



Neutrosophic imputation: a novel approach to population mean estimation under indeterminacy

Vishal Kumar¹ · Anoop Kumar¹

Received: 10 April 2025 / Revised: 4 October 2025 / Accepted: 22 October 2025

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2025

Abstract

Accurate population mean estimation is crucial in survey sampling, particularly when dealing with missing or partial data. The intrinsic uncertainty and indeterminacy found in these datasets frequently pose challenges for computational imputation approaches. In this article, by including neutrosophic logic into the imputation process, some novel neutrosophic imputation approaches and their corresponding resultant neutrosophic estimators for population mean estimation are presented along with their properties under simple random sampling (SRS). Neutrosophic logic offers a strong foundation for overcoming the drawbacks of conventional approaches because of its capacity to handle indeterminate and uncertain data. We show that the proposed novel neutrosophic imputation performs better than the conventional neutrosophic imputation approaches through comprehensive simulations and practical experiments. The simulation and practical results demonstrate how the proposed approaches may be used to increase the precision of statistical studies, particularly in domains where ambiguity and partial data are common.

Keywords Missing data · Neutrosophic Imputation · Efficiency · Neutrosophic logic · Indeterminate information

1 Introduction

In survey sampling, one of the most important tasks is the estimation of the population mean, which is crucial in many disciplines including social sciences, health, economics. The accuracy of these estimates often depends on the quantity and caliber of the data collected. However, partial or missing data is a common issue in real-life data collection, which can seriously reduce the validity of statistical inferences. To deal with this issue, the conventional imputation approaches including mean imputation, regression imputation, power transformation imputation, exponential imputation, logarithmic imputation have been developed. A number of distinguished researchers have investigated the issue of missing data by developing several imputation approaches. The use of ideal auxiliary data is also important to elevate the effectiveness of the imputation approaches that have been developed. In his fundamental work, Rubin

[40] established three primary notions, including missing at random (MAR), observed at random (OAR), and parameter distribution (PD). Heitjan and Basu [29] distinguished between MAR and missing completely at random (MCAR). Ahmed et al. [5] suggested population mean estimation utilizing several imputation approaches. Bouza and Al-Omari [23] presented ranked set sampling (RSS) estimation procedures with imputation of the missing data. To estimate the population mean utilizing SRS with a known correlation coefficient, Al-Omari et al. [6] advocated some imputation approaches for missing data. The ratio estimators for the population mean with missing values under RSS were designed by Al-Omari and Bouza [7]. Mahdizadeh and Zamanzade [36] developed a smooth estimation of a reliability function under ranked set sampling. Bhushan and Pandey [17] introduced a cost-effective computational technique under non-response on two occasions. Anas et al. [10] looked at robust quantile regression mean estimators based on compromised imputation. To cope with the missing data under SRS, Bhushan et al. [19] looked into a population mean estimation approach. In contrast, Bhushan et al. [20] presented several effective difference and ratio-type imputation approaches under RSS. A few multi-auxiliary information based imputation approaches under RSS were proposed

✉ Anoop Kumar
anoop.asy@gmail.com

¹ Department of Statistics, Central University of Haryana, Mahendergarh 123031, India

by Bhushan and Kumar [21]. Some RSS imputation approaches were proposed by Bhushan and Kumar [22] to address the correlated measurement errors. Under stratified SRS, Kumar et al. [32] constructed some novel imputation approaches. Bhushan and Pandey [18] suggested optimal random non-response for mean estimation on current occasion. For the purpose of estimating the domain mean, Kumar et al. [33] suggested direct and synthetic imputation approaches under small area. Even while these approaches offer useful tools for handling missing data, they frequently prove inadequate when confronted with the underlying ambiguity and uncertainty found in many datasets.

Classical set theory (Cantor, 19th century) is binary, where an element either belongs or does not belong to a set, leaving no scope for ambiguity. To overcome this rigidity, Zadeh [54] introduced fuzzy sets, where membership is gradual ($\mu_A(x) \in [0, 1]$), allowing representation of vagueness. Later, Atanassov [4] proposed intuitionistic fuzzy sets (IFS), extending fuzzy sets by including both membership $\mu_A(x)$, non-membership $\nu_A(x)$, and hesitation $\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$, thereby handling uncertainty more explicitly. However, both fuzzy and IFS restrict the sum of membership components, limiting their ability to represent strong indeterminacy or conflicting evidence. To generalize, Smarandache [42] introduced neutrosophic sets, where each element has three independent degrees such as truth $T(x)$, indeterminacy $I(x)$, and falsity $F(x)$, with $T(x), I(x), F(x) \subseteq [0^-, 1^+]$, and no restriction that $T + I + F = 1$.

