

# EFFICIENT IMPUTATION METHODS FOR ESTIMATING POPULATION PROPORTION IN DIAGONAL SYSTEMATIC SAMPLING

## Abstract

Accurate estimation of proportions is essential for effective decision-making, resource management, and ensuring the reliability of statistical analyses. It underpins many critical processes across various domains, making it a foundational aspect of data analysis and interpretation. In this paper, we have proposed a regression-type and exponential type imputation methods which are free of unknown parameters for estimating missing values or non-responses while estimating population proportion under diagonal systematic sampling design. Imputation methods are crucial in data analysis and machine learning because they address the issue of missing data. Some key significance includes improved data quality, preservation of sample size, bias reduction and enhanced predictive power. The estimators of the proposed imputation methods were derived. The properties (biases and MSEs) of the class of estimators of the proposed imputation methods were derived up to first order approximation. Results of numerical illustration using simulated data revealed that the proposed estimators are more efficient and practicable than existing estimators considered in the study.

**Keywords:** Imputation methods, Proportion, Diagonal Systematic sampling, Non-response.

## 1.0 Introduction

Population parameters like total, mean, variance, proportion can be estimated using efficient estimators in which the study attribute are associated with auxiliary information under the assumption that, full sample information is available in sample survey is one of the research interests. Several authors like Khan and Shabbir (2017), have worked extensively in this direction. However, surveys like medical and social science surveys often face the problem of non-response due to involvement of human in data collection. These lost values in turn create complications in data handling and analysis. Over time, many methods have been developed to address the problem of estimating unknown parameters in the presence of missing values. Imputation is a common technique used to handle situations where data is missing Hansen and Hurwitz (1946) were the first to consider the problem of nonresponse. Several authors also proposed imputation methods to deal with non-response or missing values. Recent among them include Singh and Horn (2000), Singh and Deo (2003), Wang and Wang (2006), Kadilar and Cingi (2008), Toutenburg et al. (2008), Singh (2009), Diana and Perri (2010), Al-Omari et al.

(2013), Singh et al. (2014), Gira (2015), Singh et al. (2016), Bhushan and Pandey (2016), Prasad (2017). Missing values can be completed with specific substitutes and data can be analyzed using standard methods. Information about unit of characteristic of interest observed and auxiliary attribute help improve the accuracy of demographic parameter estimates. However, some of the estimators in above aforementioned literatures depend on unknown parameters of study attribute  $y$  which make them impracticable in real life application.

Recently, Azeem (2021) suggested a new approach called diagonal systematic sampling. The approach arrange the population of size  $N = nk$  units, having  $n$  rows and  $k$  columns where  $n \leq k$ , draw a random  $r$  number where  $1 \leq r \leq k$ . The conventional estimator for the procedure was proposed and its efficiency was compared to that of conventional estimators based on simple random sampling with replacement and linear systematic sampling. However, his proposed estimator did not consider the situation of presence of outliers and non-response. Therefore, in the current study, we have considered the modification of Azeem (2021) estimator of population proportion under the framework of diagonal systematic sampling in the presence of non-response and outliers using imputation approach.

### 1.1 Notations

Consider a finite population  $U = U_1, U_2, U_3, \dots, U_N$  of size  $N$  units numbered from 1 to  $N$  in some order. A sample of size  $n$  is taken at random from the first  $k$  units and every  $k^{th}$  subsequent unit; Then,  $N = nk$  where  $n$  and  $k$  are positive integers; thus, there will be  $k$  samples (clusters) each of size  $n$  and observe the study variate  $y$  and auxiliary variate  $x$  for each and every unit selected in the sample.

Let every population unit  $(y_i, x_i)$  for  $i = 1, 2, \dots, N$  belongs to one of two mutually exclusive classes  $H$  and  $H^c$  where  $H$  is the class of units having the characteristics of interest. That is, let

$$y_i = \begin{cases} 1, & \text{if } i^{th} \text{ unit of population belongs to class } H, \\ 0, & \text{otherwise.} \end{cases} \quad (1.1)$$

$$x_i = \begin{cases} 1, & \text{if } i^{th} \text{ unit of population belongs to class } H, \\ 0, & \text{otherwise.} \end{cases} \quad (1.2)$$

Then, the systematic sample means are defined as  $p_{a(\text{sys})} = 1/n \sum_{i=1}^n y_i = \frac{a_{(\text{sys})}}{n}$  and

$p_{b(\text{sys})} = 1/n \sum_{i=1}^n x_i = \frac{b_{(\text{sys})}}{n}$  are unbiased estimators of the population means  $P_A = \frac{1}{N} \sum_{i=1}^N Y_i = \frac{A}{N}$  and

$P_B = 1/N \sum_{i=1}^N x_i = \frac{B}{N}$  for  $Y$  and  $X$  respectively.

The usual sample proportion  $p_{a(\text{sys})}$  estimator of population proportion  $P_A$  is given in (1.3) as

$$p_{a(\text{sys})} = \frac{a_{(\text{sys})}}{n} \quad (1.3)$$

The variance of sample proportion  $V(p_{a(\text{sys})})$  estimator of population proportion  $P_A$  is given as in (1.4).

$$V(p_{a(\text{sys})}) = \frac{p(1-p)}{n} p_A^* \quad (1.4)$$

Azeem (2021) adopted the steps involved in the Subramani (2000) diagonal systematic sampling method which are as follows:

1. Arrange the population having  $N = nk$  units in a table having  $n$  rows and  $k$  columns where  $n \leq k$ .
2. Draw a random number  $r$ , where  $1 \leq r \leq k$ .
3. The units are drawn in such a way that the sampled units are the entries in the diagonal/broken diagonal of the table as shown in Table 1.

The selected units in diagonal systematic sampling scheme are as follows (Subramani, 2000).

$$S_r = \{y_r, y_{(k+1)+r}, y_{2(k+1)+r}, \dots, y_{(n-1)(k+1)+r}\}, \quad \text{if } r \leq k - n + 1, \quad (1.5)$$

$$S_r = \{y_r, y_{(k+1)+r}, y_{2(k+1)+r}, \dots, y_{t(k+1)+r=(t+1)k}, y_{(t+1)k+1}, y_{(t+2)k+2}, \dots, y_{(n-1)k+(n-t-1)}\}, \quad \text{if } r > k - n + 1, \quad (1.6)$$

where  $0 \leq t \leq n-1$

Table 1: Arrangement of the Population units

| S/N | 1              | 2              | ... | K        |
|-----|----------------|----------------|-----|----------|
| 1   | $y_1$          | $y_2$          | ... | $y_k$    |
| 2   | $y_{k+1}$      | $y_{k+2}$      | ... | $y_{2k}$ |
| 3   | $y_{2k+1}$     | $y_{2k+2}$     | ... | $y_{3k}$ |
| ... | ...            | ...            | ... | ...      |
| $n$ | $y_{(n-1)k+1}$ | $y_{(n-1)k+2}$ | ... | $y_{nk}$ |

The first and second order inclusion probabilities under diagonal systematic sampling are given by:

$$\pi_i^* = \frac{1}{k} \quad (1.7)$$

$$\pi_{ij}^* = \begin{cases} \frac{1}{k}, & \text{if } i^{\text{th}} \text{ and } j^{\text{th}} \text{ units are from the same diagonal or broken diagonal,} \\ 0, & \text{otherwise} \end{cases} \quad (1.8)$$

