INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Check for updates

OPEN ACCESS

Received: 20-11-2023

Accepted: 30-01-2024

Published: 23-02-2024

Citation: Garg N, Patel A, Pachori M (2024) Calibration Estimation of Population Proportion in Probability Proportional to Size Sampling in the Presence of Non-Response. Indian Journal of Science and Technology 17(9): 841-851. https://doi.org/10.17485/IJST/v17i9.2938

* **Corresponding author**.

pachorim07@gmail.com

Funding: None

Competing Interests: None

Copyright: © 2024 Garg et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Calibration Estimation of Population Proportion in Probability Proportional to Size Sampling in the Presence of Non-Response

Neha Garg[1], Anant Patel[1], Menakshi Pachori[2]*

1 School of Sciences, Indira Gandhi National Open University, 110068, New Delhi, India
2 Sanskriti Ayurvedic Medical College & Hospital, Sanskriti University, Mathura, 281401, Uttar Pradesh, India

Abstract

Objectives: In this article, we addressed the problem of estimation of the finite population proportion under the Probability Proportional to Size (PPS) sampling technique, when the complete information is unavailable due to the presence of non-response. We developed calibrated estimators of the population proportion under PPS sampling in the presence of nonresponse based on the availability of auxiliary information. **Methods:** The expressions for the mean squared errors of the suggested estimators were developed to the first order of approximation. The developed estimators of the population proportion are compared with the design-based Horvitz-Thompson estimator and Horvitz-Thompson type calibration estimator which were obtained on the complete response units along with the design-based Hansen and Hurwitz type estimator in the presence of non-response. A Simulation study has also been conducted to support the performance of the developed estimators of population proportion with the help of two real datasets, by computing Percentage Absolute Relative Bias (%ARB) and Percentage Relative Root Mean Squared Error (%RRMSE) using R software. **Findings:** The simulation study supported the performance of the developed estimators of the finite population proportion based on %ARB and %RRMSE. The proposed calibration estimators of population proportion are more efficient than the other considered estimators in the presence of non-response. **Novelty:** The proposed new calibrated estimators have practical implications in the estimation of finite population proportions.

Keywords: Auxiliary information; Calibration Approach; Nonresponse; Population proportion; PPS sampling

1 Introduction

In many real-life sample surveys, it is not possible to obtain complete information on all the units. Over the years, researchers have been fascinated with the problem of estimating the various parameters of interest in the presence of non-response. To deal with this issue of nonresponse, the sub-sampling scheme for non-respondents was initiated [1]. The problem of nonresponse was dealt with in mailed questionnaires [2]. The calibrated estimator of population mean was suggested in the presence of unit non-response with the help of non-response adjustment factor on using different models (inverse linear, exponential, logistic) on using auxiliary variables [3].

An improved class of estimators has been obtained for the estimation of population mean based on Probability Proportional to size (PPS) sampling, with the help of two auxiliary variables [4]. The problem of estimating the population mean of the study variable with the help of auxiliary variables under varying probability techniques was focused with and without measurement errors [5].

Several approaches were explored to develop calibration estimators that were suitable for survey data affected by non-response or missing data, where auxiliary variables were available at both the population level and the panel level [6]. The popular calibration estimator of population total in PPS sampling was suggested when the auxiliary information is assumed to be known [7]. The population proportion in the case of the PPS sampling scheme under the calibration framework was estimated in [8]. There are several other contributions made continuously to enhance the estimation of parameters under different sampling techniques. Some of them are highlighted in [9–11].

In this paper, the problem of estimation of population proportion of the study attribute is considered using the information available on the auxiliary attribute in the presence of unit non-response. These suggested estimators are more efficient than the design-based HorvitzThompson type estimator and Horvitz-Thompson type calibration estimator determined on the complete response units together with the design-based Hansen and Hurwitz type estimator in the presence of non-response. Section 2 reviews the theoretical development of the estimators of the population proportion. In Section 3, the calibration estimators of population proportion are proposed in the presence of unit non-response and the expressions of mean squared errors of the suggested estimators have been derived. In Section 4, a simulation study is carried out based on two real datasets while Section 5 concludes the results.

2 Methodology and Theoretical Developments

Let $U = \{U_1, U_2, \ldots, U_N\}$ be a finite population of size N and a sample $s = \{s_1, s_2, \ldots, s_n\}$ of size n is selected from population $'U'$ with a specified sampling design $'d'$. Assume that the inclusion probabilities of the j^{th} and jk^{th} units to be included in the sample, i.e., π_j and π_{jk} are always positive.

If $'A'$ is an attribute of primary interest from population $'U'$, we define

$$A_k = \begin{cases} 1;\ when\ attribute\ occures\ in\ the\ k^{th} unit \\ 0;\ otherwise \end{cases}$$

The population proportion of attribute $'A'$ is defined as:

$$P_A = \frac{1}{N} \sum_{k \in U} A_k \tag{1}$$

A design-weighted Horvitz-Thompson estimator of population proportion P_A of the study attribute can be defined as:

$$\widehat{P}_{AHT\pi} = \frac{1}{N} \sum_{k \in s} d_k A_k \tag{2}$$

where $d_k = \frac{1}{\pi_k}$

The variance of $\widehat{P}_{AHT\pi}$ of the population proportion is given as:

$$V\left(\widehat{P}_{AHH\pi}\right) = \frac{1}{N^2} \sum_{j \in U} \sum_{k \in U} \Delta_{jk} \left(d_j A_j\right) \left(d_k A_k\right)$$

where, $\Delta_{jk} = \pi_{jk} - \pi_j \pi_k$

Suppose there exists an auxiliary attribute $'B'$ associated with the primary attribute $'A'$ and the values of attribute $'B'$ are known for each unit of the sample $'s'$. The population proportion P_B is also known and given as:

$$P_B = \frac{1}{N} \sum_{k \in U} B_k \tag{3}$$

A calibrated estimator of the population proportion of the attribute $'A'$ based on the auxiliary attribute $'B'$ [8] following[7] is given as:

$$\widehat{P}_{cAHT\pi} = \frac{1}{N}\sum_{k\in S} w_k A_k \tag{4}$$

The calibration weights w_k, are obtained after minimizing the chi-square distance function:

$$\chi^2 = \sum_{k\in s} \frac{(w_k - d_k)^2}{d_k q_k} \tag{5}$$

Subject to the calibration constraint

$$P_B = \frac{1}{N}\sum_{k\in U} B_k = \frac{1}{N}\sum_{k\in S} w_k B_k \tag{6}$$

where q_k is positive and known and $0 < P_B < 1$.