Traditional statistics deals with single-valued or crisp data, where values are assumed to be precise and certain. However, this framework becomes inadequate when the data exhibit indeterminacy or uncertainty. Such indeterminacy often arises when observations are expressed in intervals. To address this limitation, fuzzy statistics, based on fuzzy logic and fuzzy sets, was developed by Zadeh [54], who first introduced the concept of fuzziness through fuzzy sets. Since then, fuzzy logic has been extensively advanced and widely applied in decision-making contexts by several authors including, Chhokar and Nagar [27], Kumar and Dudeja [30], and Punia et al. [48] among others. Despite its utility, fuzzy statistics suffers from a critical shortcoming as it fails to explicitly account for indeterminacy in the data. However, it manages uncertainty, it does not provide a framework to capture the degree of indeterminacy, which is often inherent in real-world scenario. This gap was filled by neutrosophic statistics, which extended fuzzy statistics by incorporating measures of truth, falsity, and indeterminacy simultaneously. Neutrosophic logic, introduced by Smarandache [42], provided a more generalized and flexible framework to handle uncertainty and indeterminacy. For further details on neutrosophic sets, logic, and statistics, see

Smarandache [43–45]. The field of neutrosophic statistics has since gained significant attention by some prominent researchers. Smarandache [46] formally introduced neutrosophic statistics as a distinct discipline. Subsequent contributions have extended its scope to diverse applications. For instance, Chen et al. [26] proposed neutrosophic interval statistical numbers for evaluating rock joint roughness coefficients, while Alhabib et al. [1] explored neutrosophic probability statistics. Chen [25] examined robust single-valued neutrosophic soft aggregation operators in multi-criteria decision-making. Chakraborty et al. [24] investigated the classification of trapezoidal bipolar neutrosophic numbers, introduced a de-bipolarization technique, and demonstrated its application to cloud service-based MCGDM problems. Many authors including, Nabeeh et al. [38], Aslam et al. [14], Aslam and Algarhi [15], Aslam [11], Aslam [12], Aslam [13], Arif and Aslam [16], Almarashi and Aslam [9], Rao and Aslam [39], among others, have also enriched the literature on neutrosophic statistics by proposing innovative estimation procedures and extensions.

In the context of survey sampling, prior research has largely focused on precise and determinate data, where conventional estimation methods yield single-valued results that may over or underestimate the true population value. However, such methods are inadequate for handling neutrosophic data, which often involve uncertainty, imprecision, and interval-valued observations. The survey literature contains a few studies for mean estimation when the study and auxiliary variables are neutrosophic in character. The neutrosophic ratio-type population mean estimation method was initially developed under SRS by Tahir et al. [51]. Yadav and Smarandache [52] subsequently put up a generalized neutrosophic sampling technique for estimating population mean. Nevertheless, with supplementary data, Yadav and Prasad [53] developed the neutrosophic factor-type exponential estimators. Alqudah et al. [8] presented the neutrosophic robust ratio-type estimator for the mean estimation. Masood et al. [37] recommended to estimate the neutrosophic finite median using robust estimates of the auxiliary variable. Alomair and Ahmad [2] proposed a comprehensive neutrosophic regression-cum-exponential type estimator for the population mean. Al-Marzouki and Ahmad [3] introduced a generalized class of estimators for population mean within the neutrosophic framework. Priya and Kumar [47] developed the robust neutrosophic exponential estimators for the population mean. Kumar et al. [31] formulated an optimal neutrosophic framework for the population mean estimation under SRS. But, these studies fail to provide estimates of the population mean in the presence of missing values.

Several imputation approaches have been proposed in the literature to address the missingness in determinate data,

but they usually fall short in managing the uncertainty and indeterminacy that come with missing data. Neutrosophic imputation is a revolutionary development in survey sampling, especially for the population mean estimation in the presence of missingness under uncertain data. The special quality of neutrosophic logic makes it a viable option for developing neutrosophic imputation approaches in survey sampling. Therefore, we emphasize that the paper:

- (i) introduces a novel class of neutrosophic imputation approaches that simultaneously incorporate truth, falsity, and indeterminacy to better capture uncertainty in real datasets,
- (ii) derives the properties such as bias and mean square error (MSE) of the resultant estimators under SRS,
- (iii) establishes efficiency comparisons showing clear improvements over the adapted or existing neutrosophic imputation approaches, and
- (iv) validates the proposed methods through extensive simulation study and real data applications from environmental and financial domains.

These additions more prominently show how this study advances the literature on handling missing data under indeterminacy, thereby underlining the novelty and practical relevance of the proposed methodology.

1.1 Notations

A neutrosophic set generalizes the concepts of classical, fuzzy, and intuitionistic fuzzy sets. Each element in a neutrosophic set is characterized by three independent components such as truth-membership (T), indeterminacy-membership (I), and falsity-membership (F), where $T, I, F \subseteq [0,1]$. Unlike intuitionistic fuzzy sets, neutrosophic sets do not restrict $T + I + F = 1$; instead, each of the three values may independently vary within $[0,1]$, allowing flexible modeling of incomplete, inconsistent, and ambiguous information. However, a neutrosophic number is a numerical value expressed through neutrosophic components and is often represented in an interval form, indicating that the observed value lies within a range due to uncertainty or indeterminacy. In statistical literature, such neutrosophic data are frequently modeled as quantitative data existing in an unknown interval $[p, q]$. These interval-based neutrosophic numbers can take different forms; in this paper, we adopt the representation $W_N = W_L + W_U I_N$ with $I_N \in [I_L, I_U]$. Thus, neutrosophic data are consistently denoted as $W_N \in [p, q]$, where p and q represent the lower and upper bounds, respectively (see Smarandache [46] for further details on neutrosophic notations).

