_{040} - 1)G - \left(\sigma^4_Z\lambda_{120}C_Z + (1-g)^2\sigma^2_R\overline{Z}^2(C_Z - 2\lambda_{120})\right)C_Z\right], \tag{25}$$

$$MSE(t_{NP2}) = \frac{\theta}{G}\left[\begin{array}{l}\frac{1}{G}\left\{\sigma^4_Z(\lambda_{400} - 1)G - (1-g)^2\sigma^2_R\overline{Z}^2\left(2\lambda_{300} - (1-g)^2\sigma^2_R\overline{Z}^2\right)C_Z\right\}\\ -\sigma^2_Y\left\{\sigma^2_Z(\lambda_{220} - 1) + (1-g)^2\sigma^2_R\overline{Z}^2\lambda_{120}C_Z - \sigma^2_Z(\lambda_{040} - 1)\right\}\end{array}\right], \tag{26}$$

where

$$G = \left[g^2 + (1-g)^2\left(\sigma^2_R + 1\right)\right].$$

The proposed estimator under the proposed RRT model

The exponential estimator expressed in (7) can be generalized in the situation of two auxiliary non-sensitive variables as,

$$t_{1N} = \left[w_1 \left(\frac{s^2_{Z_{NP}} - (1-g)^2 \sigma^2_R \bar{z}^2}{\left[g^2 + (1-g)^2(\sigma^2_R + 1)\right]} - a^2\sigma^2_S \right) + w_2(\sigma^2_{x1} - s^2_{x1}) + w_3(\sigma^2_{x2} - s^2_{x2}) \right] \left[\left\{ \exp\left(\frac{\sigma^2_{x1} - s^2_{x1}}{\sigma^2_{x1} + s^2_{x1}} \right) \right\}^{v_1} \left(\frac{\sigma^2_{x2}}{s^2_{x2}} \right)^{v_2} \right]. \tag{27}$$

Rewrite (27) we get,

$$\begin{aligned} t_{1N} = &\left[w_1 \left\{ \frac{\sigma^2_Z(1-\delta_Z) - (1-g)^2\sigma^2_R \overline{Z}^2(1+e_0)^2}{\left[g^2 + (1-g)^2(\sigma^2_R+1)\right]} - a^2\sigma^2_S \right\} \right. \\ &\left. + w_2\left(\sigma^2_{x1} - \sigma^2_{x1}(1-\delta_{x1})\right) + w_3\left(\sigma^2_{x2} - \sigma^2_{x2}(1-\delta_{x2})\right) \right] \\ &\left[\left\{ \exp\left(\frac{\sigma^2_{x1} - \sigma^2_{x1}(1-\delta_{x1})}{\sigma^2_{x1} + \sigma^2_{x1}(1-\delta_{x1})} \right) \right\}^{v_1} \left(\frac{\sigma^2_{x2}}{\sigma^2_{x2}(1-\delta_{x2})} \right)^{v_2} \right]. \end{aligned} \tag{28}$$

The expression of the bias and MSE of t_{1N} to the first order of approximation is given by,

$$\begin{aligned} Bias(t_{1N}) = &\sigma^2_y(w_1 - 1) - w_1\theta \left[(1-g)^2 \overline{Z}^2 \sigma^2_R C_Z \left(\frac{C_z}{g^2 + (1-g)^2(\sigma^2_R + 1)} + \mu_{120} \right) \right. \\ &\left. + v_1 \frac{1}{2}\left(\sigma^2_z(\mu_{022} - 1) + v_1 \frac{1}{4}(\mu_{040} - 1) \right) - v_2 \sigma^2_Y(\mu_{004} - 1) \right] \\ &+ \theta \frac{1}{2}(\mu_{004} - 1)\left[w_2 \sigma^2_{x1} v_1 + v_2(v_2 - 1)\sigma^2_Y \right] - k_3 \sigma^2_{x2} \theta \frac{1}{2}(\mu_{022} - 1), \end{aligned} \tag{29}$$

$$\begin{aligned} MSE(t_{1N}) = &\sigma^4_y(w_1 - 1)^2 + w_1^2\theta B_1 + k_2^2\sigma^4_{x1}\theta(\mu_{040} - 1) \\ &+ \frac{1}{4}k_3^2\sigma^4_{x2}\theta(\mu_{004} - 1) + 2k_1k_2\sigma^2_{x1}\theta B_2 - 2k_1k_{23}\sigma^2_{x2}\theta B_3 \\ &+ k_{12}k_{23}\sigma^2_{x1}\sigma^2_{x2}\theta(\mu_{022} - 1). \end{aligned} \tag{30}$$

