Natural and Applied Sciences International Journal (NASIJ)
eISSN: 2788-4619 (online)
https://www.ideapublishers.org/index.php/nasij
Vol. 4, No. 2 (July-December), 2023, pp. 108-123
https://doi.org/10.47264/idea.nasij/4.2.7 Research Article

Median estimation in the presence of non-response using randomised response technique

Attique Hussain[1] | Muhammad Shahid Mahmood[1] | Amjad Mahmood*[2, 3] | Ayesha Ashraf [4]

1. National College of Business Administration and Economics, Lahore, Pakistan.
2. Punjab College of Information Technology, Lahore, Pakistan.
3. Hailey College of Commerce, University of the Punjab, Lahore, Pakistan.
4. Monash Business School, Monash University, Victoria, Australia.

* Corresponding Author Email: amjadmahmood502@gmail.com

Abstract:

The theory of parameter estimation was developed a long time ago and is currently a widely accepted scientific approach. In some situations where the mean has to deal with the effect of extreme values, the median can be used instead. As there is almost little literature available on median estimation in the existence of non-responsive with the help of the Randomized Response Technique (RRT), the main objective of this paper was to develop a basic theoretical framework for median estimation in the existence of non-responsive with the help of RRT. In this work, we suggested median estimation for delicate variables in auxiliary information using a randomised response model. We have suggested a basic median estimator, product, ratio, exponential product, exponential ratio, and regression cum ratio estimation of the median for non-response by utilising RRT. The mathematical derivations for optimum values of constants, biasness, and Mean Square Error (MSE) of purposed estimators result from the application of well-known Taylor and exponential expansions. The performance of mentioned estimators is evaluated through the numerical study of two populations which discovered the regression cum ratio estimate is more proficient than the remaining estimations mentioned in this article.

Article History

Received:
19-Sep-2023

Revised:
29-Oct-2023

Re-revised:
25-Dec-2023

Accepted:
26-Dec-2023

Published:
31-Dec-2023

Keywords: RRT, Non-response, regression cum ratio, exponential expansions, optimum value, relative competence percentage, theory of parameter estimation.

How to Cite: Hussain, A., Mahmood, M. S., Mahmood, A., & Ashraf, A. (2023). Median estimation in the presence of non-response using randomised response technique. *Natural and Applied Sciences International Journal (NASIJ), 4*(2), 108-123. https://doi.org/10.47264/idea.nasij/4.2.7

Copyright: © 2023 The Author(s), published by IDEA PUBLISHERS (IDEA Publishers Group).

License: This is an Open Access manuscript published under the Creative Commons Attribution 4.0 (CC BY 4.0) International License (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The assumption of parameter estimation has been established for a long time and is now a well-recognized scientific method worldwide. The mean is commonly used to measure central tendency for quantitative data. However, if the data contains large outliers, the median may be a more appropriate measure (Mohan & Su, 2022). The median is a popular and widely used measure of central tendency in statistics and research. Mutually all data like discrete and continuous are focused by this method, even though it is primarily used with continuous data. However, it may not always be the best choice for all statistical data and variables. Despite extensive research on population, quantity, variability, and other factors, little attention has been given to accurate median estimates.

In many areas of research, researchers are often concerned with variables such as salary, expenses, tolls, spending, and manufacturing, which tend to have highly skewed distributions. To address this issue, the median is often preferred over the mean, as it is less sensitive to extreme values, making it a more appropriate measure of location in survey sampling. Numerous empirical studies have focused on the issue of median estimates in simple random sampling techniques, with researchers investigating the use of auxiliary variables. It is also worth noting that in cases where sensitive questions are asked, many people may hesitate to provide honest answers or will not respond. The bias in data occurred due to non-responsiveness.

In 1965, Warner used the Randomized Response Technique (RRT) to lower biases in facts during surveys. This practice was consumed to acquire a precise estimation of susceptible variables. Various statisticians made use of a RRT model to estimate the variables that are sensitive in nature and devoid of taking into account supplementary facts. Other researchers have considered this by suggesting explanations favouring extra data. Another empirical data used the variable of auxiliary to predict the tousled replies and projected estimates of regression type (Gupta *et al.*, 2012). Arithmetical features were the results that predicted the output of the suggested estimation.

A research study was conducted using the additive model for randomized responses. Using RRT, Zamanzade and Al-Omar (2011) proposed an exponential estimate with either single or double auxiliary variables. The investigation analysed real data and demonstrated that the recommended assessors were more productive than the existing estimators. Another study by Koyuncu *et al.* (2014) used RRT to recommend estimations of exponential type with sole and double variables to progress the competence of the estimation of mean. The study obtained the expressions of MSE and bias for both estimators up to the first order. The results showed that the recommended assessor was more productive than the established one, thus indicating a gain in efficiency.

