DOI: https://doi.org/10.4038/sljas.v27i1.8211

Sri Lankan Journal of Applied Statistics, Vol (27-1)

An Improved Compromised Imputation Method Using Ranked Set Sampling

P. Garg[1,*], N. Srivastava[1], and M. K. Srivastava[3]

[1]Department of Statistics, St. John's College, Agra, U.P, India

[2]Department of Statistics (I.S.S), Dr. Bhimrao Ambedkar University, Agra, India

*Corresponding author: prachigarg2093@gmail.com

Received: 01[st] August, 2025/ Revised: 24[th] December, 2025/ Accepted: 07[th] January, 2026/
Published: 31[st] March, 2026

ABSTRACT

This paper presents a modified new imputation procedure for handling missing data using Ranked Set Sampling (RSS) design. The Proposed estimator is based on Singh and Horn (2000) imputation method. Approximate expressions for the bias and mean squared error (MSE) of the proposed estimator are derived up to the first order of approximation. The optimum conditions and minimum MSE for the proposed estimators are also obtained. The efficiency of the proposed estimator is examined through numerical investigations using six real datasets and a simulation study is also performed using artificially generated datasets. The results indicate that the proposed estimators outperform all existing estimator.

Keywords: Missing data, Imputation, Ranked set sampling, Bias, Mean squared error

1 Introduction

In many real-world studies, the primary variable of interest may be difficult, time-consuming, or expensive to measure directly. However, it is often feasible to rank such a variable with minimal resources or at no additional cost. Recognizing this, McIntyre (1952) introduced the concept of Rank Set Sampling (RSS)—a sampling technique that improves the efficiency of estimation by incorporating ranking information into the sample design. RSS has proven valuable in agricultural studies such as crop and pasture yield estimation. A practical example of RSS occurs in agricultural studies, where plants are first visually ranked by height or vigor, and only

one plant from each ranked set is measured for yield. Since ranking is inexpensive while actual measurement is costly, RSS achieves higher precision by using ranking information to select more informative units. This leads to improved estimator efficiency at a significantly lower measurement cost compared to simple random sampling.
RSS has since gained attention for its ability to enhance the precision of estimators, especially when measurement is expensive or destructive. The importance of each observation becomes particularly pronounced when dealing with small sample sizes within groups, where each unit plays a crucial role in obtaining reliable estimates of population parameters.
However, in such contexts, the presence of missing data poses a significant challenge. Missing values not only reduce the effective sample size but may also compromise statistical inference, resulting in biased or inefficient estimators. Although the issue of missing data is widely acknowledged, it has received limited attention within the framework of Ranked Set Sampling (RSS). This paper seeks to address this gap by investigating effective imputation strategies and assessing their performance under the RSS scheme using MCAR mechanism. In this regard, we propose a novel compromised imputation method tailored to the RSS context, which demonstrates improved efficiency over the imputation approach suggested by Singh and Horn (2000).

2 Sampling Methodology

McIntyre (1952) first introduced the Ranked Set Sampling (RSS) methodology, where m samples of size m are drawn by simple random sampling from the parent population. Within each sample, the m units are ranked based on the auxiliary variable. From the first sample, the unit ranked 1^{st} is selected for measurement on both the auxiliary variable and the associated study variable. From the second sample, the unit ranked 2^{nd} is chosen for measurement, and this process continues until, in the last sample, the unit ranked m^{th} is selected. This complete process is referred to as a cycle. The cycle is repeated r times, resulting in a sample of $n = rm$ units. Considering that some values of the study variable are missing, following the approach of ranked set sampling, we assume that m' units out of the available m units in each cycle respond to measurement such that $m > m'$. This entire process is repeated r times until responses are obtained from n' units out of the initially selected n units, where $n > n'$.

3 Notations

A finite population $\Omega = \{\Omega_1, \Omega_2, \ldots, \Omega_N\}$, consisting of N identifiable units. For the i^{th} unit $(i = 1, 2, \ldots, N)$,let (Y_i, X_i) denote the value of the study variable Y and the auxiliary variable X, respectively. To estimate the population mean of Y is $\mu_y = \frac{1}{N}\sum_{i=1}^{N} Y_i$. A ranked set sample (RSS) s of size $n = rm$ is selected, where m is the number of units in each cycle, and r is the number of cycles. In case of missing observations in one cycle, let P denote the proportion of units belonging to the responding group A, while $(1-P)$is the proportion of units that fall into the non-responding group $\bar{A}$. Thus, the total sample is divided as $s = A \cup \bar{A}$. The values Y_i are observed for units in A. For units in $\bar{A}$, the values of Y are missing and need to be imputed to obtain the complete dataset for valid statistical inference. The unbiased estimators of the population means μ_x and μ_y under RSS with missing observations are $\bar{X}_{r,\text{rss}} = \frac{1}{rmP}\sum_{i=1}^{m'}\sum_{j=1}^{r} X_{(i:i)j}$, and $\bar{Y}_{r,\text{rss}} = \frac{1}{rmP}\sum_{i=1}^{m'}\sum_{j=1}^{r} Y_{[i:i]j}$, where $X_{(i:i)j}$ and $Y_{[i:i]j}$ represent the i^{th} order statistics and the i^{th} judgment order statistic, respectively, in the j^{th} cycle for variables X and Y. For simplicity, we denote them as $X_{(i)}$ and $Y_{[i]}$. According to Bouza(2013), the variance of the estimator $\bar{Y}_{r,\text{rss}}$ in the presence of missing observations with response probability P is given by

$$\text{Var}(\bar{Y}_{r,\text{rss}}) = \frac{\sigma_y^2}{nP} - \frac{1}{nmP}\sum_{i=1}^{m'} D_{y[i]}^2$$

,
where, $D_{y[i]} = \mu_{y[i]} - \mu_y$ denotes the deviation of the expected i^{th} order statistic $\mu_{y[i]}$ (from a ranked set of size m) from the population mean μ_y. The term $D_{y[i]}^2$ represents the contribution of rank i to the variance reduction achieved by ranked set sampling under perfect ranking.
To derive the expressions for the bias and mean square error (MSE) of the proposed and existing estimators presented in the following sections, we define the following relative error terms:
$\bar{Y}_{r,\text{rss}} = \mu_y(1+\varepsilon_0)$, $\bar{X}_{r,\text{rss}} = \mu_x(1+\varepsilon_1)$, and $\bar{X}_{n,\text{rss}} = \mu_x(1+\varepsilon_2)$, where ε_0,ε_1,and ε_2 are the random error terms such that $E(\varepsilon_0) = E(\varepsilon_1) = E(\varepsilon_2) = 0$ and

$$E(\varepsilon_0^2) = (\frac{\sigma_y^2}{nP} - \frac{1}{nmP}\sum_{i=1}^{m} D_{y[i]}^2) = (\frac{C_y^2}{nP} - \frac{1}{nmP}\sum_{i=1}^{m}\frac{D_{y[i]}^2}{\mu_y^2}) = (\gamma^* C_y^2 - W_y^{2*})$$