**1.2 Estimator Of Population Proportion In Diagonal Systematic Sapling And Its Properties**  
Azeem (2021) suggested the sample proportion based on the diagonal systematic sampling scheme given by:

$$p_{ad(sys)} = \frac{a_{d(sys)}}{n} \quad (1.9)$$

$$\text{where } a_{d(sys)} = \begin{cases} \sum_{i=0}^{n-1} y_{i(k+1)+r}, & \text{if } r \leq k-n+1, \\ \sum_{i=0}^t y_{i(k+1)+r} + \sum_{i=1}^{n-t-1} y_{(t+i)k+i}, & \text{if } r > k-n+1. \end{cases}$$

The variance of  $p_{ad(sys)}$  is given by:

$$Var(p_{ad(sys)}) = \frac{1}{k} \sum_{i=1}^k (p_{ad(sys)} - P)^2 \quad (1.10)$$

**Theorem 1** Under diagonal systematic sampling scheme, the sample proportion can be written in the form of Horvitz–Thompson estimator pHT, suggested by Horvitz and Thompson (1952). Also,  $p_{ad(sys)}$  is unbiased for population proportion P.

**Proof** By definition

$$p_{ad(s)} = \frac{a_{d(s)}}{n} = \frac{ka_{d(s)}}{kn} = \frac{k \sum_{i \in S} y_i}{N}$$

where  $s$  denotes the sample drawn from population.

$$p_{ad(s)} = \frac{1}{N} \left( \sum_{i \in S} \frac{y_i}{1/k} \right) = \frac{1}{N} \sum_{i \in S} \frac{y_i}{\pi_i} \quad (1.11)$$

where  $\pi_i = 1/k$

Taking expectation on both sides of (1.9) yield

$$E(p_{ad(s)}) = E\left(\frac{a_{d(s)}}{n}\right) = E\left(\frac{ka_{d(s)}}{kn}\right) = \frac{k}{N} \sum_{i=1}^n E(y_i) = \frac{k}{N} \sum_{i=1}^n \left[ \sum_{i \in S} y_i \frac{1}{nk} \right] = \frac{\sum_{i \in S} y_i}{N} = \frac{A}{N} = P \quad (1.12)$$

where  $s$  denotes the sample drawn from population.

**Remark 1** Although Horvitz–Thompson estimator is usually applied to estimate the finite population mean or population total, it can also be used to estimate population proportion by simply treating the variable as binary variable having possible values 0 and 1.

Remark 2 Using Sen–Yates–Grundy approach suggested by Sen (1953) and Yates and Grundy (1953).

The variance of  $P_{ad(s)}$  can be written as:

$$\text{Var}(P_{ad(s)}) = \frac{1}{N^2} \left\{ \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N (\pi_i^* \pi_j^* - \pi_{ij}^*) \left( \frac{y_i}{\pi_i^*} - \frac{y_j}{\pi_j^*} \right)^2 \right\} \quad (1.13)$$

Also, the estimate of variance  $\text{Var}(P_{ad(s)})$  denoted by  $\overline{\text{Var}}(P_{ad(s)})$  is given below:

$$\overline{\text{Var}}(P_{ad(s)}) = \frac{1}{2N^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \left( \frac{\pi_i^* \pi_j^* - \pi_{ij}^*}{\pi_{ij}^*} \right) \left( \frac{y_i}{\pi_i^*} - \frac{y_j}{\pi_j^*} \right)^2 \quad (1.14)$$

The values of  $\pi_i^*$  and  $\pi_{ij}^*$  can be used from (1.7) and (1.8) in expression (1.13) and (1.14) to obtain the sampling variance of the sample proportion and its estimator under the diagonal systematic sampling scheme.

## 2.0 Materials and Method

### 2.1 Proposed Imputation Schemes

Having study the estimator of Azeem (2021), the following imputation schemes were presented:

Let consider  $(y, x) \in \mathfrak{R}^+$  be pair of study and associated auxiliary attributes measured on the studied population. Also, let  $J$  denotes the set of responses with  $n_1$  units,  $J^c$  denotes the set of non-responses having  $n - n_1$  units or missing units (out of  $n$ ) and  $S$  denotes the set of  $n$  units sampled without replacement from the  $N$  units in the population of interest. For each  $i \in J$ , the value of  $y_i$  is observed. However, for unit  $i \in J^c$ ,  $y_i$  is missing due to non-response and obtained using different methods of imputation.

#### 2.1.1 First Proposed Imputation

Motivated by Audu et al. (2023), the regression-type compromised imputation scheme for population proportion under diagonal systematic sampling design is proposed as in (2.1)

$$y_i = \begin{cases} y_i & \text{if } i \in J \\ \frac{p_{ad(sys)}^* + \hat{\beta}_\phi (P_B - p_{bd(sys)}^*)}{V_1 p_{bd(sys)}^* + V_2} (V_1 P_B + V_2) & \text{if } i \in J^c \end{cases} \quad (2.1)$$

where  $p_{ad(sys)}^* = \frac{a_{d(sys)}^*}{r}$ ,  $p_{bd(sys)}^* = \frac{b_{d(sys)}^*}{r}$ ,  $\hat{\beta}_\phi = s_{(ab)r} / s_{(b)}^2$ ,  $s_{(ab)r} = (r-1)^{-1} \sum_{i=1}^r (y_i - p_a^*)(x_i - p_b^*)$ ,

$$s_{(b)}^2 = (r-1)^{-1} \sum_{i=1}^r (x_i - p_b^*)^2, \quad a_{d(sys)}^* = \begin{cases} \sum_{i=0}^{n_1-1} y_{i(k+1)+r}, & \text{if } r \leq k - n_1 + 1, \\ \sum_{i=0}^t y_{i(k+1)+r} + \sum_{i=1}^{n_1-t-1} y_{(t+i)k+i}, & \text{if } r > k - n_1 + 1 \end{cases}, \quad b_{d(sys)}^* = \begin{cases} \sum_{i=0}^{n_1-1} x_{i(k+1)+r}, & \text{if } r \leq k - n_1 + 1, \\ \sum_{i=0}^t x_{i(k+1)+r} + \sum_{i=1}^{n_1-t-1} x_{(t+i)k+i}, & \text{if } r > k - n_1 + 1 \end{cases},$$

$V_1$  and  $V_2$  are either constants or known functions of auxiliary attributes like coefficients of skewness  $\beta_{1(b)}$ , kurtosis  $\beta_{2(b)}$ , variation  $C_B$ , standard deviation  $S_B$  etc, that are arbitrarily chosen to generate the members of proposed class of estimators