The scholars have conducted numerous investigations to investigate the variables, but a meticulous part of the estimation of the median has not been explored well in the presence of non-responsive with the use of RRT. This study aimed to propose some estimators that would help to predict the median of a populace in the company of variables of auxiliary. The development of these estimators included a varied variety of estimations for different occurrences in various techniques of drawing the samples. These estimators will pave the way for new methods in estimating non-response using the randomized response technique.

2. Literature

When the variable under investigation is connected to the variable of auxiliary, utilizing auxiliary variables can help improve the effectiveness of estimators. The most common methods of estimation that use auxiliary information are the classical methods. Laplace (1820) and Watson (1937) were two pioneers who used auxiliary information to estimate population and leaf area, respectively. When a variable covers a sensitive or private domain that a respondent does not want to disclose, it is known as a sensitive variable.

In recent years, various techniques have been proposed to assess the means of restricted populace with the usage of variables of auxiliary. Many previous studies projected different estimations using the optional technique of RRT (Kalucha *et al.*, 2014; Gupta *et al.*, 2016). The comparisons, both theoretically and numerically, were agreed upon by Gupta *et al.* (2010) regarding the basic mean estimator. Another research utilized stratification to list assessors of a susceptible inquired variable with the use of facts about auxiliary (Mushtaq *et al.*, 2017). The proposed class of assessors is more productive than the current assessors given by Sousa *et al.* (2014) according to exact and hypothetical outcomes through simulation. Ozgul and Cingi (2017) developed a better estimator to assess the meaning of the susceptible data of the variable under investigation using RRT with the additive of partial nature in the presence of twice variables of auxiliary of non-sensitive type. The MSE of the developed estimator is derived and compared with the estimators suggested by Sousa *et al.* (2014), Diana and Perri (2011), and Gupta et al. (2016).

Khalil *et al.* (2018) proposed a widespread estimate for sensitive study variables using benchmark and ME using two different sampling techniques. They compared their proposed estimators with the mean and ratio estimators mathematically and numerically and found that the recommended estimators performed better than the competing estimator for the two sampling schemes.

In one of the arguments, in unbiased research, a comparison was made in 4 presented quantitative RRT with the help of divide and non-divide metrics about the privacy of the sample and the effectiveness of this model. It was found that under particular circumstances, some variables performed better than others for meticulous values of parameters and constants (Azeem *et al.*, 2023). However, such models become ineffective when applied with the diverse values of the parameters. The arithmetical setting for the competence of unlike models was also considered.

Grover and Kaur (2018) also proposed a productive mixed assessor for the populace by utilizing the transformation of non-delicate outcomes. They conducted productivity examinations with the leaving assessors, both mathematically and hypothetically, and found that the mentioned estimation did better than the established assessors under specific circumstances. Khalil *et al.* (2019) improved the generalized estimator proposed by using the optional technique for calculating sensitive study variables. The results obtained showed a better efficiency of the estimator.

Zhang *et al.* (2019) presented a geometric mean proportion assessor for the limited population mean utilizing discretionary RRT models. The outcomes showed that the mean of the mathematics assessor is more effective than the standard mean RRT mean assessment in

circumstances where the linkage with coefficient among the variables of review and the assistant is higher than 0.5. This empirical research also demonstrated that the expertise of the projected mean of geometric proportion estimation and the newly introduced estimator of proportion are similar but with lesser biases.

Saleem *et al.* (2019) utilized double sampling types to estimate the variable's mean with sensitive features. They also suggested the arithmetical expressions of biases and MSE and conducted a numerical study to conclude the capability of the predictable estimator. Lastly, another research in 2022, discovered the innovative estimation to examine the t (a, b) variable that is globally practiced for the data that is auxiliary for SRWOR (Adichwal *et al.*, 2022).

In the contemporary world, a number of investigations have introduced dissimilar methods that review the researched variables of receptive type. For instance, Sanaullah *et al.* (2019) proposed the use of non-sensitive concomitant variables to develop more efficient estimators, which have rejected the estimates given by previous studies like Sousa *et al.* (2010), Gupta *et al.* (2012), and Koyuncu *et al.* (2014). Similarly, Shahzad *et al.* (2019) recommended an estimation of the regression type for the mean populace with the inquired variables of a sensitive nature and data of auxiliary under ranked set sampling. They showed that their proposed estimators performed better than other difficult assessors.

Waseem *et al.* (2020) used one or double variables of auxiliary type to generalize the exponential-type estimation for scrambled responses in SRS design. They also discussed special cases of the suggested estimator with privacy protection. Numerical results were obtained through simulation and real data illustration to support the competence of the recommended estimates. Anwar (2020) has given classical and modified forms of estimators, a class of estimates of exponential and comprehensive estimations for mean assessment of investigated variables of a susceptible nature by consuming a systematic sampling plan. They did a broad study with numerical by focusing on actual and fabricated samples.