The optimized calibrated weights w_k are given as:

$$w_k = d_k + \frac{\lambda d_k q_k B_k}{N} \tag{7}$$

where, $\lambda = \frac{N^2\left(P_B - \widehat{P}_{BH}\right)}{\sum_{k\in S} d_k q_k B_k}$, is a Lagrange multiplier.

$\widehat{P}_{BH} = \frac{1}{N}\sum_{k\in S} d_k B_k$ is the well-known Horvitz-Thompson estimator of population proportion for attribute B. The estimator given in Equation (4) is reduced to:

$$\widehat{P}_{cAHT\pi} = \widehat{P}_{AHT\pi} + \frac{\left(P_B - \widehat{P}_{BH}\right)}{\sum_{k\in S} d_k q_k B_k}\sum_{k\in S} d_k q_k B_k A_k \tag{8}$$

The expression for asymptotic variance of $\widehat{P}_{cAHT\pi}$ is given as:

$$V\left(\widehat{P}_{cAHT\pi}\right) = \frac{1}{N^2}\sum_{j\in U}\sum_{k\in U} \Delta_{jk}\left(d_j E_j\right)\left(d_k E_k\right)$$

were, $\Delta_{jk} = \pi_{jk} - \pi_j \pi_k; E_j = A_j - D_j$ and $D = \frac{\sum_{j\in U} q_j B_j A_j}{\sum_{j\in U} q_j B_j}$

The expression for an estimator of the variance is derived as:

$$\widehat{V}\left(\widehat{P}_{cAHT\pi}\right) = \frac{1}{N^2}\sum_{j\in U}\sum_{k\in U} \frac{\Delta_{jk}}{\pi_{jk}}\left(d_j e_j\right)\left(d_k e_k\right)$$

where, $e_j = A_j - \widehat{D}B_j$ and $\widehat{D} = \frac{\sum_{j\in s} d_j q_j A_j B_j}{\sum_{k\in s} d_j q_j B_j}$

3 Proposed Estimators in the Presence of Non-response

The problem of non-response occurs more commonly in sample surveys. A sub-sampling plan is adopted to deal with the problem of non-response[1]. In this plan, a sampling scheme consisting of mail surveys is considered at the first attempt while the personal interviews are at the second attempt. In this approach, a questionnaire is mailed to the randomly selected sample of size n. Then again, a sub-sample has been selected from n_{nr} non-responding units and then these selected sampling units are contacted through personal interviews. In this method, it is also assumed that the population consists of a response group, say U_r of size N_r, and the non-response group, say U_{nr} of size N_{nr}, where $N_{nr} = (N - N_r)$.

A sample $'s'$ of size $'n'$ is selected from the population of size N by probability proportion to size sampling technique with known sampling design $p(s)$. Let $\pi_j = P(j \in s)$ and $\pi_{jk} = P(j\&k \in s); j \neq k$, indicate the first-order and second-order inclusion probabilities of the j^{th} unit and pair of the $(j,k)^{th}$ units, respectively. In sample surveys, it may be difficult to obtain complete responses on all sampling units in many real-life situations. Suppose the sample s results into response subset S_r of size n_r who responded in the first attempt and non-response subset s_{nr} $(= s - s_r)$ of size n_{nr} $(= n - n_r)$ who did not provide complete response in the first attempt, such that $s = s_r \cup s_{nr}$ and $n = n_r + n_{nr}$.

To deal with the non-response, the non-response sample s_{nr} was followed up to gather the information as much as possible. Suppose it responds subsample $s_{nr}^{'}$ of size $n_{nr}^{'}$ from nonresponse sample s_{nr} (such that $s_{nr}^{'} \subseteq s_{nr}$) using the design $p(.\mid s_{nr})$ with positive inclusion probabilities denoted by $\pi_{j||s_r}$ and $\pi_{jk|_s}$; $j \neq k \in S_{nr}$.

Let $'A'$ and $'B'$ denote the study attribute and the auxiliary attribute, respectively. The HH type estimator [1] for the population proportion of the study attribute $'P_A'$ can be defined as:

$$\widehat{P}_{AHH\pi} = \frac{1}{N}\Sigma_s d_k^* A_k \tag{9}$$

where, $d_k^* = \frac{1}{\pi_k^*}$ and $\pi_k^* = \begin{cases} \pi_k \,;\, if\ k \in S_r \\ \pi_k \pi_{k|s_n};\ if\ k \in s_{nr}^{'} \end{cases}$

The estimator $\widehat{P}_{AHH\pi}$ can be rewritten as:

$$\widehat{P}_{AHH\pi} = \frac{1}{N}\left(\Sigma_{s_r} A_k^{'} + \Sigma_{s_{rr}^{'}} A_k^{''}\right) \tag{10}$$

where $A_k^{'} = \frac{A_k}{\pi_k}$ and $A_k^{''} = \frac{A_k}{\pi_k \pi_{k|s_s}}$

The expected value of the estimator $\widehat{P}_{AHH\pi}$ is given as:

$$E\left(\widehat{P}_{AHH\pi}\right) = E\left[E_r\left\{E_{nr}\left(\widehat{P}_{AHH\pi} \mid s, s_{nr}\right)\right\}\right] = P_A$$

This implies the estimator $\widehat{P}_{AHH\pi}$ is an unbiased estimator of population proportion P_A. The expression for variance of the estimator $\widehat{P}_{AHH\pi}$ of the population proportion is given as:

$$V(\widehat{P}_{AHH\pi}) = \frac{1}{N^2}\left[\sum_{j=1}^{N_1}\sum_{k=1}^{N_1}\Delta_{jk}\frac{A_j}{\pi_j}\frac{A_k}{\pi_k} + \sum_{j=1}^{N_2}\sum_{k=1}^{N_2}\Delta_{jk}\frac{A_j}{\pi_j}\frac{A_k}{\pi_k} + 2\sum_{j=1}^{N_1}\sum_{k=1}^{N_2}\Delta_{jk}\frac{A_j}{\pi_j}\frac{A_k}{\pi_k} + EE_r\left\{\sum_{j=1}^{s_{nr}}\sum_{k=1}^{s_{nr}}\Delta_{jk|s_{nr}}\frac{A_j}{\pi_j\pi_{j|s_{nr}}}\frac{A_k}{\pi_k\pi_{k|s_{nr}}}\right\}\right]$$