The estimator of the first proposed imputation can be obtained using the function defined in (2.2)

$$T = \frac{1}{n} \left( \sum_{i \in J} y_i + \sum_{i \in J^c} y_i \right) \quad (2.2)$$

$$t_{1(i)} = \frac{1}{n} \left( \sum_{i=1}^r y_i + \sum_{i=1}^{n-r} \frac{p_{ad(sys)}^* + \hat{\beta}_\phi (P_B - p_{bd(sys)}^*)}{V_1 p_{bd(sys)}^* + V_2} (V_1 P_B + V_2) \right) \quad (2.3)$$

$$t_{1(i)} = \frac{1}{n} \left( r p_{ad(sys)}^* + (n-r) \frac{p_{ad(sys)}^* + \hat{\beta}_\phi (P_B - p_{bd(sys)}^*)}{V_1 p_{bd(sys)}^* + V_2} (V_1 P_B + V_2) \right) \quad (2.4)$$

$$t_{1(i)} = \frac{r}{n} p_{ad(sys)}^* + \left(1 - \frac{r}{n}\right) \frac{p_{ad(sys)}^* + \hat{\beta}_\phi (P_B - p_{bd(sys)}^*)}{V_1 p_{bd(sys)}^* + V_2} (V_1 P_B + V_2) \quad (2.5)$$

Using error terms defined in (2.6), (2.5) can be expressed as obtained in (2.7)

$$e_0 = \frac{p_{ad(sys)}^* - P_A}{P_A}, e_1 = \frac{p_{bd(sys)}^* - P_B}{P_B} \text{ Such that } p_{ad(sys)}^* = (1 + e_0) P_A, p_{bd(sys)}^* = (1 + e_1) P_B \quad (2.6)$$

$$t_{1(i)} = \frac{r}{n} (1 + e_0) P_A + \left(1 - \frac{r}{n}\right) \left[ \frac{(1 + e_0) P_A + \hat{\beta}_\phi (P_B - (1 + e_1) P_B)}{V_1 (1 + e_1) + V_2} \right] (V_1 P_B + V_2) \quad (2.7)$$

$$t_{1(i)} = \frac{r}{n} (1 + e_0) P_A + \left(1 - \frac{r}{n}\right) \frac{P_A + P_A e_0 - \hat{\beta}_\phi P_B e_1}{(V_1 P_B + V_2) \left[1 + \frac{V_1 P_B e_1}{V_1 P_B + V_2}\right]} (V_1 P_B + V_2) \quad (2.8)$$

$$t_{1(i)} = \frac{r}{n} (1 + e_0) P_A + \left(1 - \frac{r}{n}\right) (P_A + P_A e_0 - \hat{\beta}_\phi P_B e_1) (1 + H e_1)^{-1} \quad (2.9)$$

$$\text{where } H = \frac{V_1 P_B}{V_1 P_B + V_2}$$

By simplifying (2.9) using Taylor series approximation, (2.10) is obtained

$$t_{1(i)} = \frac{r}{n} (1 + e_0) P_A + \left(1 - \frac{r}{n}\right) (P_A + P_A e_0 - \hat{\beta}_\phi P_B e_1) (1 - H e_1 + H^2 e_1^2) \quad (2.10)$$

$$t_{1(i)} = \frac{r}{n} P_A + \frac{r}{n} P_A e_0 + \left(1 - \frac{r}{n}\right) (P_A - H P_A e_1 + H^2 P_A e_1^2 + P_A e_0 - H P_A e_0 e_1 - \hat{\beta}_\phi P_B e_1 + H \hat{\beta}_\phi P_B e_1^2) \quad (2.11)$$

$$t_{1(i)} = \frac{r}{n} P_A + \frac{r}{n} P_A e_0 + P_A - \frac{r}{n} P_A + P_A e_0 - \frac{r}{n} P_A e_0 + \left(1 - \frac{r}{n}\right) \left( \begin{aligned} & - (H P_A + \hat{\beta}_\phi P_B) e_1 \\ & + (H^2 P_A + H \hat{\beta}_\phi P_B) e_1^2 - H P_A e_0 e_1 \end{aligned} \right) \quad (2.12)$$

$$t_{1(i)} - P_A = P_A e_0 + \left(1 - \frac{r}{n}\right) \left( - (H P_A + \hat{\beta}_\phi P_B) e_1 + (H^2 P_A + H \hat{\beta}_\phi P_B) e_1^2 - H P_A e_0 e_1 \right) \quad (2.13)$$

By taking the expectation of (2.13), the Bias of the first proposed estimator denoted by  $t_{1(i)}$  is obtained as in (2.14)

$$\text{Bias}(t_{1(i)}) = \left(1 - \frac{r}{n}\right) \left( (H^2 P_A + H \hat{\beta}_\phi P_B) E(e_1^2) - H P_A E(e_0 e_1) \right) \quad (2.14)$$

$$\text{Bias}(t_3) = \left(1 - \frac{r}{n}\right) \left( \begin{aligned} & (H^2 P_A + H \hat{\beta}_\phi P_B) \frac{1}{N_1^2 P_B} \left\{ \frac{1}{2} \sum_{i=1}^{N_1} \sum_{\substack{j=1 \\ j \neq i}}^{N_1} (\pi_i \pi_j - \pi_{ij}) \left( \frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right)^2 \right\} \\ & - H P_A \frac{1}{N_1^2 P_A P_B} \left\{ \frac{1}{2} \sum_{i=1}^{N_1} \sum_{\substack{j=1 \\ j \neq i}}^{N_1} (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \left( \frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right)^2 \right\} \end{aligned} \right) \quad (2.15)$$

Also, by taking expectation and squaring both sides of (2.13), the MSE of the first proposed estimator denoted by  $t_{1(i)}$  is obtained as in (2.16)

$$MSE(t_{1(i)}) = E \left[ P_A e_0 + \left(1 - \frac{r}{n}\right) \left( - (HP_A + \hat{\beta}_\phi P_B) e_1 + (H^2 P_A + H \hat{\beta}_\phi P_B) e_1^2 - HP_A e_0 e_1 \right) \right]^2 \quad (2.16)$$

$$MSE(t_{1(i)}) = P_A^2 E(e_0^2) + \left(1 - \frac{r}{n}\right)^2 (HP_A + \hat{\beta}_\phi P_B)^2 E(e_1^2) - 2 \left(1 - \frac{r}{n}\right) P_A (HP_A + \hat{\beta}_\phi P_B) E(e_0 e_1) \quad (2.17)$$

$$\begin{aligned} MSE(t_{1(i)}) = & P_A^2 \frac{1}{N_1^2 P_A} \left\{ \frac{1}{2} \sum_{i=1}^{N_1} \sum_{\substack{j=1 \\ j \neq i}}^{N_1} (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \right\} \\ & + \left(1 - \frac{r}{n}\right)^2 (HP_A + \hat{\beta}_\phi P_B)^2 \frac{1}{N_1^2 P_B} \left\{ \frac{1}{2} \sum_{i=1}^{N_1} \sum_{\substack{j=1 \\ j \neq i}}^{N_1} (\pi_i \pi_j - \pi_{ij}) \left( \frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right)^2 \right\} \\ & - 2 \left(1 - \frac{r}{n}\right) P_A (HP_A + \hat{\beta}_\phi P_B) \frac{1}{N_1^2 P_A P_B} \left\{ \frac{1}{2} \sum_{i=1}^{N_1} \sum_{\substack{j=1 \\ j \neq i}}^{N_1} (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \left( \frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right)^2 \right\} \end{aligned} \quad (2.18)$$