Differentiate with respect to w_1, w_2 and w_3, the optimum values attained are as $w_1 = \frac{\sigma^4_y}{\sigma^4_y + \theta B_8}$, $w_2 = \frac{-w_1 B_4}{\sigma^2_{x1} B_5}$ and $w_3 = \frac{w_1 B_6}{\sigma^2_{x2} B_7}$.

The MSE imputing this optimum value is given as,

$$\min MSE(t_{1N}) = 1 - \frac{\sigma^6_y}{\sigma^4_y + \theta B_8}, \tag{31}$$

where $D_1 = \frac{1}{\left[g^2 + (1-g)^2(\sigma^2_R+1)\right]^2}\left[\sigma^4_z(\mu_{400} - 1) + 4(1-g)^4\sigma^4_z \overline{Z}^4 C^2_z - 2(1-g)^2\sigma^2_z\sigma^2_R\overline{Z}^2\mu_{300}C_z \right]$, $D_2 = \frac{1}{\left[g^2 + (1-g)^2(\sigma^2_R+1)\right]}\left[\sigma^2_z\{v_1(\mu_{220} - 1) + v_2(\mu_{202} - 1)\} - 2(1-g)^2\sigma^2_R\overline{Z}^2\{v_1\mu_{120} + v_2\mu_{102}\}C_z \right]$, $D_3 = v_1^2(\mu_{040} - 1) + v_2^2(\mu_{004} - 1) - 2v_1v_2(\mu_{022} - 1)$, $D_4 = \theta\left(D_1 + \sigma^4_y D_2 - 2\sigma^2_y D_3 \right)$, $B_1 = D_1 + \sigma^4_y\left[\frac{v_1^2}{2}(\mu_{040} - 1) + v_2^2 (\mu_{004} - 1) - 2v_1v_2(\mu_{022} - 1) \right] - \sigma^2_y[v_1D_2 - v_2D_3 + 2v_1v_2(\mu_{022} - 1)]$, $B_2 = D_2 + \left[\sigma^2_y(\mu_{040} - 1) + 2(\mu_{022} - 1) \right] v_1$, $B_3 = D_3 + \left[\sigma^2_y(\mu_{022} - 1) + 2(\mu_{004} - 1) \right] v_2$, $B_4 = [B_2(\mu_{004} - 1) - B_3(\mu_{022} - 1)]$, $B_5 = \sigma^2_{X1}[(\mu_{040}$ $B_5 = \sigma^2_{X1}[(\mu_{040} - 1)(\mu_{004} - 1)$ $B_6 = [B_2(\mu_{022} - 1) - B_3(\mu_{040} - 1)]$, $B_7 = \sigma^2_{X2}[(\mu_{004} - 1)(\mu_{040} - 1) - (\mu_{022} - 1)^2]$, and,

$$B_8 = B_1 + (\mu_{040} - 1)\left[\frac{B_4}{B_5} \right]^2 + (\mu_{004} - 1)\left[\frac{B_6}{B_7} \right]^2 - 2\left[\frac{B_2B_4}{B_5} - \frac{B_3B_6}{B_7} + (\mu_{022} - 1)\frac{B_4B_6}{B_5B_7} \right].$$

Privacy levels

In the literature, many privacy protection measures are presented by different authors. For our study, privacy measure due to Yan et al.[30] is used to compute privacy for Diana and Perri's[6] model, and the proposed randomized response model.

