ISSN: 2456-799X, Vol.09, No.(2) 2024, Pg. 131-140

Oriental Journal of Physical Sciences

www.orientjphysicalsciences.org

Some Log Type Estimators for Estimation of Population Mean of Sensitive Variable using Additive Scrambling Model

MOJEED ABIODUN YUNUSA[1] and AWWAL ADEJUMOBI[2*]

[1]Department of Statistics, Usmanu Danfodiyo University, Sokoto, Nigeria.
[2]Department of Mathematics, Kebbi State University of Science and Technology, Aliero, Nigeria.

Abstract

Since it can be challenging to estimate the mean of a sensitive study changeable when direct methods of information collection are used to elicit sensitive information, through the use of the randomized response technique (RRT), protect respondents' privacy while also obtaining more valid and reliable information. The present study introduces a set of log-type estimators for calculating averages through the utilization of the additive scrambling model. The estimators' mean square error (MSE) is determined a maximum of two degrees. There are defined circumstances in which the estimators perform better. An investigation was carried out utilizing three datasets, and the findings indicated that the suggested estimators outperform estimators found in the previous research both in efficiency comparison and empirical investigations. The first table shows the Mean Square Errors (MSEs) of the proposed and existing estimators used in this research, and we found out that $\bar{z}_{p7}$ has the smallest MSE, followed by the $\bar{z}_{p1}, \bar{z}_{p2}, \bar{z}_{p3}, \bar{z}_{p4}$ and co. The second table show the same results where $\bar{z}_{p7}$ is having highest Percentage Relative Efficiecy (PRE), then other estimators follows. This explain that it is the most efficient estimator in this research.

Article History
Received: 12 July 2024
Accepted: 18 September 2024

Keywords
Mean Square Error, Estimator; Randomized Response; Sample; Sensitive Issue.

Introduction

Today's society is filled with many delicate topics, like cases of rape, sexual practices, earned income that is not legal, etc. It can be challenging to gather information on these topics when using the direct questioning technique, which will ultimately lead to unreliable information gathered from respondents and unreliable estimates from the estimation of parameter of the study variable that is sensitive in nature. We employ a covert technique to gather information from interview subjects while maintaining their privacy in order to produce more accurate data and a trustworthy estimate. When data on delicate subjects are tainted by response bias and non-

CONTACT Awwal Adejumobi awwaladejumobi@gmail.com Department of Mathematics, Kebbi State University of Science and Technology, Aliero, Nigeria.

Doi: DOI : http://dx.doi.org/10.13005/OJPS09.02.09

response, estimation becomes problematic. The tacit means of acquiring private data is employed to counter this.[18] was the one who first presented the technique. Subsequently,[13,18,19] introduced the quantitative version of their questionnaire to gather quantitative data on delicate topics for mean estimation. The use of additive models to get respondents to give randomized or scrambled responses was first introduced by[5,13] presented and applied the multiplicative scrambling model. Additional references are.[3,4,6]

Numerous authors, including[7,9,11] are among the researchers who have worked on sampling surveys and have estimated the population mean of the sensitive study variable in the presence of the non-sensitive auxiliary variable. In order to achieve precision, we plan to propose a few log-type estimators for the population mean estimation of sensitive study variables in the current study.

Literature Review

Let $z_1, z_2, \ldots\ldots, z_n$ be randomized replies from n respondents who were sampled from population of size N with units $z_1, z_2, \ldots\ldots, z_n$ using the[24] additive model Z=Y+S, where S is the jumbled variable, the distribution of which the researcher knows, with mean and variance as $\left(\bar{S}=N^{-1}\sum_{i=1}^{N}S_i=0, S_s^2=(N-1)^{-1}\sum_{i=1}^{N}(S_i-\bar{S})^2\right)$, the study variable's mean and variance $\left(\bar{Y}=N^{-1}\sum_{i=1}^{N}Y_i, S_y^2=(N-1)^{-1}\sum_{i=1}^{N}(Y_i-\bar{Y})^2\right)$ is unknown to the investigator but the mean and variance of the auxiliary variable $\left(\bar{X}=N^{-1}\sum_{i=1}^{N}X_i, S_x^2=(N-1)^{-1}\sum_{i=1}^{N}(X_i-\bar{X})^2\right)$ is provided by the respondents truthfully. The coefficient of variation for the scrambled response is given by $C_z=\frac{\left(Var(Y)+S_s^2\right)^{1/2}}{\bar{Y}}$ and the correlation coefficient between Z and X is given by $\rho_{zx}=\frac{\rho_{yx}}{\sqrt{1+\frac{S_s^2}{S_y^2}}}$ We

We established the following relationships for the relative error in order to derive the estimators' properties

$e_0=\bar{Z}^{-1}(\bar{z}-\bar{Z})$, $e_1=\bar{X}^{-1}(\bar{x}-\bar{X})$, $\bar{z}=\bar{Z}(1+e_0)$, $\bar{x}=\bar{X}(1+e_1)$, $E(e_i)=0, i=0,1$, $E(e_0^2)=\lambda C_z^2$, $E(e_1^2)=\lambda C_x^2$, $E(e_0e_1)=\lambda\rho_{zx}C_zC_x$.

Various Current Randomized Response Estimators in the Absence of Auxiliary Information

The first quantitative randomized response technique (RRT) model, known as the additive model, was described by[19] for the purpose of estimating the mean of the quantitatively sensitive study variable.[13] further examined estimation along this line.