### 2.1.2 Second Proposed Imputation

Motivated by Audu and Singh (2021), the exponential-type regression compromised imputation scheme for population proportion under diagonal systematic sampling design is proposed as in (2.19)

$$y_i = \begin{cases} y_i & \text{if } i \in J \\ \frac{p_{ad(sys)}^* + \hat{\beta}_\phi (P_B - p_{bd(sys)}^*)}{V_1 p_{bd(sys)}^* + V_2} (V_1 P_B + V_2) \exp \left( \frac{P_B - p_{bd(sys)}^*}{P_B + p_{bd(sys)}^*} \right) & \text{if } i \in J^c \end{cases} \quad (2.19)$$

Using (2.2), the estimator of the second proposed imputation can be obtained as follows:

$$t_{2(i)} = \frac{1}{n} \left( \sum_{i=1}^r y_i + \sum_{i=1}^{n-r} \frac{p_{ad(sys)}^* + \hat{\beta}_\phi (P_B - p_{bd(sys)}^*)}{V_1 p_{bd(sys)}^* + V_2} (V_1 P_B + V_2) \exp \left( \frac{P_B - p_{bd(sys)}^*}{P_B + p_{bd(sys)}^*} \right) \right) \quad (2.20)$$

$$t_{2(i)} = \frac{1}{n} \left( r p_{ad(sys)}^* + (n-r) \frac{p_{ad(sys)}^* + \hat{\beta}_\phi (P_B - p_{bd(sys)}^*)}{V_1 p_{bd(sys)}^* + V_2} (V_1 P_B + V_2) \exp \left( \frac{P_B - p_{bd(sys)}^*}{P_B + p_{bd(sys)}^*} \right) \right) \quad (2.21)$$

$$t_{2(i)} = \frac{r}{n} p_{ad(sys)}^* + \left(1 - \frac{r}{n}\right) \frac{p_{ad(sys)}^* + \hat{\beta}_\phi (P_B - p_{bd(sys)}^*)}{V_1 p_{bd(sys)}^* + V_2} (V_1 P_B + V_2) \exp \left( \frac{P_B - p_{bd(sys)}^*}{P_B + p_{bd(sys)}^*} \right) \quad (2.22)$$

Using error terms defined in (2.6), (2.11) can be expressed as obtained in (2.23)

$$t_{2(i)} = \frac{r}{n}(1+e_0)P_A + \left(1 - \frac{r}{n}\right) \frac{P_A + P_A e_0 - \hat{\beta}_\phi P_B e_1}{V_1(1+e_1)P_B + V_2} (V_1 P_B + V_2) \exp\left(\frac{P_B - (1+e_1)P_B}{P_B + (1+e_1)P_B}\right) \quad (2.23)$$

$$t_{2(i)} = \frac{r}{n}(1+e_0)P_A + \left(1 - \frac{r}{n}\right) \frac{P_A + P_A e_0 - \hat{\beta}_\phi P_B e_1}{(V_1 P_B + V_2) \left[1 + \frac{V_1 P_B e_1}{V_1 P_B + V_2}\right]} (V_1 P_B + V_2) \exp\left(\frac{-e_1}{2+e_1}\right) \quad (2.24)$$

$$t_{2(i)} = \frac{r}{n}(1+e_0)P_A + \left(1 - \frac{r}{n}\right) (P_A + P_A e_0 - \hat{\beta}_\phi P_B e_1) \left(1 + \frac{V_1 P_B e_1}{V_1 P_B + V_2}\right)^{-1} \exp\left(\frac{-e_1}{2} \left(1 + \frac{e_1}{2}\right)^{-1}\right) \quad (2.25)$$

$$t_{2(i)} = \frac{r}{n}(1+e_0)P_A + \left(1 - \frac{r}{n}\right) (P_A + P_A e_0 - \hat{\beta}_\phi P_B e_1) (1 + H e_1)^{-1} \exp\left(\frac{-e_1}{2} + \frac{e_1^2}{4}\right) \quad (2.26)$$

$$t_{2(i)} = \frac{r}{n}P_A + \frac{r}{n}P_A e_0 + \left(1 - \frac{r}{n}\right) (P_A + P_A e_0 - \hat{\beta}_\phi P_B e_1) (1 - H e_1 + H^2 e_1^2) \left(1 - \frac{e_1}{2} + \frac{e_1^2}{8}\right) \quad (2.27)$$

$$t_{2(i)} = \frac{r}{n}P_A + \frac{r}{n}P_A e_0 + \left(1 - \frac{r}{n}\right) \left( \begin{aligned} &P_A + P_A e_0 - \left(\frac{P_A}{2} + H P_A + \hat{\beta}_\phi P_B\right) e_1 - P_A e_0 e_1 \left(\frac{1}{2} + H\right) \\ &+ \left(\frac{P_A}{8} + \frac{H P_A}{2} + H^2 P_A + \frac{\hat{\beta}_\phi P_B}{2} + H \hat{\beta}_\phi P_B\right) e_1^2 \end{aligned} \right) \quad (2.28)$$

$$t_{2(i)} - P_A = P_A e_0 + \left(1 - \frac{r}{n}\right) \left( \begin{aligned} &-\left(\frac{P_A}{2} + H P_A + \hat{\beta}_\phi P_B\right) e_1 - P_A e_0 e_1 \left(\frac{1}{2} + H\right) \\ &+ \left(\frac{P_A}{8} + \frac{H P_A}{2} + H^2 P_A + \frac{\hat{\beta}_\phi P_B}{2} + H \hat{\beta}_\phi P_B\right) e_1^2 \end{aligned} \right) \quad (2.29)$$

By taking the expectation of (2.29), the Bias of the second proposed estimator denoted by  $t_{2(i)}$  is obtained as in (2.30)

$$Bias(t_{2(i)}) = \left(1 - \frac{r}{n}\right) \left( \left( \frac{P_A}{8} + \frac{H P_A}{2} + H^2 P_A + \frac{\hat{\beta}_\phi P_B}{2} + H \hat{\beta}_\phi P_B \right) E(e_1^2) - P_A \left(\frac{1}{2} + H\right) E(e_0 e_1) \right) \quad (2.30)$$

$$Bias(t_{2(i)}) = \left(1 - \frac{r}{n}\right) \left( \begin{aligned} &\left( \frac{P_A}{8} + \frac{H P_A}{2} + H^2 P_A + \frac{\hat{\beta}_\phi P_B}{2} + H \hat{\beta}_\phi P_B \right) \\ &\frac{1}{N_1^2 P_B} \left\{ \frac{1}{2} \sum_{i=1}^{N_1} \sum_{\substack{j=1 \\ j \neq i}}^{N_1} (\pi_i \pi_j - \pi_{ij}) \left( \frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right)^2 \right\} \\ &- P_A \left(\frac{1}{2} + H\right) \frac{1}{N_1^2 P_A P_B} \left\{ \frac{1}{2} \sum_{i=1}^{N_1} \sum_{\substack{j=1 \\ j \neq i}}^{N_1} (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \left( \frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right)^2 \right\} \end{aligned} \right) \quad (2.31)$$