The estimate of $V\left(\widehat{P}_{AHH\pi}\right)$ is defined as:

$$\widehat{V}\left(\widehat{P}_{AHH\pi}\right) = \frac{1}{N^2}\left[\sum_{j=1}^{s_r}\sum_{k=1}^{s_r}\frac{\Delta_{jk}}{\pi_{jk}^*}\frac{A_j}{\pi_j}\frac{A_k}{\pi_k} + \sum_{j=1}^{s_m}\sum_{k=1}^{s_n}\frac{\Delta_{jk}}{\pi_{jk}^*}\frac{A_j}{\pi_j}\frac{A_k}{\pi_k} + 2\sum_{j=1}^{s_r}\sum_{k=1}^{s_{nr}}\frac{\Delta_{jk}}{\pi_{jk}^*}\frac{A_j}{\pi_j}\frac{A_k}{\pi_k} + \sum_{j=1}^{s_{nr}^{'}}\sum_{k=1}^{s_{nr}^{'}}\frac{\Delta_{jk|s_{nr}}}{\pi_{jk|s_{nr}}^*}\frac{A_j}{\pi_j\pi_{j|s_{nr}}}\frac{A_k}{\pi_k\pi_{k|s_{nr}}}\right]$$

$$\pi_{jk}^* = \begin{cases} \pi_{jk}\pi_{jk||nr} & ;\ if\ j,k \in S_{nr} \\ \pi_{jk}\pi_{j|s_{nr}} & ;\ if\ j \in S_{nr}, k \in S_r \\ \pi_{jk}\pi_{k|s_{nr}} & ;\ if\ j \in S_r, k \in S_{nr} \\ \pi_{jk} & ;\ if\ j,k \in S_r \end{cases}$$

The auxiliary attribute which is highly correlated with the study attribute may be available for (i) only the response group, (ii) only the non-response group, and (iii) both groups. In this paper, we propose three calibration estimators in these cases. Let us discuss these cases one by one.

3.1 When auxiliary information is available for the Response Group

Let us consider the case when auxiliary information is available for the response group s_r only.

Let us define the design weights $d_k = \frac{1}{\pi_k}$, and $d_{k|n_{nr}} = \frac{1}{\pi_{k|s_{nr}}}$. The population proportion of auxiliary attributes which is assumed to be known in the response group is given as:

$$P_{Br} = \frac{B_r}{N_r} = \frac{1}{N_r}\Sigma_{k\in U_r} B_k$$

We now propose the partial calibration estimator of population proportion of the study attribute utilizing auxiliary information which is available for the response group s_r only, given as:

$$\widehat{P}_{cA\pi}^{'} = \frac{1}{N}\left(\Sigma_{k\in S_r}\psi_{kr}A_k + \Sigma_{k\in s_{nr}^{'}} d_k d_{k|s_{nr}} A_k\right) \tag{11}$$

where ψ_{kr} are the calibrated weights in the response group and $d_k.d_{k|s_{nr}} = \frac{1}{\pi_k.\pi_{k|nr}}$ are the design weights in the non-response group.

Here design weights (d_k) of the response group are modified with the help of the calibration technique, by minimizing the chi-square type distance function given as:

$\sum_{k=1}^{s_r} \frac{(\psi_{kr}-d_k)^2}{d_k q_k}$ subject to the calibration constraint defined as:

$$\frac{1}{N_r}\sum_{k\in S_r} \psi_{kr} B_k = \widehat{P}_{Br}$$

$$i.e., \sum_{k\in s_r} \psi_{kr} B_k = B_r \tag{12}$$

To obtain the calibration weights, the Lagrangian function is stated as:

$$\phi_r = \sum_{k\in S_r} \frac{(\psi_{kr}-d_k)^2}{d_k q_k} - \lambda_r \left(\sum_{k\in S_r} \psi_{kr} B_k - B_r\right) \tag{13}$$

After minimization of the Lagrangian function (Equation (13)) subject to the calibration constraint given in Equation (12), the expression of calibrated weights is obtained as follows:

$$\psi_{kr} = d_k + \left(B_r - \sum_{k\in S_r} d_k B_k\right) \frac{1}{\sum_{k\in S_r} q_k d_k B_k^2} q_k d_k B_k \tag{14}$$

After substituting the calibrated weights derived in Equation (14), the proposed calibration estimator of the population proportion of the study attribute is determined as:

$$\widehat{P}'_{cA\pi} = \frac{1}{N}\left[\sum_{k\in S_r}\left\{d_k + \left(B_r - \sum_{k\in s_r} d_k B_k\right)\frac{1}{\sum_{k\in S_r} q_k d_k B_k^2} \cdot q_k d_k B_k\right\} A_k + \sum_{k\in S'_{nr}} d_k d_{k|s_{nr}} A_k\right]$$

or,

$$\widehat{P}'_{cA\pi} = \frac{1}{N}\left[\sum_{k\in S_r} d_k A_k + \widehat{\beta}'_{c\pi} \cdot \left(B_r - \sum_{k\in S_r} d_k B_k\right) + \sum_{k\in S'_{nr}} d_k d_{k|s_{nr}} A_k\right] \tag{15}$$

where, $\widehat{\beta}'_{c\pi} = \frac{\sum_{k\in s_r} q_k d_k A_k B_k}{\sum_{k\in s_r} q_k d_k B_k^2}$

Here q_k is a suitably chosen constant and different calibration estimators can be obtained depending upon the chosen values of q_k.

Now, the expression of variance of the suggested calibration estimator $\widehat{P}'_{cA\pi}$ after neglecting second and higher-order terms can be derived as:

$$V\left(\widehat{P}'_{cA\pi}\right) = \frac{1}{N^2}\left(\sum_{j=1}^{N_r}\sum_{k=1}^{N_r} \Delta_{jk}\frac{\varepsilon'_j}{\pi_j}\frac{\varepsilon'_k}{\pi_k} + \sum_{j=1}^{N_{nr}}\sum_{k=1}^{N_{nr}} \Delta_{jk}\frac{A_j}{\pi_j}\frac{A_k}{\pi_k}\right.$$
$$\left. + 2\sum_{j=1}^{N_r}\sum_{k=1}^{N_{nr}} \Delta_{jk}\frac{\varepsilon_j}{\pi_j}\frac{A_k}{\pi_k} + EE_r\left\{\sum_{j=1}^{s_{nr}}\sum_{k=1}^{s_{nr}} \Delta_{jk|s_{nr}}\frac{A_j}{\pi_j\pi_{j|s_{nr}}}\frac{A_k}{\pi_k\pi_{k|s_{nr}}}\right\}\right)$$

where, $\varepsilon'_k = A_k - \beta'_{c\pi} B_k$ and $\beta'_{c\pi} = \frac{\sum_{k\in U_r} q_k d_k A_k B_k}{\sum_{k\in U_r} q_k d_k B_k^2}$