,

$$E(\varepsilon_1^2) = (\frac{C_x^2}{nP} - \frac{1}{nmP}\sum_{i=1}^{m}\frac{D_{x[i]}^2}{\mu_x^2}) = (\gamma^* C_x^2 - W_x^{2*})$$

,

$$E(\varepsilon_2^2) = (\frac{C_x^2}{n} - \frac{1}{nm}\sum_{i=1}^{m}\frac{D_{x[i]}^2}{\mu_x^2}) = (\gamma C_x^2 - W_x^2)$$

,

$$E(\varepsilon_0\varepsilon_1) = (\frac{\rho_{xy}C_xC_y}{nP} - \frac{1}{nmP}\sum_{i=1}^{m}\frac{D_{xy[i]}}{\mu_x\mu_y}) = (\gamma^*\rho_{xy}C_xC_y - W^*_{xy})$$

,

$$E(\varepsilon_0\varepsilon_2) = (\frac{\rho_{xy}C_xC_y}{n} - \frac{1}{nm}\sum_{i=1}^{m}\frac{D_{xy[i]}}{\mu_x\mu_y}) = (\gamma\rho_{xy}C_xC_y - W_{xy})$$

,

$$E(\varepsilon_1\varepsilon_2) = (\frac{C_x^2}{n} - \frac{1}{nm}\sum_{i=1}^{m}\frac{D^2_{x[i]}}{\mu_x^2}) = (\gamma C_x^2 - W_x^2)$$

.

where $\gamma^* = \frac{1}{nP}$, $\gamma = \frac{1}{n}$, $W_y^2 = \frac{1}{nmP}\sum_{i=1}^{m}\frac{D^2_{y[i]}}{\mu_y^2}$, $W_x^2 = \frac{1}{nm}\sum_{i=1}^{m}\frac{D^2_{x(i)}}{\mu_x^2}$, $W_x^{2*} = \frac{1}{nmP}\sum_{i=1}^{m}\frac{D^2_{x(i)}}{\mu_x^2}$, $W_{xy} = \frac{1}{nm}\sum_{i=1}^{m}\frac{D_{xy[i]}}{\mu_x\mu_y}$, $W^*_{xy} = \frac{1}{nmP}\sum_{i=1}^{m}\frac{D_{xy[i]}}{\mu_x\mu_y}$, $D_{y[i]} = (\mu_{y[i]} - \mu_y)$, $D_{x[i]} = (\mu_{x[i]} - \mu_x)$, $D_{xy[i]} = (\mu_{x[i]} - \mu_x)(\mu_{y[i]} - \mu_y)$, $C_x = \frac{S_x}{\mu_x}$, $C_y = \frac{S_y}{\mu_y}$, $\mu_y = E(Y)$, $\mu_x = E(X)$, $\mu_{y[i]} = E(Y_{(i)})$ and $\mu_{x(i)} = E(X_{(i)})$.

$S_x^2 = \dfrac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$ and $S_y^2 = \dfrac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$ represent the population variances of the auxiliary variable X and the study variable Y, respectively, and $s_x^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ and $s_y^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ are the sample variances. Similarly, C_x and C_y are the population coefficients of variation for the variables X and Y, respectively. The symbol ρ_{xy} denotes the population correlation coefficient between X and Y. Additionally, the terms $\mu_{X(i)}$ and $\mu_{Y(i)}$ refer to the expected values of the i^{th} order statistics derived from the respective probability distributions. These values are computed using the `EnvStats` package in R.

4 Review of Imputation Methods under RSS

Some of the existing methods of imputation under Ranked Set Sampling (RSS) are described below.

4.1 Mean Imputation Method The mean imputation estimator and its variance under RSS are obtained following Bouza (2013), where the imputed values are defined as

$$y_{.i} = \begin{cases} Y_i, & i \in A \text{ with probability } P, \\ \bar{Y}_{r,\text{rss}}, & i \in \bar{A} \text{ with probability } (1-P). \end{cases}$$

The point estimator of the population mean is

$$\bar{y}_{m,\text{rss}} = \bar{Y}_{r,\text{rss}}. \tag{1}$$

The mean square error (MSE) is given by

$$\text{MSE}(\bar{y}_{m,\text{rss}}) = \mu_y^2 \left(\gamma^* C_y^2 - W_y^{(2*)}\right). \tag{2}$$

4.2 Ratio Imputation Method The ratio imputation estimator and the corresponding mean squared error (MSE) under RSS are derived using the methodology of Bouza (2010, 2012) in conjunction with the classical RSS ratio estimator of McIntyre (1952) and Takahasi and Wakimoto (1968), where the imputed values are defined as

$$y_{.ir} = \begin{cases} Y_i, & i \in A \text{ with probability } P, \\ \dfrac{1}{1-f}\left[\bar{Y}_{r,\text{rss}}\left(\dfrac{\bar{X}_{n,\text{rss}}}{\bar{X}_{r,\text{rss}}}\right) + \bar{Y}_{r,\text{rss}}\right], & i \in \bar{A} \text{ with probability } (1-P), \end{cases}$$

here $f = r/n$.

The point estimator is

$$\bar{y}_{\text{RAT,RSS}} = \bar{Y}_{r,\text{rss}}\left(\frac{\bar{X}_{n,\text{rss}}}{\bar{X}_{r,\text{rss}}}\right). \tag{3}$$

The MSE is

$$\begin{aligned} \text{MSE}(\bar{y}_{\text{RAT,RSS}}) = \mu_y^2 \Big[& (\gamma^* C_y^2 - W_y^{(2*)}) + (\gamma^* C_x^2 - W_x^{(2*)} - \gamma C_x^2 + W_x^2) \\ & - 2(\gamma^* \rho_{xy} C_x C_y - W_{xy}^* - \gamma \rho_{xy} C_x C_y + W_{xy}) \Big]. \end{aligned} \tag{4}$$

4.3 Al-Omari and Bouza Imputation Method To enhance the efficiency of estimators in the presence of missing data, Kadilar and Cingi (2008) and Al-Omari and Bouza (2014) proposed regression-cum-ratio type estimators within the Ranked Set Sampling (RSS) framework.