Besides that, another study suggested a pack of mean estimators of in-type to predict variables for study with data of a sensitive nature with disparity of the variables of auxiliary with sampling methods both simple and stratified (Yousaf, 2020). They found that the projected types of estimation outnumbered present-day estimators for double-kind populations. To conclude, Anwar (2020) recommended employing generalized estimations for susceptible variables by applying data of non-sensitive features in diverse techniques of sampling. It derived the terms of biasness and MSE for suggested estimation and completed a comparative investigation with conventional estimates. The observational evaluation showed that the widespread method is much better than the previous estimations, with diverse techniques used to draw samples from the population.

Rana (2021) proposed the generalized estimators for confidential study variables using auxiliary attributes in double sampling phases of simple and stratified. They also made comparisons and derivations of the expressions of biases and MSE for the projected estimate with mean, ratio, and exponential estimators. The outcomes of the numerical answers revealed that the estimation of generalized estimations behaved better than the established formula for sampling designs with a double stage of simple and stratified sampling. Tiwari *et al.* (2022) checked the influence of non-response on estimations of mean by applying RRT. Another investigation was replicated to check the efficiency of the proposed estimates.

Generally, survey sampling deals with the development of literature and methods relating to the estimation of means and totals. However, with the advancement in survey sampling design, many authors have worked on median estimation depending on the nature of the variable. Such as, when we have ordinal variables or ordered observations, then finding the appropriate central tendency median is proposed. Also, the use of existing information has been an orthodox method since the birth of sampling used for surveys. Helping statistics is commonly utilized to get the modified methods and high precision for estimating unknown parameters.

Numerous investigations have been carried out to inquire about the variables, but little research has been done on a meticulous part of the estimation of the median in the presence of non-responsive with the use of RRT. To cover up this vacuum in research, this research was done to develop a basic theoretical framework for median estimation in the presence of nonresponsive using a randomized response technique. Some suggestions about median estimation for delicate variables in the existence of data that is of auxiliary in nature by using a model of RRT were also specified. The basic median estimator, product, ratio, exponential product, exponential ratio, and regression cum ratio estimators of the median for no response using RRT were also introduced.

3. Methods

In this study, the estimators for sensitive study variables were proposed while taking into account auxiliary information for estimating the population median using RRT model. Warner was the pioneer in initiating this model in 1965, and diverse statisticians have since modified it. This technique aims to obtain the most accurate answers to susceptible queries that respondents may be cautious to respond truly. In this paper, we use the additive model introduced in 1976 by Pollock and Beck for the quantitative replies, which is the most commonly used model.

$$Z = Y + S,$$

Where Z is the reported response, Y is the variable of attention and S is the variable that is scrambled. Further, it is implicated as S~N (0, 10% of Sx)

The mean and variance of sensitive study variables is:
``

$$E(Z) = E(Y)$$

$$S_z^2 = S_y^2 + S_s^2$$

3.1. Terminology

Here, the investigated variable that is sensitive is specified as Y. Allow X to be an auxiliary variable of non-sensitive that correlates to Y. Suppose the tangled variable is S and is unrelated to Y and X. To respond, the respondent was given the scrambled formula Z=Y+S. but is required to give an accurate answer for X. Let us say that for the i^{th} component (i=1, 2, 3, 4, ..., N), a sample that is of random nature with n size taken without substitution from a restricted populace U= (U1, U2, U3, ..., UN). The yi and xi represent the corresponding variables' sample data. Let M_y and M_x now represent, respectively, the variable under research and the variable

of auxiliary population medians. Also take into account the sample medians M_y, M_x, and M_z. Assume that the X median is accessible, that S belongs to a normal distribution with a zero mean, and that its normal error is 10% of the X median's standard error. Let the error terms be denoted as:

$$e_o = \frac{\hat{M}_z - M_z}{M_z} \qquad e_1 = \frac{\hat{M}_x - M_x}{M_x} \qquad \mathrm{e}_2 = \frac{\hat{M}_{s_x} - M_{s_x}}{M_{s_x}}$$

Using these notations, we can define expectations like

$$E(e_i) = 0 \qquad i=0,1,2$$

$$E(e_o^2) = \gamma C_{M_z}^2 \qquad E(e_x^2) = \gamma C_{M_x}^2 \qquad E(e_o e_1) = \gamma \rho_{M_x M_z} C_{M_x} C_{M_z}$$

$$\text{where} \quad \gamma = \frac{(1-f)}{4n} \quad \text{and} \quad f = \frac{n}{N}$$