In a classical random sample, the sample size is fixed (say n), and each observation is recorded as a precise, determinate value. The randomness comes solely from the sampling design. However, in a neutrosophic random sample, the sample size itself may be uncertain and is represented as an interval $n_N \in [n_L, n_U]$. Consider a neutrosophic random sample of size $n_N \in [n_L, n_U]$, drawn from a finite population $\Omega = (\Omega_1, \Omega_2, \dots, \Omega_N)$ of N units. Let $y_N(i) \in [y_L, y_U]$ and $x_N(i) \in [x_L, x_U]$ denote the neutrosophic observations on the i^{th} unit for the study variable and the auxiliary variable, respectively. Suppose a_N units respond from the drawn n_N sample. Let A_N and $\bar{A}_N$ be the sets of responding and non-responding units, respectively. Thus, the values $y_N(i)$ are available for all $i \in A_N$, while the values for $i \in \bar{A}_N$ are missing and require imputation. The imputation is performed using the additional neutrosophic auxiliary information $x_N(i)$, which is assumed to be available and strictly positive for all $i \in S_N$. Let the population means of the study and auxiliary variables be denoted as $\bar{Y}_N \in [\bar{Y}_L, \bar{Y}_U]$ and $\bar{X}_N \in [\bar{X}_L, \bar{X}_U]$, with their corresponding sample means $\bar{y}_N \in [\bar{y}_L, \bar{y}_U]$ and $\bar{x}_N \in [\bar{x}_L, \bar{x}_U]$. The neutrosophic population standard deviations are denoted by $S_{y_N} \in [S_{y_L}, S_{y_U}]$ and $S_{x_N} \in [S_{x_L}, S_{x_U}]$. The neutrosophic coefficients of variation are represented by $C_{y_N} \in [C_{y_L}, C_{y_U}]$ and $C_{x_N} \in [C_{x_L}, C_{x_U}]$, while the neutrosophic correlation coefficient between x_N and y_N is denoted by ρ_{xy_N} . Further, the flow chart of the methodology used is presented in Fig. 1.

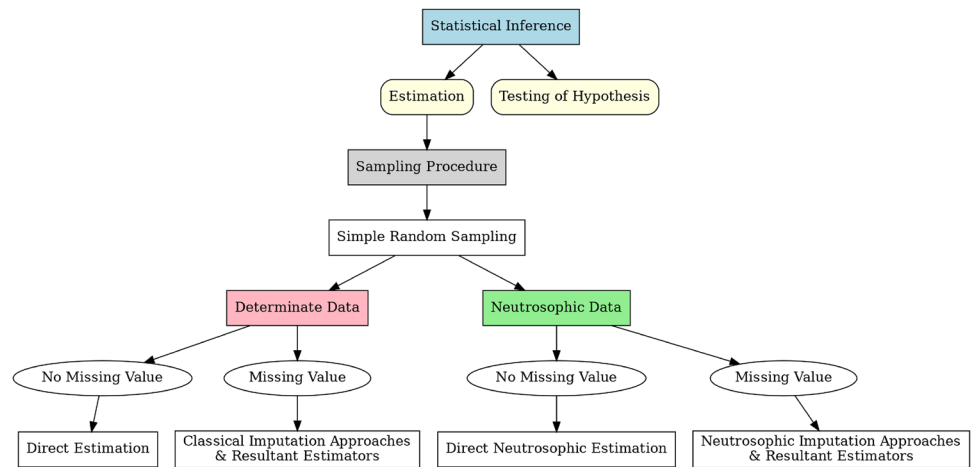
To deduce the $Bias_N \in [Bias_L, Bias_U]$ and $MSE_N \in [MSE_L, MSE_U]$ of the resultant neutrosophic estimators, we define the neutrosophic error terms $e_{0N} \in [e_{0L}, e_{0U}]$, $e_{1N} \in [e_{1L}, e_{1U}]$, and $e_{2N} \in [e_{2L}, e_{2U}]$ with their expected values as follows:

$$\left. \begin{aligned} e_{0N} &= \frac{(\bar{y}_N - \bar{Y}_N)}{\bar{Y}_N}, \quad e_{1N} = \frac{(\bar{x}_N - \bar{X}_N)}{\bar{X}_N}, \\ e_{2N} &= \frac{(\bar{x}_N - \bar{X}_N)}{\bar{X}_N}, \\ E(e_{0N}^2) &= f_{rN} C_{y_N}^2, \quad E(e_{1N}^2) = f_{rN} C_{x_N}^2, \\ E(e_{2N}^2) &= f_{nN} C_{x_N}^2, \\ E(e_{0N}e_{1N}) &= f_{rN} \rho_{xy_N} C_{x_N} C_{y_N}, \quad E(e_{0N}e_{2N}) = f_{nN} \rho_{xy_N} C_{x_N} C_{y_N}, \\ E(e_{1N}e_{2N}) &= f_{nN} C_{x_N}^2, \end{aligned} \right\} \quad (1.1)$$

where $f_{nN} = (n_N^{-1} - N^{-1})$, $f_{rN} = (r_N^{-1} - N^{-1})$, $f_{rnN} = (r_N^{-1} - n_N^{-1})$, $C_{y_N} = S_{y_N}/\bar{Y}_N$, $C_{x_N} = S_{x_N}/\bar{X}_N$, and $\rho_{xy_N} = S_{xy_N}/S_{x_N}S_{y_N}$.

Since the computations in this paper are carried out on interval-valued neutrosophic numbers, the basic arithmetic operators on intervals are briefly recalled below. Let $C = [c_L, c_U]$ and $D = [d_L, d_U]$ be two closed intervals of real numbers, where $c_L \leq c_U$ and $d_L \leq d_U$. Then:

1. **Addition** $C + D = [c_L + d_L, c_U + d_U]$
2. **Subtraction** $C - D = [c_L - d_U, c_U - d_L]$

Fig. 1 Flow chart of the methodology used

3. **Multiplication** $C \times D = [\min(c_L d_L, c_L d_U, c_U d_L, c_U d_U), \max(c_L d_L, c_L d_U, c_U d_L, c_U d_U)]$

4. **Division** (when $0 \notin [d_L, d_U]$) $C \div D = C \times \left[\frac{1}{d_U}, \frac{1}{d_L} \right]$

These rules allow consistent manipulation of neutrosophic quantities such as means, variances, bias, and MSE, which are all expressed in interval form in this study.

There will be numerous sections in the forthcoming article. In Sect. 2, we adapt some prominent neutrosophic imputation approaches along with their properties. In Sect. 3, we propose some novel neutrosophic imputation approaches and resultant neutrosophic estimators to handle missingness in the indeterminate data. The efficiency comparison of the proposed neutrosophic imputations with the adapted neutrosophic imputations is provided in Sect. 4. An extensive simulation and a real data application that contrast the effectiveness of our suggested neutrosophic imputations with the conventional neutrosophic imputations are then presented in Sect. 5 and Sect. 6, respectively. Lastly, we talk about the significance of our results and possible directions for further investigation in Sect. 7.

2 Adapted neutrosophic imputation approaches

The versatility of adapted neutrosophic imputation approaches to manage the inherent uncertainty and ambiguity associated with the missing data in survey sampling makes them significant. In order to fill in the gaps in the literature study and assess the effectiveness of the suggested novel neutrosophic imputation approaches and the resultant neutrosophic estimators, we adapt certain well-known neutrosophic imputation approaches and the resultant neutrosophic estimators along with their characteristics.

2.1 Neutrosophic mean imputation approach

Referring Lee et al. [35], when the i^{th} sample unit is missing and require imputation then the neutrosophic mean imputation approach is prescribed below:

$$y_{i_{mN}} = \begin{cases} y_{i_N} & \text{for } i \in A_N \\ \bar{y}_{r_N} & \text{for } i \in \bar{A}_N \end{cases}$$

The resultant neutrosophic mean estimator is obtained as $T_{mN} = \bar{y}_{r_N}$ having variance as

$$V(T_{mN}) = f_{r_N} \bar{Y}_N^2 C_{y_N}^2. \quad (2.1)$$

In the neutrosophic mean imputation approach, missing values of the study variable y_N are imputed using available information. If no auxiliary data are considered, the missing values are simply replaced by the neutrosophic sample mean of the responding units, i.e., $\bar{y}_{r_N}$. However, when neutrosophic auxiliary information is available, the efficiency of the imputation procedure can be improved by incorporating it. Depending on whether the neutrosophic population mean $\bar{X}_N$ of the auxiliary variable is known, the following three situations arise:

1. **Situation I: $\bar{X}_N$ is available and $\bar{x}_{n_N}$ is considered.**

- In this case, the neutrosophic population mean of the auxiliary variable is known, and the neutrosophic sample mean of the entire sample of size n_N is used.
- The imputation utilizes the relationship between the known $\bar{X}_N$ and the sample mean $\bar{x}_{n_N}$, thereby improving the precision of the estimator compared to using only $\bar{y}_{r_N}$.

2. **Situation II: $\bar{X}_N$ is available and $\bar{x}_{r_N}$ is considered.**

- In this case, the neutrosophic population mean of the auxiliary variable is known. But instead of the full sample mean $\bar{x}_{n_N}$, we only rely on the neutrosophic mean of the responding units, $\bar{x}_{r_N}$.
- This situation arises when auxiliary information is only partially observed for the responding units, and it exploits the known population mean $\bar{X}_N$ for adjustment.