$$y_{.ir} = \begin{cases} Y_i, & i \in A \text{ with probability } P, \\ \dfrac{\bar{Y}_{r,\text{rss}} + B(X_i - \bar{X}_{r,\text{rss}})}{\bar{X}_{r,\text{rss}}}\bar{X}_{r,\text{rss}}, & i \in \bar{A} \text{ with probability } (1-P), \end{cases}$$

where $B = S_{xy}/S_x^2$ is the regression coefficient of Y on X, computed from the observed units in the RSS sample. The Al-Omari and Bouza estimator of the population mean is

$$\bar{y}_{KC} = \frac{\bar{Y}_{r,\text{rss}} + B(\bar{X}_{n,\text{rss}} - \bar{X}_{r,\text{rss}})}{\bar{X}_{r,\text{rss}}} \bar{X}_{r,\text{rss}}. \tag{5}$$

The approximate MSE is

$$\begin{aligned} \text{MSE}(\bar{y}_{KC}) \approx \mu_y^2(\gamma^* C_y^2 - W_y^{(2*)}) + (R+B)^2 \mu_x^2(\gamma^* C_x^2 - W_x^{(2*)}) \\ - 2(R+B)\mu_x \mu_y(\gamma^* \rho_{xy} C_x C_y - W_{xy}^*). \end{aligned} \tag{6}$$

4.4 Sohail, Shabbir and Ahmed (2018) Imputation Methods Building on the work of Bouza (2013), Al-Omari and Bouza (2014), Sohail, Shabbir, and Ahmed (2018) adapted the ratio-type estimator proposed by Ahmed et al. (2006) using Ranked Set Sampling (RSS) for imputing missing values. These imputation methods are

$$y_{i.s_1} = \begin{cases} Y_i, & i \in A, \\ \dfrac{1}{n-r}\left[n\bar{Y}_{r,\text{rss}}\left(\dfrac{\bar{X}_{n,\text{rss}}}{\bar{X}_{r,\text{rss}}}\right)^{\beta_1} - r\bar{Y}_{r,\text{rss}}\right], & i \in \bar{A}, \end{cases}$$

$$y_{i.s_2} = \begin{cases} Y_i, & i \in A, \\ \dfrac{1}{n-r}\left[n\bar{Y}_{r,\text{rss}}\left(\dfrac{\bar{X}_{n,\text{rss}}}{\beta_2\bar{X}_{r,\text{rss}} + (1-\beta_2)\bar{X}_{n,\text{rss}}}\right) - r\bar{Y}_{r,\text{rss}}\right], & i \in \bar{A}. \end{cases}$$

The estimators are

$$\bar{y}_{s_1,\text{RSS}} = \bar{Y}_{r,\text{rss}}\left(\frac{\bar{X}_{n,\text{rss}}}{\bar{X}_{r,\text{rss}}}\right)^{\beta_1}, \tag{7}$$

$$\bar{y}_{s_2,\text{RSS}} = \bar{Y}_{r,\text{rss}}\left\{\frac{\bar{X}_{n,\text{rss}}}{\beta_2\bar{X}_{r,\text{rss}} + (1-\beta_2)\bar{X}_{n,\text{rss}}}\right\}. \tag{8}$$

The MSE is given by

$$\begin{aligned} \text{MSE}(\bar{y}_{s_i,\text{RSS}}) = \mu_y^2\Big[\gamma^* C_y^2 - W_y^{(2*)} + \beta_i^2(\gamma^* C_x^2 - W_x^{(2*)} - \gamma C_x^2 + W_x^2) \\ - 2\beta_i(\gamma^* \rho_{xy} C_x C_y - W_{xy}^* - \gamma \rho_{xy} C_x C_y + W_{xy})\Big], \quad i = 1, 2. \end{aligned} \tag{9}$$

and the minimum mean square errors are

$$\min \mathrm{MSE}(\bar{y}_{s_i \cdot \mathrm{RSS}}) =$$

$$\mu_y^2 \left\{ \gamma^* C_y^2 - W_y^{(2*)} - \frac{\left(\gamma^* \rho_{xy} C_x C_y - W_{xy}^* - \gamma \rho_{xy} C_x C_y + W_{xy}\right)^2}{\gamma^* C_x^2 - W_x^{(2*)} - \gamma C_x^2 + W_x^2} \right\}, \qquad i = 1, 2. \tag{10}$$

The optimum value is

$$\beta_{1(opt)} = \beta_{2(opt)} = \frac{\gamma^* \rho_{xy} C_x C_y - W_{xy}^* - \gamma \rho_{xy} C_x C_y + W_{xy}}{\gamma^* C_x^2 - W_x^{(2*)} - \gamma C_x^2 + W_x^2}.$$

4.5 The Singh and Horn (2000) Compromised Imputation Method

Building upon the methodology introduced by Singh and Horn (2000), Bouza (2013) developed a compromised estimation approach under Ranked Set Sampling (RSS) in the presence of missing data. The imputation method is defined as

$$y_{.i,\mathrm{COMP}} = \begin{cases} \dfrac{\alpha\, nY_i}{r} + (1-\alpha)\, \hat{b}\, \bar{x}_{n,\mathrm{rss}}, & i \in A \text{ with probability } P, \\ (1-\alpha)\, \hat{b}\, \bar{x}_{n,\mathrm{rss}}, & i \in \bar{A} \text{ with probability } (1-P), \end{cases}$$

where $\hat{b} = \frac{\bar{Y}_{r,\mathrm{rss}}}{\bar{X}_{r,\mathrm{rss}}}$.

The resulting estimator of the population mean is

$$\bar{y}_{\mathrm{COMP,RSS}} = \alpha\, \bar{Y}_{r,\mathrm{rss}} + (1-\alpha) \frac{\bar{Y}_{r,\mathrm{rss}}}{\bar{X}_{r,\mathrm{rss}}} \bar{X}_{n,\mathrm{rss}}. \tag{11}$$

The mean square error (MSE) of $\bar{y}_{\mathrm{COMP,RSS}}$ is given by

$$\begin{aligned} \mathrm{MSE}(\bar{y}_{\mathrm{COMP,RSS}}) = \mu_y^2 \Big[& \gamma^* C_y^2 - W_y^{(2*)} + \alpha^2 (\gamma^* C_x^2 - W_x^{(2*)} - \gamma C_x^2 + W_x^2) \\ & - 2\alpha (\gamma^* \rho_{xy} C_x C_y - W_{xy}^* - \gamma \rho_{xy} C_x C_y + W_{xy}) \Big]. \end{aligned} \tag{12}$$

The minimum mean square error is

$$\min \mathrm{MSE}(\bar{y}_{\mathrm{COMP,RSS}}) =$$

$$\mu_y^2\left\{\gamma^* C_y^2 - W_y^{(2*)} - \frac{\left(\gamma^*\rho_{xy}C_xC_y - W_{xy}^* - \gamma\rho_{xy}C_xC_y + W_{xy}\right)^2}{\gamma^* C_x^2 - W_x^{(2*)} - \gamma C_x^2 + W_x^2}\right\}. \tag{13}$$