Also, by taking expectation and squaring both sides of (2.29), the MSE of the second proposed estimator denoted by  $t_{2(i)}$  is obtained as in (2.32)

$$MSE(t_{2(i)}) = E \left[ P_A e_0 + \left(1 - \frac{r}{n}\right) \left( - \left( \frac{P_A}{2} + HP_A + \hat{\beta}_\phi P_B \right) e_3 - P_A e_0 e_1 \left( \frac{1}{2} + H \right) + \left( \frac{P_A}{8} + \frac{HP_A}{2} + H^2 P_A + \frac{\hat{\beta}_\phi P_B}{2} + H \hat{\beta}_\phi P_B \right) e_1^2 \right) \right]^2 \quad (2.32)$$

$$MSE(t_{2(i)}) = P_A^2 E(e_0^2) + \left(1 - \frac{r}{n}\right)^2 \left( \frac{P_A}{2} + HP_A + \hat{\beta}_\phi P_B \right)^2 E(e_1^2) - 2 \left(1 - \frac{r}{n}\right) P_A \left( \frac{P_A}{2} + HP_A + \hat{\beta}_\phi P_B \right) E(e_0 e_1) \quad (2.33)$$

$$MSE(t_{2(i)}) = P_A^2 \frac{1}{N_1^2 P_A} \left\{ \frac{1}{2} \sum_{i=1}^{N_1} \sum_{\substack{j=1 \\ j \neq i}}^{N_1} (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \right\} + \left(1 - \frac{r}{n}\right)^2 \left( \frac{P_A}{2} + HP_A + \hat{\beta}_\phi P_B \right)^2 \frac{1}{N_1^2 P_B} \left\{ \frac{1}{2} \sum_{i=1}^{N_1} \sum_{\substack{j=1 \\ j \neq i}}^{N_1} (\pi_i \pi_j - \pi_{ij}) \left( \frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right)^2 \right\} - 2 \left(1 - \frac{r}{n}\right) P_A \left( \frac{P_A}{2} + HP_A + \hat{\beta}_\phi P_B \right) \frac{1}{N_1^2 P_A P_B} \left\{ \frac{1}{2} \sum_{i=1}^{N_1} \sum_{\substack{j=1 \\ j \neq i}}^{N_1} (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right) \left( \frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right) \right\} \quad (2.34)$$

### 3.0 Empirical Study

In this section, simulation studies to assess the performance of the proposed estimators  $t_{1(i)}$  and  $t_{2(i)}$  with respect to Azeem (2021) estimator under the effect of non-response and outliers were conducted. Data of size 500 units were generated for the study population using binomial distribution. A sample of size 100 was selected by the method of diagonal systematic sampling 500 times. The Biases, MSEs and PREs of the considered estimators were computed using (3.1), (3.2) and (3.3)

$$Bias(T) = \frac{1}{500} \sum_{j=1}^{500} (p_{ad} - P_A) \quad (3.1)$$

$$MSE(T) = \frac{1}{500} \sum_{j=1}^{500} (p_{ad} - P_A)^2 \quad (3.2)$$

$$PREs(T) = \frac{MSE(t_0)}{MSE(T)} \times 100 \quad (3.3)$$

The simulation procedure is described in the following steps

Step1: Population of size 500 is generated

Step2: Compute the required parameters of the auxiliary attribute

Step3: Take a sample of size 100 from Step1 and compute the statistics needed

Step4: Use the results of 2 and 3 to compute the biases, MSEs and PREs.

Step5: Repeat step 3 and 4 500 times

Step6: Compute the averages of the results of step5

#### 4.0 Results and Discussion

**Table 2: Biases, MSEs and PREs of Estimators  $p_{ad(sys)}$ ,  $t_{1(i)}$  and  $t_{2(i)}$  when  $P=0.5$**

| ESTIMATOR                     | BIAS    | MSE    | PRE      | ESTIMATOR                     | BIAS    | MSE    | PRE      |
|-------------------------------|---------|--------|----------|-------------------------------|---------|--------|----------|
| $p_{ad(sys)}$<br>Azzem(2021)  | -0.0009 | 0.0015 | 100      | $p_{ad(sys)}$<br>Azzem(2021)  | -0.0009 | 0.0015 | 100      |
| Proposed Estimator $t_{1(i)}$ |         |        |          | Proposed Estimator $t_{2(i)}$ |         |        |          |
| $t_{1(1)}$                    | 0.0003  | 0.0013 | 125.1558 | $t_{2(1)}$                    | -0.0022 | 0.0011 | 145.9438 |
| $t_{1(2)}$                    | -0.0009 | 0.0015 | 108.1937 | $t_{2(2)}$                    | -0.0036 | 0.0012 | 133.5281 |
| $t_{1(3)}$                    | -0.0009 | 0.0015 | 105.095  | $t_{2(3)}$                    | -0.0037 | 0.0012 | 130.5208 |
| $t_{1(4)}$                    | -0.0006 | 0.0017 | 92.08923 | $t_{2(4)}$                    | -0.0039 | 0.0013 | 116.735  |
| $t_{1(5)}$                    | 0.00032 | 0.0012 | 131.9072 | $t_{2(5)}$                    | -0.0012 | 0.0011 | 147.008  |
| $t_{1(6)}$                    | -0.0009 | 0.0015 | 105.095  | $t_{2(6)}$                    | -0.0037 | 0.0012 | 130.5208 |
| $t_{1(7)}$                    | -0.0006 | 0.0017 | 103.0892 | $t_{2(7)}$                    | -0.0039 | 0.0013 | 116.735  |
| $t_{1(8)}$                    | 0.00032 | 0.0012 | 131.9072 | $t_{2(8)}$                    | -0.0012 | 0.0011 | 147.008  |
| $t_{1(9)}$                    | -0.0009 | 0.0014 | 112.0605 | $t_{2(9)}$                    | -0.0034 | 0.0011 | 137.0513 |
| $t_{1(10)}$                   | 0.0005  | 0.0019 | 108.1483 | $t_{2(10)}$                   | -0.0035 | 0.0015 | 102.0924 |
| $t_{1(11)}$                   | -4.8208 | 0.0012 | 128.4142 | $t_{2(11)}$                   | -0.0017 | 0.0010 | 146.9423 |
| $t_{1(12)}$                   | -0.0650 | 0.0243 | 6.480046 | $t_{2(12)}$                   | -0.0679 | 0.0245 | 6.420426 |
| $t_{1(13)}$                   | 0.0011  | 0.0021 | 74.64178 | $t_{2(13)}$                   | -0.0031 | 0.0016 | 96.91431 |
| $t_{1(14)}$                   | -0.0005 | 0.0013 | 122.6186 | $t_{2(14)}$                   | -0.0025 | 0.0010 | 144.7304 |
| $t_{1(15)}$                   | -0.0009 | 0.0016 | 111.5792 | $t_{2(15)}$                   | -0.0039 | 0.0013 | 123.7993 |
| $t_{1(16)}$                   | -0.0009 | 0.0016 | 102.2689 | $t_{2(16)}$                   | -0.0039 | 0.0013 | 124.5315 |
| $t_{1(17)}$                   | -0.0009 | 0.0016 | 100.6587 | $t_{2(17)}$                   | -0.0039 | 0.0012 | 125.9936 |

Table 2 displays the outcomes of biases, mean squared errors (MSEs), and percentage relative efficiency (PREs) for the existing and proposed estimators when  $p = 0.5$ . The findings indicate that, with the exception of  $(t_{1(4)}, t_{1(12)}, t_{1(13)}, t_{2(12)}$  and  $t_{2(13)})$ , all the proposed estimators exhibit lower MSEs and higher PREs in comparison to the existing estimator considered in this investigation. Consequently, the proposed estimators in this context are more efficient than their competitor in the study and are likely to yield superior estimates of the population proportion, particularly in scenarios involving non-response.