3. Situation III: $\bar{X}_N$ is not available, and $\bar{x}_{n_N}$ and $\bar{x}_{r_N}$ are considered.

- When the neutrosophic population mean of the auxiliary variable is not known, both the neutrosophic sample mean from all selected units $\bar{x}_{n_N}$ and from responding units $\bar{x}_{r_N}$ are used.
- In this case, the imputation relies entirely on the observed auxiliary data at the sample level. The adjustment is made by combining $\bar{x}_{n_N}$ and $\bar{x}_{r_N}$ to approximate the role of the unavailable $\bar{X}_N$.

2.2 Neutrosophic ratio imputation approaches

Referring the work of Singh and Horn [50], we adapt the conventional neutrosophic ratio imputation approaches under situations I-III as

Situation I

$$y_{i_{r_1N}} = \begin{cases} y_{i_N} & \text{for } i \in A_N \\ \frac{1}{n_N - r_N} \left[n_N \bar{y}_{r_N} \left(\frac{\bar{X}_N}{\bar{x}_{n_N}} \right) - r_N \bar{y}_{r_N} \right] & \text{for } i \in \bar{A}_N \end{cases}$$

Situation II

$$y_{i_{r_2N}} = \begin{cases} y_{i_N} & \text{for } i \in A_N \\ \frac{1}{n_N - r_N} \left[n_N \bar{y}_{r_N} \left(\frac{\bar{X}_N}{\bar{x}_{r_N}} \right) - r_N \bar{y}_{r_N} \right] & \text{for } i \in \bar{A}_N \end{cases}$$

Situation III

$$y_{i_{r_3N}} = \begin{cases} y_{i_N} & \text{for } i \in A_N \\ \frac{1}{n_N - r_N} \left[n_N \bar{y}_{r_N} \left(\frac{\bar{x}_{n_N}}{\bar{x}_{r_N}} \right) - r_N \bar{y}_{r_N} \right] & \text{for } i \in \bar{A}_N \end{cases}$$

The resultant neutrosophic ratio estimators under situations I-III are

$$T_{r_1N} = \bar{y}_{r_N} \left(\frac{\bar{X}_N}{\bar{x}_{n_N}} \right), \quad (2.2)$$

$$T_{r_2N} = \bar{y}_{r_N} \left(\frac{\bar{X}_N}{\bar{x}_{r_N}} \right), \quad (2.3)$$

$$\text{and } T_{r_3N} = \bar{y}_{r_N} \left(\frac{\bar{x}_{n_N}}{\bar{x}_{r_N}} \right). \quad (2.4)$$

The MSE of the neutrosophic ratio estimators is given below.

$$MSE(T_{r_1N}) = f_{r_N} \bar{Y}_N^2 C_{y_N}^2 + f_{n_N} \bar{Y}_N^2 (C_{x_N}^2 - 2\rho_{xy_N} C_{x_N} C_{y_N}), \quad (2.5)$$

$$MSE(T_{r_2N}) = f_{r_N} \bar{Y}_N^2 (C_{y_N}^2 + C_{x_N}^2 - 2\rho_{xy_N} C_{x_N} C_{y_N}), \quad (2.6)$$

$$\text{and } MSE(T_{r_3N}) = f_{n_N} \bar{Y}_N^2 C_{y_N}^2 + f_{r_N} \bar{Y}_N^2 (C_{y_N}^2 + C_{x_N}^2 - 2\rho_{xy_N} C_{x_N} C_{y_N}). \quad (2.7)$$

2.3 Neutrosophic regression imputation approaches

Referring the work of Diana and Perri [28], we adapt the neutrosophic regression imputation approaches under situations I-III as

Situation I

$$y_{i_{lr_1N}} = \begin{cases} y_{i_N} & \text{for } i \in A_N \\ \bar{y}_{r_N} + \frac{n_N \beta_1}{n_N - r_N} (\bar{X}_N - \bar{x}_{n_N}) & \text{for } i \in \bar{A}_N \end{cases}$$

Situation II

$$y_{i_{lr_2N}} = \begin{cases} y_{i_N} & \text{for } i \in A_N \\ \bar{y}_{r_N} + \frac{n_N \beta_2}{n_N - r_N} (\bar{X}_N - \bar{x}_{r_N}) & \text{for } i \in \bar{A}_N \end{cases}$$

Situation III

$$y_{i_{lr_3N}} = \begin{cases} y_{i_N} & \text{for } i \in A_N \\ \bar{y}_{r_N} + \frac{n_N \beta_3}{n_N - r_N} (\bar{x}_{n_N} - \bar{x}_{r_N}) & \text{for } i \in \bar{A}_N \end{cases}$$

The resultant neutrosophic regression estimators under situations I-III are

$$T_{lk_1N} = \bar{y}_{r_N} + \beta_1 (\bar{X}_N - \bar{x}_{n_N}), \quad (2.8)$$

$$T_{lk_2N} = \bar{y}_{r_N} + \beta_2 (\bar{X}_N - \bar{x}_{r_N}), \quad (2.9)$$

$$\text{and } T_{lk_3N} = \bar{y}_{r_N} + \beta_3 (\bar{x}_{n_N} - \bar{x}_{r_N}), \quad (2.10)$$

where β_i , $i = 1, 2, 3$ are neutrosophic regression coefficients of y_N on x_N .