The optimum value of α is obtained as

$$\alpha_{\text{opt}} = \frac{\gamma^*\rho_{xy}C_xC_y - W_{xy}^* - \gamma\rho_{xy}C_xC_y + W_{xy}}{\gamma^* C_x^2 - W_x^{(2*)} - \gamma C_x^2 + W_x^2}.$$

5 The Proposed Imputation Methods

The estimator proposed here is inspired by Singh and Horn's (2000) compromised estimator under Ranked Set Sampling and is defined as follows

$$y_{.i,\text{rss}} = \begin{cases} \dfrac{\alpha_1 n y_{i,\text{rss}}}{r} + \alpha_2 \hat{b} x_{i,\text{rss}}, & i \in A, \\ \alpha_2 \hat{b} x_{i,\text{rss}}, & i \in \bar{A}, \end{cases}$$

where α_1 and α_2 are real constants (Singh and Horn's (2000) method is a particular case when α_1+α_2=1, whose values are chosen optimally to minimize the variance of the resulting estimator. With the above imputation method, the resulting estimator of the population mean $\bar{y}$ is obtained as

$$\begin{aligned}\bar{y}_{\text{INP,RSS}} &= \frac{1}{n}\sum_{i=1}^{n} y_{.i,\text{rss}} \\ &= \frac{1}{n}\left[\sum_{i=1}^{r}\frac{\alpha_1 n\, y_{i,\text{rss}}}{r} + \sum_{i=1}^{r}\alpha_2 \hat{b}\, x_{i,\text{rss}} + \sum_{i=r+1}^{n}\alpha_2 \hat{b}\, x_{i,\text{rss}}\right]. \qquad (14) \\ &= \alpha_1 \bar{y}_{r,\text{rss}} + \alpha_2 \frac{\bar{y}_{r,\text{rss}}}{\bar{x}_{r,\text{rss}}}\bar{x}_{n,\text{rss}}.\end{aligned}$$

Theorem 5.1 The bias of the estimator $\bar{y}_{\text{INP,RSS}}$ up to the first order of approximation is given as

$$\begin{aligned}\text{Bias}(\bar{y}_{\text{INP,RSS}}) = \mu_y\Big[&(\alpha_1 - 1) + \alpha_2\{(\gamma^* C_x^2 - W_x^{(2*)} - \gamma C_x^2 + W_x^2) \\ &-(\gamma^*\rho_{xy}C_xC_y - W_{xy}^* - \gamma\rho_{xy}C_xC_y + W_{xy})\}\Big]. \end{aligned} \tag{15}$$

Proof. The estimator $\bar{y}_{\text{INP,RSS}}$ in terms of the error variables ε can be written as

$$\bar{y}_{\text{INP,RSS}} = \alpha_1\mu_y(1+\varepsilon_0) + \alpha_2\frac{\mu_y(1+\varepsilon_0)}{\mu_x(1+\varepsilon_1)}\mu_x(1+\varepsilon_2).$$

or,

$$\bar{y}_{\text{INP,RSS}} - \bar{Y} = \alpha_1\mu_y(1+\varepsilon_0) + \alpha_2\mu_y(1+\varepsilon_0)(1+\varepsilon_1)^{-1}(1+\varepsilon_2) - \mu_y.$$

Expanding the above expression up to the first order of approximation and neglecting higher order terms, we obtain

$$\bar{y}_{\text{INP,RSS}} - \bar{Y} \approx \mu_y\Big[\alpha_1(1+\varepsilon_0) + \alpha_2\big\{1-\varepsilon_0-\varepsilon_1+\varepsilon_2-\varepsilon_0\varepsilon_1+\varepsilon_0\varepsilon_2-\varepsilon_1\varepsilon_2+\varepsilon_1^2\big\}-1\Big]. \tag{16}$$

Hence, the bias of $\bar{y}_{\text{INP,RSS}}$ is

$$\begin{aligned}\text{Bias}(\bar{y}_{\text{INP,RSS}}) &= \mathbb{E}(\bar{y}_{\text{INP,RSS}} - \mu_y)\\ &= \mu_y\Big[(\alpha_1-1) + \alpha_2\big\{(\gamma^* C_x^2 - W_x^{(2*)} - \gamma C_x^2 + W_x^2)\\ &\qquad - (\gamma^*\rho_{xy}C_xC_y - W_{xy}^* - \gamma\rho_{xy}C_xC_y + W_{xy})\big\}\Big]. \end{aligned}\tag{17}$$

Theorem 5.2 The minimum mean square error of the proposed estimator $\bar{y}_{\text{INP,RSS}}$ is given by

$$\min_{\alpha_1,\alpha_2} \text{MSE}(\bar{y}_{\text{INP,RSS}}) = \mu_y^2\left[1 - \frac{A_2 - 2A_3A_4 + A_1A_4^2}{A_1A_2 - A_3^2}\right], \tag{18}$$

where

$$\begin{aligned}
A_1 &= 1 + (\gamma^* C_y^2 - W_y^{(2*)}),\\
A_2 &= 1 + (\gamma^* C_y^2 - W_y^{(2*)}) + 3(\gamma^* C_x^2 - W_x^{(2*)} - \gamma C_x^2 + W_x^2)\\
&\quad - 4(\gamma^*\rho_{xy}C_xC_y - W_{xy}^* - \gamma\rho_{xy}C_xC_y + W_{xy}),\\
A_3 &= 1 + (\gamma^* C_y^2 - W_y^{(2*)}) + (\gamma^* C_x^2 - W_x^{(2*)} - \gamma C_x^2 + W_x^2)\\
&\quad - 2(\gamma^*\rho_{xy}C_xC_y - W_{xy}^* - \gamma\rho_{xy}C_xC_y + W_{xy}),\\
A_4 &= 1 + (\gamma^* C_x^2 - W_x^{(2*)} - \gamma C_x^2 + W_x^2)\\
&\quad - (\gamma^*\rho_{xy}C_xC_y - W_{xy}^* - \gamma\rho_{xy}C_xC_y + W_{xy}).
\end{aligned}$$

Proof. The mean square error (MSE) of $\bar{y}_{\text{INP,RSS}}$ is defined as

$$\text{MSE}(\bar{y}_{\text{INP,RSS}}) = \mathbb{E}(\bar{y}_{\text{INP,RSS}} - \mu_y)^2.$$

From equation (16), we have

$$\text{MSE}(\bar{y}_{\text{INP,RSS}}) = \mu_y^2\mathbb{E}\Big[\alpha_1(1+\varepsilon_0) + \alpha_2(1+\varepsilon_0-\varepsilon_1+\varepsilon_2-\varepsilon_0\varepsilon_1+\varepsilon_0\varepsilon_2-\varepsilon_1\varepsilon_2+\varepsilon_1^2)-1\Big]^2.$$