Now expression for the estimator of variance of the proposed estimator is obtained as follows:

$$\widehat{V}\left(\widehat{P}'_{cA\pi}\right) = \frac{1}{N^2}\left(\sum_{j=1}^{s_r}\sum_{k=1}^{s_r} \frac{\Delta_{jk}}{\pi'_{jk}}\frac{e'_j}{\pi_j}\frac{e'_k}{\pi_k} + \sum_{j=1}^{s_{nr}}\sum_{k=1}^{s_{nr}} \frac{\Delta_{jk}}{\pi'_{jk}}\frac{A_j}{\pi_j}\frac{A_k}{\pi_k}\right.$$
$$\left. + 2\sum_{j=1}^{s_r}\sum_{k=1}^{s_{nr}} \frac{\Delta_{jk}}{\pi'_{jk}}\frac{e'_j}{\pi_j}\frac{A_k}{\pi_k} + \left\{\sum_{j=1}^{s'_{nr}}\sum_{k=1}^{s'_{nr}} \frac{\Delta_{jk|snr}}{\pi_{jk||nr}}\frac{A_j}{\pi_j\pi_{j|s_{nr}}}\frac{A_k}{\pi_k\pi_{k|s_{nr}}}\right\}\right)$$

where, $e'_k = A_k - \widehat{\beta}'_{c\pi} B_k$ and $\pi'_{jk} = \begin{cases} \pi_{jk}\pi_{jk|_{nr}} & ; if\ j,k \in S_{nr} \\ \pi_{jk}\pi_{j|_{nr}} & ; if\ j \in S_{nr}, k \in S_r \\ \pi_{jk}\pi_{k|_{s_{nr}}} & ; if\ j \in S_r, k \in S_{nr} \\ \pi_{jk} & ; if\ j,k \in S_r \end{cases}$

3.2 When auxiliary information is available for Non-Response Group

Let us consider the case when the complete auxiliary information is available for only the nonresponse group s_{nr}. Let us define the design weights $d_{k|sr} = \frac{1}{\pi_{k|s_{nr}}}$. We define the population proportion of the auxiliary attribute which is assumed to be known in the non-response group as:

$$P_{Bnr} = \frac{B_{nr}}{N_{nr}} = \frac{1}{N_{nr}} \sum_{k \in U_{nr}} B_k$$

The partially calibrated estimator of population proportion of the study attribute utilizing auxiliary attribute only on the non-response group s_{nr} can be proposed as:

$$\widehat{P}''_{cA\pi} = \frac{1}{N} \left(\sum_{k \in S_r} d_k A_k + \sum_{k \in s'_{nr}} \psi_{knr} A_k \right) \tag{16}$$

Since we have auxiliary information available on the non-response group only, we modify the design weights $d_k d_{k|s_{nr}}$, with the help of the calibration technique. Here the chi-square type distance function $\sum_{k=1}^{s'_{nr}} \frac{\left(\psi_{knr} - d_k d_{k|s_{nr}}\right)^2}{q_k d_k d_{k|s_{nr}}}$ is minimized concerning the calibration constraint defined as:

$$\frac{1}{N_{nr}} \sum_{k \in s'_{nr}} \psi_{knr} B_k = P_{Bnr}$$

$$i.e., \sum_{k \in s'_{nr}} \psi_{knr} B_k = B_{nr} \tag{17}$$

To determine the calibrated weights, the Lagrangian function is formulated as:

$$\phi_{nr} = \sum_{k \in s'_{nr}} \frac{\left(\psi_{knr} - d_k d_{k|s_{nr}}\right)^2}{q_k d_k d_{k|s_{nr}}} - \lambda_{nr} \left(\sum_{k \in s'_{nr}} \psi_{knr} B_k - B_{nr} \right) \tag{18}$$

The calibrated weights after minimization of the Lagrangian function (Equation (18)) subject to calibration constraint given in Equation (17) are obtained as follows:

$$\psi_{knr} = d_k d_{k|s_{nr}} + \left(B_{nr} - \sum_{k \in s'_{nr}} d_k d_{k|s_{nr}} B_k \right) \frac{1}{\sum_{k \in s'_{nr}} q_k d_k d_{k|s_{nr}} B_k^2} \cdot q_k d_k d_{k|s_{nr}} B_k \tag{19}$$

After substituting calibrated weights derived in Equation (19), the proposed calibration estimator (given in Equation (16)) of the population proportion of the study attribute is obtained as:

$$\widehat{P}''_{cA\pi} = \frac{1}{N} \left[\sum_{k \in s'_{nr}} d_k A_k + \sum_{k \in S'_{nr}} \left\{ d_k d_{k|s_{nr}} + \left(B_{nr} - \sum_{k \in s'_{nr}} d_k d_{k||_{nr}} B_k \right) \frac{1}{\sum_{k \in s'_{nr}} q_k d_k d_{k|s_{nr}} B_k^2} \cdot q_k d_k d_{k||_{nr}} B_k \right\} A_k \right]$$

$$\widehat{P}''_{cA\pi} = \frac{1}{N} \left[\sum_{k \in S_r} d_k A_k + \sum_{k \in s'_{nr}} d_k d_{k|s_{nr}} A_k + \widehat{\beta}''_{c\pi} \cdot \left(\widehat{B}_{nr} - \sum_{k \in s'_{nr}} d_k d_{k|s_{nr}} B_k \right) \right] \tag{20}$$

where $\widehat{\beta}''_{cr} = \frac{\sum_{k \in s'_{nr}} q_k d_k d_{k|s_{nr}} A_k B_k}{\sum_{k \in s'_{nr}} q_k d_k d_{k|s_{nr}} B_k^2}$

and q_k is a suitably chosen constant. The different forms of calibration estimators can be obtained for different values of q_k. The expression of variance of $\widehat{P}''_{AA\pi}$ after neglecting the second and higher-order terms is obtained as:

$$V\left(\widehat{P}''_{cA\pi}\right) = \frac{1}{N^2} \left[\sum_{j=1}^{N_r} \sum_{k=1}^{N_r} \Delta_{jk} \frac{A_j}{\pi_j} \frac{A_k}{\pi_k} + \sum_{j=1}^{N_{nr}} \sum_{k=1}^{N_{nr}} \Delta_{jk} \frac{\varepsilon''_j}{\pi_j} \frac{\varepsilon''_k}{\pi_k} + 2 \sum_{j=1}^{N_r} \sum_{k=1}^{N_{nr}} \Delta_{jk} \frac{A_j}{\pi_j} \frac{\varepsilon''_k}{\pi_k} + E_r \left\{ \sum_{j=1}^{s_{nr}} \sum_{k=1}^{s_{nr}} \Delta_{jk|s_{nr}} \frac{\varepsilon''_j}{\pi_j \pi_{j|s_{nr}}} \frac{\varepsilon''_k}{\pi_k \pi_{k|s_{nr}}} \right\} \right]$$

where $\varepsilon''_k = A_k - \beta''_{c\pi} B_k$ and $\beta''_{c\pi} = \frac{\sum_{k \in U_{nr}} q_k d_k d_{k|s_{nr}} A_k B_k}{\sum_{k \in U_{nr}} q_k d_k d_{k|s_{nr}} B_k^2}$