The minimum MSE of the resultant neutrosophic regression estimators at the optimum values of $\beta_{i(opt)} = (\bar{Y}_N / \bar{X}_N) \rho_{xy_N} (C_{y_N} / C_{x_N})$, $i = 1, 2, 3$ is given below.

$$\min.MSE(T_{lr_1N}) = \bar{Y}_N^2 C_{y_N}^2 (f_{r_N} - f_{n_N} \rho_{xy_N}^2), \quad (2.11)$$

$$\min.MSE(T_{lr_2N}) = f_{r_N} \bar{Y}_N^2 C_{y_N}^2 (1 - \rho_{xy_N}^2), \quad (2.12)$$

$$\text{and } \min.MSE(T_{l_{r_{3N}}}) = f_{n_N} \bar{Y}_N^2 C_{y_N}^2 + f_{r_{n_N}} C_{y_N}^2 \bar{Y}_N^2 (1 - \rho_{xy_N}^2). \quad (2.13)$$

2.4 Neutrosophic logarithmic imputation approaches

Referring the work of Bhushan et al. [19], we adapt some neutrosophic logarithmic imputation approaches under situations I-III as

Situation I

$$y_{i_{l_{1N}}} = \begin{cases} \frac{y_{i_N}}{n_N - r_N} \left[n_N \bar{y}_{r_N} \left\{ 1 + \log \left(\frac{\bar{x}_{n_N}}{\bar{X}_N} \right) \right\}^{\varphi_1} \right] & \text{for } i \in A_N \\ -r_N \bar{y}_{r_N} & \text{for } i \in \bar{A}_N \end{cases}$$

Situation II

$$y_{i_{l_{2N}}} = \begin{cases} \frac{y_{i_N}}{n_N - r_N} \left[n_N \bar{y}_{r_N} \left\{ 1 + \log \left(\frac{\bar{x}_{r_N}}{\bar{X}_N} \right) \right\}^{\varphi_2} \right] & \text{for } i \in A_N \\ -r_N \bar{y}_{r_N} & \text{for } i \in \bar{A}_N \end{cases}$$

Situation III

$$y_{i_{l_{3N}}} = \begin{cases} \frac{y_{i_N}}{n_N - r_N} \left[n_N \bar{y}_{r_N} \left\{ 1 + \log \left(\frac{\bar{x}_{n_N}}{\bar{X}_N} \right) \right\}^{\varphi_3} \right] & \text{for } i \in A_N \\ -r_N \bar{y}_{r_N} & \text{for } i \in \bar{A}_N \end{cases}$$

The resultant neutrosophic logarithmic estimators under situations I-III are

$$y_{i_{e_{1N}}} = \begin{cases} \frac{y_{i_N}}{n_N - r_N} \left(\left[n_N \left[\Gamma_1 \bar{y}_{r_N} + \Delta_1 \bar{y}_{r_N} \left\{ \frac{\bar{X}_N}{\theta \bar{x}_{n_N} + (1 - \theta) \bar{X}_N} \right\}^g \right] \times \left\{ 1 + \log \left(\frac{\bar{x}_{n_N}}{\bar{X}_N} \right) \right\}^{\varphi_1} \right] \right) & \text{for } i \in A_N \\ -r_N \bar{y}_{r_N} & \text{for } i \in \bar{A}_N \end{cases}$$

Situation II

$$y_{i_{e_{2N}}} = \begin{cases} \frac{y_{i_N}}{n_N - r_N} \left(\left[n_N \left[\Gamma_2 \bar{y}_{r_N} + \Delta_2 \bar{y}_{r_N} \left\{ \frac{\bar{X}_N}{\theta \bar{x}_{r_N} + (1 - \theta) \bar{X}_N} \right\}^g \right] \times \left\{ 1 + \log \left(\frac{\bar{x}_{r_N}}{\bar{X}_N} \right) \right\}^{\varphi_2} \right] \right) & \text{for } i \in A_N \\ -r_N \bar{y}_{r_N} & \text{for } i \in \bar{A}_N \end{cases}$$

$$T_{l_{1N}} = \bar{y}_{r_N} \left[1 + \log \left(\frac{\bar{x}_{n_N}}{\bar{X}_N} \right) \right]^{\varphi_1}, \quad (2.14)$$

$$T_{l_{2N}} = \bar{y}_{r_N} \left[1 + \log \left(\frac{\bar{x}_{r_N}}{\bar{X}_N} \right) \right]^{\varphi_2}, \quad (2.15)$$

$$\text{and } T_{l_{3N}} = \bar{y}_{r_N} \left[1 + \log \left(\frac{\bar{x}_{n_N}}{\bar{X}_N} \right) \right]^{\varphi_3}, \quad (2.16)$$

where $\varphi_i, i=1,2,3$ are appropriate scalars to optimize MSE .