When asking respondents for sensitive information,[13] thought that an additive RRT model would be the best option.

$$Z=Y+S \quad \text{...(2.1)}$$

An estimator of μ_y and its variance are given by

$$\bar{z}_{PB}=\mu_y+\mu_s \quad \text{...(2.2)}$$

$$Var(\bar{z}_{PB})=\frac{\sigma_y^2+\sigma_s^2}{n} \quad \text{...(2.3)}$$

In this case, the distribution of scramble variable S is known beforehand. An RRT model with multiplicative properties was presented by[5] to extract sensitive data from respondents as

$$Z=YS \quad \text{...(2.4)}$$

An estimator of μ_y and it's variance are given by

$$\bar{z}_{EH}=\mu_y\mu_s \quad \text{...(2.5)}$$

$$Var(\bar{z}_{EH})=\frac{1}{n}\left(\sigma_y^2+(\sigma_y^2+\sigma_y^2)\left(\frac{\sigma_s}{\mu_s}\right)^2\right) \quad \text{...(2.6)}$$

An optional RRT model with one stage multiplicative was proposed by.[8] Respondents will give a jumbled response (YS) in this model if they believe the question is sensitive; otherwise, they will directly respond to the sensitive survey question (Y) with their genuine response. In this model, they assume that both Y and S are positive valued random variables and that the mean of the scrambling variable $\mu_s=1$ and variance σ_s^2. Under this model, the reported response Z is given by

$$Z=\begin{cases} Y & \text{with probability } 1-W \\ YS & \text{with probability } W \end{cases} \quad \text{...(2.7)}$$

The mean of Z is given by

$$\hat{\mu}_z=\hat{\mu}_y=\frac{1}{n}\sum_{i=1}^{n}Z_i \quad \text{...(2.8)}$$

The variance of this unbiased estimator of the population mean μ_y is given by:

$$Var(\hat{\mu}_y)=\frac{1}{n}\left(\sigma_y^2+W\sigma_s^2\left(\sigma_y^2+\mu_y^2\right)\right) \quad \text{...(2.9)}$$

It should be noted that W increases as $Var(\hat{\mu}_y)$ increases and hence there is gain in efficiency compared to non-optional model when W=1. They also gave an estimator for the sensitivity level W as:

$$\hat{W} = \frac{\frac{1}{n}\sum_{i=1}^{n}\log(z_i) - \log\left(\frac{1}{n}\sum_{i=1}^{n} z_i\right)}{E(\log(s))},\ 0 \le \hat{W} \le 1 \qquad ...(2.10)$$

[5]Multiplicative model was modified by[3] to propose the forced quantitative RRT (FQRRT) model (1983). According to this model, each respondent is asked to use a randomization tool like a spinner to complete a Bernoulli trial. Let P represent the percentage of the delicate question that is answered. In all honesty, the researcher fixes this proportion if the spinner stops in the shaded area; if it stops in the unshaded area, (1-P) represents the proportion of reporting a scrambled response YS. The responses are distributed as follows.

$$Z_i = \begin{cases} Y_i & with\ probability\ P \\ Y_i S_i & with\ probability\ 1-P \end{cases} \qquad ...(2.11)$$

The unbiased estimator for μ_y and its variance are given by

$$\hat{\mu}_y = \frac{1}{n\left[(1-P)\mu_s + P\right]}\sum_{i=1}^{n} Z_i \qquad ...(2.12)$$

$$Var(\hat{\mu}_y) = \frac{\mu_y^2}{n}\left(C_y^2 + \left(1 + C_y^2\right)C_s^2(P)\right) \qquad ...(2.13)$$

Where, $C_s^2(P) = \frac{(1-P)\mu_s^2\left(1+C_s^2\right)}{\left[(1-P)\mu_s + P\right]^2} - 1$

An information-gathering model for Y, the sensitive variable, was presented by 15 using the randomized response technique (RRT). This model is similar to 5 model, but he considered incorporating two scrambled variables and whose distributions are assumed to be known and independent of Y in the model. According to this model, each respondent is requested to provide the randomized answer as.

$$Z = S_1\left(Y + S_2\right) \qquad ...(2.14)$$

The unbiased estimator of μ_y is given by

$$\hat{\mu}_y = \frac{\overline{z}}{\mu_{s_1}} - \mu_{s_2} \qquad ...(2.15)$$

Where $\overline{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$. The variance of the estimator of μ_y is given by

$$Var(\hat{\mu}_y) = \frac{\mu_{s_2}^2\sigma_{s_1}^2 + \mu_{s_1}^2\left(\sigma_{s_2}^2 + \sigma_y^2\right) + \mu_{s_1}^2\left(\mu_{2,y} + \sigma_{s2}^2\right) + 2\mu_y\mu_{s_2}\sigma_{s_1}^2}{n\mu_{s_1}^2} \qquad ...(2.16)$$

Where, $\mu_y,\ \mu_i\ (i=1,2), \sigma_y^2,\ \sigma_{s_i}^2\ (i=1,2), \mu_{2,y}$ are the population means and population variances of the variable of interest Y and scrambled variables and population second moment.

The optional scrambling model was presented by,[6] they identify Y as the quantitative sensitive variable, S as an additive random variable, and the observed responses from respondents are expressed as

$$Z_i = \begin{cases} Y_i + \alpha S & with\ probability\ P = \frac{\beta}{\alpha+\beta} \\ Y_i - \beta S & with\ probability\ 1-P = \frac{\alpha}{\alpha+\beta} \end{cases} \qquad ...(2.17)$$

Where, α and β denote the constants are determined by the interviewer.

An estimator for μ_y using the[6] scrambling procedure may be expressed as

$$\hat{\mu}_{GS} = \frac{1}{n}\sum_{i=1}^{n} Z_i \qquad ...(2.18)$$

The variance of $\hat{\mu}_{GS}$ can be derived as

$$Var(\hat{\mu}_{GS}) = \frac{1}{n}\left(\alpha\beta\left(\sigma_s^2 + \theta^2\right) + \sigma_y^2\right) \qquad ...(2.19)$$

$E(Y_i) = \mu_y, E(S) = \theta, Var(Y_i) = \sigma_y^2, Var(S) = \sigma_s^2$ are the population means and population variances of the variable of interest Y and random variable S.