The estimator of $V\left(\widehat{P}''_{c\pi}\right)$ can be written as:

$$\widehat{V}\left(\widehat{P}''_{cA\pi}\right)=\frac{1}{N^2}\left[\sum_{j=1}^{s_r}\sum_{k=1}^{s_r}\frac{\Delta_{jk}}{\pi''_{jk}}\frac{A_j}{\pi_j}\frac{A_k}{\pi_k}+\sum_{j=1}^{s_{nr}}\sum_{k=1}^{s_{nr}}\frac{\Delta_{jk}}{\pi''_{jk}}\frac{e''_j}{\pi_j}\frac{e''_k}{\pi_k}\right.$$
$$\left.+2\sum_{j=1}^{s_r}\sum_{k=1}^{s_{nr}}\frac{\Delta_{jk}}{\pi''_{jk}}\frac{A_j}{\pi_j}\frac{e''_k}{\pi_k}+\left\{\sum_{j=1}^{s'_{nr}}\sum_{k=1}^{s'_{nr}}\frac{\Delta_{jk|s_{nr}}}{\pi''_{jk|s_{nr}}}\frac{e''_j}{\pi_j\pi_{j|s_{nr}}}\frac{e''_k}{\pi_k\pi_{k|s_{nr}}}\right\}\right]$$

where

$$e''_k=A_k-\widehat{\beta}''_{c\pi}B_k \; and \pi''_{jk}=\begin{cases}\pi_{jk}\pi_{jk|s_{nr}} & ;\, if\ j,k\in S_{nr}\\ \pi_{jk}\pi_{j|s_{nr}} & ;\, if\ j\in S_{nr},k\in S_r\\ \pi_{jk}\pi_{k|s_{nr}} & ;\, if\ j\in S_r,k\in S_{nr}\\ \pi_{jk} & ;\, if\ j,k\in S_r\end{cases}$$

3.3 When auxiliary information is available for Response Group and Non-Response Group

Let us consider the case when an auxiliary information is available for both the response group S_r and the non-response group S_{nr}. Let us define the design weights for response and nonresponse groups, respectively. The population proportions of auxiliary attributes are P_{Br} and P_{Bnr} which are assumed to be known in both response and non-response groups, respectively.

We now propose the calibration estimator of population proportion of the study attribute utilizing known available auxiliary information in response as well as non-response groups given as:

$$\widehat{P}^{*}_{cA\pi}=\frac{1}{N}\left(\sum_{k\in S_r}\psi^{*}_{kr}A_k+\sum_{k\in S'_{nr}}\psi^{*}_{knr}A_k\right)\tag{21}$$

where ψ^{*}_{kr} and ψ^{*}_{knr} are the calibration weights for the response and non-response groups, respectively.

Here we consider two chi-square type distance functions corresponding to both response and non-response groups:

For the response group:

$$\sum_{k=1}^{s_r}\frac{\left(\psi^{*}_{kr}-d_k\right)^2}{d_kq_k}\tag{22}$$

For the non-response group:

$$\sum_{k=1}^{s'_{nr}}\frac{\left(\psi^{*}_{knr}-d_kd_{k|s_{nr}}\right)^2}{q_kd_kd_{k|s_{nr}}}\tag{23}$$

Let us define the calibration constraints corresponding to response and non-response groups as:

For response group:

$$\sum_{k\in s_r}\psi^{*}_{kr}B_k=B_r\tag{24}$$

For -non-response group:

$$\sum_{k\in s'_{nr}}\psi^{*}_{knr}B_k=B_{nr}\tag{25}$$

The Lagrangian function considering both distance functions (given in Equations (22) and (23)) as well as calibration constraints defined in Equations (24) and (25), can be formulated as:

$$\phi^{*}=\sum_{k\in S_r}\frac{\left(\psi^{*}_{kr}-d_k\right)^2}{d_kq_k}+\sum_{k\in s'_{nr}}\frac{\left(\psi^{*}_{knr}-d_kd_{k|s_{nr}}\right)^2}{q_kd_kd_{k|s_{nr}}}-\lambda^{*}_r\left(\sum_{k\in s'_{nr}}\psi^{*}_{kr}B_k-B_r\right)-\lambda^{*}_{nr}\left(\sum_{k\in s'_{nr}}\psi^{*}_{knr}B_k-B_{nr}\right)\tag{26}$$

After minimization of Lagrange's Equation (26) subject to calibration constraints given in Equations (24) and (25), respectively, the calibration weights are obtained as follows:

$$\psi^{*}_{kr}=d_k+\left(B_r-\sum_{k\in S_r}d_kB_k\right)\frac{1}{\sum_{k\in S_r}q_kd_kB_k^2}q_kd_kB_k\tag{27}$$

$$\psi_{knr}^{*} = d_k d_{k|s_{nr}} + \left(B_{nr} - \sum_{k\in s'_{nr}} d_k d_{k|s_{nr}} B_k\right) \frac{1}{\sum_{k\in s'_{nr}} q_k d_k d_{k|s_{nr}} B_k^2} \cdot q_k d_k d_{k|s_{nr}} B_k \tag{28}$$