The minimum MSE of the resultant neutrosophic logarithmic estimators at the optimum values of $\varphi_{i(opt)} = -\rho_{xy_N} (C_{y_N}/C_{x_N})$, $i = 1, 2, 3$ is given below.

$$\min.MSE(T_{l_{1N}}) = \bar{Y}_N^2 C_{y_N}^2 (f_{r_N} - f_{n_N} \rho_{xy_N}^2), \quad (2.17)$$

$$\min.MSE(T_{l_{2N}}) = f_{r_N} \bar{Y}_N^2 C_{y_N}^2 (1 - \rho_{xy_N}^2), \quad (2.18)$$

$$\text{and } \min.MSE(T_{l_{3N}}) = f_{n_N} \bar{Y}_N^2 C_{y_N}^2 + f_{r_{n_N}} C_{y_N}^2 \bar{Y}_N^2 (1 - \rho_{xy_N}^2). \quad (2.19)$$

Remark 2.1 Notably, for situations I-III, the minimum MSE of the resultant neutrosophic logarithmic estimators is comparable to that of the neutrosophic regression estimators.

2.5 Kumar and Kumar [34] neutrosophic imputation approaches

Kumar and Kumar [34] proposed some neutrosophic imputation approaches under situations I-III as

Situation I

Situation III

$$y_{i_{e3N}} = \begin{cases} y_{i_N} & \text{for } i \in A_N \\ \frac{1}{n_N - r_N} \left(\left[n_N \left[\Gamma_3 \bar{y}_{r_N} + \Delta_3 \bar{y}_{r_N} \left\{ \frac{\bar{x}_{n_N}}{\theta \bar{x}_{r_N} + (1-\theta) \bar{x}_{n_N}} \right\}^g \right] \right. \right. \\ \quad \times \left. \left. \left\{ 1 + \log \left(\frac{\bar{x}_{r_N}}{\bar{x}_{n_N}} \right) \right\}^{\varphi_3} \right] \right) & \text{for } i \in \bar{A}_N \\ -r_N \bar{y}_{r_N} \end{cases}$$

The resultant neutrosophic estimators under situations I-III are given by

$$T_{e1N} = [\Gamma_1 \bar{y}_{r_N} + \Delta_1 \bar{y}_{r_N} \left\{ \frac{\bar{X}_N}{\theta \bar{x}_{n_N} + (1-\theta) \bar{X}_N} \right\}^g] \left\{ 1 + \log \left(\frac{\bar{x}_{n_N}}{\bar{X}_N} \right) \right\}^{\varphi_1} \quad (2.20)$$

$$T_{e2N} = [\Gamma_2 \bar{y}_{r_N} + \Delta_2 \bar{y}_{r_N} \left\{ \frac{\bar{X}_N}{\theta \bar{x}_{r_N} + (1-\theta) \bar{X}_N} \right\}^g] \left\{ 1 + \log \left(\frac{\bar{x}_{r_N}}{\bar{X}_N} \right) \right\}^{\varphi_2} \quad (2.21)$$

$$T_{e3N} = [\Gamma_3 \bar{y}_{r_N} + \Delta_3 \bar{y}_{r_N} \left\{ \frac{\bar{x}_{n_N}}{\theta \bar{x}_{r_N} + (1-\theta) \bar{x}_{n_N}} \right\}^g] \left\{ 1 + \log \left(\frac{\bar{x}_{r_N}}{\bar{x}_{n_N}} \right) \right\}^{\varphi_3} \quad (2.22)$$

where Γ_j , $j = 1, 2, 3$, Δ_j , and φ_j are optimizing scalars. Also, g and θ are constants that take real values to design different neutrosophic imputation approaches and the corresponding resultant estimators.

The MSE of the resultant neutrosophic estimators T_{ej} , $j = 1, 2, 3$ of the neutrosophic estimators are given by

$$MSE(T_{ejN}) = \bar{Y}_N^2 (1 + \Gamma_j^2 A_j + \Delta_j^2 B_j + 2\Gamma_j \Delta_j C_j - 2\Gamma_j D_j - 2\Delta_j E_j) \quad (2.23)$$