[9]Proposed a two stage additive optional RRT model with two scrambling variables $S_i\ (i=1,2)$ and are uncorrelated with mean $\mu_{s_i} = 0$ and variance $\sigma_{s_i}^2$. Two independent subsample approach of size $n_i\ (i=1,2)$, are drawn from the population, such that $n = n_1 + n_2$. In the i[th] sample, a fixed predetermined proportion (T_0) of respondent is instructed to tell the truth and the remaining proportion $(1-T_0)$ of respondents have an option to scramble their response additively as $Y + S_i$ if they consider the sensitive question as sensitive or report truthfully if they consider the sensitive question non-sensitive. The distribution of their responses is given by

$$Z_i = \begin{cases} Y & \text{with probability } T_0 + (1-T_0)(1-W) \\ Y + S_i & \text{with probability } (1-T_0)W \end{cases} \quad ...(2.20)$$

The mean and variance of Z are

$$\mu_{z_i} = \mu_y + \mu_{s_i} W(1-T_0) \quad ...(2.21)$$

$$\sigma_{z_i}^2 = \sigma_y^2 + \sigma_{s_i}^2 W(1-T_0) + W(1-T_0)\left(1 - W(1-T_0)\right)\mu_{s_i}^2 \quad ...(2.22)$$

The unbiased estimator for μ_y and W are given as well as their variances as

$$\hat{\mu}_y = \frac{\mu_{s_2}\bar{z}_1 - \mu_{s_1}\bar{z}_2}{\mu_{s_2} - \mu_{s_1}} \quad ...(2.23)$$

$$\hat{W} = \frac{\bar{z}_2 - \bar{z}_1}{(\mu_{s2} - \mu_{s_1})(1-T_0)} \quad ...(2.24)$$

$$Var(\hat{\mu}_y) = \frac{1}{(\mu_{s_2} - \mu_{s_1})^2}\left(\mu_{s_2}^2 \frac{\sigma_{z_1}^2}{n_1} + \mu_{s_1}^2 \frac{\sigma_{z_2}^2}{n_2}\right) \quad ...(2.25)$$

$$Var(\hat{W}) = \frac{1}{(\mu_{s_1} - \mu_{s_2})^2 (1-T_0)}\left(\frac{\sigma_{z_1}^2}{n_1} + \frac{\sigma_{z_2}^2}{n_2}\right) \quad ...(2.26)$$

[9]Proposed a two stage multiplicative optional RRT model with two scrambling variables $S_i\,(i=1,2)$ and are uncorrelated with mean $\mu_{s_i} = 1$ and variance $\sigma_{s_i}^2$. Two independent subsample approach of size $n_i\,(i=1,2)$, are drawn from the population, such that $n = n_1 + n_2$. In the i[th] sample, a fixed predetermined proportion (T_0) of respondent is instructed to tell the truth and the remaining proportion $(1-T_0)$ of respondents have an option to scramble their response additively as YS_i if they consider the sensitive question as sensitive or report truthfully if they consider the sensitive question non-sensitive. The distribution of their responses is given by

$$Z_i = \begin{cases} Y & \text{with probability } T_0 + (1-T_0)(1-W) \\ YS_i & \text{with probability } (1-T_0)W \end{cases} \quad ...(2.27)$$

The mean and variance of Z are

$$\mu_{z_i} = \mu_y\left[\mu_{s_i} W(1-T_0) + \left(1 - W(1-T_0)\right)\right] \quad ...(2.28)$$

$$\sigma_{z_i}^2 = \sigma_y^2 + \sigma_{s_i}^2 W(1-T_0) + W(1-T_0)\left(1 - W(1-T_0)\right)\mu_{s_i}^2 \quad ...(2.29)$$

The unbiased estimator for μ_y and W are given as well as their variances as

$$\hat{\mu}_y = \frac{(\mu_{s_2} - 1)\bar{z}_1 - (\mu_{s_1} - 1)\bar{z}_2}{\mu_{s_2} - \mu_{s_1}} \quad ...(2.30)$$

$$\hat{W} = \frac{\bar{z}_2 - \bar{z}_1}{\left[\bar{z}_1(\mu_{s_2} - 1) - (\mu_{s_2} - 1)\right](1-T_0)} \quad ...(2.31)$$

$$Var(\hat{\mu}_y) = \frac{1}{(\mu_{s_2} - \mu_{s_1})^2}\left((\mu_{s_2} - 1)^2 \frac{\sigma_{z_1}^2}{n_1} + (\mu_{s_1} - 1)^2 \frac{\sigma_{z_2}^2}{n_2}\right) \quad ...(2.32)$$

$$Var(\hat{W}) = \frac{1}{\left[\bar{z}_1(\mu_{s_2} - 1) - (\mu_{s_2} - 1)\right]^2 (1-T_0)^2}\left(\frac{\sigma_{z_1}^2}{n_1} + \frac{\sigma_{z_2}^2}{n_2}\right) \quad ...(2.33)$$

[4]Presented a randomization strategy that combines additive and multiplicative approaches. They believe that this combination can increase respondents' confidence regarding privacy protection because their model takes into account two scrambling variables, S and T. The observed responses based on this model are provided by:

$$Z = TY + S \quad ...(2.34)$$

They also assumed that $E(T) = \mu_T = 1$ and $E(S) = \mu_S = 0,$ then the mean and variance of Z are given by

$$\mu_z = \mu_y \mu_T + \mu_s \quad ...(2.35)$$

$$\sigma_z^2 = \sigma_T^2\left(\mu_y^2 + \sigma_y^2\right) + \sigma_y^2 + \sigma_s^2 \quad ...(2.36)$$

