Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 10(1): 1061–1084.
DOI: 10.3934/math.2025051
Received: 22 September 2024
Revised: 19 December 2024
Accepted: 02 January 2025
Published: 17 January 2025

Research article

An improved family of unbiased ratio estimators for a population distribution function

Sohail Ahmad[1], **Moiz Qureshi**[2,3], **Hasnain Iftikhar**[3,*] , **Paulo Canas Rodrigues**[4] **and Mohd Ziaur Rehman**[5]

[1] School of Mathematics and Statistics, Central South University, Changsha 410083, China

[2] Government Degree College, Tando Jam, Hyderabad 70060, Pakistan

[3] Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan

[4] Department of Statistics, Federal University of Bahia, Salvador 40170-110, Brazil

[5] Department of Finance, College of Business Administration, King Saud University, P.O. Box 71115, Riyadh 11587, Saudi Arabia

* **Correspondence:** Email: hasnain@stat.qau.edu.pk.

Abstract: This study discusses a novel family of unbiased ratio estimators using the Hartley-Ross (HR) method. The estimators are designed to estimate the population distribution function (PDF) in the context of simple random sampling with non-response. To assess their performance, expressions for variance are obtained up to the initial (first) approximation order. The efficiency of the proposed estimators is evaluated analytically and numerically compared to existing estimators. In addition, the accuracy of the estimators is assessed using four real-world datasets and a simulation analysis. The proposed estimator demonstrates exceptional performance for the distribution function under simple random sampling, achieving percentage relative efficiencies of 272.052, 301.279, 214.1214, and 280.9528 across four distinct populations, significantly outperforming existing estimators. For the distribution function under non-response using different weights, the proposed estimator exhibits remarkable efficiency, with percentage relative efficiencies of w_1=339.7875, w_2=334.6623, w_3=337.7393 in Population 1, w_1=257.0119, w_2=274.7351, w_3=316.0341 in Population 2, w_1=231.8627, w_2=223.0608, w_3=219.9059 in Population 3, and w_1=261.3122, w_2=242.7319, w_3=240.0694 in Population 4, validating its robustness and superiority.

Keywords: Hartley-Ross; simple random sampling; distribution function; non-response; sampling schemes
Mathematics Subject Classification: 68T07, 03H10, 68T09, 37N40, 62P20, 91G15, 91G30

1. Introduction

Auxiliary information in the sampling technique is used to improve estimator accuracy at the discovery level, the design stage, or both stages [1, 2]. If the research and the auxiliary variables are positively correlated, the ratio estimation method is used; otherwise, the product estimation method is used. Different forms of impartial estimators for population mean were used by [3–5]. [6] used the coefficient of variance in conjunction with the population mean of the auxiliary variate. [7] used many different independent population mean estimators. In simple random sampling (SRS), the ratio cum-product estimator of the finite population mean was suggested by [8] and [9, 10] using the same data.

The proportionate size of sample variable Y, (i.e., the number of population units with values less than or equal to the threshold value t_y), is also of concern to statisticians. For example, one may be curious to seek a ratio between the percentage of people in the country who have lung cancer and the proportion who have another disease. The population distribution function (PDF) plays a significant role in this situation. Depending on auxiliary details, [11] and [12] investigated specific methods and characteristics for estimating the finite PDF under quantiles. To assess the PDF using survey data, [13] proposed a prediction and classical system. Several estimators of the finite population distribution function are derived by taking into account sampling with probability proportional to size or aggregate size [14–16]. The model-based and design-based methods of sampling inference are used to build rival estimators of the distribution function of a finite population. [17–19]. In [20] and [21] two novel estimators for modified FDP were proposed. However, two novel families of estimators based on SRS were presented by [22] to estimate the FPD function under the presence of non response. The authors of [23–25] work introduced advanced estimators using exponential transformations to improve the precision of ratio and product-type estimations. Studies [26–28] present a regression-based framework for estimating the finite population mean under conditions of non response. In [29, 30], the authors found the characteristics of various estimators using a generalized exponential-type estimator. [31, 32] analyzed the properties of various estimators using a generalized exponential-type estimator. [33] proposed estimators for population mean under non response outperform existing estimators in terms of bias and mean square error, demonstrating improved relative efficiency. [34] introduced the proposed general class of estimators for population distribution functions that demonstrate superior performance, validated through theoretical comparisons, empirical data, and simulation studies. [35] proposed exponential estimators for population mean under non response, showing superior efficiency over existing methods through theoretical and numerical validation.

This study discusses a novel family of unbiased ratio estimators using the Hartley-Ross (HR) method. The estimators are designed to estimate the PDF in the context of SRS with non response. To assess their performance, expressions for variance are obtained up to the initial (first) approximation order. The proposed estimators' efficiency is evaluated analytically and numerically compared to existing estimators. Additionally, the accuracy of the estimators is assessed using four real-world datasets and a simulation analysis.

The rest of the work is organized as follows. Section 2 revises the existing estimators. Section 3 proposes a new, improved class of estimators under non-response. Sections 4–7 present the theoretical comparison, data comparison, conditional values, and simulation study, respectively. Section 8 gives an overview of the existing estimators under non response. Section 9 discusses the estimation of the PDF under non response. Section 10 proposes a class of estimators under non response. Sections 11–13 present the theoretical comparison, data comparison, and conditional values, respectively. The paper concludes in Section 14 with the conclusion.

2. Existing estimators

2.1. Notations and symbols

Let us consider finite population $\Omega = \{1, 2, \dots, i, \dots, N\}$ having N independent and identifiable units. Here, Y is the study variable X is the auxiliary variable, R is rank of X, PV is population variance, PC is population covariance, PCn is population correlation, PCV is population coefficient of variance, SDF is sample distribution function, and PDF is population distribution function.

$$Y_f = I(Y_i \leq t_y) = \begin{cases} 1 & \text{if } Y_i \leq t_y \\ 0 & \text{otherwise} \end{cases}$$

$$X_f = I(X_i \leq t_y) = \begin{cases} 1 & \text{if } X_i \leq t_y \\ 0 & \text{otherwise} \end{cases}$$

$$Z_f = I(Z_i \leq t_y) = \begin{cases} 1 & \text{if } Z_i \leq t_y \\ 0 & \text{otherwise} \end{cases}$$

The PDF of Z is $P(t_z) = \frac{1}{N} \sum_{i=1}^{N} Z_f$, where $Z_i = \frac{Y_i}{X_i}$.

Let $\hat{P}(t_y) = \frac{1}{n} \sum_{i=1}^{n} Y_f$, $\hat{P}(t_x) = \frac{1}{n} \sum_{i=1}^{n} X_f$, and $\hat{P}(t_z) = \frac{1}{n} \sum_{i=1}^{n} Z_f$ be the SDFs correspondingly. Let $P(t_y)$, $P(t_x)$ and $P(t_z)$ be the PDFs respectively.

Let $S^2_{P(t_y)} = \sum_{i=1}^{N} \frac{\{Y_f - P(t_y)\}^2}{N-1}$, $S^2_{P(t_x)} = \sum_{i=1}^{N} \frac{\{X_f - P(t_x)\}^2}{N-1}$ and $S^2_{P(t_z)} = \sum_{i=1}^{N} \frac{\{Z_f - P(t_z)\}^2}{N-1}$ be the PV of Y_f, X_f, and Z_f.

Let $C_{P(t_y)} = \frac{S_{P(t_y)}}{P(t_y)}$, $C_{P(t_x)} = \frac{S_{P(t_x)}}{P(t_x)}$ and $C_{P(t_z)} = \frac{S_{P(t_z)}}{P(t_z)}$ be the PCV of Y_f, X_f, and Z_f.

Let $S_{P(t_y)P(t_x)} = \sum_{i=1}^{N} \frac{\left[\left(Y_f - P(t_y)\right)\left(X_f - P(t_x)\right)\right]}{N-1}$ be the PC between Y_f and X_f.

Let $\rho_{P(t_y)P(t_x)} = \frac{S_{P(t_y)P(t_x)}}{S_{P(t_y)} S_{P(t_x)}}$ be the PCV between Y_f and X_f.

The following error terms are used below. Let $e_0 = \frac{\hat{P}(t_y) - P(t_y)}{P(t_y)}$, $e_1 = \frac{\hat{P}(t_x) - P(t_x)}{P(t_x)}$ and $e_2 = \frac{\hat{P}(t_z) - P(t_z)}{P(t_z)}$, where, $E(e_i) = 0$, $(i = 0, 1, 2)$, $E(e_0^2) = \lambda C^2_{P(t_y)}$, $E(e_1^2) = \lambda C^2_{P(t_x)}$, $E(e_2^2) = \lambda C^2_{P(t_z)}$, $E(e_0 e_1) = \lambda \rho_{P(t_y)P(t_x)} C_{P(t_y)} C_{P(t_x)}$, $E(e_0 e_2) = \lambda \rho_{P(t_y)P(t_z)} C_{P(t_y)} C_{P(t_z)}$, and $E(e_1 e_2) = \lambda \rho_{P(t_x)P(t_z)} C_{P(t_x)} C_{P(t_z)}$, where $\lambda = (\frac{1}{n} - \frac{1}{N})$.