**Table 3: Biases, MSEs and PREs of Estimators  $p_{ad(sys)}$ ,  $t_{1(i)}$  and  $t_{2(i)}$  when  $P=0.6$**

| ESTIMATOR                     | BIAS    | MSE    | PRE      | ESTIMATOR                     | BIAS    | MSE    | PRE      |
|-------------------------------|---------|--------|----------|-------------------------------|---------|--------|----------|
| $p_{ad(sys)}$<br>Azzem(2021)  | -0.0009 | 0.0015 | 100      | $p_{ad(sys)}$<br>Azzem(2021)  | -0.0009 | 0.0015 | 100      |
| Proposed Estimator $t_{1(i)}$ |         |        |          | Proposed Estimator $t_{2(i)}$ |         |        |          |
| $t_{1(1)}$                    | 0.0001  | 0.0018 | 108.9821 | $t_{2(1)}$                    | 0.0014  | 0.0017 | 112.8677 |
| $t_{1(2)}$                    | -0.0003 | 0.0019 | 103.0852 | $t_{2(2)}$                    | 0.0003  | 0.0018 | 110.1613 |
| $t_{1(3)}$                    | -0.0004 | 0.0019 | 101.8556 | $t_{2(3)}$                    | 0.0002  | 0.0018 | 109.3862 |
| $t_{1(4)}$                    | -8.9443 | 0.0020 | 96.80879 | $t_{2(4)}$                    | -8.2330 | 0.0019 | 105.8678 |
| $t_{1(5)}$                    | 0.0001  | 0.0018 | 108.8499 | $t_{2(5)}$                    | 0.0014  | 0.0018 | 112.8367 |
| $t_{1(6)}$                    | -0.0004 | 0.0019 | 101.8556 | $t_{2(6)}$                    | 0.0002  | 0.0018 | 109.3862 |
| $t_{1(7)}$                    | -8.9443 | 0.0020 | 96.80879 | $t_{2(7)}$                    | -8.2330 | 0.0019 | 105.8678 |
| $t_{1(8)}$                    | -0.0001 | 0.0018 | 108.8499 | $t_{2(8)}$                    | 0.0014  | 0.0018 | 112.8367 |
| $t_{1(9)}$                    | -0.0004 | 0.0018 | 104.6066 | $t_{2(9)}$                    | 0.0005  | 0.0018 | 111.0486 |
| $t_{1(10)}$                   | 0.0010  | 0.0022 | 89.64492 | $t_{2(10)}$                   | 0.0005  | 0.0019 | 100.4655 |
| $t_{1(11)}$                   | 0.0001  | 0.0018 | 108.9158 | $t_{2(11)}$                   | 0.0014  | 0.0017 | 112.8524 |
| $t_{1(12)}$                   | -0.0684 | 0.0060 | 32.39762 | $t_{2(12)}$                   | -0.0683 | 0.0061 | 32.46881 |
| $t_{1(13)}$                   | 0.0009  | 0.0022 | 89.74117 | $t_{2(13)}$                   | 0.0005  | 0.0019 | 100.5394 |
| $t_{1(14)}$                   | 0.0002  | 0.0018 | 109.0494 | $t_{2(14)}$                   | 0.0015  | 0.0017 | 112.8826 |
| $t_{1(15)}$                   | -0.0003 | 0.0019 | 100.0373 | $t_{2(15)}$                   | 4.6749  | 0.0018 | 108.1691 |
| $t_{1(16)}$                   | -0.0003 | 0.0019 | 100.0186 | $t_{2(16)}$                   | 4.5844  | 0.0018 | 108.1562 |
| $t_{1(17)}$                   | -0.0003 | 0.0019 | 99.98126 | $t_{2(17)}$                   | 4.4057  | 0.0018 | 108.1304 |

Table 3 displays the outcomes of biases, mean squared errors (MSEs), and percentage relative efficiency (PREs) for the existing and proposed estimators when  $p = 0.6$ . The findings indicate

that, with the exception of  $t_{1(4)}, t_{1(7)}, t_{1(10)}, t_{1(12)}, t_{1(13)}, t_{1(14)}, t_{1(17)}$  and  $t_{2(12)}$ , all the proposed estimators exhibit lower MSEs and higher PREs in comparison to the existing estimator considered in this investigation. Consequently, the proposed estimators in this context are more efficient than their competitor in the study and are likely to yield superior estimates of the population proportion, particularly in scenarios involving non-response.

**Table 4: Biases, MSEs and PREs of Estimators  $p_{ad(sys)}, t_{1(i)}$  and  $t_{2(i)}$  when  $P=0.7$**

| ESTIMATOR                     | BIAS    | MSE    | PRE      | ESTIMATOR                     | BIAS    | MSE    | PRE      |
|-------------------------------|---------|--------|----------|-------------------------------|---------|--------|----------|
| $p_{ad(sys)}$<br>Azzem(2021)  | -0.0006 | 0.0016 | 100      | $p_{ad(sys)}$<br>Azzem(2021)  | -0.0006 | 0.0016 | 100      |
| Proposed Estimator $t_{1(i)}$ |         |        |          | Proposed Estimator $t_{2(i)}$ |         |        |          |
| $t_{1(1)}$                    | -6.3700 | 0.0016 | 100.7015 | $t_{2(1)}$                    | -0.0002 | 0.0016 | 98.5387  |
| $t_{1(2)}$                    | -0.0005 | 0.0016 | 100.5071 | $t_{2(2)}$                    | -0.0011 | 0.0016 | 100.5984 |
| $t_{1(3)}$                    | -0.0005 | 0.0016 | 100.3393 | $t_{2(3)}$                    | -0.0012 | 0.0016 | 100.8379 |
| $t_{1(4)}$                    | -0.0004 | 0.0016 | 99.2362  | $t_{2(4)}$                    | -0.0014 | 0.0016 | 101.3943 |
| $t_{1(5)}$                    | 3.8626  | 0.0016 | 100.6312 | $t_{2(5)}$                    | -7.8749 | 0.0016 | 98.1435  |
| $t_{1(6)}$                    | -0.0006 | 0.0016 | 100.3393 | $t_{2(6)}$                    | -0.0012 | 0.0016 | 100.8379 |
| $t_{1(7)}$                    | -0.0004 | 0.0016 | 99.2362  | $t_{2(7)}$                    | -0.0014 | 0.0016 | 101.3943 |
| $t_{1(8)}$                    | 3.8626  | 0.0016 | 100.6312 | $t_{2(8)}$                    | -7.8749 | 0.0016 | 98.1435  |
| $t_{1(9)}$                    | -0.0005 | 0.0016 | 100.6604 | $t_{2(9)}$                    | -0.0009 | 0.0016 | 100.2353 |
| $t_{1(10)}$                   | 0.0002  | 0.0017 | 96.95981 | $t_{2(10)}$                   | -0.0012 | 0.0016 | 101.228  |
| $t_{1(11)}$                   | -1.4861 | 0.0016 | 100.6703 | $t_{2(11)}$                   | -0.0002 | 0.0016 | 98.3497  |
| $t_{1(12)}$                   | 0.0339  | 0.0914 | 1.7666   | $t_{2(12)}$                   | 0.0329  | 0.0915 | 1.7642   |
| $t_{1(13)}$                   | 0.0003  | 0.0017 | 96.5466  | $t_{2(13)}$                   | -0.0011 | 0.0016 | 101.1309 |
| $t_{1(14)}$                   | -0.0001 | 0.0016 | 100.7245 | $t_{2(14)}$                   | -0.0003 | 0.0016 | 98.7024  |
| $t_{1(15)}$                   | -0.0006 | 0.0016 | 99.9608  | $t_{2(15)}$                   | -0.0014 | 0.0016 | 101.15   |
| $t_{1(16)}$                   | -0.0006 | 0.0016 | 99.9803  | $t_{2(16)}$                   | -0.0014 | 0.0016 | 101.1385 |
| $t_{1(17)}$                   | -0.0006 | 0.0016 | 100.0186 | $t_{2(17)}$                   | -0.0013 | 0.0016 | 101.1148 |