The unbiased estimator $\hat{\mu}_y$ and its variance are given by

$$\hat{\mu}_y = (\mu_z - \mu_s)/\mu_T \quad ...(2.37)$$

$$Var(\hat{\mu}_y) = \frac{1}{n}\left(\sigma_T^2\left(\mu_y^2 + \sigma_y^2\right) + \sigma_y^2 + \sigma_s^2\right) \quad ...(2.38)$$

[11]Proposed a three stage additive optional RRT model with two scrambling variables $S_i\,(i=1,2)$ and are uncorrelated. Two independent subsample approach of size $n_i\,(i=1,2)$, are drawn from the population, such that $n = n_1 + n_2$. In the i[th] sample, a fixed predetermined proportion (T_0) of respondent is instructed to tell the truth and a fixed predetermined proportion (F_0) of respondents is instructed to scrambled additively $Y + S_i$ and the remaining proportion $(1 - T_0 - F_0)$ of respondents have an option to scramble their response additively if they consider the sensitive question as sensitive or report truthfully if they consider the sensitive question non-sensitive. The distribution of their responses is given by

$$Z_i = \begin{cases} Y & \text{with probability } T_0 + (1-T_0-F_0)(1-W) \\ Y + S_i & \text{with probability } F_0 + (1-T_0-F_0)W \end{cases} \quad ...(2.30)$$

The mean and variance of Z are

$$\mu_{z_i} = \mu_y + \left[(1-T_0-F_0)W\right]\mu_{s_i} \quad ...(2.39)$$

$$\sigma_{z_i}^2 = \sigma_y^2 + \begin{bmatrix} F_0 \\ +(1-F_0-F_0)W \end{bmatrix}\left[1-\begin{pmatrix} F_0 \\ +(1-F_0-F_0)W \end{pmatrix}\right]\mu_{s_i}^2 + \begin{bmatrix} F_0 \\ +(1-F_0-F_0)W \end{bmatrix}\sigma_{s_i}^2 \quad ...(2.40)$$

The unbiased estimator for μ_y and W are given as well as their variances as

$$\hat{\mu}_{ym} = \frac{\mu_{s_1}\bar{z}_2 - \mu_{s_2}\bar{z}_1}{\mu_{s_1} - \mu_{s_2}} \quad ...(2.41)$$

$$\hat{W}_m = \frac{\bar{z}_1 - \bar{z}_2}{\mu_{s_1} - \mu_{s_2}} \quad ...(2.42)$$

$$Var(\hat{\mu}_{ym}) = \frac{1}{(1-T_0-F_0)^2(\mu_{s_1}-\mu_{s_2})^2}\left(\mu_{s_2}^2\frac{\sigma_{z_1}^2}{n_1} + \mu_{s_1}^2\frac{\sigma_{z_2}^2}{n_2}\right) \quad ...(2.43)$$

$$Var(\hat{W}_m) = \frac{1}{(1-T_0-F_0)^2(\mu_{s_1}-\mu_{s_2})^2}\left(\frac{\sigma_{z_1}^2}{n_1} + \frac{\sigma_{z_2}^2}{n_2}\right) \quad ...(2.44)$$

When gathering data on a quantitatively sensitive variable for the purpose of estimating the population mean,[10] took into consideration the subtractive randomized response technique (RRT) model. The model asks respondents to deduct the value of their random or scrambled variable (S) from a known distribution from the true value of the sensitive response or true response. As indicated by Z, Y is the observed/scrambled response and is provided by:

$$Z = Y - S \quad ...(2.45)$$

The scrambled variable S is distributed independently of the sensitive variable Y, whose mean $\mu_s = 0$ and variance σ_s^2 are known, then the mean and variance of Z are given by

$$\mu_z = \bar{z} = \mu_y - \mu_s \quad ...(2.46)$$

$$\sigma_z^2 = \sigma_y^2 + \sigma_s^2 \quad ...(2.47)$$

If $z_1, z_2, ... z_n$ are the observed responses of sample size n then an unbiased estimator for μ_y is given by:

$$\hat{\mu}_y = \mu_z + \mu_s \quad ...(2.48)$$

The variance of $\hat{\mu}_y$ is given by

$$Var(\hat{\mu}_y) = \frac{\sigma_y^2 + \sigma_s^2}{n} \quad ...(2.49)$$

Following their analysis of the quantitative model put forth by[6,12] proposed an optional scrambling model. The observed responses can be expressed as:

$$Z_i = \begin{cases} Y_i - \beta S & \text{with probability } P_1 = \dfrac{\alpha}{\alpha+\beta+\gamma} \\ Y_i + \alpha S & \text{with probability } P_2 = \dfrac{\beta}{\alpha+\beta+\gamma} \\ Y_i & \text{with probability } P_3 = \dfrac{\gamma}{\alpha+\beta+\gamma} \end{cases} \quad ...(2.50)$$

Where, α, β and γ denote the constants and are determined by the interviewer before the survey is conducted.

An estimator for μ_y using the[12] scrambling procedure may be expressed as

$$\hat{\mu}_{NS} = \frac{1}{n}\sum_{i=1}^{n} Z_i \quad ...(2.51)$$

The variance of $\hat{\mu}_{NS}$ can be derived as:

$$Var(\hat{\mu}_{NS}) = \frac{1}{n}\left(\frac{\alpha\beta(\alpha+\beta)(\sigma_s^2+\theta^2)}{\alpha+\beta+\gamma} + \sigma_y^2\right) \quad ...(2.52)$$

$E(Y_i) = \mu_y, E(S) = \theta, Var(Y_i) = \sigma_y^2, Var(S) = \sigma_s^2$ are the population means and population variances of the variable of interest Y and random variable S.