The privacy protection measure presented by Yan et al. is given by,

$$\Delta = E(Z_i - Y_i)^2. \tag{32}$$

a. Diana and Perri[6] model is given by,

$$Z = RY + S. \tag{33}$$

The privacy protection level is given by

$$\begin{aligned}\Delta_D =& E(Z_i - Y_i)^2 \\ =& \sigma_R^2\left(\mu_Y^2 + \sigma_y^2\right) + \sigma_s^2.\end{aligned} \tag{34}$$

b. The proposed Randomized response model's privacy protection mlevel is as,

$$\begin{aligned}\Delta_{PN} =& E(Z_i - Y_i)^2 \\ =& \left(\mu_Y^2 + \sigma_y^2 + a^2\sigma_s^2\right)\left(1 + \left(1-g\right)^2\sigma_R^2\right) - \left(\mu_Y^2 + \sigma_y^2\right).\end{aligned} \tag{35}$$

c. Comparison of privacy protection levels for Diana and Perris'[6] model and proposed model. Using (34) and (35), we have

$$\Delta_{PN} - \Delta_D = \left(\mu_Y^2 + \sigma_y^2\right)\sigma_R^2 g\left(g-2\right) - \sigma_s^2\left[1 - a^2\left(1 + \left(1-g\right)^2\sigma_R^2\right)\right] > 0, \tag{36}$$

$$\text{iff}\, g\left(g-2\right) > \frac{\sigma_s^2\left[1 - a^2\left(1+\sigma_R^2\right)\right]}{\left(\sigma_R^2\left(\mu_Y^2 + \sigma_y^2\right) - a^2\sigma_s^2\sigma_R^2\right)}.$$

An application of the proposed model

In this section, motivated by Saleem and Sanaullah[17] a real-life application is presented to analyze the efficiency of the proposed RRT model compared to the existing models.

A survey is organized to collect real data for the problem of the estimation of the true variance of the Grade Point Average (GPA) of the students of the Department of Statistics, in Forman Christian College University Lahore, who have studied the Course: *Statistical Methods* in Spring 2023. Ninety students registered in three sections in this statistics course are considered as our population. In this application, the variable of interest Y is the CGPA of students, and the two auxiliary variables i.e., X_1 is the weekly study hours, and X_2 is the number of courses studied in recent semesters. For the scrambling variables, S is a normal random variable with a mean equal to zero and a standard deviation equal to 2, and R is a normal random variable with a mean equal to 1 and a standard deviation equal to 0.02. The following are some characteristics of the population:

$$\mathrm{N} = 90, \mu_{X_1} = 27.61;\ \mu_{X_2} = 19.88;\ \sigma_{X_1} = 8.66;\ \sigma_{X_2} = 18.83.$$

For model, $Z = RY + S$,

$$\mu_Z = 3.889;\ \sigma_Z = 2.59;\ \rho_{ZX_1} = -0.053;\ \rho_{ZX_2} = 0.017.$$

For model, $Z_{NP} = g(Y + aS) + \left(1-g\right)R(Y + aS)$,

		$Z = R\times Y + S$		$Z_{NP} = g(Y + aS) + \left(1-g\right)R(Y + aS)$		
n	**Estimators**	**Mean**	**MSE**	**Estimators**	**Mean**	**MSE**
20	t_0	5.3726	8.2592	t_{NP1}	73.5542	229.6445
	t_{ratio}	7.9481	4.2433	t_{NP2}	107.2465	78.2292
	t_{1D}	3.4357	3.3778	t_{1N}	47.0329	64.1311
38	t_0	5.3279	3.4451	t_{NP1}	73.5636	96.6005
	t_{ratio}	6.4387	2.3176	t_{NP2}	87.7502	96.1805
	t_{1D}	3.4072	1.4090	t_{1N}	47.0429	39.5865

Table 2. The MSEs of the estimators for real population.

Model	Population I	Population II
Diana and Perri[6]	10.1472	7.1942
Proposed model	5.6528	4.9476

Table 3. Privacy level for two populations.