(1) The usual sample distribution function HP is given by:

$$\hat{P}(t_{y_0}) = \hat{P}(t_y). \tag{2.1}$$

The variance of $\hat{P}(t_{y_0})$ is given by:

$$\text{Var}(\hat{P}(t_{y_0})) \cong \lambda P^2(t_y) C^2_{P(t_y)}. \tag{2.2}$$

(2) The ratio estimator in the form of distribution function HP is given by:

$$\hat{P}(t_{y_R}) = \hat{P}(t_y)\Big(\frac{P(t_x)}{\hat{P}(t_x)}\Big). \tag{2.3}$$

Following equation provides the bias and mean square error of $\hat{P}(t_{y_R})$ as

$$\text{Bias}(\hat{P}(t_{y_R})) \cong \lambda P(t_y)\Big[C^2_{P(t_x)} - \rho_{P(t_y)P(t_x)} C_{P(t_y)} C_{P(t_x)}\Big] \tag{2.4}$$

and

$$\text{MSE}(\hat{P}(t_{y_R})) \cong \lambda P^2(t_y)\Big[C^2_{P(t_y)} + C^2_{P(t_x)} - 2\rho_{P(t_y)P(t_x)} C_{P(t_y)} C_{P(t_x)}\Big]. \tag{2.5}$$

(3) Using HP [24], we obtain:

$$\hat{P}(t_{y_{\exp(R)}}) = \hat{P}(t_y)\exp\Big(\frac{P(t_x) - \hat{P}(t_x)}{P(t_x) + \hat{P}(t_x)}\Big). \tag{2.6}$$

The following equation provides the bias and MSE of $\hat{P}(t_{y_{\exp(R)}})$ as:

$$\text{Bias}(\hat{P}(t_{y_{\exp(R)}})) \cong \lambda P(t_y)\Big[\frac{3}{8}C^2_{P(t_x)} - \frac{1}{2}\rho_{P(t_y)P(t_x)} C_{P(t_y)} C_{P(t_x)}\Big] \tag{2.7}$$

and

$$\text{MSE}(\hat{P}(t_{y_{\exp(R)}})) \cong \lambda P^2(t_y)\Big[C^2_{P(t_y)} + \frac{1}{4}C^2_{P(t_x)} - \rho_{P(t_y)P(t_x)} C_{P(t_y)} C_{P(t_x)}\Big]. \tag{2.8}$$

(4) The ratio type estimator [25] is given as:

$$\hat{P}(t_{y_{\text{SK}(R)}}) = \hat{P}(t_y)\Big(\frac{P(t_x)}{\hat{P}(t_x)}\Big)^2. \tag{2.9}$$

Bias and MSE of $\hat{P}(t_{y_{\text{SK}(R)}})$ is given as

$$\text{Bias}\Big(\hat{P}(t_{y_{\text{SK}(R)}})\Big) \cong \lambda P(t_y)\Big[3C^2_{P(t_x)} - 2\rho_{P(t_y)P(t_x)} C_{P(t_y)} C_{P(t_x)}\Big] \tag{2.10}$$

and

$$\text{MSE}\Big(\hat{P}(t_{y_{\text{SK}(R)}})\Big) \cong \lambda P^2(t_y)\Big[C^2_{P(t_y)} + 4C^2_{P(t_x)} - 4\rho_{P(t_y)P(t_x)} C_{P(t_y)} C_{P(t_x)}\Big]. \tag{2.11}$$

3. New improved class of estimators

We have proposed a new family of HR type unbiased ratio estimators for estimating the PDF under SRS and developed the HR type estimators for the population mean along side their variance equations. Thus, the HR type unbiased estimator for estimating DF is given as:

$$\hat{P}(t_z) = \frac{1}{n}\sum_{i=1}^{n} Z_f,$$

where $Z_i = \frac{Y_i}{X_i}$.

Now, the HR estimator:

$$\hat{P}(t_{y_{HR}}) = P(t_x)\hat{P}(t_z). \tag{3.1}$$

Taking expectations on both sides:

$$E\Big(\hat{P}(t_{y_{HR}})\Big) = P(t_x)E\Big(\hat{P}(t_z)\Big),$$

$$E\Big(\hat{P}(t_{y_{HR}})\Big) = P(t_x)E\Big(\frac{1}{n}\sum_{i=1}^{n} Z_f\Big)$$

or

$$E\Big(\hat{P}(t_{y_{HR}})\Big) = P(t_x)\Big(\frac{1}{N}\sum_{i=1}^{N} Z_f\Big). \tag{3.2}$$

The following equation provides the bias of $\hat{P}(t_{y_{HR}})$, as:

$$\text{Bias}\Big(\hat{P}(t_{y_{HR}})\Big) = E(\hat{P}(t_{y_{HR}})) - P(t_y),$$

$$\text{Bias}\Big(\hat{P}(t_{y_{HR}})\Big) = P(t_x)\Big(\frac{1}{N}\sum_{i=1}^{N} Z_f\Big) - (\frac{1}{N}\sum_{i=1}^{N} Y_f)$$

or

$$\widehat{\text{Bias}}\Big(\hat{P}(t_{y_{HR}})\Big) = -\Big(\frac{N-1}{N}\Big)\Big(\frac{n}{n-1}\Big)\Big(\hat{P}(t_y) - \hat{P}(t_z)\hat{P}(t_x)\Big). \tag{3.3}$$

Now adjust the bias estimator as follows:

$$\hat{P}(t_{y_{HR}}) = P(t_x)\hat{P}(t_z) - \widehat{\text{Bias}}\Big(P(t_{\bar{y}_{HR}})\Big),$$

$$\hat{P}(t_{y_{HR}}) = P(t_x)\hat{P}(t_z) + \Big(\frac{N-1}{N}\Big)\Big(\frac{n}{n-1}\Big)\Big(\hat{P}(t_y) - \hat{P}(t_z)\hat{P}(t_x)\Big). \tag{3.4}$$

Under the assumption $\Big(\frac{N-1}{N}\Big)\Big(\frac{n}{n-1}\Big) \approx 1$.

$$\hat{P}(t_{y_{HR}}) = P(t_x)\hat{P}(t_z) + \Big(\hat{P}(t_y) - \hat{P}(t_z)\hat{P}(t_x)\Big) \tag{3.5}$$

or

$$\hat{P}(t_{y_{HR}}) = \hat{P}(t_y) + \hat{P}(t_z)\left[P(t_x) - \hat{P}(t_x)\right].$$

Regarding to errors,

$$\hat{P}(t_{y_{HR}}) = P(t_y)(1+e_0) + P(t_z)(1+e_2)\Big(-e_1 P(t_x)\Big). \tag{3.6}$$

So the estimator becomes:

$$E\Big(\hat{P}(t_{y_{HR}})\Big) = P(t_y). \tag{3.7}$$

Now, the variance of the proposed estimator is given by:

$$\text{Var}(\hat{P}(t_{y_{HR}})) \cong \lambda\Big[S^2_{P(t_y)} + P^2(t_z)S^2_{P(t_x)} - 2P(t_z)S_{P(t_y)P(t_x)}\Big]. \tag{3.8}$$

4. Theoretical comparison

In this section, we compare the variance of the proposed HR type unbiased ratio estimators in Eq (3.8) with the MSE of the existing estimators in (Section 2); we have the following conditions:

(i) By Eqs (2.2) and (3.8),
$\text{Var}(\hat{P}(t_{y_0})) - \text{Var}(\hat{P}(t_{y_{HR}})) > 0$ if

$$[-P^2(t_z)S^2_{P(t_x)} + 2P(t_z)S_{P(t_y)P(t_x)}] > 0.$$

(ii) By Eqs (2.5) and (3.8),
$\text{MSE}(\hat{P}(t_{y_R})) - \text{Var}(\hat{P}(t_{y_{HR}})) > 0$ if

$$[P^2(t_y)C^2_{P(t_x)} - 2P^2(t_y)\rho_{P(t_y)P(t_x)}C_{P(t_y)}C_{P(t_x)} - P^2(t_z)S^2_{P(t_x)} + 2P(t_z)S_{P(t_y)P(t_x)}] > 0.$$

(iii) By Eqs (2.8) and (3.8),
$\text{MSE}(\hat{P}(t_{y_{\exp(R)}})) - \text{Var}(\hat{P}(t_{y_{HR}})) > 0$ if

$$[\frac{1}{4}P^2(t_y)C^2_{P(t_x)} - 2P^2(t_y)\rho_{P(t_y)P(t_x)}C_{P(t_y)}C_{P(t_x)} - P^2(t_z)S^2_{P(t_x)} + 2P(t_z)S_{P(t_y)P(t_x)}] > 0.$$

(iv) By Eqs (2.11) and (3.8),
$\text{MSE}(\hat{P}(t_{y_{SK(R)}})) - \text{Var}(\hat{P}(t_{y_{HR}})) > 0$ if

$$[4P^2(t_y)C^2_{P(t_x)} - 4P^2(t_y)\rho_{P(t_y)P(t_x)}C_{P(t_y)}C_{P(t_x)} - P^2(t_z)S^2_{P(t_x)} + 2P(t_z)S_{P(t_y)P(t_x)}] > 0.$$

If requirements (i) through (iv) are satisfied, the suggested estimator $\hat{P}(t_{y_{HR}})$ performs better than other existing estimators.

5. Data comparison

Four natural populations are utilized to assess the effectiveness of the suggested estimator. $\hat{P}(t_{y_{HR}})$ in relation to other existing ones. The following are the summary statistics of the population.

Pop I. Source: [26]

Using $t_y = \bar{Y}$, $t_x = \bar{X}$ and $t_z = \bar{Z}$:

Let Y represent the sum of apple production in 1999 and X represent the number of apple trees in 1999.

$N = 106$, $n = 30$, $P(t_y) = 0.8301887$, $P(t_x) = 0.7641509$, $C_{P(t_y)} = 0.4544156$, $C_{P(t_x)} = 0.5581948$, $P(t_z) = 0.5660377$, $S_{P(t_y)} = 0.3772507$, $S_{P(t_x)} = 0.4265451$, $S_{P(t_y)P(t_x)} = 0.1309973$, $\rho_{P(t_y)P(t_x)} = 0.8140806$, $\bar{Y} = 1536.774$, $\bar{X} = 24375.59$ and $\bar{Z} = 0.03927598$.