With the help of updated sets of the calibrated weights ψ'_{kr} and ψ''_{kn}, the calibrated estimator $\widehat{P}^{*}_{cA\pi}$ of population proportion of the study attribute can be articulated as:

$$\begin{aligned}\widehat{P}^{*}_{cA\pi} = \frac{1}{N}\Bigg[\sum_{s_r} &\left\{d_k A_k + \left(B_r - \sum_{s_r} d_k B_k\right)\frac{1}{\sum_{s_r} d_k q_k B_k^2} d_k q_k A_k B_k\right\} \\ + \sum_{s'_n r} &\left\{d_k d_{k|s_{nr}} A_k + \left(B_{nr} - \sum_{s'_n r} d_k d_{k|s_{nr}} B_k\right)\frac{1}{\sum_{s'_n r} q_k d_k d_{k|s_{nr}} B_k^2} q_k d_k d_{k|s_{nr}} A_k B_k\right\}\Bigg]\end{aligned} \tag{29}$$

$$\widehat{P}^{*}_{cA\pi} = \frac{1}{N}\left[\sum_{s_f} d_k A_k + \widehat{\beta}^{*}_{c}\sum_{s_f}\left(B_r - \sum_{s_f} d_k B_k\right) + \sum_{s'_{ns}} d_k d_{k|s_r} A_k + \widehat{\beta}^{*}_{crr}\sum_{s'_{nf}}\left(B_{rr} - \sum_{s'_r} d_k d_{k|s_{nr}} B_k\right)\right] \tag{30}$$

where, $\widehat{\beta}^{*}_{cr} = \frac{d_k q_k A_k B_k}{\sum_{s_r} d_k q_k B_k^2}$ and $\widehat{\beta}^{*}_{cnr} = \frac{1}{\sum_{s'_{nr}} d_k d_{k|s_{nr}} q_k B_k^2} d_k d_{k|s_{nr}} q_k A_k B_k$

Since $E\left[E_r\left\{E_{nr}\left(P^{*}_{cA\pi} \mid s_r, s_{nr}\right)\right\}\right] = P$, this shows that the estimator $\widehat{P}^{*}_{cA\pi}$ is an unbiased estimator of population proportion.

Now, the variance of $\widehat{P}^{*}_{cA\pi}$ after neglecting the second and higher-order terms is expressed as follows:

$$\begin{aligned}V\left(\widehat{P}^{*}_{cA\pi}\right) = \frac{1}{N^2}\Bigg(&\sum_{j=1}^{N_r}\sum_{k=1}^{N_r}\Delta_{jk}\frac{\varepsilon^{*}_{rj}}{\pi_j}\frac{\varepsilon^{*}_{rk}}{\pi_k} + \sum_{j=1}^{N_{nr}}\sum_{k=1}^{N_{nr}}\Delta_{jk}\frac{\varepsilon^{*}_{nrj}}{\pi_j}\frac{\varepsilon^{*}_{nrk}}{\pi_k} \\ &+ 2\sum_{j=1}^{N_r}\sum_{k=1}^{N_{nr}}\Delta_{jk}\frac{\varepsilon^{*}_{rj}}{\pi_j}\frac{\varepsilon^{*}_{nrk}}{\pi_k} + E_r\left\{\sum_{j=1}^{s_{nr}}\sum_{k=1}^{s_{nr}}\Delta_{jk|s_{nr}}\frac{\varepsilon^{*}_{nrj}}{\pi_j\pi_{j|s_{nr}}}\frac{\varepsilon^{*}_{nrk}}{\pi_k\pi_{k|s_{nr}}}\right\}\Bigg)\end{aligned}$$

where, $\varepsilon^{*}_{rk} = (A_k - \beta^{*}_{cr}B)_k$, $\varepsilon^{*}_{nrk} = (A_k - \beta^{*}_{crr}B_k)$, $\beta^{*}_{cr} = \frac{\sum_{k\in U_r} q_k d_k d_{k|s_{nr}} A_k B_k}{\sum_{k\in U_r} q_k d_k d_{k|s_{nr}} B_k^2}$ and $\beta^{*}_{cnr} = \frac{\sum_{k\in U_{nr}} q_k d_k d_{k|s_{nr}} A_k B_k}{\sum_{k\in U_{nr}} q_k d_k d_{k|s_{nr}} B_k^2}$

Now the expression for the estimator of variance of the proposed estimator is as follows:

$$\begin{aligned}\widehat{V}\left(\widehat{P}_{c\pi}\right) = \frac{1}{N^2}\Bigg(&\sum_{j=1}^{s_r}\sum_{k=1}^{s_r}\frac{\Delta_{jk}}{\pi^{*}_{jk}}\frac{e^{*}_{rj}}{\pi_j}\frac{e^{*}_{rk}}{\pi_k} + \sum_{j=1}^{s_{nr}}\sum_{k=1}^{s_{nr}}\frac{\Delta_{jk}}{\pi^{*}_{jk}}\frac{e^{*}_{nrj}}{\pi_j}\frac{e^{*}_{nrk}}{\pi_k} \\ &+ 2\sum_{j=1}^{s_r}\sum_{k=1}^{s_{nr}}\frac{\Delta_{jk}}{\pi^{*}_{jk}}\frac{e^{*}_{rj}}{\pi_j}\frac{e^{*}_{nrk}}{\pi_k} + \sum_{j=1}^{s'_{nr}}\sum_{k=1}^{s'_{nrr}}\frac{\Delta_{jk|s_{nr}}}{\pi_{jk|s_{nr}}}\frac{e^{*}_{nrj}}{\pi_j\pi_{j||nr}}\frac{e^{*}_{nkk}}{\pi_k\pi_{k|s_{nr}}}\Bigg)\end{aligned}$$

where $e^{*}_{rk} = A_k - \widehat{\beta}^{*}_{cr}B_k, e^{*}_{nrk} = A_k - \widehat{\beta}^{*}_{cnr}B_k$ and $\pi^{*}_{jk} = \begin{cases}\pi_{jk}\pi_{jk||nr} & ; if\ j,k\in S_{nr} \\ \pi_{jk}\pi_{j|s_{nr}} & ; if\ j\in S_{nr}, k\in S_r \\ \pi_{jk}\pi_{k|s_{nr}} & ; if\ j\in S_r, k\in S_{nr} \\ \pi_{jk} & ; if\ j,k\in S_r\end{cases}$

4 Simulation Study

The performance of the proposed estimators is examined by the simulation study carried out on two real datasets. The absolute relative bias (%ARB) and percentage relative root mean squared error (%RRMSE) for the estimators of population proportion are computed as:

$\%ARB(\alpha) = \frac{1}{R}\sum_{i=1}^{R}\left|\frac{(\alpha_i - P_A)}{P_A}\right| \times 100;\ \alpha_i = \widehat{P}_{AHT\pi}, \widehat{P}_{cAHT\pi}, \widehat{P}_{AHH\pi}, \widehat{P}'_{cA\pi}, \widehat{P}''_{cA\pi}$ and $\widehat{P}^{*}_{cA\pi}$

$\%RRMSE(\alpha) = \sqrt{\frac{1}{R}\sum_{i=1}^{R}\left(\frac{(\alpha_i - P_A)}{P_A}\right)^2} \times 100;\ \alpha_i = \widehat{P}_{AHT\pi}, \widehat{P}_{CAHT\pi}, \widehat{P}_{AHH\pi}, \widehat{P}'_{cA\pi}, \widehat{P}''_{cA\pi}$ and $\widehat{P}^{*}_{cA\pi}$

Dataset I: We have taken the MU284 population given in Appendix B [12] ($N = 284$ units). The study and auxiliary variables are converted into categorical variables from the P85 and ME84 variables, respectively, of the MU284 data. The study attribute $'A'$ and auxiliary attribute $'B'$ are considered as:

$$A_k = \begin{cases}1;\ \text{if } P85 > 30 \\ 0;\ \text{otherwise}\end{cases} \quad \text{and } B_k = \begin{cases}1;\ \text{if ME84} > 800 \\ 0;\ \text{otherwise}\end{cases}$$

We have considered P75 as the Z variable used to compute the inclusion probabilities. The non-response units were followed up to get a response on some units.