Four randomized response models were compared by[2] in order to extract sensitive information. The models are provided by

$$Z_i = \begin{cases} Y_i + \alpha S & \text{with probability } P = \dfrac{\beta}{\alpha+\beta} \\ Y_i - \beta S & \text{with probability } 1-P = \dfrac{\alpha}{\alpha+\beta} \end{cases} \quad ...(2.53)$$

$$Z = TY + S \quad ...(2.54)$$

$$Z_i = \begin{cases} Y_i - \beta S & \text{with probability } P_1 = \dfrac{\alpha}{\alpha+\beta+\gamma} \\ Y_i + \alpha S & \text{with probability } P_2 = \dfrac{\beta}{\alpha+\beta+\gamma} \\ Y_i & \text{with probability } P_3 = \dfrac{\gamma}{\alpha+\beta+\gamma} \end{cases} \quad ...(2.55)$$

$$Z = \begin{cases} Y & \text{with probability } \dfrac{\gamma}{\alpha+\beta+\gamma} \\ Y + S & \text{with probability } \dfrac{\beta}{\alpha+\beta+\gamma} \\ TY + S & \text{with probability } \dfrac{\alpha}{\alpha+\beta+\gamma} \end{cases} \quad ...(2.56)$$

An expression for the unbiased estimator is

$$\hat{\mu}_{A1} = \frac{1}{n}\sum_{i=1}^{n} Z_i \quad ...(2.57)$$

$$\hat{\mu}_{A2} = \frac{1}{n}\sum_{i=1}^{n} Z_i \quad ...(2.58)$$

$$\hat{\mu}_{A3} = \frac{1}{n}\sum_{i=1}^{n} Z_i \quad ...(2.59)$$

$$\hat{\mu}_{A4} = \frac{1}{n}\sum_{i=1}^{n} Z_i \quad ...(2.60)$$

The variance expressions at $E(S)=\theta=0$ for all the models are reformulated as:

$$Var(\hat{\mu}_{A1}) = \frac{1}{n}\left(\alpha\beta\sigma_s^2 + \sigma_y^2\right) \quad ...(2.61)$$

$$Var(\hat{\mu}_{A2}) = \frac{1}{n}\left(\sigma_s^2\left(\mu_y^2 + \sigma_y^2\right) + \sigma_y^2 + \sigma_T^2\right) \quad ...(2.62)$$

$$Var(\hat{\mu}_{A3}) = \frac{1}{n}\left(\frac{\alpha\beta(\alpha+\beta)\sigma_s^2}{\alpha+\beta+\gamma} + \sigma_y^2\right) \quad ...(2.63)$$

$$Var(\hat{\mu}_{A4}) = \frac{1}{n}\left(\frac{\alpha+\beta}{\alpha+\beta+\gamma}\sigma_s^2 + \sigma_y^2 + \frac{\alpha}{\alpha+\beta+\gamma}\sigma_T^2\left(\sigma_y^2 + \mu_y^2\right)\right) \quad ...(2.64)$$

Comparing the likelihoods found in[12] with those found in.[8] For the study variable that is sensitive, the mean estimator is provided by

$$\bar{z} = n^{-1}\sum_{i=1}^{n} z_i \quad ...(2.65)$$

The Variance of $\bar{z}$ is given by

$$Var(\bar{z}) = \lambda \bar{Z}^2 C_z^2 \quad ...(2.66)$$

Some Existing Randomized Response Estimators with Auxiliary Information

In order to obtain information about Y indirectly,[14] employed the additive model and presented the ratio method of estimation as

$$\bar{z}_s = \bar{z}\left(\frac{\bar{X}}{\bar{x}}\right) \quad ...(2.67)$$

The mean square error of $\bar{z}_s$ is given by

$$Var(\bar{z}_s) = \lambda \bar{Z}^2\left(C_z^2 + C_x^2 - 2\rho_{zx}C_z C_x\right) \quad ...(2.68)$$

In order to obtain information about Y indirectly,[16] employed the additive model and presented the exponential ratio estimation method as

$$\bar{z}_{sh} = \bar{z}\exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right) \quad ...(2.69)$$

The mean square error of $\bar{z}_{sh}$ is given by

$$Var(\bar{z}_{sh}) = \lambda \bar{Z}^2\left(C_z^2 + 0.25C_x^2 - \rho_{zx}C_z C_x\right) \quad ...(2.70)$$

Materials and Methods

Proposed Estimator for Sensitive Study Variable

In this segment, we modified and suggested the estimator created by[1] when sensitive study variables were present, along with alternative estimator classes as.

$$\bar{z}_{p0} = \frac{\left[\bar{z} - Ln\left(\frac{\bar{x}}{\bar{X}}\right)\right]\bar{X}}{\bar{x}} \quad ...(3.1)$$

$$\bar{z}_{p1} = 2^{-1}\left[u_1\frac{\bar{z}}{2}\left(\frac{\bar{X}}{\bar{x}} + \frac{\bar{x}}{\bar{X}}\right) - v_1\bar{X}Ln\left(\frac{\bar{x}}{\bar{X}}\right)\right]\left[1 + \exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right)\right] \quad ...(3.2)$$

$$\bar{z}_{p2} = 2^{-1}\left[u_2\frac{\bar{z}}{2}\left(\frac{\bar{X}}{\bar{x}} + \frac{\bar{x}}{\bar{X}}\right) - v_2\bar{X}Ln\left(\frac{\bar{x}}{\bar{X}}\right)\right]\left[1 + \left(\frac{\bar{X}}{\bar{x}}\right)\exp\left(\frac{\bar{x}-\bar{X}}{\bar{x}+\bar{X}}\right)\right] \quad ...(3.3)$$

$$\bar{z}_{p3} = 2^{-1}\left[u_3\frac{\bar{z}}{2}\left(\frac{\bar{X}}{\bar{x}} + \frac{\bar{x}}{\bar{X}}\right) - v_3\bar{X}Ln\left(\frac{\bar{x}}{\bar{X}}\right)\right]\left[\exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right) + \left(\frac{\bar{X}}{\bar{x}}\right)\right] \quad ...(3.4)$$