Let Y_f=1 for $Y_i \le 1536.774$, Y_f=0 for all $Y_i > 1536.774$,
X_f=1 for $X_i \le 24375.59$, X_f=0 for all $X_i > 24375.59$ and
Z_f=1 for $Z_i \le 0.03927598$, Z_f=0 for all $Z_i > 0.03927598$.

Pop II. Source: [27]

Using $t_y = \bar{Y}$, $t_x = \bar{X}$ and $t_z = \bar{Z}$:

Assume that Y is the total amount of fish that recreational marine fishermen caught in 1995 and X is the total amount of fish that recreational marine fishermen caught in 1992.

$N = 69$, $n = 25$, $P(t_y) = 0.7246377$, $P(t_x) = 0.7101449$, $C_{P(t_y)} = 0.6209575$, $C_{P(t_x)} = 0.643557$, $P(t_z) = 0.7246377$, $S_{P(t_y)} = 0.4499692$, $S_{P(t_x)} = 0.4570188$, $S_{P(t_y)P(t_x)} = 0.1690111$, $\rho_{P(t_y)P(t_x)} = 0.821861$, $\bar{Y} = 4514.899$, $\bar{X} = 4230.174$ and $\bar{Z} = 1.251484$.

Let Y_f=1 for $Y_i \leq 4514.899$, Y_f=0 for all $Y_i > 4514.899$,
X_f=1 for $X_i \leq 4230.174$, X_f=0 for all $X_i > 4230.174$ and
Z_f=1 for $Z_i \leq 1.251484$, Z_f=0 for all $Z_i > 1.251484$.

Pop III. Source: [26]

Using $t_y = \bar{Y}$, $t_x = \bar{X}$ and $t_z = \bar{Z}$:

X denotes the quantity of apple trees in 1999, and Y represents the total amount of apples produced in that year.

$N = 171$, $n = 95$, $P(t_y) = 0.9005848$, $P(t_x) = 0.8479532$, $C_{P(t_y)} = 0.3332251$, $C_{P(t_x)} = 0.4246941$, $P(t_z) = 0.5964912$,
$S_{P(t_y)} = 0.3000975$, $S_{P(t_x)} = 0.3601208$, $S_{P(t_y)P(t_x)} = 0.07891297$, $\rho_{P(t_y)P(t_x)} = 0.7301934$, $\bar{Y} = 5588.012$, $\bar{X} = 74364.68$ and $\bar{Z} = 0.04436282$.

Let Y_f=1 for $Y_i \leq 5588.012$, Y_f=0 for all $Y_i > 15588.012$,
X_f=1 for $X_i \leq 74364.68$, X_f=0 for all $X_i > 74364.68$ and
Z_f=1 for $Z_i \leq 0.04436282$, Z_f=0 for all $Z_i > 0.04436282$.

Pop IV. Source: [26]

Using $t_y = \bar{Y}$, $t_x = \bar{X}$ and $t_z = \bar{Z}$:

Let Y be the total amount of apples produced in 1999, and let X be the total number of apple trees in 1998.

$N = 95$, $n = 13$, $P(t_y) = 0.8191489$, $P(t_x) = 0.7765957$, $C_{P(t_y)} = 0.4723909$, $C_{P(t_x)} = 0.53922581$, $P(t_z) = 0.6489362$, $S_{P(t_y)} = 0.3869585$, $S_{P(t_x)} = 0.4187605$, $S_{P(t_y)P(t_x)} = 0.1312057$, $\rho_{P(t_y)P(t_x)} = 0.809697$, $\bar{Y} = 9384.309$, $\bar{X} = 72409.95$ and $\bar{Z} = 0.06322204$.

Let Y_f=1 for $Y_i \leq 9384.309$, Y_f=0 for all $Y_i > 9384.309$,
X_f=1 for $X_i \leq 72409.95$, X_f=0 for all $X_i > 72409.95$ and
Z_f=1 for $Z_i \leq 0.06322204$, Z_f=0 for all $Z_i > 0.06322204$.

The (PRE) can be found using the equation that follows:

$$PRE = \frac{\text{Var}(\hat{P}(t_{y_0}))}{\text{Var}(\hat{P}(t_{y_k}))\ or\ \text{MSE}(\hat{P}(t_{y_k}))} \times 100,$$

where k= 0, R, exp(R), SK(R) and HR.

6. Conditional values

By adjusting for possible bias and inflated variance, the modified HR estimators conditionally provide more accurate estimates in the presence of simple random sampling.

6.1. Simple random sampling

Every person has an equal chance of being chosen by simple random sampling, which encourages representatives in the sample. For example, in a diet-related study, researchers may randomly select participants from a list of people in the public. Even with this random selection, there is a chance that some participants would omit questions regarding certain foods they consume, such as sugary snacks, which could leave gaps in the data. If specific dietary practices are underrepresented, these gaps may impact the analysis as a whole. Researchers rely on the selection process's inherent unpredictability in these situations to minimize bias and guarantee that the sample still accurately represents the larger population.

When using a SRS, the HR type estimators conditionally produce unbiased estimates of the population mean if the sample conditions ensure each unit has an equal probability of selection. In these circumstances, the estimators efficiently utilize the additional data, which lowers variance and improves precision.

The study thoroughly evaluates the performance of several estimators, highlighting the superior effectiveness of the proposed estimator. As shown in Table 1, all the evaluators demonstrate positive outcomes, indicating that they are generally reliable and capable of performing estimation tasks to a satisfactory standard. However, the proposed estimator consistently outperforms its counterparts, demonstrating its ability to provide more accurate and efficient estimates of the expected values. This distinguishes it as the most reliable choice among those examined.

Table 1. Conditional values for (IV) populations.

Conditions	Pop I	Pop II	Pop III	Pop IV
		$t_y, t_x, t_z = \bar{Y}, \bar{X}, \bar{Z}$		
i	0.0900	0.1353	0.0479	0.0964
ii	0.0201	0.0078	0.0267	0.0148
iii	0.0014	0.0172	0.0008	0.0068
iv	0.3797	0.3153	0.2979	0.3232

To further support the findings, additional analyses presented in Tables 2 and 3 provide a deeper examination of the comparative performance of the estimators. These tables utilize various performance metrics, including MSE and possibly other criteria, such as bias, variance, or computational efficiency, to deliver a comprehensive assessment. The results in these tables clearly indicate that the proposed estimator consistently outperforms the others across different conditions, demonstrating superior accuracy and precision. This sustained advantage not only highlights the robustness of the proposed estimator but also its adaptability to various estimation scenarios.

In conclusion, the conditional results presented in Table 1, alongside the thorough comparisons in Tables 2 and 3, strongly support the effectiveness of the proposed estimator. Its superior performance across multiple metrics demonstrates its efficiency and reliability in estimating expected values, outperforming alternative estimators. These findings reinforce the conclusion that the proposed estimator is the best choice for obtaining accurate and dependable estimates in the contexts examined. It is a valuable tool for both practical applications and future research efforts.

Table 2. Using PRE, and MSE with a population means.

	$t_y, t_x, t_z = \bar{Y}, \bar{X}, \bar{Z}$			
	Pop I		Pop II	
Estimators	MSE	PRE	MSE	PRE
$\hat{P}(t_{y_0})$	0.00474	100.000	0.00810	100.000
$\hat{P}(t_{y_R})$	0.00241	196.496	0.00300	269.855
$\hat{P}(t_{y_{\exp}(R)})$	0.00179	265.091	0.00338	239.949
$\hat{P}(t_{y_{SK}(R)})$	0.01440	32.942	0.01530	52.928
$\hat{P}(t_{y_{HR}})$	**0.00174**	**272.052**	**0.00269**	**301.279**

Table 3. Utilizing population mean, MSE, and PRE of several estimators.

	$t_y, t_x, t_z = \bar{Y}, \bar{X}, \bar{Z}$			
	Pop III		Pop IV	
Estimators	MSE	PRE	MSE	PRE
$\hat{P}(t_{y_0})$	0.00095	100.000	0.0115	100.000
$\hat{P}(t_{y_R})$	0.00072	131.0473	0.0052	220.0355
$\hat{P}(t_{y_{\exp}(R)})$	0.00046	210.3242	0.0047	249.0718
$\hat{P}(t_{y_{SK}(R)})$	0.00356	26.49113	0.0289	39.7629
$\hat{P}(t_{y_{HR}})$	**0.00044**	**214.1214**	**0.0041**	**280.9528**

Tables 2 and 3 provide a detailed comparison of the performance of various estimators, including ratio and exponential-type estimators. The results indicate that both ratio and exponential-type estimators generally perform well across different scenarios, demonstrating their effectiveness in many estimation tasks. However, a notable observation is that dual ratio-type estimators perform relatively weaker compared to the other estimators evaluated. This suggests that while dual ratio-type estimators may have some utility, their efficiency and accuracy do not match those of the ratio, exponential-type, or other estimators included in the analysis. This finding underscores the importance of selecting the appropriate estimator based on the specific requirements and characteristics of the estimation problem at hand.

Researchers and practitioners can obtain accurate and dependable estimations by using these estimators in their statistical analyses due to their better performance. Nevertheless, prior to choosing the best estimator, it is also crucial to thoroughly evaluate the particular context and features of the data that is being studied.

7. Simulation study

A detailed simulation study is carried out to evaluate the usefulness of the suggested type of unbiased estimators. This study produced 1000 iterations utilizing SRS, and the formula given in

Eq (20) was utilized to get the MSE values for the unbiased ratio estimators. The simulation results are shown in Tables 4 and 5.

The suggested estimator surpassed all other known estimators when the results were analyzed. t_y, t_x, and t_z, the mean values of population Y, X, and Z, respectively, were used to illustrate this degree of superiority. The suggested estimator performed better in estimating the required quantities than other estimators.