Dataset II: To examine the performance of the proposed estimators, a simulation study has also been carried out on another real dataset (UC Irvine Machine Learning Repository) [13]. This heart failure clinical records dataset contains the medical records of 299 patients who had heart failure which was collected during their follow-up period, where each patient profile has 13 clinical features. The variables considered here are $zi = age, x = sex$ and $y =$ anemia, and varying sizes of non-response in population (20, 25, 30 and 35) in the sample.

The simulation study was conducted in R software generating 20,000 samples of various sizes. We have considered 20% to 35% of the non-responses. Several samples of varying sizes are drawn from the population of the response group S_r and non-response group S_{nr}.

The values of %ARB and %RRMSE for the suggested estimators ($\widehat{P}'_{cA\pi}$, $\widehat{P}''_{cA\pi}$ and $\widehat{P}^{*}_{cA\pi}$), design based Horvitz-Thompson type estimator ($\widehat{P}_{AHT\pi}$), and calibration Horvitz-Thompson type estimator ($\widehat{P}_{cAHT\pi}$) computed on the complete response units as well as design-based Hansen and Hurwitz type estimator in the presence of non-response ($\widehat{P}_{AHH\pi}$) of population proportion are computed and given in Tables 1, 2, 3 and 4 for both datasets.

Table 1. Absolute Relative Bias (%ARB) for 20% to 35% non-response for Dataset I

$N_2\%$	$n_{a2}\%$	$\widehat{P}_{AHT\pi}$	$\widehat{P}_{cAHT\pi}$	$\widehat{P}_{AHH\pi}$	$\widehat{P}'_{cA\pi}$	$\widehat{P}''_{cA\pi}$	$\widehat{P}_{cAHT\pi}$
20	50	0.2412	0.4644	0.3319	0.1235	0.3188	0.1874
	40	0.3935	0.5724	0.3308	0.1616	0.4512	0.2116
	30	0.5467	0.6804	0.3303	0.2047	0.5887	0.2411
	20	0.7001	0.7885	0.3305	0.2508	0.7272	0.2735
25	50	0.2673	0.4862	0.3537	0.1431	0.3314	0.1998
	40	0.4128	0.5880	0.3531	0.1836	0.4664	0.2325
	30	0.5585	0.6901	0.3525	0.2260	0.5973	0.2612
	20	0.7043	0.7913	0.3504	0.2691	0.6024	0.2836
30	50	0.2611	0.4976	0.3711	0.1422	0.3215	0.1961
	40	0.4020	0.5984	0.3709	0.1803	0.4545	0.2281
	30	0.5514	0.6991	0.3705	0.2276	0.5873	0.2605
	20	0.7016	0.7993	0.3682	0.2757	0.7248	0.2966
35	50	0.2604	0.5169	0.3868	0.1387	0.3215	0.1926
	40	0.4065	0.6115	0.3866	0.1863	0.4543	0.2305
	30	0.5518	0.7064	0.3859	0.2350	0.5870	0.2677
	20	0.7030	0.8058	0.3842	0.2858	0.7247	0.3057

Table 2. Percentage Relative Root Mean Squared Error (%RRMSE) for 20% to 35% non-response for Dataset I

$N_2\%$	$n_{a2}\%$	$\widehat{P}_{AHT\pi}$	$\widehat{P}_{cAHT\pi}$	$\widehat{P}_{AHH\pi}$	$\widehat{P}'_{cA\pi}$	$\widehat{P}''_{cA\pi}$	$\widehat{P}^{*}_{cA\pi}$
20	50	25.6920	46.8177	33.9841	14.6836	12.7171	10.4805
	40	40.2746	57.5364	34.3000	18.7583	16.6988	13.2425
	30	55.2556	68.2630	34.8921	23.4014	19.2510	16.6607
	20	70.3528	78.9954	36.0493	28.6748	22.9483	20.7419
25	50	28.0560	48.9615	36.1174	16.5920	13.9439	11.6792
	40	42.1005	59.0727	36.4333	20.8162	17.1657	15.1779
	30	56.3841	69.2181	36.9707	25.3670	21.0834	17.5091
30	50	27.4382	50.0771	37.7926	16.4657	14.9914	12.3009
	40	41.0566	60.1056	38.1236	20.5033	18.0105	16.7597
	30	55.6724	70.1031	38.6254	25.4485	22.1112	18.3929
	20	70.4755	80.0654	39.4459	30.9303	27.7145	22.8242
35	50	27.3895	51.9867	39.3107	16.1425	17.0120	13.0017
	40	41.4798	61.3944	39.6070	21.0175	19.0025	17.9615
	30	55.7052	70.8284	40.0542	26.1121	23.0777	19.0443
	20	70.6087	80.7009	40.8817	31.8102	28.7064	23.6132

Table 3. Absolute Relative Bias (%ARB) for20% to 35% non-response for Dataset II

$N_2\%$	$n_{a2}\%$	$\widehat{P}_{AHT\pi}$	$\widehat{P}_{cAHT\pi}$	$\widehat{P}_{AHH\pi}$	$\widehat{P}'_{cA\pi}$	$\widehat{P}^{*}_{cA\pi}$	$\widehat{P}^{*}_{cA\pi}$
20	50	0.3054	0.2046	0.0876	0.0805	0.0602	0.0675
	40	0.4447	0.2646	0.2647	0.2294	0.0964	0.0805
	30	0.5831	0.3236	0.4461	0.3884	0.1879	0.1387
	20	0.7227	0.3847	0.6387	0.3011	0.3010	0.2201
25	50	0.3542	0.2614	0.1120	0.1080	0.0604	0.0631
	40	0.4814	0.3190	0.2921	0.2584	0.1333	0.1087
	30	0.6135	0.3779	0.4789	0.4174	0.2435	0.1855
	20	0.7404	0.4351	0.6512	0.5665	0.3459	0.2632
30	50	0.3904	0.3009	0.1148	0.1049	0.0627	0.0654
	40	0.5076	0.3538	0.2844	0.2420	0.1340	0.1039
	30	0.6305	0.4093	0.4616	0.3900	0.2405	0.1733
	20	0.7535	0.4646	0.6383	0.5374	0.3495	0.2509
35	50	0.4320	0.3490	0.1126	0.0999	0.0634	0.0655
	40	0.5429	0.3973	0.2841	0.2316	0.1412	0.1032
	30	0.6606	0.4491	0.4652	0.3786	0.2538	0.1722
	20	0.7714	0.4973	0.6387	0.5175	0.3647	0.2466