$$\bar{z}_{p4} = 2^{-1}\left[u_4\frac{\bar{z}}{2}\left(\frac{\bar{X}}{\bar{x}} + \frac{\bar{x}}{\bar{X}}\right) - v_4\bar{X}Ln\left(\frac{\bar{x}}{\bar{X}}\right)\right]\left[\left(\frac{\bar{X}}{\bar{x}}\right)\exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right) + \left(\frac{\bar{x}}{\bar{X}}\right)\right] \quad ...(3.5)$$

$$\bar{z}_{p5} = 2^{-1}\left[u_5\frac{\bar{z}}{2}\left(\frac{\bar{X}}{\bar{x}} + \frac{\bar{x}}{\bar{X}}\right) - v_5\bar{X}Ln\left(\frac{\bar{x}}{\bar{X}}\right)\right]\left[\left(\frac{\bar{X}}{\bar{x}}\right)^{\frac{1}{2}}\exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right) + \left(\frac{\bar{x}}{\bar{X}}\right)^{\frac{1}{2}}\exp\left(\frac{1}{2}\left(\frac{\bar{x}-\bar{X}}{\bar{x}+\bar{X}}\right)\right)\right] \quad ...(3.6)$$

$$\bar{z}_{p6} = 2^{-1}\left[u_6\frac{\bar{z}}{2}\left(\frac{\bar{X}}{\bar{x}} + \frac{\bar{x}}{\bar{X}}\right) - v_6\bar{X}Ln\left(\frac{\bar{x}}{\bar{X}}\right)\right]\left[1 + \exp\left(-\frac{1}{2}\left(\frac{\bar{x}-\bar{X}}{\bar{x}+\bar{X}}\right)\right)\right] \quad ...(3.7)$$

$$\bar{z}_{p7} = 2^{-1}\left[u_7\frac{\bar{z}}{2}\left(\frac{\bar{X}}{\bar{x}} + \frac{\bar{x}}{\bar{X}}\right) - v_7\bar{X}Ln\left(\frac{\bar{x}}{\bar{X}}\right)\right]\left[1 + \exp\left(-\frac{1}{3}\left(\frac{\bar{x}-\bar{X}}{\bar{x}+\bar{X}}\right)\right)\right] \quad ...(3.8)$$

The general form of the estimators (3.2) through (3.8) is as follows:

$$\bar{z}_{pi} = 2^{-1}\left[u_i\frac{\bar{z}}{2}\left(\frac{\bar{X}}{\bar{x}} + \frac{\bar{x}}{\bar{X}}\right) - v_i\bar{X}Ln\left(\frac{\bar{x}}{\bar{X}}\right)\right]\left[\left(\frac{\bar{X}}{\bar{x}}\right)^a\exp\left(b\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right) + \left(\frac{\bar{x}}{\bar{X}}\right)^c\exp\left(d\frac{\bar{x}-\bar{X}}{\bar{x}+\bar{X}}\right)\right] \quad ...(3.9)$$

Where, i =1,2,3,4,5,6,7 and a,b,c,d are real numbers. Applying the relative definitions from section one to (3.2) and (3.9), we get

$$\bar{z}_{p0} = \frac{\left[\bar{Z}(1+e_0) - Ln\left(\frac{\bar{X}(1+e_1)}{\bar{X}}\right)\right]\bar{X}}{\bar{X}(1+e_1)} \quad ...(3.10)$$

$$\bar{z}_{pi} = 2^{-1}\left[u_i\frac{\bar{Z}(1+e_0)}{2}\left(\frac{\bar{X}}{\bar{X}(1+e_1)} + \frac{\bar{X}(1+e_1)}{\bar{X}}\right) - v_i\bar{X}Ln\left(\frac{\bar{X}(1+e_1)}{\bar{X}}\right)\right]\left[1 - G_1e_1 + G_2e_1^2\right] \quad ...(3.11)$$

By streamlining 3.10 and 3.11, we arrive at (3.12) and (3.13).

$$\bar{z}_{p0}=\left[\bar{Z}\left(1+e_0\right)-\left(e_1-\frac{e_1^2}{2}\right)\right]\left(1-e_1+e_1^2\right) \qquad ...(3.12)$$

$$\bar{z}_{pi}=\left[u_i\bar{Z}\left(1+e_0-G_1e_1+\left(G_2+\frac{1}{2}\right)e_1^2-G_1e_0e_1\right)-v_i\bar{X}\left(e_1-\left(G_2+\frac{1}{2}\right)e_1^2\right)\right] \qquad ...(3.13)$$

Subtracting $\bar{z}$ from both sides of (3.12) and (3.13), we obtained,

$$\bar{z}_{p0}-\bar{Z}=\bar{Z}\left[e_0-\left(\frac{\bar{Z}+1}{\bar{Z}}\right)e_1+\left(\frac{\bar{Z}+3}{2\bar{Z}}\right)e_1^2-e_0e_1\right] \qquad ...(3.14)$$

$$\bar{z}_{pi}-\bar{Z}=\bar{Z}\left[u_i\left(1+e_0-G_1e_1+\left(G_2+\frac{1}{2}\right)e_1^2-G_1e_0e_1\right)-v_i\frac{\bar{X}}{\bar{Z}}\left(e_1-\left(G_2+\frac{1}{2}\right)e_1^2\right)-1\right] \qquad ...(3.15)$$