On simplification, we obtain

$$\begin{aligned}\mathrm{MSE}(\bar{y}_{\text{INP,RSS}}) = \mu_y^2\,\mathbb{E}\Big[&(\alpha_1-1)^2+\alpha_1^2\varepsilon_0^2+\alpha_2^2(1+\varepsilon_0-\varepsilon_1+\varepsilon_2-\varepsilon_0\varepsilon_1+\varepsilon_0\varepsilon_2\\ &-\varepsilon_1\varepsilon_2+\varepsilon_1^2)^2\\ &+2(\alpha_1-1)\alpha_1\varepsilon_0+2(\alpha_1-1)\alpha_2(1+\varepsilon_0-\varepsilon_1+\varepsilon_2-\varepsilon_0\varepsilon_1\\ &+\varepsilon_0\varepsilon_2-\varepsilon_1\varepsilon_2+\varepsilon_1^2)\\ &+2\alpha_1\alpha_2\varepsilon_0(1+\varepsilon_0-\varepsilon_1+\varepsilon_2-\varepsilon_0\varepsilon_1+\varepsilon_0\varepsilon_2-\varepsilon_1\varepsilon_2+\varepsilon_1^2)\Big].\end{aligned} \tag{19}$$

Neglecting higher order terms and simplifying further, we get

$$\begin{aligned}\mathrm{MSE}(\bar{y}_{\text{INP,RSS}}) \approx \mu_y^2\,\mathbb{E}\Big[&\alpha_1^2(1+\varepsilon_0^2+2\varepsilon_0)+\alpha_2^2(1+\varepsilon_0^2+3\varepsilon_1^2+\varepsilon_2^2+4\varepsilon_0\varepsilon_2\\ &-4\varepsilon_0\varepsilon_1-4\varepsilon_1\varepsilon_2)\\ &+2\alpha_1\alpha_2(1+\varepsilon_0+\varepsilon_0^2-2\varepsilon_0\varepsilon_1+2\varepsilon_0\varepsilon_2-\varepsilon_1\varepsilon_2+\varepsilon_1^2)-2\alpha_1(1+\varepsilon_0)\\ &-2\alpha_2(1-\varepsilon_0\varepsilon_1+\varepsilon_0\varepsilon_2-\varepsilon_1\varepsilon_2+\varepsilon_1^2)+1\Big].\end{aligned} \tag{20}$$

Taking expectations, we obtain

$$\begin{aligned}&\mathrm{MSE}(\bar{y}_{\text{INP.RSS}}) = \mu_y^2\Bigg[1+\alpha_1^2\left\{1+\left(\gamma^*C_y^2-W_y^{(2*)}\right)\right\}\\ &+\alpha_2^2\Big\{1+\left(\gamma^*C_y^2-W_y^{(2*)}\right)+3\left(\gamma^*C_x^2-W_x^{(2*)}-\gamma C_x^2+W_x^2\right)\\ &-4\left(\gamma^*\rho_{xy}C_xC_y-W_{xy}^*-\gamma\rho_{xy}C_xC_y+W_{xy}\right)\Big\}\\ &+2\alpha_1\alpha_2\Big\{1+\left(\gamma^*C_y^2-W_y^{(2*)}\right)+\left(\gamma^*C_x^2-W_x^{(2*)}-\gamma C_x^2+W_x^2\right)\\ &-2\left(\gamma^*\rho_{xy}C_xC_y-W_{xy}^*-\gamma\rho_{xy}C_xC_y+W_{xy}\right)\Big\}\\ &-2\alpha_1-2\alpha_2\Big\{1+\left(\gamma^*C_x^2-W_x^{(2*)}-\gamma C_x^2+W_x^2\right)\\ &-\left(\gamma^*\rho_{xy}C_xC_y-W_{xy}^*-\gamma\rho_{xy}C_xC_y+W_{xy}\right)\Big\}\Bigg].\end{aligned}$$

$$\mathrm{MSE}(\bar{y}_{\text{INP.RSS}}) = \mu_y^2\left[1+\alpha_1^2A_1+\alpha_2^2A_2+2\alpha_1\alpha_2A_3-2\alpha_1-2\alpha_2A_4\right]. \tag{21}$$

where $A_1 = 1+(\gamma^*C_y^2-W_y^{(2*)})$,
$A_2 = 1+(\gamma^*C_y^2-W_y^{(2*)})+3(\gamma^*C_x^2-W_x^{(2*)}-\gamma C_x^2+W_x^2)$
$-4(\gamma^*\rho_{xy}C_xC_y-W_{xy}^*-\gamma\rho_{xy}C_xC_y+W_{xy})$,
$A_3 = 1+(\gamma^*C_y^2-W_y^{(2*)})+(\gamma^*C_x^2-W_x^{(2*)}-\gamma C_x^2+W_x^2)$
$-2(\gamma^*\rho_{xy}C_xC_y-W_{xy}^*-\gamma\rho_{xy}C_xC_y+W_{xy})$,

$A_4 = 1 + (\gamma^* C_x^2 - W_x^{(2*)} - \gamma C_x^2 + W_x^2)$
$- (\gamma^* \rho_{xy} C_x C_y - W_{xy}^* - \gamma \rho_{xy} C_x C_y + W_{xy}).$

Differentiating equation (21) with respect to α_1 and α_2 and equating to zero, the optimum values are obtained as

$$\alpha_{1,\text{opt}} = \frac{A_2 - A_3 A_4}{A_1 A_2 - A_3^2}, \qquad \alpha_{2,\text{opt}} = \frac{-A_3 + A_1 A_4}{A_1 A_2 - A_3^2}.$$

Substituting these values of $\alpha_1 opt$ and $\alpha_2 opt$ in equation (21), we get the minimum mean square error

$$\min_{\alpha_1,\alpha_2} \text{MSE}(\bar{y}_{\text{INP,RSS}}) = \mu_y^2 \left[1 - \frac{A_2 - 2A_3A_4 + A_1 A_4^2}{A_1 A_2 - A_3^2} \right].$$

6 Computational Study

To compare the efficiency of the proposed estimator with the other existing imputation methods, a numerical study was conducted. A theoretical comparison based on the mean squared error (MSE) was carried out using six real populations obtained from different sources. A simulation study is also performed on artificially generated bivariate population drawn from normal distribution.

6.1 Numerical Study In this subsection, a numerical study is carried out in the to evaluate the performance of the proposed imputation methods of in comparison with competing methods, using MSE and percent relative efficiency(PRE) as performance measures. The study is based on six real data sets. which are described as follows:

Population 1: This dataset, taken from Kadilar and Cingi (2003), consists of information from 69 villages in the South Anatolia region of Turkey in 1999. The study variable is the level of apple production, and the auxiliary variable is the number of apple trees.