The two populations previously mentioned were used to model a real-life dataset to further confirm the suggested estimators' practical utility. By including these populations in the analysis, the suggested estimators demonstrated their capacity to offer precise estimations in practical situations. The simulation analysis and real-world dataset model verified that the suggested class of unbiased estimators in the Hartley-Ross style has much practical use. Tables 4 and 5 exhibit how well they perform, which implies that they are useful tools for estimating population parameters. They may also be applied practically in various statistical analyses and real-world settings.

$$\mathrm{Var}(\hat{P}(t_{y_{HR}})) = \frac{1}{1000}\sum_{i=1}^{1000}[\hat{P}(t_{y_{ik}}) - \hat{P}(t_{y_0})]^2. \tag{7.1}$$

The values of k are 0, R, exp(R), SK(R), and HR.

Table 4. PRE of different estimators.

Pop I			
Estimators	n=40	n=50	n=60
$\hat{P}(t_{y_0})$	100.000	100.000	100.000
$\hat{P}(t_{y_R})$	259.5124	280.8869	265.6376
$\hat{P}(t_{y_{\exp(R)}})$	190.1314	195.3926	188.3682
$\hat{P}(t_{y_{SK(R)}})$	50.82763	75.24619	71.41701
$\hat{P}(t_{y_{HR}})$	**340.6476**	**483.2862**	**605.6851**

Table 5. PRE of different estimators.

Pop II			
Estimators	n=40	n=50	n=60
$\hat{P}(t_{y_0})$	100.000	100.000	100.000
$\hat{P}(t_{y_R})$	254.1994	263.7851	296.0213
$\hat{P}(t_{y_{\exp(R)}})$	224.886	231.225	230.7736
$\hat{P}(t_{y_{SK(R)}})$	43.33509	43.88207	58.32776
$\hat{P}(t_{y_{HR}})$	**817.8522**	**716.9396**	**444.0827**

Among all the evaluated estimators, the HR unbiased estimator shows the least variation, as demonstrated by the results in Tables 4 and 5. Notably, the proposed unbiased estimator outperforms alternative methods in terms of variation, yielding results that are both consistent and precise. When specifically compared to the ratio of exponential estimators, the proposed estimator consistently provides more reliable estimates, reinforcing its robustness.

The performance of the proposed estimator improves as the sample size increases, demonstrating its scalability and adaptability to larger datasets. This scalability indicates that the estimator not only maintains its precision with smaller samples but also achieves greater accuracy with larger ones. Such features make it a promising choice for practical applications where the availability of data may vary. Overall, the HR unbiased estimator stands out as an excellent option for obtaining reliable and consistent results, especially in situations that require low variation and high accuracy.

8. Existing estimators under non response

8.1. Notations and symbols

Suppose population of $\Omega = 1, 2, \ldots, N$ that consists of N units. To determine the size of the sample, this study employed simple random sampling without replacement (SRSWOR) sample n from the N.

Let $P(t_y) = \dfrac{1}{N}\sum_{i=1}^{N} Y_f$: the PDF of Y_f,

$\hat{P}(t_y) = \dfrac{1}{n}\sum_{i=1}^{n} Y_f$: the SDF of Y_f when there is a full response,

$P(t_x) = \dfrac{1}{N}\sum_{i=1}^{N} X_f$: the PDF of X_f,

$\hat{P}(t_x) = \dfrac{1}{n}\sum_{i=1}^{n} X_f$: the SDF of X_f when there is a full response,

$P(t_z) = \dfrac{1}{N}\sum_{i=1}^{N} Z_f$: PDF of Z_f and

$\hat{P}(t_z) = \dfrac{1}{n}\sum_{i=1}^{n} Z_f$: SDF of Z_f when there is a full response.

Let $P^*(t_y) = \dfrac{1}{N}\sum_{i=N_1+1}^{N} Y_f^*$: the PDF of Y_f^* for non response group,

$P^*(t_x) = \dfrac{1}{N}\sum_{i=N_1+1}^{N} X_f^*)$: the PDF of X_f^* for non response group and

$P^*(t_z) = \dfrac{1}{N}\sum_{i=1}^{N} Z_f^*$: the PDF of Z_f^* for non response group.

Let $S^2_{P(t_y)} = \sum_{i=1}^{N} \dfrac{(Y_f - P(t_y))^2}{N-1}$: the PV of Y_f and

$S^2_{P(t_x)} = \sum_{i=1}^{N} \dfrac{(X_f - P(t_x))^2}{N-1}$: the PV of X_f.

Let $S^2_{P^*(t_y)} = \sum_{i=N_1+1}^{N} \dfrac{(Y^*_f - P^*(t_y))^2}{N_2 - 1}$: the PV of Y^*_f for the non response group and

$S^2_{P^*(t_x)} = \sum_{i=N_1+1}^{N} \dfrac{(X^*_f - P^*(t_x))^2}{N_2 - 1}$: the PV of X^*_f for the non-response group.

Let $C_{P(t_y)} = \dfrac{S_{P(t_y)}}{P(t_y)}$: the PCV of $\{Y_f\}$ and

$C_{P(t_x)} = \dfrac{S_{P(t_x)}}{P(t_x)}$: the PCV of Y_f.

Let $C_{P^*(t_y)} = \dfrac{S_{P^*(t_y)}}{P^*(t_y)}$: the PCV of $\{Y^*_f\}$ for the non-response group and

$C_{P^*(t_x)} = \dfrac{S_{P^*(t_x)}}{P^*(t_x)}$: the PCV of $I(X^*_i \leq t_x)$ for the non-response group.

Let $S_{P(t_y)P(t_x)} = \sum_{i=1}^{N} \dfrac{[(Y_f - P(t_y))(X_f - P(t_x))]}{N-1}$: the PC between Y_f and X_f.

Let $S_{P^*(t_y)P^*(t_x)} = \sum_{i=N_1+1}^{N} \dfrac{[(Y^*_f - P^*(t_y))(X^*_f - P^*(t_x))]}{N_2 - 1}$: the PC between Y^*_f and X^*_f for the non-response group.

Let $\rho_{P(t_y)P(t_x)} = \dfrac{S_{P(t_y)P(t_x)}}{S_{P(t_y)} S_{P(t_x)}}$: the PC between Y_f and X_f.

Let $\rho_{P^*(t_y)P^*(t_x)} = \dfrac{S_{P^*(t_y)P^*(t_x)}}{S_{P^*(t_y)} S_{P^*(t_x)}}$: the PC between Y^*_f and X^*_f for the non-response group.

Let

$$U_{rst} = \frac{E[(P(t_y) - \hat{P}(t_y)^r (P(t_x) - \hat{P}(t_x)^s (P(t_z) - \hat{P}(t_z)^t]}{P(t_y)^r P(t_x)^s P(t_z)^t}$$

and

$$U^*_{rst} = \frac{E[(P(t_y) - \hat{P}^*(t_y)^r (P(t_x) - \hat{P}^*(t_x)^s (P(t_z) - \hat{P}^*(t_z)^t]}{P(t_y)^r P(t_x)^s P(t_z)^t},$$

where $r, s, t = 0, 1, 2$.

Let $e_0 = \frac{\hat{P}(t_y)-P(t_y)}{P(t_y)}$, $e_1 = \frac{\hat{P}(t_x)-P(t_x)}{P(t_x)}$, $e_2 = \frac{\hat{P}(t_z)-P(t_z)}{P(t_z)}$, $e^*_0 = \frac{\hat{P}^*(t_y)-P(t_y)}{P(t_y)}$, $e^*_1 = \frac{\hat{P}^*(t_x)-P(t_x)}{P(t_x)}$ and $e^*_2 = \frac{\hat{P}^*(t_z)-P(t_z)}{P(t_z)}$ such that $E(e_i) = 0$, $(i = 0, 1, 2)$ and $E(e^*_i) = 0$, $(i^* = 0, 1, 2)$. $E(e_0^2) = \lambda C^2_{P(t_y)} \cong U_{200}$, $E(e_1^2) = \lambda C^2_{P(t_x)} \cong U_{020}$, $E(e_2^2) = \lambda C^2_{P(t_z)} \cong U_{002}$, $E(e_0 e_1) = \lambda \rho_{P(t_y)P(t_x)} C_{P(t_y)} C_{P(t_x)} \cong U_{110}$, $E(e_0 e_2) = \lambda \rho_{P(t_y)P(t_z)} C_{P(t_y)} C_{P(t_z)} \cong U_{101}$, and $E(e_1 e_2) = \lambda \rho_{P(t_z)P(t_z)} C_{P(t_x)} C_{P(t_z)} \cong U_{011}$. Let $E(e_0^{*2}) = \frac{\lambda S^2_{P(t_y)} + \lambda^* S^2_{P^*(t_y)}}{P^2(t_y)} \cong U^*_{200}$, $E(e_1^{*2}) = \frac{\lambda S^2_{P(t_x)} + \lambda^* S^2_{P^*(t_x)}}{F^2(t_x)} \cong U^*_{020}$, $E(e^*_0 e^*_1) =$ $\frac{\lambda S^2_{P(t_y)P(t_x)} + \lambda^* S^2_{P^*(t_y)P^*(t_x)}}{P(t_y)P(t_x)} \cong U^*_{110}$, $E(e^*_0 e_1) = \lambda \rho_{P(t_y)P(t_x)} C_{P(t_y)} C_{P(t_x)} \cong U_{110}$, where $\lambda = (\frac{1}{n} - \frac{1}{N})$, $\lambda^* = (\frac{W_2(K-1)}{n})$, $W_2 = \frac{N_2}{N}$, N_2 be the number of population corresponding to the non response group.

9. Estimating population distribution function under non response

Let us assume that there are two categories within the underlying population, (i) the response (N_1 units) and (ii) the non response (N_2 units), and N_1, N_2 represents the number of units in the response and non response groups, where $N_1 + N_2 = N$ is the number of units in the response and non response groups. The finite population distribution function CDF, $P(t_y)$, is given by:

$$P(t_y) = W_1 P(t_y) + W_2 P^*(t_y) \tag{9.1}$$

where $W_1 = \frac{N_1}{N}$ and $W_2 = \frac{N_2}{N}$.