Taking the square root of (3.14) and (3.15) and using expectation on both sides to get the estimators' mean square as

$$MSE\left(\bar{z}_{p0}\right)=\lambda\bar{Z}^2\left[C_z^2+\left(\frac{\bar{Z}+1}{\bar{Z}}\right)^2C_x^2-2\rho_{zx}\left(\frac{\bar{Z}+1}{\bar{Z}}\right)C_zC_x\right] \qquad ...(3.16)$$

$$MSE\left(\bar{z}_{pi}\right)=\bar{Z}^2\left[1+u_i^2A+v_i^2B-2u_iv_iC-2u_iD+2v_iE\right] \qquad ...(3.17)$$

Where

$A=1+\lambda\left(C_z^2+\left(G_1^2+2G_2+1\right)C_x^2-4G_1\rho_{zx}C_zC_x\right)$, $G_1=\dfrac{\left(a+\frac{b}{2}\right)-\left(c+\frac{d}{2}\right)}{2}$,

$G_2=\dfrac{\left(a+\frac{b}{2}\right)^2+\left(a+\frac{b}{2}\right)+\left(c+\frac{d}{2}\right)^2-\left(c+\frac{d}{2}\right)}{4}$, $B=\lambda\left(\dfrac{\bar{X}}{\bar{Z}}\right)^2C_x^2$,

$E=-\lambda\left(\dfrac{\bar{X}}{\bar{Z}}\right)\left(G_1+\dfrac{1}{2}\right)C_x^2$, $C=\lambda\left(\dfrac{\bar{X}}{\bar{Z}}\right)\left(\rho_{zx}C_zC_x-\left(2G_1+\dfrac{1}{2}\right)C_x^2\right)$,

$D=1+\lambda\left[\left(G_2+\dfrac{1}{2}\right)C_x^2-G_1\rho_{zx}C_zC_x\right]$, $\lambda=\dfrac{N-n}{Nn}$.

Differentiating (3.8) with respect to u_i and v_i, equate to zero and solve, we have,

$u_i=\dfrac{CE-BD}{C^2-AB}$ and $v_i=\dfrac{AE-CD}{C^2-AB}$

Efficiency Comparisons

Provided certain conditions are met, the suggested estimators outperform estimators found in the literature.

$$MSE\left(\bar{z}_0\right)<MSE\left(\bar{z}\right) \qquad ...(4.1)$$

$$MSE\left(\bar{z}_{pi}\right)<MSE\left(\bar{z}\right) \qquad ...(4.2)$$

$$\left[\left(\frac{\bar{Z}+1}{\bar{Z}}\right)^2C_x^2-2\rho_{zx}\left(\frac{\bar{Z}+1}{\bar{Z}}\right)C_zC_x\right]<0 \qquad ...(4.3)$$

$$\left[1+u_i^2A+v_i^2B-2u_iv_iC-2u_iD-2v_iE\right]-\lambda C_z^2<0 \qquad ...(4.4)$$

$$MSE\left(\bar{z}_0\right)<MSE\left(\bar{z}_s\right) \qquad ...(4.5)$$

$$MSE\left(\bar{z}_{pi}\right)<MSE\left(\bar{z}_s\right) \qquad ...(4.6)$$

$$\left[\left(\left(\frac{\bar{Z}+1}{\bar{Z}}\right)^2-1\right)C_x^2-2\rho_{zx}\left(\left(\frac{\bar{Z}+1}{\bar{Z}}\right)-1\right)C_zC_x\right]<0 \qquad ...(4.7)$$

$$\left[1+u_i^2A+v_i^2B-2u_iv_iC-2u_iD-2v_iE\right]-\lambda\left(C_z^2+C_z^2-2\rho_{zx}C_zC_x\right)<0 \qquad ...(4.8)$$

$$MSE\left(\bar{z}_0\right)<MSE\left(\bar{z}_{sh}\right) \qquad ...(4.9)$$

$$MSE\left(\bar{z}_{pi}\right)<MSE\left(\bar{z}_{sh}\right) \qquad ...(4.10)$$

$$\left[\left(\left(\frac{\bar{Z}+1}{\bar{Z}}\right)^2-\frac{1}{4}\right)C_x^2-2\rho_{zx}\left(\left(\frac{\bar{Z}+1}{\bar{Z}}\right)-\frac{1}{2}\right)C_zC_x\right]<0 \qquad ...(4.11)$$

$$\left[1+u_i^2A+v_i^2B-2u_iv_iC-2u_iD-2v_iE\right]-\lambda\left(C_z^2+\frac{C_x^2}{4}-\rho_{zx}C_zC_x\right)<0 \qquad ...(4.12)$$

Results and Discussion

This section presents empirical investigations that were carried out using three data sets to show how well the suggested estimators performed in comparison to some existing ones.

Population 1: [Source:[17]]

$N=250, n=80, \bar{Z}=20.0074, \bar{X}=40.0029, S_z=12.4384, S_x=8.6517, \rho_{zx}=0.5557$

Population 2: [Source:[17]]

$N=1000, n=150, \bar{Z}=200.0065, \bar{X}=2300.0016, S_z=41.5169, S_x=109.5119, \rho_{zx}=0.637$

Population 3: [Source:[17]]

$N=100, n=25, \bar{Z}=3.0047, \bar{X}=15.0018, S_z=4.2342, S_x=4.461, \rho_{zx}=0.6304$

Using the three data sets, Tables 1 and 2 present the empirical results of the MSE and PRE of the suggested and current estimators. The suggested estimators exhibit lower PRE and minimum MSE across all populations. This illustrates how the highly efficient estimators that have been suggested can yield more accurate population mean estimates when sensitive surveys are present than the estimators that have been taken into consideration in this study.