Population 2: Sourced from Sarndal (2003), this dataset uses population size (in millions) from the year 1983 as the study variable, with exports (in millions of U.S. dollars) as the auxiliary variable.

Population 3: Also taken from Sarndal (2003), this dataset contains data from 1982. The study variable is the total number of municipal council seats, while the auxiliary variable is the number of conservative seats.

Population 4: This dataset is selected from Kadilar and Cingi (2003), comprising data from 106 villages in the Marmara region of Turkey during 1999. The study variable is apple production, and the auxiliary variable is the number of apple trees.

Population 5: Sourced from Murthy (1967), this dataset includes output as the auxiliary variable (X) and fixed capital as the study variable (Y).

Population 6: Taken from Samawi and Muttlak (1996), this dataset uses the serial number of patients as the auxiliary variable (X), and the number of infected patients as the study variable (Y).

The necessary parameter values for all six populations are reported accordingly.
The mean squared errors (MSEs) are computed to compare the proposed

Table 1: Description of the Population Parameters

Parameters	**Pop 1**	**Popn 2**	**Pop 3**	**Pop 4**	**Pop5**	**Pop 6**
N	69	124	284	106	80	295
n	12	12	12	12	12	12
m	3	3	3	3	3	3
r	4	4	4	4	4	4
μ_y	71.34	36.65	47.50	1536.77	1126.463	2.63678
μ_x	3165.02	14276.03	9.05	24375.59	5182.637	620.5508
S_y	110.85	116.80	11.06	6425.08	845.6097	1.53940
S_x	3965.24	31431.81	4.95	49189.08	1835.6585	85.3034
ρ_{xy}	0.91	0.23	0.66	0.8156	0.97	0.92

imputation methods with the existing ones. Further, the percentage relative efficiencies (PREs) of the proposed estimators with respect to the conventional estimator $\bar{y}_{m\cdot\mathrm{RSS}}$ are calculated using the following formula:

$$\mathrm{PRE}(\bar{y}_t) = \frac{\mathrm{MSE}(\bar{y}_{m\cdot\mathrm{RSS}})}{\mathrm{MSE}(\bar{y}_{t\cdot\mathrm{RSS}})}\times 100\%, \quad t = m,\ \mathrm{RAT},\ \mathrm{KC},\ \mathrm{COMP},\ s_i\ (i = 1, 2),\ \mathrm{INP}.$$

To compute the MSE, the values of n, m, and r are kept the same for all real populations so that the performance of the estimators can be compared under a uniform sampling framework.

Table 2: PRE of Proposed Estimators for Population 1

Estimator	$P = 0.5$	$P = 0.6$	$P = 0.7$	$P = 0.8$
$\bar{y}_{m \cdot \text{RSS}}$	100	100	100	100
$\bar{y}_{\text{RAT,RSS}}$	152.06	137.7206	125.8525	115.8676
$\bar{y}_{\text{KC}}$	90.5562	90.5562	90.5562	90.5562
$\bar{y}_{\text{COMP,RSS}} = \bar{y}_{s_i \cdot \text{RSS}},\ i = 1, 2$	152.1134	137.7555	125.8744	115.88
$\bar{y}_{\text{INP,RSS}}$	172.5768	154.9369	140.6929	128.9138

Table 3: PRE of Proposed Estimators for Population 2

Estimator	$P = 0.5$	$P = 0.6$	$P = 0.7$	$P = 0.8$
$\bar{y}_{m \cdot \text{RSS}}$	100	100	100	100
$\bar{y}_{\text{RAT,RSS}}$	63.86473	68.83982	74.65552	81.54453
$\bar{y}_{\text{KC}}$	36.76257	36.76257	36.76257	36.76257
$\bar{y}_{\text{COMP,RSS}} = \bar{y}_{s_i \cdot \text{RSS}},\ i = 1, 2$	112.6305	109.8554	107.2139	104.6963
$\bar{y}_{\text{INP,RSS}}$	242.1525	185.3050	166.2289	156.0007

6.2 Simulation Analysis In addition to the numerical study based on real populations, a simulation study is also conducted to further examine the performance of the proposed imputation methods in comparison with existing imputation methods. These simulations were performed on four artificial population which are described below:

Population 7. A population of size $N = 500$ with a study variable y and one auxiliary variable x is generated from a bivariate normal distribution. The study variable y is correlated with the auxiliary variable x with correlation coefficients $\rho_{yx} = 0.5, 0.6, 0.7, 0.8$, and 0.9. The bivariate variables (y, x) are generated using the `MVNORM` package in R software. From this population, a sample of size $n = 51$ is drawn, out of which $r = 30$ responding units.

Population 8. An artificial population of size $N = 200$ with study variable y and auxiliary variable x is generated from a bivariate normal distribution. The study variable y is correlated with the auxiliary variable x with correlation coefficients $\rho_{yx} = 0.5, 0.6, 0.7, 0.8$, and 0.9. The bivariate variables (y, x) are generated using the `MVNORM` package in R software. From

Table 4: PRE of Proposed Estimators for Population 3

Estimator	$P = 0.5$	$P = 0.6$	$P = 0.7$	$P = 0.8$
$\bar{y}_{m \cdot \text{RSS}}$	100	100	100	100
$\bar{y}_{\text{RAT,RSS}}$	34.0313	39.2038	46.2304	56.3258
$\bar{y}_{\text{KC}}$	12.57516	12.57516	12.57516	12.57516
$\bar{y}_{\text{COMP,RSS}} = \bar{y}_{s_i \cdot \text{RSS}},\ i = 1, 2$	106.4979	105.1316	103.8	102.5016
$\bar{y}_{\text{INP,RSS}}$	107.3957	105.7946	104.3020	102.8882

Table 5: PRE of Proposed Estimators for Population 4

Estimator	$P = 0.5$	$P = 0.6$	$P = 0.7$	$P = 0.8$
$\bar{y}_{m\cdot\text{RSS}}$	100	100	100	100
$\bar{y}_{\text{RAT,RSS}}$	124.3552	118.5792	113.3159	108.5
$\bar{y}_{\text{KC}}$	99.47661	99.47661	99.47661	99.47661
$\bar{y}_{\text{COMP,RSS}} = \bar{y}_{s_i\cdot\text{RSS}},\ i = 1, 2$	126.4818	120.1198	114.3672	109.1404
$\bar{y}_{\text{INP,RSS}}$	262.4532	236.2527	216.3375	200.4289

Table 6: PRE of Proposed Estimators for Population 5

Estimator	$P = 0.5$	$P = 0.6$	$P = 0.7$	$P = 0.8$
$\bar{y}_{m\cdot\text{RSS}}$	100	100	100	100
$\bar{y}_{\text{RAT,RSS}}$	150.0219	136.3781	125.0091	115.3898
$\bar{y}_{\text{KC}}$	177.1816	177.1816	177.1816	177.1816
$\bar{y}_{\text{COMP,RSS}} = \bar{y}_{s_i\cdot\text{RSS}},\ i = 1, 2$	179.9324	155.1322	136.3403	121.6093
$\bar{y}_{\text{INP,RSS}}$	181.7213	157.2919	138.6963	124.0542

this population, a sample of size $n = 24$ is drawn, with $r = 14$ responding units.