Var(S)	*N*	Estimators	Mean $\left(\widehat{\sigma}_y^2\right)$	MSE	PRE	ϑ
0.2	200	t_0	12.6563	4.1872	100.00	0.4126
		t_{ratio}	12.7547	4.1573	100.72	0.4097
		t_{gratio}	12.6568	4.1013	102.09	0.4042
		t_{1D}	12.0520	3.9767	105.29	0.3919
	500	t_0	12.6540	1.2985	100.00	0.1280
		t_{ratio}	12.6910	1.0253	126.65	0.1010
		t_{gratio}	12.6539	1.0065	129.01	0.0980
		t_{1D}	12.0586	0.2382	545.13	0.0235
0.5	200	t_0	14.6764	4.1979	100.00	0.4137
		t_{ratio}	14.8676	4.0264	104.26	0.3968
		t_{gratio}	14.6795	3.9598	106.01	0.3902
		t_{1D}	13.0802	2.9901	140.39	0.0976
	500	t_0	14.7046	1.4063	100.00	0.1386
		t_{ratio}	14.7376	1.0221	137.59	0.1007
		t_{gratio}	14.7047	0.9945	141.41	0.0980
		t_{1D}	13.0697	0.2428	579.20	0.0239
1	200	t_0	13.1945	3.2376	100.00	0.3191
		t_{ratio}	13.3712	3.2163	100.66	0.3170
		t_{gratio}	13.1975	3.1964	101.29	0.3150
		t_{1D}	12.0749	0.9605	337.07	0.0947
	500	t_0	13.1984	1.3192	100.00	0.1300
		t_{ratio}	13.2229	0.8057	163.73	0.0794
		t_{gratio}	13.1979	0.7942	166.10	0.0783
		t_{1D}	12.0569	0.2420	545.12	0.0238

Table 4. The MSEs and PREs of the estimators for Population I with $\sigma_T^2 = 0.5$ using Z = YT + S.

Var(S)	N	Estimators	Mean $\left(\widehat{\sigma}_y^2\right)$	MSE	PRE	ϑ
0.2	200	t_0	8.1658	1.5367	100.00	0.2136
		t_{ratio}	8.2248	1.4245	107.88	0.1980
		t_{gratio}	8.1710	1.3286	115.66	0.1847
		t_{1D}	8.0373	1.1215	137.02	0.1559
	500	t_0	8.1555	0.3809	100.00	0.0529
		t_{ratio}	8.1822	0.3570	106.69	0.0496
		t_{gratio}	8.1582	0.3324	114.59	0.0462
		t_{1D}	8.0517	0.2871	132.67	0.0399
0.5	200	t_0	8.9991	2.1105	100.00	0.2934
		t_{ratio}	9.0607	1.8331	115.13	0.2548
		t_{gratio}	9.0032	1.7227	122.51	0.2395
		t_{1D}	8.0269	1.1276	187.17	0.1567
	500	t_0	8.9905	0.5165	100.00	0.0718
		t_{ratio}	9.0144	0.4564	113.17	0.0634
		t_{gratio}	8.9923	0.4291	120.37	0.0596
		t_{1D}	8.0471	0.2798	184.60	0.0389
1	200	t_0	8.7447	2.7373	100.00	0.3805
		t_{ratio}	8.8042	2.6687	102.57	0.3710
		t_{gratio}	8.7519	2.5569	107.06	0.3554
		t_{1D}	8.0489	1.1238	243.58	0.1562
	500	t_0	8.7367	0.6678	100.00	0.0928
		t_{ratio}	8.7636	0.6414	104.12	0.0892
		t_{gratio}	8.7390	0.6144	108.69	0.0854
		t_{1D}	8.0495	0.2794	239.01	0.0388

Table 5. The MSEs and PREs of the estimators for Population II with $\sigma_T^2 = 0.5$ using Z = YT + S.

Var(S)	N	Estimators	Population I				Population II			
			Mean $\left(\hat{\sigma}_y^2\right)$	MSE	PRE	ϑ	Mean $\left(\hat{\sigma}_y^2\right)$	MSE	PRE	ϑ
0.2	200	t_{NP1}	12.8756	3.3689	100.00	0.5960	7.9867	1.3649	100.00	0.2759
		t_{NP2}	13.0140	3.5686	94.40	0.6313	8.0386	1.2930	105.56	0.2613
		t_{1N}	11.0777	1.0035	335.71	0.1775	7.0332	1.1664	117.02	0.2358
	500	t_{NP1}	12.8432	0.8166	100.00	0.1445	7.5139	0.3972	100.00	0.0803
		t_{NP2}	12.8931	1.1165	73.14	0.1975	7.5327	0.3666	108.35	0.0741
		t_{1N}	11.0683	0.2599	314.20	0.0460	7.0452	0.2466	161.07	0.0498
0.5	200	t_{NP1}	13.4345	4.4759	100.00	0.7918	8.0543	1.8112	100.00	0.3661
		t_{NP2}	13.5454	5.0172	89.21	0.8876	8.1007	1.5241	118.84	0.3080
		t_{1N}	12.0603	1.0186	439.42	0.1802	7.4002	1.1855	152.78	0.2396
	500	t_{NP1}	13.4065	1.0886	100.00	0.1926	8.0695	0.4521	100.00	0.0914
		t_{NP2}	13.4385	1.2013	90.62	0.2125	8.0792	0.3813	118.57	0.0771
		t_{1N}	12.0664	0.2485	438.07	0.0440	7.0388	0.2986	151.41	0.0604
1	200	t_{NP1}	13.9783	4.5249	100.00	0.8005	7.6827	1.9927	100.00	0.4028
		t_{NP2}	14.0973	4.8623	93.06	0.8602	7.7180	1.6920	117.77	0.3420
		t_{1N}	13.0412	1.3838	326.99	0.2448	7.1079	1.2108	164.58	0.2447
	500	t_{NP1}	13.9681	1.0967	100.00	0.1940	7.7117	0.5027	100.00	0.1016
		t_{NP2}	13.9922	1.3800	79.47	0.2441	7.7162	0.4298	116.96	0.0869
		t_{1N}	13.0686	0.3347	327.67	0.0592	7.0337	0.2945	170.70	0.0595