Also, $C_y = \left[M_y f_y(M_y)\right]^{-1} C_x = [M_x f_x(M_x)]^{-1}$

$$\rho_c = 4(p11 - 1) \quad with \qquad p11 = P(M_x M_y)$$

$$E(Y) = M_y \qquad E(Z) = M_z \qquad \text{and} \qquad E(S) = 0 \qquad \text{so} \qquad E(Z) = E(Y)$$

Hence $M_y = M_z$ and $C_{M_z}^2 = C_{M_y}^2 + \frac{S_{M_s}^2}{M_y^2}$

When the information of X is ignored, the basic median estimator for scrambled response is an unbiased estimator given by

$$t_1 = \hat{M}_z \tag{1}$$

The variance of t_1 is specified by:

$$\mathrm{var}(t_1) = M_y^2 \gamma C_{M_z}^2 \tag{2}$$

Where

$$\gamma = \frac{(1-f)}{4n} \quad \text{and} \quad f = n/N \tag{3}$$

3.1.1. Proposed estimators

a. Product estimator

The product estimator of the median for scrambled response is defined as

$$t_p = \hat{M}_Z \frac{\hat{M}_x}{M_x} \tag{4}$$

The biasness and MSE of t_p up to first-order guess is shown as

$$Bias\left(t_p\right) = \gamma M_y \rho_{M_x M_z} C_{M_x} C_{M_z} . \tag{5}$$

And

$$\mathrm{MSE}\left(t_p\right) = \gamma M_y^2 \left(C_{M_z}^2 + C_{M_x}^2 + 2\rho_{M_x M_z} C_{M_x} C_{M_z}\right). \tag{6}$$

b. The ratio estimator

The ratio median estimator for replies of a scrambled nature is illustrated as

$$t_r = \hat{M}_Z \frac{M_x}{\hat{M}_x} \tag{7}$$

The bias and MSE of t_r are derived below:

$$Bias\left(t_p\right) = \gamma M_y \left(C_{M_x}^2 - \rho_{M_x M_z} C_{M_x} C_{M_z}\right) \tag{8}$$

And

$$\mathrm{MSE}\left(t_r\right) = \gamma M_y^2 \left(C_{M_z}^2 + C_{M_x}^2 - 2\rho_{M_x M_z} C_{M_x} C_{M_z}\right). \tag{9}$$

c. The exponential product median estimator

The exponential product median estimation for the scrambled response is given by Equation 10.

$$t_{ep} = \hat{M}_Z \exp\left(\frac{\hat{M}_x - M_x}{\hat{M}_x + M_x}\right) \tag{10}$$

After important mathematical computations, the bias and MSE of t_{ep} are derived as below.

$$Bias(t_{ep}) = M_y^2 \gamma \left(\frac{1}{2} \rho_{M_x M_z} C_{M_x} C_{M_z}\right) \tag{11}$$

$$MSE\left(t_{ep}\right) = M_y^2 \gamma \left(C_{M_z}^2 + \frac{1}{4} C_{M_x}^2 + \rho_{M_x M_z} C_{M_x} C_{M_z}\right). \tag{12}$$

d. The exponential ratio median estimator

The exponential ratio median estimate for the variable of susceptible in nature is specified as

$$t_{er} = \hat{M}_Z \exp\left(\frac{M_x - \hat{M}_x}{M_x + \hat{M}_x}\right) \tag{13}$$

Using the Taylor series and the first order of estimation of the biasness and MSE are results under.

$$Bias(t_{er}) = -M_y^2 \gamma\left(\frac{1}{2}\rho_{M_x M_z} C_{M_x} C_{M_z}\right) \tag{14}$$

And

$$MSE\left(t_{er}\right) = M_y^2 \gamma\left(C_{M_z}^2 + \frac{1}{4}C_{M_x}^2 - \rho_{M_x M_z} C_{M_x} C_{M_z}\right) \tag{15}$$

e. Generalized regression-cum-ratio median estimator

Motivated by Gupta et al. (2012) who gave this estimate and proposed regression-cum-ratio generalized median estimator as given in Equation (16)

$$t_{GRR} = \left[K_1 \hat{M}_z + K_2\left(M_x - \hat{M}_x\right)\right]\left(\frac{M_x}{\hat{M}_x}\right) \tag{16}$$

3.1.2. Some specific cases

Using the Taylor series up to the 2nd order of estimation, it is given as

$$t_{GRR} - M_y = \left[M_y\left(K_1 - 1\right) + K_1 M_y\left(e_o - e_1 + e_1^2 - e_o e_1\right) + K_2 M_x\left(e_1^2 - e_1\right)\right] \tag{17}$$

For Bias, we have

$$Bias\left(t_{GRR}\right) = \left[M_y\left(K_1 - 1\right) + K_1 M_y \gamma\left(C_{M_x}^2 - \rho_{M_x M_z} C_{M_x} C_{M_z}\right) + K_2 M_x \gamma C_{M_x}^2\right] \tag{18}$$