Population 9. A population of size $N = 2000$ with study variable y and auxiliary variable x is generated from a multivariate normal distribution with correlation coefficients $\rho_{yx} = 0.5, 0.6, 0.7, 0.8$, and 0.9. The bivariate variables (y, x) are generated using the `MVNORM` package in R software. From this population, a sample of size $n = 256$ is drawn, out of which $r = 204$ units respond.

Population 10. A population of size $N = 2000$ with study variable y and auxiliary variable x is generated from a multivariate normal distribution with correlation coefficients $\rho_{yx} = 0.5, 0.6, 0.7, 0.8$, and 0.9. The bivariate variables (y, x) are generated using the `MVNORM` package in R software. From this population, a sample of size $n = 255$ is drawn, with $r = 155$ responding units.

Table 7: PRE of Proposed Estimators for Population 6

Estimator	$P = 0.5$	$P = 0.6$	$P = 0.7$	$P = 0.8$
$\bar{y}_{m\cdot\text{RSS}}$	100	100	100	100
$\bar{y}_{\text{RAT,RSS}}$	120.7275	115.922	111.4844	107.374
$\bar{y}_{\text{KC}}$	150.3562	150.3562	150.3562	150.3562
$\bar{y}_{\text{COMP,RSS}} = \bar{y}_{s_i\cdot\text{RSS}},\ i = 1, 2$	155.9138	140.232	127.4165	116.7471
$\bar{y}_{\text{INP,RSS}}$	156.9786	141.4838	128.7846	118.1832

The relationship between the variables was defined as follows:

$$Y = 2.9 + \sqrt{1-\rho_{xy}^2}\,Y^* + \rho_{xy}\left(\frac{S_y}{S_x}\right)X^*,$$

$$X = 2.5 + X^*.$$

Here, X^* and Y^* represent independent variables from the corresponding parent population, and ρ_{xy} denotes the correlation coefficient between them. The quantities S_x and S_y represent the standard deviations of X and Y, respectively. The variables X and Y are generated for the values of $\rho_{xy} = 0, 0.5, 0.6, 0.7, 0.8,$ and 0.9.

Using the sampling methodology described in Section 2, a Ranked Set Sample (RSS) of size 12 (with set size 3) was drawn from each population. Missing observations were then introduced in the study variable. The proposed imputation methods, along with the competing imputation methods, were subsequently applied to handle the missing data. The simulation experiment was replicated $1{,}000$ times. The empirical mean squared errors (MSEs) and percentage relative efficiencies (PREs) of the proposed estimators were computed in comparison with the conventional mean estimator. The MSE are calculated using the formula:

$$\text{MSE}(t_m) = \frac{1}{1000}\sum_{i=1}^{1000}\left(t_m - \mu_y\right)^2.$$

PRE is defined as the ratio of the MSE of the standard estimator to that of the modified estimator, expressed as a percentage.

$$\text{PRE}(t_m) = \frac{\text{MSE}(T)}{\text{MSE}(t_m)} \times 100.$$

where:
t_m is the estimate produced by the proposed method,
T denotes the conventional sample mean, and
μ_y is the true population mean of the study variable Y.

The outcomes of these simulations are summarized in Tables 8–11, which present the PRE values for each combination of the correlation coefficient $\rho_{xy} = 0, 0.5, 0.6, 0.7, 0.8,$ and 0.9.

Table 8: PRE of the proposed estimators for Population 7 ($X^* \sim N(20, 25)$, $Y^* \sim N(30, 35)$; $N = 500$, $n = 51$, $r = 30$, $P = 0.4$)

Correlation ρ_{yx}	0.5	0.6	0.7	0.8	0.9
$\bar{y}_{(m\cdot\text{RSS})}$	100	100	100	100	100
$\bar{y}_{(\text{RAT,RSS})}$	85.5193	92.2621	100.0593	109.7248	128.3454
$\bar{y}_{\text{KC}}$	75.5056	78.8770	82.8667	87.1631	97.3327
$\bar{y}_{(\text{COMP,RSS})}$ $=\bar{y}_{(s_i\cdot\text{RSS})},\ i=1,2$	102.0854	104.1194	107.2481	112.9200	128.3695
$\bar{y}_{(\text{INP,RSS})}$	105.5448	107.4949	110.3071	115.9034	131.2728

Table 9: PRE OF THE PROPOSED ESTIMATORS FOR POPULATION 8 ($X^* \sim N(20, 25), Y^* \sim N(30, 35)$; $N = 200$, $n = 24$, $r = 14$, $P = 0.4$)

Correlation ρ_{yx}	0.5	0.6	0.7	0.8	0.9
$\bar{y}_{(m\cdot\text{RSS})}$	100	100	100	100	100
$\bar{y}_{(\text{RAT,RSS})}$	83.3543	90.8511	97.5748	110.0237	130.6380
$\bar{y}_{\text{KC}}$	71.5724	75.0670	76.8751	83.3607	90.6031
$\bar{y}_{(\text{COMP,RSS})}$ $=\bar{y}_{(s_i\cdot\text{RSS})},\ i=1,2$	102.4627	104.8400	107.3591	114.0499	130.2264
$\bar{y}_{(\text{INP,RSS})}$	109.2898	111.4921	114.2973	121.9163	137.2337

Table 10: PRE OF THE PROPOSED ESTIMATORS FOR POPULATION 9 ($X^* \sim N(20, 25), Y^* \sim N(30, 35)$; $N = 2000, n = 256, r = 204$, $P = 0.2$)

Correlation ρ_{yx}	0.5	0.6	0.7	0.8	0.9
$\bar{y}_{(m\cdot\text{RSS})}$	100	100	100	100	100
$\bar{y}_{(\text{RAT,RSS})}$	97.1107	99.6719	103.3056	108.3639	117.2199
$\bar{y}_{\text{KC}}$	90.4453	91.5421	93.3596	96.3177	101.6550
$\bar{y}_{(\text{COMP,RSS})}$ $=\bar{y}_{(s_i\cdot\text{RSS})},\ i=1,2$	102.1777	103.1581	105.2039	103.1849	117.2702