Table 4 displays the outcomes of biases, mean squared errors (MSEs), and percentage relative efficiency (PREs) for the existing and proposed estimators when  $p = 0.7$ . The findings indicate that, with the exception of  $t_{1(4)}, t_{1(7)}, t_{1(10)}, t_{1(12)}, t_{1(13)}, t_{1(14)}, t_{1(15)}, t_{1(16)}, t_{2(1)}, t_{2(5)}, t_{2(8)}, t_{2(11)}, t_{2(12)}$  and  $t_{2(14)}$ , all the proposed estimators exhibit lower MSEs and higher PREs in comparison to the

existing estimator considered in this investigation. Consequently, the proposed estimators in this context are more efficient than their competitor in the study and are likely to yield superiorestimates of the population proportion, particularly in scenarios involving non-response.

**Table 5: Biases, MSEs and PREs of Estimators  $p_{ad(sys)}$ ,  $t_{1(i)}$  and  $t_{2(i)}$  when P=0.8**

| ESTIMATO<br>R                 | BIAS    | MSE    | PRE      | ESTIMATO<br>R                 | BIAS    | MSE    | PRE      |
|-------------------------------|---------|--------|----------|-------------------------------|---------|--------|----------|
| $P_{ad(sys)}$<br>Azzem(2021)  | -0.0035 | 0.0039 | 100      | $P_{ad(sys)}$<br>Azzem(2021)  | -0.0035 | 0.0039 | 100      |
| Proposed Estimator $t_{1(i)}$ |         |        |          | Proposed Estimator $t_{2(i)}$ |         |        |          |
| $t_{1(1)}$                    | -0.0036 | 0.0038 | 104.3494 | $t_{2(1)}$                    | -0.0039 | 0.0037 | 106.4786 |
| $t_{1(2)}$                    | -0.0039 | 0.0039 | 102.3417 | $t_{2(2)}$                    | -0.0046 | 0.0037 | 105.6342 |
| $t_{1(3)}$                    | -0.0039 | 0.0039 | 101.8392 | $t_{2(3)}$                    | -0.0046 | 0.0037 | 105.3624 |
| $t_{1(4)}$                    | -0.0037 | 0.0039 | 99.3582  | $t_{2(4)}$                    | -0.0047 | 0.0038 | 103.8639 |
| $t_{1(5)}$                    | -0.0036 | 0.0038 | 104.5048 | $t_{2(5)}$                    | -0.0039 | 0.0037 | 106.5135 |
| $t_{1(6)}$                    | -0.0039 | 0.0039 | 101.8392 | $t_{2(6)}$                    | -0.0046 | 0.0037 | 105.3624 |
| $t_{1(7)}$                    | -0.0037 | 0.0039 | 99.3582  | $t_{2(7)}$                    | -0.0047 | 0.0038 | 103.8639 |
| $t_{1(8)}$                    | -0.0036 | 0.0038 | 104.5048 | $t_{2(8)}$                    | -0.0039 | 0.0037 | 106.5135 |
| $t_{1(9)}$                    | -0.0039 | 0.0038 | 102.9154 | $t_{2(9)}$                    | -0.0045 | 0.0037 | 105.9235 |
| $t_{1(10)}$                   | -0.0029 | 0.0041 | 95.3261  | $t_{2(10)}$                   | -0.0044 | 0.0039 | 101.1446 |
| $t_{1(11)}$                   | -0.0036 | 0.0039 | 104.4269 | $t_{2(11)}$                   | -0.0039 | 0.0037 | 106.497  |
| $t_{1(12)}$                   | 0.0139  | 0.0503 | 7.8559   | $t_{2(12)}$                   | 0.0129  | 0.0504 | 7.8408   |
| $t_{1(13)}$                   | -0.0028 | 0.0042 | 94.7915  | $t_{2(13)}$                   | -0.0043 | 0.0039 | 100.7688 |
| $t_{1(14)}$                   | -0.0037 | 0.0038 | 104.2761 | $t_{2(14)}$                   | -0.0040 | 0.0037 | 106.4596 |
| $t_{1(15)}$                   | -0.0038 | 0.0039 | 100.9115 | $t_{2(15)}$                   | -0.0047 | 0.0038 | 104.8271 |
| $t_{1(16)}$                   | -0.0038 | 0.0039 | 100.9476 | $t_{2(16)}$                   | -0.0047 | 0.0038 | 104.8486 |
| $t_{1(17)}$                   | -0.0038 | 0.0039 | 101.0193 | $t_{2(17)}$                   | -0.0047 | 0.0038 | 104.8912 |

Table 5 displays the outcomes of biases, mean squared errors (MSEs), and percentage relative efficiency (PREs) for the existing and proposed estimators when  $p = 0.8$ . The findings indicate that, with the exception of  $t_{1(4)}$ ,  $t_{1(7)}$ ,  $t_{1(10)}$ ,  $t_{1(12)}$ ,  $t_{1(13)}$  and  $t_{2(12)}$ , all the proposed estimators exhibit lower MSEs and higher PREs in comparison to the existing estimator considered in this

investigation. Consequently, the proposed estimators in this context are more efficient than their competitor in the study and are likely to yield superioestimates of the population proportion, particularly in scenarios involving non-response.