Table 1: MSE of proposed estimator and the existing ones using the populations

Estimators	Population 1	Population 2	Population 3
$\overline{Z}$	1.315067	9.767367	0.5378535
$\overline{z}_s$	0.9657649	7.426973	0.4187068
$\overline{z}_{sh}$	1.00627	8.468694	0.4722927
$\overline{z}_{p0}$	0.96566587	7.417854	0.3896769
$\overline{z}_{p1}$	0.294805	5.449369	0.1054512
$\overline{z}_{p2}$	0.294805	5.449369	0.1054512
$\overline{z}_{p3}$	0.294805	5.449369	0.1054512
$\overline{z}_{p4}$	0.294805	5.449369	0.1054512
$\overline{z}_{p5}$	0.09347122	2.772475	0.0713895
$\overline{z}_{p6}$	0.09347122	2.772475	0.0713895
$\overline{z}_{p7}$	0.02633714	1.880148	0.0599886

Table 2: PRE of proposed estimator and the existing ones using the populations

Estimators	Population 1	Population 2	Population 3
$\overline{Z}$	100	100	100
$\overline{z}_s$	136.17	131.51	128.46
$\overline{z}_{sh}$	119.98	113.34	113.88
$\overline{z}_{p0}$	136.18	131.67	138.03
$\overline{z}_{p1}$	446.08	179.24	510.05
$\overline{z}_{p2}$	446.08	179.24	510.05
$\overline{z}_{p3}$	446.08	179.24	510.05
$\overline{z}_{p4}$	446.08	179.24	510.05
$\overline{z}_{p5}$	1406.92	353.30	753.41
$\overline{z}_{p6}$	1406.92	353.30	753.41
$\overline{z}_{p7}$	4993.21	518.50	896.59

Conclusion

In this work, we proposed several log-type mean estimators for research variables that are sensitive, and we computed the mean square errors of these estimators. Based on the empirical results, the proposed estimators were seen to be more effective than the alternatives considered for this investigation. Accordingly, in situations where a sensitive issue is present, the suggested estimators ought to be applied when estimating the population mean.

Acknowledgment

The authors would like to thank Statistics Department of Usman Danfodiyo University for the successful completion of this research.

Funding Sources

The authors received no financial support for the research, authorship and publication of this article.

Conflict of Interest

The author(s) declare no conflict of interest.

Data Availability Statement

This statement does not apply to this article

Ethics Statement

This research did not involve human participants, animal subjects, or any material that requires ethical approval

Informed Consent Statement

This study did not involve human participants, and therefore, informed consent was not required.

Author Contributions

Each author mentioned has significantly and directly contributed intellectually to the project and has given their approval for its publication.

- **Mojeed Abiodun Yunusa:** conceptualization, methodology, writing-original draft.
- **Awwal Adejumobi:** data collection analysis, writing-review and editing.

References

1. Adejumobi A., Yunusa M. A., Erinola Y. A. and Abubakar K. An efficient logarithmic ratio type estimator of finite population mean under simple random sampling. *International Journal of Engineering and Applied Physics.* Vol. 3(2), pp. 700-705, (2023).
2. Azeem, M., Shabbir, J., Salahuddin, N., Hussain, S. and Salam, A. (2023) A comparative study of randomized response techniques using separate and combined metrics of efficiency and privacy. *PLOS ONE,* 19(1), 24-42. (2023).
3. Bar-Lev S.K., Bobovitch E. and Bouka B. A note on randomized response models for quantitative data, *Metrika,* 60, 255-260. (2004).
4. Diana G. and Perri P.F. New scrambled response models for estimating the mean of a sensitive quantitative character. Journal of Applied Statistics, 37(11), 1875-1890. (2010).
5. Eichhorn B.H. and Hayre L.S. Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and inference,* 7(4), 307-316. (1983).
6. Gjestvang C.R. and Singh S. An improved randomized response model: Estimation of mean. *Journal of Applied Statistics,* 36(12), 1361-1367. (2009).
7. Grover L.K. and Kaur P. An improved estimator of the finite population mean in simple random sampling. *Models Assisted Statistics and Application*, 6(1), 47-55. (2011).
8. Gupta S., Gupta B. and Singh S. Estimation of sensitivity level of personal interview survey questions. *Journal of Statistical Planning and Inference,* 100(2), 239-247. (2002).
9. Gupta S., Shabbir J. and Sehra S. Mean and sensitivity estimation in optional randomized response models. *Journal of Statistical Planning and inference,* 14(110), 2870-2874. (2010).
10. Hussain Z. Improvement of the Gupta and Thornton scrambling model through double use of randomization device. *International journal Academic Research Business and Social Sciences,* 291-297. (2012).
11. Mehta S., Dass B.K., Shabbir J. and Gupta S. A third stage optional randomized response model. *Journal of Statistical theory and Practice,* 6(3), 412-427. (2012).
12. Narjis G. and Shabbir J. An efficient new scrambled response model for estimating sensitive population mean in successive sampling. *Communications in Statistics-Simulation and Computation,* 1-18. (2021).
13. Pollock K.H. and Bek Y. A comparison of three randomized response models for quantitative data. *Journal of American Statistical Application,* 71, 884-886. (1976).
14. Sousa R., Shabbir J., Real P.C. and Gupta S. Ratio estimation of the mean of a sensitive variable in the presence of auxiliary information. *Journal of Statistics theory and Practice,* 4(3), 495-507. (2010).
15. Sahai A. A simple randomized response technique in complex surveys, *Metron,* 65, 59-66. (2007).
16. Shahzad U., Hanif M., Koyuncu N. and Luengo A.V.G. A new family of estimators for mean estimation alongside the sensitivity issues. *Journal of reliability and Statistical Studies,* 10(2), 43-63. (2017).
17. Tiwari K.K., Bhougal S., Kumar S. and Rather K.U. Using randomized response to estimate the population mean of a sensitive variable under measurement error. *Journal of Statistical Theory and Practice,* 16-28. (2022).

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

19. Warner S.L. The linear randomized response model. *Journal of the American Statistical Association,* 66(336), 884-888. (1971).