Table 11: PRE OF THE PROPOSED ESTIMATORS FOR POPULATION 10 ($X^* \sim N(20,25)$, $Y^* \sim N(30,35)$; $N = 2000$, $n = 255$, $r = 155$, $P = 0.4$)

Correlation ρ_{yx}	0.5	0.6	0.7	0.8	0.9
$\bar{y}_{(m\cdot\text{RSS})}$	100	100	100	100	100
$\bar{y}_{(\text{RAT,RSS})}$	92.9995	98.2934	105.5728	115.4018	134.7492
$\bar{y}_{\text{KC}}$	81.4296	83.7460	87.0355	91.7301	101.9212
$\bar{y}_{(\text{COMP,RSS})} = \bar{y}_{(s_i\cdot\text{RSS})}$, i=1,2	103.6397	105.7200	109.8157	116.7968	134.8287
$\bar{y}_{(\text{INP,RSS})}$	104.2385	106.1774	110.1754	117.1900	135.0721

7 Interpretations of the Computational Results

The following interpretations can be drawn from the above tables:

Table 1 presents the parametric values for the six real populations considered in the analysis. The results of the numerical investigations, based on theoretical comparisons, are reported in Tables 2–7 for all six populations. It is evident from Tables 2–7 that the proposed imputation method, $\bar{y}_{(\text{INP,RSS})}$, outperforms both the conventional imputation methods and the adapted compromised imputation method proposed by Singh and Horn (2000) in terms of Percent Relative Efficiency (PRE).

As shown in Tables 2–7, the proposed imputation technique, $\bar{y}_{(.i,\text{RSS})}$, consistently outperforms the conventional mean imputation method, the ratio method of imputation, as well as the imputation procedures suggested by Al-Omari and Bouza (2014), and Sohail, Shabbir, and Ahmed (2018), across all real-world populations. Furthermore, it is observed from Tables 2–7 that as the proportion of non-response decreases, the PRE of the proposed estimator consistently increases across all six populations.

To establish the superiority of our proposed estimator, simulations were also conducted using data generated from Normal distributions. The results are presented in Tables 8–11, which provide a comprehensive summary of performance across all population types. The findings clearly demonstrate that the proposed imputation strategies outperform the mean imputation method, the ratio imputation method, the regression-cum-ratio type imputation methods suggested by Al-Omari and Bouza (2014), the ratio-type estimators proposed by Sohail, Shabbir, and Ahmed (2018), and the adapted compromised imputation method proposed by Singh and Horn (2000). This superiority holds across a wide range of correlation coefficients ρ_{xy}.

The simulation results presented in Tables 8–15 reflect similar patterns in PRE as observed with the real data sets. Both theoretical and empirical evaluations confirm the superior performance of our proposed estimator.‘

8 Conclusion

This paper proposed a modified version of the compromised estimator of Singh and Horn (2000) using Ranked Set Sampling (RSS). Our proposed estimator is judged superior because it achieves the lowest mean squared error (MSE) and the highest percent relative efficiency (PRE) in every theoretical, real-data, and simulated scenario examined.

The results from numerical comparisons show that the proposed method consistently outperforms conventional imputation techniques—including mean and ratio imputation—as well as the imputation procedures suggested by Al-Omari and Bouza (2014), Sohail, Shabbir, and Ahmed (2018), and the compromised estimator adapted from Singh and Horn (2000), in terms of PRE. Moreover, the efficiency of the proposed estimator improves as the non-response rate decreases.

Simulation findings further validate the superiority of the proposed imputation method across a range of correlation coefficients ρ_{xy}, confirming its robustness and reliability for practical applications.

References

Ratio estimators of the population mean with missing values using ranked set sampling. *Environmetrics*, 26, 67–76.

Bouza, C. N. (2010). Imputation methods in ranked set sampling for finite population estimation. *SORT*, 34(1), 79–96.

Bouza, C. N. (2012). Ratio and regression estimators in ranked set sampling with missing data. *Journal of Applied Statistics*, 39(9), 1921–1934.

Bouza, C. N. (2013a). Imputation methods in ranked set sampling. *Journal of Statistical Computation and Simulation*, **(**8), 1341–1354.

Bouza, C. N. (2013b). *Handling Missing Data in Ranked Set Sampling*. Springer, Berlin/Heidelberg.

Bouza, C. N., Al-Omari, A. I. (2011). Ranked set estimation with imputation of the missing observations: The median estimator. *Revista Investigación Operacional*, 32, 30–37.

Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). John Wiley & Sons, New York.

Diana, G., Perri, P. F. (2010). Improved estimators of the population mean for missing data. *Communications in Statistics–Theory and Methods*, 39, 3245–3251.

Heitjan, D. F., Basu, S. (1996). Distinguishing missing at random and missing completely at random. *The American Statistician*,50, 207–213.

Horvitz, D. G., Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663–685.

Kadılar, C., Cingi, H. (2003). Yield estimation in stratified sampling using auxiliary information. *International Journal of Agricultural and Statistical Sciences*, 1(1), 3–10.

Kadılar, C., Cingi, H. (2008). Estimators for the population mean in the case of missing data. *Communications in Statistics–Theory and Methods*,37, 2226–2236.

Lee, H., Rancourt, E., Särndal, C. E. (1994). Experiments with variance estimation from survey data with imputed values. *Journal of Official Statistics*, 10(3), 231–243.

McIntyre, G. A. (1952). A method of unbiased selective sampling using ranked sets. *Australian Journal of Agricultural Research*, 3, 385–390.

Mohamed, C., Sedory, S. A., Singh, S. (2016). Imputation using higher order moments of an auxiliary variable. *Communications in Statistics–Simulation and Computation*, 46, 6588–6617.

Murthy, M. N. (1967). *Sampling Theory and Methods*. Statistical Publishing Society, Calcutta.

Rubin, D. B. (1978). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, New York.

Samawi, H. M., Muttlak, H. A. (1996). Estimation of the population mean using ratio and product estimators in simple random sampling. *Communications in Statistics–Theory and Methods*, 25(2), 651–665.

Särndal, C. E., Swensson, B., Wretman, J. (2003). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

Singh, R., Mangat, N. S. (1996). *Elements of Survey Sampling*. Kluwer Academic Publishers, Boston.

Singh, S. (2003). *Advanced Sampling Theory with Applications*, Vols. 1–2. Kluwer Academic Publishers, Dordrecht.

Singh, S., Horn, S. (2000). Compromised imputation in survey sampling. *Metrika*, 51, 267–276.

Sohail, M. U., Shabbir, J., Ahmed, S. (2018). A class of ratio type estimators for imputing the missing values under ranked set sampling. *Journal of Statistical Theory and Practice*, 12, 704–717.