Now to get expressions for MSE, consider Equation 9.2 and squaring and applying expectation, we get.

$$\left(t_{GRR} - M_y\right)^2 = \begin{bmatrix} M_y^2\left(K_1 - 1\right)^2 + K_1^2 M_y^2\left(e_o - e_1 + e_1^2 - e_o e_1\right)^2 + K_2^2 M_x^2\left(e_1^2 - e_1\right)^2 + \\ 2M_y K_1\left(K_1 - 1\right)\left(e_o - e_1 + e_1^2 - e_o e_1\right) - 2M_x M_y K_2\left(K_1 - 1\right)\left(e_1^2 - e_1\right) - \\ 2M_x M_y K_1 K_2\left(e_o - e_1 + e_1^2 - e_o e_1\right)\left(e_1^2 - e_1\right) \end{bmatrix},$$

After applying expectations, we have

$$\left(t_{GRR} - M_y \right)^2 = \begin{bmatrix} M_y^2 \left(K_1 - 1 \right)^2 + K_1^2 M_y^2 \gamma \left(C_{M_z}^2 + C_{M_x}^2 - 2\rho_{M_x M_z} C_{M_x} C_{M_z} \right) + \\ K_2^2 M_x^2 \gamma C_{M_x}^2 + 2M_y K_1^2 \gamma \left(C_{M_x}^2 - \rho_{M_x M_z} C_{M_x} C_{M_z} \right) - \\ 2M_y K_1 \gamma \left(C_{M_x}^2 - \rho_{M_x M_z} C_{M_x} C_{M_z} \right) - 2M_x M_y K_2 \left(K_1 - 1 \right) \gamma \left(-C_{M_x}^2 \right) - \\ 2M_x M_y K_1 K_2 \gamma \left(\rho_{M_x M_z} C_{M_x} C_{M_z} - C_{M_x}^2 \right) \end{bmatrix},$$

Or

$$MSE\left(t_{GRR} \right) = \begin{bmatrix} M_y^2 \left(K_1 - 1 \right)^2 + K_1^2 M_y^2 \gamma \left(C_{M_z}^2 + 3C_{M_x}^2 - 4\rho_{M_x M_z} C_{M_x} C_{M_z} \right) + \\ K_2^2 M_x^2 \gamma C_{M_x}^2 - 2M_y K_1 \gamma \left(C_{M_x}^2 - \rho_{M_x M_z} C_{M_x} C_{M_z} \right) - \\ 2M_x M_y K_2 \gamma C_{M_x}^2 - 2M_x M_y K_1 K_2 \gamma \left(\rho_{M_x M_z} C_{M_x} C_{M_z} - 2C_{M_x}^2 \right) \end{bmatrix} \tag{19}$$

Now, to obtain the minimum mean square error and the optimum values of K_1 and K_2, the researcher distinguished the Equation 19 with the values of K_1 and K_2 and have set it equal to nil.

$$\begin{aligned} \frac{\partial MSE\left(t_{GRR} \right)}{\partial \mathrm{K}_1} = {} & 2\left(K_1 - 1 \right) M_y^2 + 2K_1 M_y^2 \gamma \left(C_{M_z}^2 + 3C_{M_x}^2 - 4\rho_{M_x M_z} C_{M_x} C_{M_z} \right) - \\ & 2M_y \gamma \left(C_{M_x}^2 - \rho_{M_x M_z} C_{M_x} C_{M_z} \right) - 2M_x M_y K_2 \gamma \left(\rho_{M_x M_z} C_{M_x} C_{M_z} - 2C_{M_x}^2 \right) \end{aligned} \tag{20}$$

$$\frac{\partial MSE\left(t_{GRR} \right)}{\partial \mathrm{K}_2} = 2K_2 M_x^2 \gamma C_{M_x}^2 - 2M_x M_y \gamma C_{M_x}^2 - 2M_x M_y K_1 \gamma \left(\rho_{M_x M_z} C_{M_x} C_{M_z} - 2C_{M_x}^2 \right) \tag{21}$$

After solving the above equations, we obtained the following optimum values.

$$K_1 = \frac{1 - \gamma C_{M_x}^2}{\left[1 - \gamma \left\{ C_{M_x}^2 + C_{M_z}^2 \left(1 - \rho_{M_x M_z}^2 \right) \right\} \right]}$$

And

$$K_2 = \frac{M_y}{M_x} \left[1 + K_{1(opt)} \left(\frac{\rho_{M_x M_z} C_{M_z}}{C_{M_x}} - 2 \right) \right]$$

After putting the value of K_1 and K_2 in Equation 19 and doing the necessary mathematical computations we have our minimum MSE as

$$MSE\left(t_{GRR} \right)_{\min} = \frac{M_y^2 C_{M_z}^2 \left(1 - \rho_{M_x M_z}^2 \right) \gamma \left\{ 1 - \gamma C_{M_x}^2 \right\}}{C_{M_z}^2 \left(1 - \rho_{M_x M_z}^2 \right) \gamma + \left\{ 1 - \gamma C_{M_x}^2 \right\}} \tag{22}$$

4. Numerical study

Here, we have conducted a numerical experiment to check the validity of our derived formulas. We consider two populations, the descriptions of which are as follows. The formula Z = Y + S provides the reported response.