Let this study select n units from a population of N units where n_1 units respond and n_2 units do not respond, where $n_1 + n_2 = n$. Through a personal interview, the non respondents are again reached to receive a response from n_2. Then subsamples with a size of $r = \frac{n_2}{k}$ for ($k > 1$) are obtained from non responding units of n_2. Both r units are presumed to respond. Let $P(t_y)$ and $P^{(*r)}(t_y)$ be a DF based on sensitive units of n_2 and r. [28] proposed a non response unbiased estimator for $P(t_y)$, given as:

$$\hat{P}^*(t_y) = w_1 P(t_y) + w_2 P^{(*r)}(t_y) \tag{9.2}$$

where, $w_1 = \frac{n_1}{n}$ and $w_2 = \frac{n_2}{n}$.

The MSE of $P(t_y)$ is given by:

$$\text{MSE}(\hat{P}^*(t_y)) = \lambda S^2_{P(t_y)} + \lambda^* S^2_{P^*(t_y)}. \tag{9.3}$$

Similarly, the estimator (unbaised) of $P(t_x)$ for non response:

$$\hat{P}^*(t_x) = w_1 P(t_x) + w_2 P^{(*r)}(t_x). \tag{9.4}$$

The MSE of $P(t_x)$ is given by:

$$\text{MSE}(\hat{P}^*(t_x)) = \lambda S^2_{P(t_x)} + \lambda^* S^2_{P^*(t_x)}. \tag{9.5}$$

The non response in this study affects both the auxiliary and study variables.

(1) The ratio type estimator of distribution function under non-response is given by:

$$\hat{P}_2(t_{y_R}) = \hat{P}^*(t_y)\left(\frac{P(t_x)}{\hat{P}^*(t_x)}\right). \tag{9.6}$$

The bias and MSE of $\hat{P}_2(t_{y_R})$ are given by:

$$\text{Bias}(\hat{P}_2(t_{y_R})) = P(t_y)[U^*_{020} + U^*_{110}] \tag{9.7}$$

and

$$\text{MSE}(\hat{P}_2(t_{y_R})) = P^2(t_y)[U^*_{200} + U^*_{020} - 2U^*_{110}]. \tag{9.8}$$

(2) [24] suggested ratio estimators of finite population distribution function under non response, which are

$$\hat{P}_2(t_{y_{\exp(R)}}) = \hat{P}^*(t_y)\exp\left(\frac{P(t_x) - \hat{P}^*(t_x)}{P(t_x) + \hat{P}^*(t_x)}\right), \tag{9.9}$$

The bias and MSE of $\hat{P}_2(t_{y_{\exp(R)}})$ are given by:

$$\text{Bias}(\hat{P}_2(t_{y_{\exp(R)}})) = P(t_y)[\frac{3}{8}U^*_{020} - \frac{1}{2}U^*_{110}] \tag{9.10}$$

and

$$\text{MSE}(\hat{P}_2(t_{y_{\exp(R)}})) = P^2(t_y)[U^*_{200} + \frac{1}{4}U^*_{020} - U^*_{110}]. \tag{9.11}$$

(3) [25] proposed ratio type estimators of finite population distribution function under non response, is given by:

$$\hat{P}_2(t_{y_{\text{SK}(R)}}) = \hat{P}^*(t_y)\Big(\frac{P(t_x)}{\hat{P}^*(t_x)}\Big)^2, \tag{9.12}$$

The bias and MSE of $\hat{P}_2(t_{y_{\text{SK}(R)}})$, is given by:

$$\text{Bias}\Big(\hat{P}_2(t_{y_{\text{SK}(R)}})\Big) = P(t_y)[3U^*_{020} - 2U^*_{110}] \tag{9.13}$$

and

$$\text{MSE}\Big(\hat{P}_2(t_{y_{\text{SK}(R)}})\Big) = P^2(t_y)[U^*_{200} + 4U^*_{020} - 4U^*_{110}]. \tag{9.14}$$

10. Improved class of estimators

HR type unbiased estimator is used to estimate population distribution function in non response presence.

Suppose that this study takes n units as a sample from a population of N units, this study may define the ratio as follows.

$$\hat{P}^*(t_z) = \frac{1}{n}\sum_{i=1}^{n} Z^*_f,$$

where $Z^*_i = \frac{Y_i}{X_i}$.

Now, the ratio estimator of the population mean given as:

$$\hat{P}_2(t_{y_{HR}}) = P(t_x)\hat{P}^*(t_z). \tag{10.1}$$

Now, apply expectations on both sides of the equation.

$$E\Big(\hat{P}_2(t_{y_{HR}})\Big) = P(t_x)E\Big(\hat{P}^*(t_z)\Big),$$

$$E\Big(\hat{P}_2(t_{y_{HR}})\Big) = P(t_x)E\Big(\frac{1}{n}\sum_{i=1}^{n} Z^*_f\Big)$$

and

$$E\Big(\hat{P}_2(t_{y_{HR}})\Big) = P(t_x)\Big(\frac{1}{N}\sum_{i=1}^{N} Z^*_f\Big). \tag{10.2}$$

It is a biased estimator. To find its bias, we use

$$\text{Bias}\Big(\hat{P}_2(t_{y_{HR}})\Big) = E(P(t_{y_{HR}})) - P(t_y),$$

$$\text{Bias}\Big(\hat{P}_2(t_{y_{HR}})\Big) = P(t_x)\Big(\frac{1}{N}\sum_{i=1}^{N} Z_f^*\Big) - (\frac{1}{N}\sum_{i=1}^{N} Y_f^*),$$

$$\text{Bias}\Big(\hat{P}_2(t_{y_{HR}})\Big) = -\Big(\frac{N_2-1}{N_2}\Big)\Big(\frac{n_2}{n_2-1}\Big)\Big(\hat{P}^*(t_y) - \hat{P}^*(t_z)\hat{P}(t_x)\Big). \tag{10.3}$$

Now, the estimator becomes

$$\hat{P}_2(t_{y_{HR}}) = P(t_x)\hat{P}^*(t_z) - \text{Bias}\Big(P(t_{y_{HR}})\Big),$$

$$\hat{P}_2(t_{y_{HR}}) = P(t_x)\hat{P}^*(t_z) + \Big(\frac{N_2-1}{N_2}\Big)\Big(\frac{n_2}{n_2-1}\Big)\Big(\hat{P}^*(t_y) - \hat{P}^*(t_z)\hat{P}^*(t_x)\Big). \tag{10.4}$$

Under the assumption $\left(\frac{N_2-1}{N_2}\right)\left(\frac{n_2}{n_2-1}\right) \approx 1$.

$$\hat{P}_2(t_{y_{HR}}) = P(t_x)\hat{P}^*(t_z) + \Big(\hat{P}^*(t_y) - \hat{P}^*(t_z)\hat{P}^*(t_x)\Big), \tag{10.5}$$

$$\hat{P}_2(t_{y_{HR}}) = P(t_y)(1+e_0^*) + P(t_z)(1+e_2^*)(-e_1^* P(t_x)). \tag{10.6}$$

Now, the variance of the proposed estimators under non-response is given by:

$$\text{Var}(\hat{P}_2(t_{y_{HR}})) = [P^2(t_y)U_{200}^* + P^2(t_z)P^2(t_x)U_{020}^* - 2P(t_z)P(t_y)P(t_x)U_{110}^*]. \tag{10.7}$$

11. Theoretical comparison

Now, the suggested estimator to the following estimators:

(i) From (9.3) and (10.7),
MSE($\hat{P}_2(t_{y_0})$) − Var($\hat{P}_2(t_{y_{HR}})$) > 0 or if

$$[\lambda S^2_{P(t_y)} + \lambda^* S^2_{P^*(t_y)}] - [P^2(t_y)U_{200}^* + P^2(t_z)P^2(t_x)U_{020}^* - 2P(t_z)P(t_y)P(t_x)U_{110}^*)] > 0.$$

(ii) From (9.8) and (10.7),
MSE($\hat{P}_2(t_{y_R})$) − Var($\hat{P}_2(t_{y_{HR}})$) > 0 or if

$$U_{020}^*[P^2(t_y) - P^2(t_z)P^2(t_x)] - 2U_{110}^* P(t_y)[P(t_y) - P(t_z)P(t_x)] > 0.$$

(iii) From (9.11) and (10.7),
MSE($\hat{P}_2(t_{y_{\exp(R)}})$) − Var($\hat{P}_2(t_{y_{HR}})$) > 0 or if

$$U_{020}^*[\frac{1}{4}P^2(t_y) - P^2(t_z)P^2(t_x)] - U_{110}^* P(t_y)[P(t_y) + 2P(t_z)P(t_x)] > 0.$$

(iv) From (9.14) and (10.7),
MSE($\hat{P}_2(t_{y_{SK(R)}})$) − Var($\hat{P}_2(t_{y_{HR}})$) > 0 or if

$$U_{020}^*[4P^2(t_y) - P^2(t_z)P^2(t_x)] - 2U_{110}^* P(t_y)[2P(t_y) - P(t_z)P(t_x)] > 0.$$

The conditions under which the proposed estimator $\hat{P}_2(t_{y_{HR}})$ exhibits lower mean squared error, indicating greater accuracy compared to both simple and usual estimators, can be derived from the expressions provided above. By analyzing the mathematical expressions and considering the properties of the proposed estimator, this study can determine the specific conditions or scenarios where $\hat{P}_2(t_{y_{HR}})$ outperforms the other estimators in terms of accuracy. These conditions will enable researchers and practitioners to identify the appropriate circumstances for utilizing $\hat{P}_2(t_{y_{HR}})$ to achieve more precise estimations in their statistical analyses.