Table 4. Percentage Relative Root Mean Squared Error (%RRMSE) for20% to 35% nonresponse for Dataset II

$N_2\%$	$n_{a2}\%$	$\widehat{P}_{AHT\pi}$	$\widehat{P}_{cAHT\pi}$	$\widehat{P}_{AHH\pi}$	$\widehat{P}'_{cA\pi}$	$\widehat{P}''_{cA\pi}$	$\widehat{P}^{*}_{cA\pi}$
20	50	30.9316	21.3881	10.2869	9.6940	7.5382	8.4942
	40	44.7198	27.3857	27.1228	23.9545	11.5677	9.9529
	30	58.4800	33.3855	44.9486	39.4804	20.6376	16.2303
	20	72.3698	39.7831	64.0470	31.8349	31.8556	24.7367
25	50	35.7279	26.8265	12.6672	12.4241	7.5497	7.9005
	40	48.3599	32.6256	29.8422	26.8003	15.1916	12.9597
	30	61.5013	38.6184	48.2097	42.4030	25.8440	20.7691
30	50	34.1392	44.5936	65.2989	57.2248	36.0735	28.8058
	40	50.9469	35.9862	29.0946	25.2903	15.2398	12.4752
	30	63.1902	41.6372	46.5139	39.7581	25.5402	19.6624
	20	75.4372	47.4165	64.0223	54.4009	36.3556	27.6612
35	50	43.4131	35.3450	12.7959	11.7711	7.9075	8.2046
	40	54.4589	40.2273	29.0971	24.3947	15.9155	12.4450
	30	66.1812	45.5271	46.8754	38.7064	26.7723	19.5882
	20	77.2190	50.5630	64.0700	52.5212	37.7709	27.3106

5 Conclusion

In this study, we proposed three calibrated estimators of finite population proportion of the study attributes based on the auxiliary information, which can be available for (i) response group only, (ii) non-response group, (iii) both response as well as non-response groups. The proposed estimators are compared with the existing estimators. A simulation study has been performed on two-real datasets to evaluate the performance of the proposed estimators. The percentage absolute relative biases and percentage relative root mean squared errors of these estimators are shown in Tables 1, 2, 3 and 4. It is observed that %ARB and %RRMSE of the suggested estimators ($\widehat{P}'_{cA\pi}$, $\widehat{P}''_{cA\pi}$ and $\widehat{P}^{*}_{cA\pi}$) are less than the estimators ($\widehat{P}_{AHT\pi}$, $\widehat{P}_{cAHT\pi}$ and $\widehat{P}_{AHH\pi}$) for varying sample sizes of non-response groups. Hence, it can be concluded that the proposed calibrated estimators of the population proportion are more precise in comparison to already existing estimators and hence have more practical implications.

References

1) Hansen MH, Hurwitz WN. The Problem of Non-Response in Sample Surveys. *Journal of the American Statistical Association*. 1946;41(236):517–529. Available from: https://www.tandfonline.com/doi/abs/10.1080/01621459.1946.10501894.
2) El-Badry MA. A Sampling Procedure for Mailed Questionnaires. *Journal of the American Statistical Association*. 1956;51(274):209–227. Available from: https://doi.org/10.2307/2281342.
3) Jaiswal AK, Usman M, Singh GN. New calibration estimation procedure in the presence of unit non response. *Ain Shams Engineering Journal*. 2023;14(3):1–12. Available from: https://doi.org/10.1016/j.asej.2022.101910.
4) Ahmad S, Zahid E, Shabbir J, Aamir M, Onyango R. Enhanced Estimation of the Population Mean Using Two Auxiliary Variables under Probability Proportional to Size Sampling. *Mathematical Problems in Engineering*. 2023;2023:1–14. Available from: https://doi.org/10.1155/2023/5564360.
5) Sinha RR, Khanna B. Estimation of population mean under probability proportional to size sampling with and without measurement errors. *Concurrency and Computation: Practice and Experience*. 2022;34(18). Available from: https://doi.org/10.1002/cpe.7023.
6) Arcos A, Rueda MDM, Pasadas-Del-Amo S. Treating Nonresponse in Probability-Based Online Panels through Calibration: Empirical Evidence from a Survey of Political Decision-Making Procedures. *Mathematics*. 2020;8(3):1–16. Available from: https://doi.org/10.3390/math8030423.
7) Deville JC, Särndal CE. Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*. 1992;87(418):376–382. Available from: https://doi.org/10.2307/2290268.
8) Martínez S, Arcos A, Martínez H, Singh S. Estimating population proportions employing calibration estimators. *Revista Colombiana de Estadística*. 2015;38(1):267–293. Available from: https://www.redalyc.org/articulo.oa?id=89938627015.
9) Babatunde OT, Oladugba AV, Ude IO, Adubi AS. Calibration estimation of population mean in stratified sampling using standard deviation. *Quality & Quantity*. 2023. Available from: https://doi.org/10.1007/s11135-023-01737-1.
10) Ahmad S, Ullah K, Zahid E, Shabbir J, Aamir M, Alshanbari HM, et al. A new improved generalized class of estimators for population distribution function using auxiliary variable under simple random sampling. *Scientific Reports*. 2023;13:1–14. Available from: https://doi.org/10.1038/s41598-023-30150-9.
11) Ozgul N. New improved calibration estimator based on two auxiliary variables in stratified two-phase sampling. *Journal of Statistical Computation and Simulation*. 2021;91(6):1243–1256. Available from: https://doi.org/10.1080/00949655.2020.1844702.
12) Sarndal CE, Swensson B, Wretman J. Model Assisted Survey Sampling. New York, USA. Springer Verlag. 2003.
13) Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Medical Informatics and Decision Making*. 2020;20(1):1–16. Available from: https://doi.org/10.1186/s12911-020-1023-5.