Table 6. The MSEs of the estimators for population I and II with σ_T^2 using the proposed model.

$$\mu_{Z_{NP}} = 52.31;\ \sigma_{Z_{NP}} = 6.78;\ \rho_{Z_{NP}X_1} = -0.044;\ \rho_{Z_{NP}X_2} = 0.140.$$

Table 2 shows the results for MSE estimates of the model given by Diana and Perri[6] and the proposed model. The resuts are obtained byusing two different sample sizes n = 20 and 38. One can notice that the proposed estimator provides minimum and better results as compared to the other estimators under both models.

Simulation study

In this section, we conduct a simulation study to evaluate the performance of the proposed generalized exponential-type estimators by comparing some existing variance estimators.

Population I:

$$\Sigma = \begin{bmatrix} 10 & 3 & 2.9 \\ 3 & 2 & 1.1 \\ 2.9 & 1.1 & 2 \end{bmatrix}, \rho_{x1y} = 0.6817, \rho_{x2y} = 0.6705.$$

Population II:

$$\Sigma = \begin{bmatrix} 6 & 3 & 2.9 \\ 3 & 2 & 1.1 \\ 2.9 & 1.1 & 2 \end{bmatrix}, \rho_{x1y} = 0.8706, \rho_{x2y} = 0.8706.$$

For both populations, we ruminate three different samples of sizes 200, 300, and 500. The variance of S i.e. Var (S) and variance of T, i.e.Var (T) choose different values for simulation.

Table 3 provides the privacy protection level of the RRT models discussed in this study, we follow Gupta et al.[31] unified measure of the estimator and is given by

$$\vartheta = \frac{MSE(t_i)}{\Delta_j} \times 100,$$

where $i = 0$, $ratio$, $1D$, $NP1$, $NP2$, 1; $j = D$ and PN, $MSE(t_i)$ is the theoretical MSE of the various estimators and Δ_j is the privacy level for Diana and Perri's[6] model and the proposed model as discussed in "Privacy levels" section.

Tables 4, 5 and 6 give the MSE and percent relative efficiency (PRE) results for the proposed estimator and existing estimators discussed in this article. The following expression is implied to get the PRE,

$$PRE = \frac{MSE(t_0)}{MSE(t_i)} \times 100,$$

where i = ratio, gratio, and gep.

The results are presented in Tables 4, 5 and 6. The Tables 4 and 5 provides the numerical results of estimators dicussed in "Sampling strategy for scrambled response model" and "The proposed estimator and its class of estimators" sections whereas the Table 6 presented the results of estimators discussed in "The proposed RRT model and estimator-II" section based on proposed model. The values from Tables 4, 5 and 6 confirm that the existing estimators presented by Gupta et al.[28] are less efficient as compared to the generalized estimator. Also while comparing the proposed model and existing model estimator results in these tables on mayobseve that the proposed model provides more efficient MSE values as compared to the model presented by Diana and Perri[6]. As we can see as the variance of S increases the MSE decreases.

A smaller value of ϑ is to be preferred. Tables 3, 4 and 5 presents the unified measure along with the PRE of the estimators. It is observed that the proposed generalized estimator using two auxiliary variables efficiently performs either using Diana and Perri's model or the proposed RRT model. One can notice that the values of ϑ are smaller for the proposed generalized model.