12. Data comparison

The performance of the proposed $\hat{P}_2(t_{y_{HR}})$ estimator is compared to the performance of other current estimators using four natural population data sets. The population's summary statistics are listed below.

Pop I. Source: [25]

Suppose Y is the total sum of apples produced, and X is ratio of apple tree.

$N = 106\,, n = 30\,, P(t_y) = 0.8301887\,, P(t_x) = 0.7641509\,, P(t_z) = 1.08642\,, S_{P(t_y)} = 0.3772507\,, S_{P(t_x)} = 0.4265451\,, S_{P(t_y)P(t_x)} = 0.1309973$ and $\rho_{P(t_y)P(t_x)} = 0.8140806$.

Let Y_f=1 for $Y \leq 1536.774$ or (mean of Y), Y_f=0 for all $Y > 1536.774$ or (mean of Y),
X_f=1 for $X \leq 24375.59$ or (mean of X), X_f=0 for all $X > 24375.59$ or (mean of X) and
Z_f=1 for $Z \leq 0.03927598$ or (mean of Z), Z_f=0 for all $Z > 0.03927598$ or (mean of Z).

(1) The non response ratio is calculated by the last 26 units of N and multiplying it by 25 percent.
$N_2 = 26\,, P^*(t_y) = 0.6923077\,, P^*(t_x) = 0.6923077\,, P^*(t_z) = 0.6153846$
Let Y_f^*=1 for $Y \leq 848.5$ or (mean of Y), Y_f^*=0 for all $Y > 848.5$ or (mean of Y),
X_f^*=1 for $X \leq 20660$ or (mean of X), X_f^*=0 for all $X > 20660$ or (mean of X) and
Z_f^*=1 for $Z \leq 0.03261708$ or (mean of Z), Z_f^*=0 for all $Z > 0.03261708$ or (mean of Z).

(2) The non response ratio is calculated by the last 32 units of N and multiplying it by 30 percent.
$N_2 = 32\,, P^*(t_y) = 0.71875\,, P^*(t_x) = 0.71875\,, P^*(t_z) = 0.59375$
Let Y_f^*=1 for $Y \leq 768.9688$ or (mean of Y), Y_f^*=0 for all $Y > 768.9688$ or (mean of Y),
X_f^*=1 for $X \leq 21605.53$ or (mean of X), X_f^*=0 for all $X > 21605.53$ or (mean of X) and
Z_f^*=1 for $Z \leq 0.03243408$ or (mean of Z), Z_f^*=0 for all $Z > 0.03243408$ or (mean of Z).

(3) The non response ratio is calculated by the last 37 units of N and multiplying it by 35 percent.
$N_2 = 37\,, P^*(t_y) = 0.7567568\,, P^*(t_x) = 0.7567568\,, P^*(t_z) = 0.5945946$
Let Y_f^*=1 for $Y \leq 685.3784$ or (mean of Y), Y_f^*=0 for all $Y > 685.3784$ or (mean of Y),
X_f^*=1 for $X \leq 19064.3$ or (mean of X), X_f^*=0 for all $X > 19064.3$ or (mean of X) and
Z_f^*=1 for $Z \leq 0.03676974$ or (mean of Z), Z_f^*=0 for all $Z > 0.03676974$ or (mean of Z).

Pop II. Source: [27]

Let Y represent the estimated number of fish caught by fishermen and X represent the estimated number of fish caught by fishermen.

$N = 69\,, n = 25\,, P(t_y) = 0.7246377\,, P(t_x) = 0.7101449\,, P(t_z) = 0.7246377\,, S_{P(t_y)} = 0.4499692\,, S_{P(t_x)} = 0.4570188\,, S_{P(t_y)P(t_x)} = 0.1690111$ and $\rho_{P(t_y)P(t_x)} = 0.821861$.

Let Y_f=1 for $Y \leq 4514.899$ or (mean of Y), Y_f=0 for all $Y > 4514.899$ or (mean of Y),
X_f=1 for $X \leq 4230.174$ or (mean of X), X_f=0 for all $X > 4230.174$ or (mean of X) and
Z_f=1 for $Z \leq 1.251484$ or (mean of Z), Z_f=0 for all $Z > 1.251484$ or (mean of Z).

(1) The non response ratio is calculated by the last 17 units of N and multiplying it by 25 percent.
$N_2 = 17$, $P^*(t_y) = 0.7647059$ $P^*(t_x) = 0.7647059$, , $P^*(t_z) = 0.6470588$
Let Y_f^*=1 for $Y \leq 3155.059$ or (mean of Y), Y_f^*=0 for all $Y > 3155.059$ or (mean of Y),
X_f^*=1 for $X \leq 2872.118$ or (mean of X), X_f^*=0 for all $X > 2872.118$ or (mean of X) and
Z_f^*=1 for $Z \leq 1.099427$ or (mean of Z), Z_f^*=0 for all $Z > 1.099427$ or (mean of Z).

(2) The non response ratio is calculated by the last 21 units of N and multiplying it by 30 percent.
$N_2 = 21$, $P^*(t_y) = 0.7142857$, $P^*(t_x) = 0.7142857$, $P^*(t_z) = 0.6666667$
Let Y_f^*=1 for $Y \leq 3356.095$ or (mean of Y), Y_f^*=0 for all $Y > 3356.095$ or (mean of Y),
X_f^*=1 for $X \leq 3135.667$ or (mean of X), X_f^*=0 for all $X > 3135.667$ or (mean of X) and
Z_f^*=1 for $Z \leq 1.08177$ or (mean of Z), Z_f^*=0 for all $Z > 1.08177$ or (mean of Z).

(3) The non response ratio is calculated by the last 24 units of N and multiplying it by 35 percent.
$N_2 = 24$, $P^*(t_y) = 0.75$, $P^*(t_x) = 0.7083333$, $P^*(t_z) = 0.6666667$
Let Y_f^*=1 for $Y \leq 4338.792$ or (mean of Y), Y_f^*=0 for all $Y > 4338.792$ or (mean of Y),
X_f^*=1 for $X \leq 4231.417$ or (mean of X), X_f^*=0 for all $X > 4231.417$ or (mean of X) and
Z_f^*=1 for $Z \leq 1.070156$ or (mean of Z), Z_f^*=0 for all $Z > 1.070156$ or (mean of Z).

Pop III. Source: [26]

Assume Y denotes the total sum of apples produced, and X is the proportion of apple trees.

$N = 171$, $n = 95$, $P(t_y) = 0.9005848$, $P(t_x) = 0.8479532$, $P(t_z) = 0.5964912$, $S_{P(t_y)} = 0.3000975$, $S_{P(t_x)} = 0.3601208$, $S_{P(t_y)P(t_x)} = 0.07891297$ and $\rho_{P(t_y)P(t_x)} = 0.7301934$.
Let Y_f=1 for $Y \leq 5588.012$ or (mean of Y), Y_f=0 for all $Y > 5588.012$ or (mean of Y),
X_f=1 for $X \leq 74364.68$ or (mean of X), X_f=0 for all $X > 74364.68$ or (mean of X) and
Z_f=1 for $Z \leq 0.04436282$ or (mean of Z), Z_f=0 for all $Z > 0.04436282$ or (mean of Z).

(1) The non response ratio is calculated by the last 43 units of N and multiplying it by 25 percent.
$N_2 = 43$, $P^*(t_y) = 0.9069767$, $P^*(t_x) = 0.8837209$, $P^*(t_z) = 0.5813953$.
Let Y_f^*=1 for $Y \leq 8266.698$ or (mean of Y), Y_f^*=0 for all $Y > 8266.698$ or (mean of Y),
X_f^*=1 for $X \leq 98138.7$ or (mean of X), X_f^*=0 for all $X > 98138.7$ or (mean of X) and
Z_f^*=1 for $Z \leq 0.04413647$ or (mean of Z), Z_f^*=0 for all $Z > 0.04413647$ or (mean of Z).

(2) The non response ratio is calculated by the last 51 units of N and multiplying it by 30 percent.
$N_2 = 51$, $P^*(t_y) = 0.9215686$, $P^*(t_x) = 0.8823529$, $P^*(t_z) = 0.5882353$
Let Y_f^*=1 for $Y \leq 7255.412$ or (mean of Y), Y_f^*=0 for all $Y > 7255.412$ or (mean of Y),
X_f^*=1 for $X \leq 87643.02$ or (mean of X), X_f^*=0 for all $X > 87643.02$ or (mean of X) and
Z_f^*=1 for $Z \leq 0.04491961$ or (mean of Z), Z_f^*=0 for all $Z > 0.04491961$ or (mean of Z).

(3) The non response ratio is calculated by the last 60 units of N and multiplying it by 35 percent.
$N_2 = 60$, $P^*(t_y) = 0.9166667$, $P^*(t_x) = 0.8666667$, $P^*(t_z) = 0.5945946$
Let Y_f^*=1 for $Y \leq 7012.967$ or (mean of Y), Y_f^*=0 for all $Y > 7012.967$ or (mean of Y),
X_f^*=1 for $X \leq 90549.07$ or (mean of X), X_f^*=0 for all $X > 90549.07$ or (mean of X) and
Z_f^*=1 for $Z \leq 0.04419811$ or (mean of Z), Z_f^*=0 for all $Z > 0.04419811$ or (mean of Z).

Pop IV. Source: [26]

Suppose Y shows the total sum of apples produced, and X represents the proportion of apple trees. $N = 95$, $n = 55$, $P(t_y) = 0.8191489$, $P(t_x) = 0.7765957$, $P(t_z) = 0.6489362$, $S_{P(t_y)} = 0.3869585$, $S_{P(t_x)} = 0.4187605$, $S_{P(t_y)P(t_x)} = 0.1312057$ and $\rho_{P(t_y)P(t_x)} = 0.809697$.
Assume Y_f=1 for $Y \leq 9384.309$ or (mean of Y), Y_f=0 for all $Y > 9384.309$ or (mean of Y),
X_f=1 for $X \leq 72409.95$ or (mean of X), X_f=0 for all $X > 72409.95$ or (mean of X) and
Z_f=1 for $Z \leq 0.06322204$ or (mean of Z), Z_f=0 for all $Z > 0.06322204$ or (mean of Z).