Population I. (Source: Singh & Tailor, 2003)

y: The number of fishes caught in 1995

x: The number of fishes caught in 1964

Different parameters for this population are:

N= 69, M_y= 2068, M_x= 2011, $f_y(M_y)$ = 0.00014, $f_x(M_x)$ = 0.00014 $\rho_c = 0.1505$, R = 0.97244

Population II. (Source: Singh & Tailor, 2003)

y: The number of fishes caught in 1995

x: The number of fish caught in 1964

Different parameters for this population are:

N= 69, M_y= 2068, M_x= 2307, $f_y(M_y)$ = 0.00014, $f_x(M_x)$ = 0.0013, $\rho_c = 0.3166$, R = 1.11557

For both populations, the sizes of sample are n= 5, 10, 20 and 30 percent of 69. R Language is used to process the data. Table-1 gives the MSEs for both populations of seven estimators which are derived. The relative effectiveness using percentage PRE is utilized to compare the estimators.

The formula of PRE is given below:

$$\text{PRE} = \frac{MSE\left(\hat{M}\right)}{MSE\left(\hat{M}_{\beta}\right)} \times 100$$

Where β= P, R, EP, ER, GRR, exp

4.1. Results

The findings from both the simulation study and the real data illustrations are in agreement with each other, as expected. The results presented in Table No. 1 indicate that the ratio estimation is a more competent estimator than the basic median estimate, particularly in the presence of non-response. Furthermore, the exponential ratio median estimator is found to contain a lessen MSE value than the ratio median estimate. The regression-associated ratio

median equations exhibit greater efficiency than the other estimators. However, the product median and exponential product median estimators do not perform well in estimating the parameters.

Table-1: MSE and PRE of all the estimations for real people

	n	Population I	PRE	Population II	PRE
M	17	393.0439	100	346.7305	100
M_p		506.0439	77.66992	393.1067	88.20264
M_r		346.7305	113.3572	319.7359	108.4428
M_{ep}		429.4556	91.52143	385.8387	89.86411
M_{er}		315.6209	124.5304	312.4678	110.9652
M_{GRR}		289.0325	135.9861	270.9405	127.9729
M_{epxI}		**287.4678**	**136.7262**	**269.7895**	**128.5189**
M	25	272.5342	100	240.4208	100
M_p		351.4665	77.54201	272.5778	88.20263
M_r		240.4208	113.3572	221.7028	108.4428
M_{ep}		297.7818	91.52144	267.5381	89.86414
M_{er}		218.8495	124.5304	216.6632	110.9652
M_{GRR}		210.2358	129.6326	195.7964	122.7912
M_{epxI}		**209.1443**	**130.3092**	**194.6975**	25+F1
M	30	228.6655	100	201.7213	100
M_p		294.8924	77.54201	228.7021	88.20264
M_r		201.7213	113.3571	186.0163	108.4428
M_{ep}		249.8492	91.52141	224.4737	89.86411
M_{er}		183.6223	124.5304	181.7879	110.9652
M_{GRR}		179.5292	127.3695	166.8378	120.9086
M_{epxI}		**178.5292**	**128.083**	**165.7607**	**121.6943**

5. Discussion

Researchers are increasingly examining sensitive issues such as drug and alcohol use, reproductive and mental health, and illegal activities. Participants often request confidentiality, anonymity, and protection when sharing such information, suspecting that data collectors may be intruding into their personal lives. The model of RRT is a legitimate process that offers greater anonymity and privacy for obtaining sensitive information.

However, the RR technique has both advantages and disadvantages. The estimation of variance is enlarged if randomly selected, and the reply to patterns of the sample under investigation individual cannot be deduced unswervingly. Nonetheless, combining RR and direct answers in the sample makes information more reliable and convincing.

Real data and simulation studies show that an estimator with greater efficiency value is better equipped to handle non-response in certain situations. The regression-cum-ratio median

estimator has been found to perform better than all other estimators mentioned in the research. This is supported by Mutembei *et al.* (2014), who proposed a regression-cum-ratio estimator that uses data from multiple auxiliary variables and characteristics simultaneously to estimate the population mean (Koyuncu *et al.*, 2019).