**Table 6: Biases, MSEs and PREs of Estimators  $p_{ad(sys)}$ ,  $t_{1(i)}$  and  $t_{2(i)}$  when P=0.9**

| ESTIMATOR                     | BIAS    | MSE    | PRE      | ESTIMATOR                     | BIAS    | MSE    | PRE      |
|-------------------------------|---------|--------|----------|-------------------------------|---------|--------|----------|
| $p_{ad(sys)}$<br>Azzem(2021)  | 0.0016  | 0.0015 | 100      | $p_{ad(sys)}$<br>Azzem(2021)  | 0.0016  | 0.0015 | 100      |
| Proposed Estimator $t_{1(i)}$ |         |        |          | Proposed Estimator $t_{2(i)}$ |         |        |          |
| $t_{1(1)}$                    | -0.0015 | 0.0013 | 109.5815 | $t_{2(1)}$                    | -0.0004 | 0.0012 | 116.7586 |
| $t_{1(2)}$                    | -0.0016 | 0.0014 | 103.3322 | $t_{2(2)}$                    | -0.0008 | 0.0013 | 112.109  |
| $t_{1(3)}$                    | -0.0016 | 0.0014 | 101.9937 | $t_{2(3)}$                    | -0.0008 | 0.0013 | 111.0008 |
| $t_{1(4)}$                    | -0.0015 | 0.0015 | 6.57065  | $t_{2(4)}$                    | -0.0008 | 0.0014 | 106.2582 |
| $t_{1(5)}$                    | -0.0015 | 0.0013 | 109.1763 | $t_{2(5)}$                    | -0.0004 | 0.0013 | 116.4925 |
| $t_{1(6)}$                    | -0.0016 | 0.0014 | 101.9937 | $t_{2(6)}$                    | -0.0008 | 0.0013 | 111.0008 |
| $t_{1(7)}$                    | -0.0015 | 0.0015 | 96.57065 | $t_{2(7)}$                    | -0.0008 | 0.0014 | 106.2582 |
| $t_{1(8)}$                    | -0.0015 | 0.0013 | 109.1763 | $t_{2(8)}$                    | -0.0004 | 0.0013 | 116.4925 |
| $t_{1(9)}$                    | -0.0016 | 0.0014 | 104.9822 | $t_{2(9)}$                    | -0.0007 | 0.0013 | 113.4304 |
| $t_{1(10)}$                   | -0.0011 | 0.0016 | 89.16261 | $t_{2(10)}$                   | -0.0007 | 0.0015 | 99.34166 |
| $t_{1(11)}$                   | -0.0015 | 0.0013 | 109.3771 | $t_{2(11)}$                   | -0.0004 | 0.0013 | 116.6252 |
| $t_{1(12)}$                   | -0.1184 | 0.0332 | 4.391399 | $t_{2(12)}$                   | -0.1179 | 0.0331 | 4.399962 |
| $t_{1(13)}$                   | -0.0012 | 0.0016 | 89.62094 | $t_{2(13)}$                   | -0.0007 | 0.0015 | 99.77996 |
| $t_{1(14)}$                   | -0.0015 | 0.0013 | 109.7969 | $t_{2(14)}$                   | -0.0004 | 0.0012 | 116.8973 |
| $t_{1(15)}$                   | -0.0016 | 0.0015 | 100.1201 | $t_{2(15)}$                   | -0.0008 | 0.0013 | 109.4034 |
| $t_{1(16)}$                   | -0.0016 | 0.0015 | 100.0597 | $t_{2(16)}$                   | -0.0008 | 0.0013 | 109.3511 |
| $t_{1(17)}$                   | -0.0016 | 0.0015 | 99.93893 | $t_{2(17)}$                   | -0.0008 | 0.0013 | 109.2464 |

Table 6 displays the outcomes of biases, mean squared errors (MSEs), and percentage relative efficiency (PREs) for the existing and proposed estimators when  $p = 0.9$ . The findings indicate that, with the exception of  $t_{1(4)}$ ,  $t_{1(7)}$ ,  $t_{1(10)}$ ,  $t_{1(12)}$ ,  $t_{1(13)}$ ,  $t_{1(17)}$ ,  $t_{2(10)}$ ,  $t_{2(12)}$  and  $t_{2(13)}$ , all the proposed estimators exhibit lower MSEs and higher PREs in comparison to the existing estimator considered

in this investigation. Consequently, the proposed estimators in this context are more efficient than their competitor in the study and are likely to yield superiorestimates of the population proportion, particularly in scenarios involving non-response.

## 5.0 Conclusion

In Surveys involving medical fields, social surveys, household surveys it is common to experience the problem of non-response for a particular unit or units in the population. This non-sampling error may creep and lead to erroneous results due to missing values in the data set. In this study, we proposed two imputation class of estimators for estimating population means in situations when the study variables are characterized with non-responses or missing values. From the results of the empirical study, the proposed estimators performed more efficiently than other estimators considered in the study. Therefore, the proposed procedure for estimating non-responses or missing values in surveys or experiments, holding to its advantages over other considered procedures, is recommended for use for academic or practical purposes.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

## References

- Al-Omari, A.I., Bouza, C.N. and Herrera, C. (2013) "Imputation Methods of Missing Data for Estimating the Populations Mean Using Simple Random Sampling with Known Correlation Coefficient" *Qual. Quant.* 47(1):353–365
- Audu, A., Singh, R. and Supriya, K. (2021) "New Regression-Type Compromised Imputation Class of Estimators with Known Parameters of Auxiliary Variable" *Communications in Statistics -Simulation and Computation*, 52(10), 4789-4801.
- Audu, A. & Singh, R. V. K. (2021) "Exponential-Type Regression Compromised Imputation Class of Estimators" *Journal of Statistics and Management Systems*, 24(6), 1253-1266.
- Azeem, M. (2021) "On Estimation of Population Proportion in Diagonal Systematic Sampling" *Life Cycle Reliability and Safety Engineering*, 10, 249-254.
- Bhushan, S. and Pandey, A.P. (2016) "Optimal imputation of missing data for estimation of Population mean." *J. Stat. Manag. Syst.* 19(6):755–769
- Diana, G. and Francesco Perri, P. (2010) "Improved Estimators of the Population Mean for Missing Data" *Communication in Statistics- Theory Methods* 39(18):3245–3251
- Gira, A.A. (2015) Estimation of population mean with a New Imputation Methods. *Appl. Math. Sci.* 9(34):1663–1672
- Wang, H. and Wang, S. (2006) "Visualization of the Critical Patterns of Missing Values in Classification data" *International Conference on Advances in Visual Information Systems, Springer*, pp. 267–274
- Hansen, M.H. and Hurwitz, W.N. (1946) "The Problem of Non-Response in Sample Surveys." *Journal of the American Statistical Association*, 41, 517-529.
- Kadilar, C. and Cingi, H. (2008) "Estimators for the Population Mean in the Case of Missing Data" *Communications in Statistics- Theory Methods* 37(14):2226–2236
- Khan, Z., Shabbir, J. and Gupta, S. (2017) "New Sampling Design for Systematic Sampling" *Communications in Statistics – Theory and Methods*, 42(18), 3359-3370.
- Prasad, S. (2017) "An exponential imputation in the case of missing data." *J. Stat. Manag Syst.* 20(6):1127–1140

- Singh, G.N; Maurya, S; Khetan, M; Kadilar, C. (2016) "Some Imputation Methods for Missing Data in Sample Surveys" *Hacet J. Math. Stat.* 45(6):1865–1880
- Singh, K., Singh, A., Priyanka and Singh, K.V. (2014) "Exponential-Type Compromised Imputation in Survey Sampling," *Journal of Statistics Application and probability*: Vol. 3:Iss. 2, Article 11.
- Singh, S. (2009) "A New Method of Imputation in Survey Sampling" *Statistics* 43(5):499–511
- Singh, S. and Horn, S. (2000) "Compromised Imputation in Survey Sampling" *Metrika*, 51, 266–276.
- Singh, S. and Deo, B. (2003) "Imputation by Power Transformation" *Stat Pap* 44(4):555–579
- Toutenburg, H. and Srivastava, V.K. (2008) "Amputation versus Imputation of Missing Values Through Ratio Method in Sample Surveys" *Stat. Pap.* 49(2):237–247