Conclusion

This study addressed the estimation of population variance for sensitive study variables using a non-sensitive auxiliary variable. A generalized exponential-type estimator, based on Diana and Perri's[6] randomized response model, was introduced and evaluated against estimators proposed by Gupta et al.[28], as detailed in Tables 4, 5 and 6. The comparative analysis indicated that the proposed estimator consistently demonstrated superior efficiency in variance estimation. Additionally, we introduced a novel generalized scrambled response model and applied it to conventional variance and ratio estimators, along with the proposed estimator. In "An application of the proposed model" section, a real survey-based study was presented, applying the proposed RRT model. The results, obtained under both our novel model and the model presented by Diana and Perri[6], revealed that the proposed estimator consistently outperformed conventional mean and ratio estimators in minimizing MSE. Notably, as the sample size increased, the efficiency of the estimator further improved. Moreover, a simulation study was conducted, and the findings are summarized in Tables 3, 4, 5 and 6, comparing expected variances, MSE, and the precision (PRE). The results indicated that the generalized proposed estimator under the proposed randomized response model consistently provided the minimum MSE for both populations, outperforming the estimator's MSE results using Diana and Perri's[6] model. This research study contributes valuable insights into variance estimation for sensitive variables. The proposed generalized estimator, underpinned by the innovative scrambled response model, demonstrated robustness, scalability, and superior performance in both real and simulated scenarios. These findings underscore the potential of this approach in advancing the precision and reliability of population variance estimation in sensitive contexts.

Data availability

The data set used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 30 May 2023; Accepted: 19 October 2023
Published online: 14 November 2023

References

1. Warner, S. L. Randomized response: A survey technique for eliminating evasive answer bias. *J. Am. Stat. Assoc.* **60**(309), 63–69 (1965).
2. Pollock, K. H. & Bek, Y. A comparison of three randomized response models for quantitative data. *J. Am. Stat. Assoc.* **71**(356), 884–886 (1976).
3. Himmelfarb, S. & Edgell, S. E. Additive constants model: A randomized response technique for eliminating evasiveness to quantitative response questions. *Psychol. Bull.* **87**(3), 525 (1980).
4. Eichhron, B. H. & Hayre, L. S. Scrambled randomized response methods for obtaining sensitive quantitative data. *J. Stat. Plan. Inference* **7**, 307–316 (1983).
5. Gupta, S., Shabbir, J. & Sehra, S. Mean and sensitivity estimation in optional randomized response models. *J. Stat. Plan. Inference* **140**(10), 2870–2874 (2010).
6. Diana, G. & Perri, P. F. A class of estimators for quantitative sensitive data. *Stat. Pap.* **52**(3), 633–650 (2011).
7. Hussain, Z. & Khan, K. On estimation of sensitive mean using scrambled data. *World Appl. Sci. J.* **23**(9), 1201–1206 (2013).
8. Zaman, Q., Ijaz, M. & Zaman, T. A randomization tool for obtaining efficient estimators through focus group discussion in sensitive surveys. *Commun. Stat. Theory Methods* **52**(10), 3414–3428 (2023).
9. Azeem, M. Using the exponential function of scrambling variable in quantitative randomized response models. *Math. Methods Appl. Sci.* **1**, 1 (2023).
10. Sousa, R., Shabbir, J., Real, P. C. & Gupta, S. Ratio estimation of the mean of a sensitive variable in the presence of auxiliary information. *J. Stat. Theory Pract.* **4**(3), 495–507 (2010).
11. Koyuncu, N., Gupta, S. & Sousa, R. Exponential-type estimators of the mean of a sensitive variable in the presence of non-sensitive auxiliary information. *Commun. Stat. Simul. Comput.* **43**(7), 1583–1594 (2014).
12. Gupta, S., Kalucha, G. & Shabbir, J. A regression estimator for finite population mean of a sensitive variable using an optional randomized response model. *Commun. Stat. Simul. Comput.* **46**(3), 2393–2405 (2017).
13. Saleem, I., Sanaullah, A. & Hanif, M. Double-sampling regression-cum-exponential estimator of the mean of a sensitive variable. *Math. Popul. Stud.* **26**(3), 163–182 (2019).
14. Shahzad, U., Perri, P. F. & Hanif, M. A new class of ratio-type estimators for improving mean estimation of nonsensitive and sensitive variables by using supplementary information. *Commun. Stat. Simul. Comput.* **48**(9), 2566–2585 (2019).
15. Sanaullah, A., Saleem, I. & Shabbir, J. Use of scrambled response for estimating mean of the sensitivity variable. *Commun. Stat. Theory Methods* **49**(11), 2634–2647 (2020).
16. Sanaullah, A., Saleem, I., Gupta, S. & Hanif, M. Mean estimation with generalized scrambling using two-phase sampling. *Commun. Stat. Simul. Comput.* **51**, 1–15 (2020).
17. Saleem, I. & Sanaullah, A. Estimation of mean of a sensitive variable using efficient exponential-type estimators in stratified sampling. *J. Stat. Comput. Simul.* **92**(2), 232–248 (2022).