6. Conclusion

To conclude, it can be observed that there are issues in actual existence where we need to address sensitive variables, and respondents usually refuse to answer such types of variables or give fabricated answers to the questions. The results gathered by using direct approach on delicate questions can lead the researcher towards nonresponse or biasness. The issue of non-response and biasness may be tackled by using non-sensitive variables. In the past, mean estimation was considered the best approach to dealing with such cases. Now, with the development of a theoretical framework of median estimation in this paper, researchers have significantly new methods to deal with the data with many outliers or extreme values. When we talk about our proposed estimators, it can be seen that both the estimation of ratio and exponential ratio work in a better way than the basic median estimator for scrambled response derived in this paper, but the outcomes of the estimate of regression-cum-ratio are much greater than the remaining. From the numerical findings, it is concluded that regression-cum-ratio estimator performs outstandingly on different sample sizes for the assessment of scrambled response under certain conditions. So, with the help of numerical results, to conclude, it is justified to say that the estimations of regression-cum-ratio is superior then all remaining estimators mentioned in this research with different sample sizes. Also, the efficiency of purposed estimators will increase with the big size of a sample. In the future, researchers can estimate the median in the occurrence of non-responsive variables by applying multiple models of randomized response approach and double variables of scrambled nature.

Declaration of conflict of interest

The author(s) declared no potential conflicts of interest(s) with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Attique Hussain https://orcid.org/0009-0004-2183-7739

M. Shahid Mahmood https://orcid.org/0009-0003-6577-9991

Amjad Mahmood https://orcid.org/0009-0006-6396-2305

Ayesha Ashraf

Publisher's Note

IDEA PUBLISHERS (IDEA Publishers Group) stands neutral with regard to jurisdictional claims in the published maps and institutional affiliations.

References

Adichwal, N. K., Ahmadini, A. A. H., Raghav, Y. S., Singh, R., & Ali, I. (2022). Estimation of General parameters using auxiliary information in simple random sampling without replacement. *Journal of King Saud University-Science, 34*, 101754. https://doi.org/10.1016/j.jksus.2021.101754

Anwar, M. (2020). *Generalized Estimators for sensitive study variable using non sensitive auxiliary attribute in simple and stratified random sampling*. Thesis submitted to National College of Business Administration & Economics, Lahore. https://prr.hec.gov.pk/jspui/bitstream/123456789/16893/1/Muhammad%20Anwar%20Mughal%20statistics%202020%20ncbae%20lhr.pdf

Azeem, M., Shabbir, J., Salahuddin, N., Hussain, S., & Ijaz, M. (2023). A comparative study of randomized response techniques using separate and combined metrics of efficiency and privacy. *PLOS One*, *18*(10), e0293628. https://doi.org/10.1371/journal.pone.0293628

Diana, G. & Perri, P. F. (2011). A class of estimators for quantitative sensitive data. *Stat Papers, 52*, 633–650. https://doi.org/10.1007/s00362-009-0273-1

Grover, L. K. & Kaur, A. (2019). An efficient scrambled estimator of population mean of quantitative sensitive variable using general linear transformation of non-sensitive auxiliary variable. *Communications in Mathematics and Statistics*, 7(4),401–415.

Gupta, S., Shabbir, J. & Sehra, S. (2010). Mean and sensitivity estimation in optional randomized response models. *Journal of Statistical Planning and Inference, 140*(10), 2870–2874.

Gupta, S., Shabbir, J., Sousa, R., & Corte-Real, P. (2016). Improved exponential type estimators of the mean of a sensitive variable in the presence of non-sensitive auxiliary information. *Communications in Statistics - Simulation and Computation*, *45*(9), 3317–3328. https://doi.org/10.1080/03610918.2014.941487

Gupta, S., Shabbir, J., Sousa, R., and Real, P. C., (2012). Estimation of the mean of a sensitive variable in the presence of auxiliary information. *Communications in Statistics-Theory and Methods, 41*, 1–12.

Khalil, S., Noor-ul-Amin, M. & Hanif, M. (2018). Estimation of population mean for a sensitive variable in the presence of measurement error, *Journal of Statistics and Management Systems, 21*(1), 81–91.

Khalil, S., Zhang, Q., & Gupta, S. (2019). Mean estimation of sensitive variables under measurement errors using optional RRT models. *Communications in Statistics-Simulation and Computation*, 1–10. https://doi.org/10.1080/03610918.2019.1584298

Koyuncu, N., Gupta, S., & Sousa, R. (2014). Exponential type estimators of the mean of a sensitive variable in the presence of non-sensitive auxiliary information. *Communications in Statistics–Simulation and Computation*, 43(7), 1583–94.