(1) The non response ratio is calculated by the last 70 units of N and multiplying it by 25 percent.
$N_2 = 70$, $P^*(t_y) = 0.8333333$, $P^*(t_x) = 0.75$, $P^*(t_z) = 0.5833333$.
Let Y^*_f=1 for $Y \leq 1437.042$ or (mean of Y), Y^*_f=0 for all $Y > 1437.042$ or (mean of Y),
X^*_f=1 for $X \leq 25656.46$ or (mean of X), X^*_f=0 for all $X > 25656.46$ or (mean of X) and
Z^*_f=1 for $Z \leq 0.04199972$ or (mean of Z), Z^*_f=0 for all $Z > 0.04199972$ or (mean of Z).

(2) The non response ratio is calculated by the last 28 units of the N and multiplying it by 30 percent.
$N_2 = 28$, $P^*(t_y) = 0.8214286$, $P^*(t_x) = 0.7857143$, $P^*(t_z) = 0.6071429$
Let Y^*_f=1 for $Y \leq 2390.429$ or (mean of Y), Y^*_f)=0 for all $Y > 2390.429$ or (mean of Y),
X^*_f=1 for $X \leq 37522.71$ or (mean of X), X^*_f=0 for all $X > 37522.71$ or (mean of X) and
Z^*_f=1 for $Z \leq 0.04519338$ or (mean of Z), Z^*_f=0 for all $Z > 0.04519338$ or (mean of Z).

(3) The non response ratio is calculated by the last 33 units of the N and multiplying it by 35 percent.
$N_2 = 33$, $P^*(t_y) = 0.7878788$, $P^*(t_x) = 0.7575758$ $P^*(t_z) = 0.5151515$
Let Y^*_f=1 for $Y \leq 2689.788$ or (mean of Y), Y^*_f=0 for all $Y > 2689.788$ or (mean of Y),
X^*_f=1 for $X \leq 40433.21$ or (mean of X), X^*_f=0 for all $X > 40433.21$ or (mean of X) and
Z^*_f=1 for $Z \leq 0.04833449$ or (mean of Z), Z^*_f=0 for all $Z > 0.04833449$ or (mean of Z).

Now, the following expression is used for real data to obtain the PRE:

$$PRE = \frac{\text{MSE}(\hat{P}_2(t_{y_0}))}{\text{MSE}(\hat{P}_2(t_{y_i}))} \times 100,$$

where $\hat{P}_2(t_{y_i})$=$\hat{P}_2(t_{y_{HR}})$, $\hat{P}_2(t_{y_R})$, $\hat{P}_2(t_{y_{\exp(R)}})$ and $\hat{P}_2(t_{y_{\text{SK}(R)}})$.

13. Conditional values

By adjusting for possible bias and inflated variance, the modified HR estimators conditionally provide more accurate estimates in the presence of non response. Even when non response causes some of the data to be missing, these estimators can increase the accuracy of population mean estimations.

13.1. Non response

Non-response is a common problem in real-world survey applications that can seriously impair the quality of information received. For instance, those with more severe symptoms may be less inclined to engage in healthcare surveys that look at mental health or chronic disorders because of stigma, cognitive difficulties, or physical restrictions. In the same way, people with higher incomes may decline to participate in income and wealth surveys due to privacy concerns. In contrast, respondents with lower incomes may find it difficult to disclose their financial situation. When this non response is consistently connected to the unobserved outcome variables, bias is introduced and it becomes non ignorable. To ensure the validity of the survey results, addressing this issue calls for advanced methods like multiple imputation or weighting changes to reduce the influence of non response bias.

The experimental results presented in Tables 6–9 demonstrate the strong performance of the proposed estimators across all tested conditions. Notably, the reaction times were consistently efficient, and the MSE was significantly lower than that of alternative estimators. This consistent reduction in MSE highlights the superior accuracy and reliability of the proposed estimators, showcasing their effectiveness across a variety of scenarios and datasets.

Table 6. Conditional values using Population I.

	W_1=0.25	W_2=0.30	W_3=0.35
i	0.00465	0.00479	0.00489
ii	0.00049	0.00039	0.00037
iii	0.00024	0.00018	0.00019
iv	0.01501	0.01523	0.01538

Table 7. Conditional values using Population II.

	W_1=0.25	W_2=0.30	W_3=0.35
i	0.00612	0.00679	0.00740
ii	0.00039	0.00034	0.00017
iii	0.00058	0.00072	0.00083
iv	0.01603	0.01666	0.01662

Table 8. Conditional values using Population III.

	W_1=0.25	W_2=0.30	W_3=0.35
i	0.00069	0.00065	0.00067
ii	0.00029	0.00033	0.00036
iii	0.00014	0.00012	0.00010
iv	0.00361	0.00384	0.00410

Table 9. Conditional values using Population IV.

	W_1=0.25	W_2=0.30	W_3=0.35
i	0.0021	0.00115	0.00217
ii	0.00034	0.00041	0.00022
iii	0.00081	0.0010	0.00023
iv	0.00767	0.00805	0.00004

A more detailed summary of these findings is provided in Tables 10–13, which compile the performance metrics and offer a comparative overview. These summarized results further confirm the robustness and efficiency of the proposed estimators, highlighting their advantages over existing methods. The superior performance shown in both the detailed and summarized results reinforces the appropriateness of these estimators for practical applications, especially in situations that require high precision and lower error rates.

Table 10. PREs of estimators using Population I.

	W_1=0.25	W_2=0.30	W_3=0.35
Estimators	*PRE*	*PRE*	*PRE*
$\hat{P}_2(t_y)$	100	100	100
$\hat{P}_2(t_{y_R})$	271.4145	281.1132	285.9669
$\hat{P}_2(t_{y_{\exp(R)}})$	303.0255	307.0009	308.9259
$\hat{P}_2(t_{y_{SK(R)}})$	38.90143	39.54889	39.86403
$\hat{P}_2(t_{y_{HR}})$	**339.7875**	**334.6623**	**337.7393**

Table 11. PREs of estimators using Population II.

	W_1=0.25	W_2=0.30	W_3=0.35
Estimators	*PRE*	*PRE*	*PRE*
$\hat{P}_2(t_y)$	100	100	100
$\hat{P}_2(t_{y_R})$	234.0302	252.4467	300.773
$\hat{P}_2(t_{y_{\exp(R)}})$	223.8694	231.7421	254.3968
$\hat{P}_2(t_{y_{SK(R)}})$	50.2447	51.59956	54.0492
$\hat{P}_2(t_{y_{HR}})$	**257.0119**	**274.7351**	**316.0341**

Table 12. PREs of estimators using Population III.

	W_1=0.25	W_2=0.30	W_3=0.35
Estimators	*PRE*	*PRE*	*PRE*
$\hat{P}_2(t_y)$	100	100	100
$\hat{P}_2(t_{y_R})$	148.2967	137.3221	133.6888
$\hat{P}_2(t_{y_{\exp(R)}})$	225.6373	218.2278	215.9063
$\hat{P}_2(t_{y_{SK(R)}})$	28.56793	27.00226	26.43597
$\hat{P}_2(t_{y_{HR}})$	**231.8627**	**223.0608**	**219.9059**

Table 13. PREs of estimators using Population IV.

Estimators	W_1=0.25	W_2=0.30	W_3=0.35
	PRE	*PRE*	*PRE*
$\hat{P}_2(t_y)$	100	100	100
$\hat{P}_2(t_{y_R})$	206.8332	189.095	204.2619
$\hat{P}_2(t_{y_{\exp(R)}})$	246.1225	227.1769	235.4009
$\hat{P}_2(t_{y_{SK(R)}})$	37.72911	37.71416	39.38638
$\hat{P}_2(t_{y_{HR}})$	**261.3122**	**242.7319**	**240.0694**

The response rates for different populations are presented in Tables 10–13, revealing the effectiveness of the proposed estimators compared to existing ones. Based on the study, the suggested estimators routinely capture greater response rates across a variety of the population than the present estimators. This suggests that the proposed estimators may be able to produce estimates that are more reliable and precise, thereby improving the standard of decision-making and data analysis procedures.

14. Conclusions

The present study proposes a class of unbiased estimators for non-response (SRS) that are modeled after the HR approach for estimating the distribution function (DF). Using both mathematical modeling and statistical computations, this study establishes the efficiency criterion for proposed estimators by taking into account both the maximum response and failure to respond (non response).

It is important to note that all the conditional outcomes in SRS with failure to respond (non response) are positive, consistent with the positive value conditions found in simulated and statistical studies. Tables 2–5 and Tables 10–13 present these findings. Through a comparison of the percent relative utility of several estimators with the normal estimators, our analysis founds that the $\hat{P}(t_{y_{HR}})$ estimators we presented perform consistently better than the competing estimators.

In light of these promising findings, the paper suggests that future investigations focus on applying the suggested estimators for practical PDF estimation in the context of auxiliary variables in SRS and non-response frameworks. The proposed estimators' proven better performance indicates that they have the potential to increase statistical accuracy and dependability, which will improve decision-making in real-world applications.

Future work will enhance this research by incorporating two-phase sampling methods and leveraging distribution functions along with probability proportional to size (PPS) sampling techniques. By integrating these advanced sampling strategies, the proposed estimators can be refined to better account for varying population characteristics, thereby improving their applicability to real-world scenarios. The use of distribution functions will allow for more robust modeling of the underlying data structure. Additionally, PPS sampling will ensure that more significant or more relevant units have a greater likelihood of being selected, which will enhance the precision and efficiency of the estimators. This extension will provide a comprehensive framework for addressing complex estimation problems across diverse fields. Future research will also examine how different sample sizes affect estimator performance to improve scalability and accuracy.