18. Khalid, A., Sanaullah, A., Almazah, M. M. A. & Al-Duais, F. S. An efficient ratio-cum-exponential estimator for estimating the population distribution function in the existence of non-response using an SRS design. *Mathematics* **11**, 1312. https://doi.org/10.3390/math11061312 (2023).
19. Juárez-Moreno, P. O., Santiago-Moreno, A., Sautto-Vallejo, J. M. & Bouza-Herrera, C. N. Scrambling reports: New estimators for estimating the population mean of sensitive variables. *Mathematics* **11**(11), 2572 (2023).
20. Gupta, S. & Shabbir, J. Variance estimation in simple random sampling using auxiliary information. *Hacettepe J. Math. Stat.* **37**(1), 57 (2008).
21. Asghar, A., Sanaullah, A. & Hanif, M. Generalized exponential type estimator for population variance in survey sampling. *Rev. Colomb. Estad.* **37**(1), 213–224 (2014).
22. Sanaullah, A., Hanif, M. & Asghar, A. Generalized exponential estimators for population variance under two-phase sampling. *Int. J. Appl. Comput. Math.* **2**(1), 75–84 (2016).
23. Niaz, I., Sanaullah, A., Saleem, I. & Shabbir, J. An improved efficient class of estimators for the population variance. *Concurr. Comput. Pract. Exp.* **34**(4), e6620 (2022).
24. Zaman, T. & Bulut, H. A new class of robust ratio estimators for finite population variance. *Sci. Iran.* https://doi.org/10.24200/sci.2022.57175.5100 (2022).
25. Singh, S., Sedory, S. A. & Arnab, R. Estimation of finite population variance using scrambled responses in the presence of auxiliary information. *Commun. Stat. Simul. Comput.* **44**(4), 1050–1065 (2015).
26. Das, A. K. & Tripathi, T. P. Use of auxiliary information in estimating the finite population variance. *Sankhya* **40**, 139–148 (1978).
27. Isaki, C. T. Variance estimation using auxiliary information. *J. Am. Stat. Assoc.* **78**(381), 117–123 (1983).
28. Gupta, S., Qureshi, M. N. & Khalil, S. Variance estimation using randomized response technique. *REVSTAT Stat. J.* **18**(2), 165–176 (2020).
29. Aloraini, B., Khalil, S., Qureshi, M. N. & Gupta, S. Estimation of population variance for a sensitive variable in stratified sampling using randomized response technique. *REVSTAT Stat. J.* **1**, 1 (2022).
30. Yan, Z., Wang, J. & Lai, J. An efficiency and protection degree-based comparison among the quantitative randomized response strategies. *Commun. Stat. Theory Methods* **38**(3), 400–408 (2008).
31. Gupta, S., Mehta, S., Shabbir, J. & Khalil, S. A unified measure of respondent privacy and model efficiency in quantitative RRT models. *J. Stat. Theory Pract.* **12**, 506–511 (2018).

Acknowledgements
Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2023R368), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author contributions
I.S.: conceptualization, theoretical and mathematical framework; software handling, data visualization, and analysis; writing the first draft. A.S.: Concept, supervision, data visualization analysis, editing and reviewing the final draft and wrote the first draft. L.A.A.-E & A.A.-M..: Editing, reviewed and proofread the final draft, Project management. resources and funding. S.B.: editing, formatting, and reviewing the final draft of the manuscript. All authors reviewed the final drafts of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to A.S.