Koyuncu, N., Saleem, I., Sanaullah, A., & Hanif, M. (2019). Estimation of mean of a sensitive quantitative variable in complex survey: improved estimator and scrambled randomized response model. *Gazi University Journal of Science, 32*(3), 1021–1043.

Laplace, P. S. (1820). *A philosophical essay on probabilities*. English translation, Dover, 1951. https://bayes.wustl.edu/Manual/laplace_A_philosophical_essay_on_probabilities.pdf

Mohan, S., & Su, M. K. (2022). Biostatistics and Epidemiology for the Toxicologist: Measures of Central Tendency and Variability-Where Is the "Middle?" and What Is the "Spread?". *Journal of Medical Toxicology: official journal of the American College of Medical Toxicology*, *18*(3), 235–238. https://doi.org/10.1007/s13181-022-00901-7

Mushtaq, N., Amin, N. U., & Hanif, M. (2017). A Family of Estimators of a Sensitive Variable Using Auxiliary Information in Stratified Random Sampling. *Pakistan Journal of Statistics and Operation Research*, *13*(1), 141–155. https://doi.org/10.18187/pjsor.v13i1.1532

Mutembei, T., Kung'u, J., & Ouma, C. (2014) Mixture Regression-Cum-Ratio Estimator Using Multi-Auxiliary Variables and Attributes in Single-Phase Sampling. *Open Journal of Statistics*, *4*, 367–376. https://doi.org/10.4236/ojs.2014.45036

Ozgul, N. İ. L. G. Ü. N., & ÇINGI, H. (2017). A new estimator based on auxiliary information through quantitative randomized response techniques. *Journal of Modern Applied Statistical Methods, 16*(1), 364-387.

Pollock, K. H. & Bek, Y. (1976). A comparison of three randomized response models for quantitative data. *Journal of the American Statistical Association, 71*(356), 884–886. https://doi.org/10.1080/01621459.1976.10480963

Rana, Q. (2021). Generalized estimators for sensitive study variable using auxiliary attribute in two-phase and stratified sampling. *Thesis submitted to National College of Business Administration & Economics, Lahore*.

Saleem, I., Sanaullah, A. & Hanif, M. (2019). Double-sampling regression-cum-exponential estimator of the mean of a sensitive variable. *Mathematical Population Studies*, *26*(3), 163–182.

Sanaullah, A., Ayaz, A., & Hanif, M. (2019). A multivariate exponential estimator for vector of population means in two-phase sampling. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, 1–13.

Shahzad, U., Hanif, M., Koyuncu, N., & Garcia Luengo, A.V. (2018). A new estimator of mean under ranked set sampling alongside the sensitivity issue. *Journal of Statistics and Management Systems, 21*(8), 1553–1564.

Singh, H. P., & Tailor, R. (2003). Use of Known Correlation Coefficient in Estimating the Finite Population Mean. *Statistics in Translation, 6*, 553–560. https://www.scirp.org/reference/referencespapers?referenceid=1212659

Sousa, R., Shabbir, J., Real, P. C., & Gupta, S. (2010). Ratio estimation of the mean of a sensitive variable in the presence of auxiliary information. *Journal of Statistical Theory and Practice*, *4*(3), 495–507.

Sousa, R., Gupta, S., Shabbir, J., & Corte-Real, P. (2014). Improved mean estimation of a sensitive variable using auxiliary information in stratified sampling. *Journal of Statistics and Management Systems*, *17*(5-6), 503–518. https://doi.org/10.1080/09720510.2013.878826

Tiwari, K. K., Bhougal, S., Kumar, S., & Rather, K. U. I. (2022). Using randomized response to estimate the population mean of a sensitive variable under the influence of measurement error. *Journal of Statistical Theory and Practice*, *16*(2), 28. https://link.springer.com/article/10.1007/s42519-022-00251-1

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

Waseem, Z., Khan, H., & Shabbir, J. (2020). Generalized exponential type estimator for the mean of sensitive variable in the presence of non-sensitive auxiliary variable. *Communications in Statistics-Theory and Methods*, 1–12.

Watson, D. J. (1937). The estimation of leaf areas. *Journal of Agriculture Sciences, 27,* 474. http://dx.doi.org/10.1017/S002185960005173X

Yousaf, M. (2020). *Class of LN type estimators for sensitive study variable in survey sampling*. Thesis submitted to National College of Business Administration and Economics, Lahore.

Zamanzade, E., & Al-Omar, A. I. (2011). New ranked set sampling for estimating the population mean and variance. https://doi.org/10.15672/HJMS.2015921316

Zhang, Q., Gupta, S., Kalucha, G., & Khalil, S. (2019). Ratio estimation of the mean under RRT models. *Journal of Statistics and Management Systems, 22*(1), 97–113. https://doi.org/10.1080/09720510.2018.1533513