Authors Contributions

Sohail Ahmad, Moiz Qureshi, and Hasnain Iftikhar: Conceptualization; Sohail Ahmad, Moiz Qureshi, and Hasnain Iftikhar: Methodology and software; Sohail Ahmad, Moiz Qureshi, Hasnain Iftikhar, Paulo Canas Rodrigues: Validation, formal analysis, and investigation; Sohail Ahmad, Moiz Qureshi, and Hasnain Iftikhar: Resources, Data Duration; Sohail Ahmad, Moiz Qureshi, Hasnain Iftikhar, Paulo Canas Rodrigues, and Mohd Ziaur Rehman: Writing-original draft preparation, writing-review and editing; Hasnain Iftikhar, Paulo Canas Rodrigues, and Mohd Ziaur Rehman: Supervision, Project Administration, and Funding acquisition. All authors have read and approved the final version of the manuscript for publication.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in creating this article.

Acknowledgments

The authors extend their sincere appreciation to the Researchers Supporting Project number (RSPD2025R1038), King Saud University, Riyadh, Saudi Arabia.

Funding

Researchers Supporting Project number (RSPD2025R1038), King Saud University, Riyadh, Saudi Arabia.

Conflict of interest

The authors declare no conflicts of interest.

References

1. H. O. Hartley, A. Ross, Unbiased ratio estimators, *Nature*, **174** (1954), 270–271. http://dx.doi.org/10.1038/174270a0
2. D. S. Robson, Applications of multivariate polykays to the theory of unbiased ratiotype estimation, *J. Amer. Statist. Assoc.*, **52** (1957), 511–522. http://dx.doi.org/10.1080/01621459.1957.10501407
3. Y. Rong, Z. Xu, J. Liu, H. Liu, J. Ding, X. Liu, et al., Du-bus: A real-time bus waiting time estimation system based on multi-source data, *IEEE Trans. Intell. Transp. Syst.*, **23** (2022), 24524–24539. https://doi.org/10.1109/TITS.2022.3210170
4. M. Murthy, N. Nanjamma, Almost unbiased ratio estimates based on interpenetrating sub-sample estimates, *Sankhya: Indian J. Stat.*, 1959, 381–392. Available from: https://www.jstor.org/stable/25048435
5. Z. Zhou, X. Zhou, H. Qi, N. Li, C. Mi, Near miss prediction in commercial aviation through a combined model of grey neural network, *Expert Syst. Appl.*, **255** (2024), 124690. http://dx.doi.org/10.1016/j.eswa.2024.124690

6. B. Sisodia, V. Dwivedi, Modified ratio estimator using coefficient of variation of auxiliary variable, *J. Indian Soc. Agric. Stat.*, 1981.
7. R. Biradar, H. Singh, On a class of almost unbiased ratio estimators, *Biom. J.*, **34** (1992), 937–944. Available from: `https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.4710340807`
8. L. N. Upadhyaya, H. P. Singh, Use of transformed auxiliary variable in estimating the finite population mean, *J. Math. Meth. Biosci.*, **41** (1999), 627–636.
9. R. Tailor, B. Sharma, A modified ratio-cum-product estimator of finite population mean using known coefficient of variation and coefficient of kurtosis, *Population*, **10** (2009).
10. B. Xu, Y. Guo, A novel DVL calibration method based on robust invariant extended Kalman filter, *IEEE Trans. Veh. Technol.*, **71** (2022), 9422–9434. http://dx.doi.org/10.1109/TVT.2022.3182017
11. R. L. Chambers, R. Dunstan, Estimating distribution functions from survey data, *Biometrika*, **73** (1986), 597–604. https://doi.org/10.1093/biomet/73.3.597
12. R. Chambers, A. Dorfman, P. Hall, Properties of estimators of the finite population distribution function , *Biometrika*, **79** (1992), 577–582. https://doi.org/10.1093/biomet/79.3.577
13. A. Y. Kuk, Estimation of distribution functions and medians under sampling with unequal probabilities, *Biometrika*, **75** (1988), 97–103. https://doi.org/10.1093/biomet/75.1.97
14. A. Y. Kuk, T. Mak, Median estimation in the presence of auxiliary information, *J. R. Stat. Soc. Ser. B. Meth.*, **51** (1989), 261–269. https://doi.org/10.1111/j.2517-6161.1989.tb01763.x
15. M. Rueda, A. Arcos, Improving ratio-type quantile estimates in a finite population, *Stat. Pap.*, **45** (2004), 231–248. https://doi.org/10.1007/BF02777225
16. M. Rueda, A. Arcos, M. Martínez-Miranda, Y. Román, Some improved estimators of finite population quantile using auxiliary information in sample surveys, *Comput. Stat. Data Anal.*, **45** (2004), 825–848. https://doi.org/10.1016/S0167-9473(03)00097-5
17. A. Dorfman, A comparison of design-based and model-based estimators of the finite population distribution function, *Austral. J. Stat.*, **35** (1993), 29–41. https://doi.org/10.1111/j.1467-842X.1993.tb01310.x
18. M. Ahmed, W. Abu-Dayyeh, Estimation of finite-population distribution function using multivariate auxiliary information, *Stat. Transit.*, **5** (2001), 501–507.
19. S. Ahmad, J. Shabbir, Use of extreme values to estimate finite population mean under PPS sampling scheme, *J. Reliab. Stat. Stud.*, 2018, 99–112. Available from: `https://journals.riverpublishers.com/index.php/JRSS/article/view/20877`
20. S. Hussain, M. Zichuan, S. Hussain, A. Iftikhar, M. Asif, S. Akhtar, S. Ahmad, On estimation of distribution function using dual auxiliary information under nonresponse using simple random sampling , *J. Probab. Stat.*, **2020** (2020). https://doi.org/10.1155/2020/1693612
21. S. Hussain, S. Ahmad, M. Saleem, S. Akhtar, Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling, *PLoS One*, **15** (2020). https://doi.org/10.1371/journal.pone.0239098
22. S. Hussain, S. Ahmad, S. Akhtar, A. Javed, U. Yasmeen, Estimation of finite population distribution function with dual use of auxiliary information under non-response, *PLoS One*, **15** (2020). https://doi.org/10.1371/journal.pone.0243584

23. S. Ahmad, M. Arslan, A. Khan, J. Shabbir, A generalized exponential-type estimator for population mean using auxiliary attributes, *PLoS One*, **16** (2021), e0246947. https://doi.org/10.1371/journal.pone.0246947
24. S. Bahl, R. Tuteja, Ratio and product type exponential estimators, *J. Inform. Optim. Sci.*, **12** (1991), 159–164. https://doi.org/10.1080/02522667.1991.10699058
25. H. P. Singh, S. Kumar, A regression approach to the estimation of the finite population mean in the presence of non-response, *Aust. N. Z. J. Stat.*, **50** (2008), 395–408. https://doi.org/10.1111/j.1467-842X.2008.00525.x
26. C. Kadilar, H. Cingi, Ratio estimators in stratified random sampling, *Biometrical J.* **45** (2003), 218–225. https://doi.org/10.1002/bimj.200390007
27. S. Singh, *Advanced sampling theory with applications: How michael "Selected"*, Springer Science & Business Media, **2** (2003). https://doi.org/10.1007/978-94-007-0789-4
28. M. H. Hansen, W. N. Hurwitz, The problem of non-response in sample surveys, *J. Amer. Statist. Assoc.*, **41** (1946), 517–529. https://doi.org/10.1080/01621459.1946.10501894
29. C. Long, W. Chen, R. Yang, D. Yao, Ratio estimation of the population mean using auxiliary information under the optimal sampling design, *Probab. Engrg. Inform. Sci.*, **36** (2022), 449–460. https://doi.org/10.1017/S0269964820000625
30. L. Chen, T. Chen, T. Lan, C. Chen, J. Pan, The contributions of population distribution, healthcare resourcing, and transportation infrastructure to spatial accessibility of health care, *INQUIRY*, **60** (2023), 1438227527. https://doi.org/10.1177/00469580221146041
31. I. Hussain, M. Qureshi, M. Ismail, H. Iftikhar, J. Zywiołek, J. L. Lopez-Gonzales, Optimal features selection in the high dimensional data based on robust technique: Application to different health databases, *Heliyon*, **10** (2024), 17. https://doi.org/10.1016/j.heliyon.2024.e37241
32. S. M. Gonzales, H. Iftikhar, J. L. Lopez-Gonzales, Analysis and forecasting of electricity prices using an improved time series ensemble approach: An application to the Peruvian electricity market, *AIMS Math.* **9** (2024), 21952–21971. http://dx.doi.org/10.3934/math.20241067
33. F. A. Almulhim, H. M. Aljohani, R. Aldallal, M. S. Mustafa, M. M. Alsolmi, A. Elshenawy, et al., Estimation of finite population mean using dual auxiliary information under non-response with simple random sampling, *Alexandria Engrg. J.*, **100** (2024), 286–299. https://doi.org/10.1016/j.aej.2024.04.058
34. M. Usman, G. N. Singh, J. Kumar, A. B. SM, An improved class of predictive estimators of population distribution function in the presence of non-response, *Heliyon*, **10** (2024), 18. https://doi.org/10.1016/j.heliyon.2024.e37232
35. M. Hussain, Q. Zaman, L. Khan, A. E. Metawa, F. A. Awwad, E. A. Ismail, et al., Improved exponential-type mean estimators for non-response case using two concomitant variables in simple random sampling, *Heliyon*, **10** (2024), 6. https://doi.org/10.1016/j.heliyon.2024.e27535

AIMS Press