The optimum values of scalars Γ_j , Δ_j , and φ_j , $j = 1, 2, 3$ are given by

$$\Gamma_{j(opt)} = \frac{B_j D_j - C_j E_j}{A_j B_j - C_j^2}, \quad \Delta_{j(opt)} = \frac{A_j E_j - C_j D_j}{A_j B_j - C_j^2}, \quad (2.24)$$

and $\varphi_{j(opt)} = -\rho_{xyN} \frac{C_{yN}}{C_{xN}}.$

The minimum MSE can be obtained by putting optimum values of Γ_j and Δ_j , $j = 1, 2, 3$ in (2.23) as

$$\min.MSE(T_{ejN}) = \bar{Y}_N^2 \left(1 - \frac{A_j E_j^2 + B_j D_j^2 - 2C_j D_j E_j}{A_j B_j - C_j^2} \right). \quad (2.25)$$

where,

$$A_1 = 1 + f_{rN} C_{yN}^2 + 2\varphi_1(\varphi_1 - 1)f_{nN} C_{xN}^2 + 4\varphi_1 f_{nN} \rho_{xyN} C_{xN} C_{yN},$$

$$B_1 = \left[1 + f_{rN} C_{yN}^2 + 2\varphi_1(\varphi_1 - 1) + g^2 \theta^2 + g(g+1)\theta^2 - 4g\theta\varphi_1 f_{nN} C_{xN}^2 \right. \\ \left. + 4(\varphi_1 - g\theta)f_{nN} \rho_{xyN} C_{xN} C_{yN} \right]$$

$$C_1 = \left[1 + f_{rN} C_{yN}^2 + \left\{ \frac{g(g+1)}{2} \theta^2 + 2\varphi_1(\varphi_1 - 1) - 2\varphi_1 g\theta \right\} f_{nN} C_{xN}^2 \right. \\ \left. + 2(2\varphi_1 - g\theta)f_{nN} \rho_{xyN} C_{xN} C_{yN} \right],$$

$$D_1 = 1 + \frac{\varphi_1(\varphi_1 - 2)}{2} f_{nN} C_{xN}^2 + \varphi_1 f_{nN} \rho_{xyN} C_{xN} C_{yN},$$

$$E_1 = 1 + \left\{ \frac{g(g+1)}{2} \theta^2 - \varphi_1 g\theta + \frac{\varphi_1(\varphi_1 - 2)}{2} \right\} f_{nN} C_{xN}^2 \\ + (\varphi_1 - g\theta)f_{nN} \rho_{xyN} C_{xN} C_{yN},$$

$$A_2 = 1 + f_{rN} C_{yN}^2 + 2\varphi_2(\varphi_2 - 1)f_{rN} C_{xN}^2 \\ + 4\varphi_2 f_{rN} \rho_{xyN} C_{xN} C_{yN},$$

$$B_2 = \left[1 + f_{rN} C_{yN}^2 + \{2\varphi_2(\varphi_2 - 1) + g^2 \theta^2 + g(g+1)\theta^2 - 4g\theta\varphi_2\} f_{rN} C_{xN}^2 \right. \\ \left. + 4(\varphi_2 - g\theta)f_{rN} \rho_{xyN} C_{xN} C_{yN} \right],$$

$$C_2 = \left[1 + f_{rN} C_{yN}^2 + \left\{ \frac{g(g+1)}{2} \theta^2 + 2\varphi_2(\varphi_2 - 1) - 2\varphi_2 g\theta \right\} f_{rN} C_{xN}^2 \right. \\ \left. + 2(2\varphi_2 - g\theta)f_{rN} \rho_{xyN} C_{xN} C_{yN} \right],$$

$$D_2 = 1 + \frac{\varphi_2(\varphi_2 - 2)}{2} f_{rN} C_{xN}^2 + \varphi_2 f_{rN} \rho_{xyN} C_{xN} C_{yN},$$

$$E_2 = 1 + \left\{ \frac{g(g+1)}{2} \theta^2 - \varphi_2 g\theta + \frac{\varphi_2(\varphi_2 - 2)}{2} \right\} f_{rN} C_{xN}^2 \\ + (\varphi_2 - g\theta)f_{rN} \rho_{xyN} C_{xN} C_{yN},$$

$$A_3 = 1 + f_{rN} C_{yN}^2 + 2\varphi_3(\varphi_3 - 1)f_{rN} C_{xN}^2 \\ + 4\varphi_3 f_{rN} \rho_{xyN} C_{xN} C_{yN},$$

$$B_3 = \left[1 + f_{rN} C_{yN}^2 + \{2\varphi_3(\varphi_3 - 1) + g^2 \theta^2 + g(g+1)\theta^2 - 4g\theta\varphi_3\} f_{rN} C_{xN}^2 \right. \\ \left. + 4(\varphi_3 - g\theta)f_{rN} \rho_{xyN} C_{xN} C_{yN} \right]$$

$$C_3 = \left[1 + f_{rN} C_{yN}^2 + \left\{ \frac{g(g+1)}{2} \theta^2 + 2\varphi_3(\varphi_3 - 1) - 2\varphi_3 g\theta \right\} f_{rN} C_{xN}^2 \right. \\ \left. + 2(2\varphi_3 - g\theta)f_{rN} \rho_{xyN} C_{xN} C_{yN} \right]$$

$$D_3 = 1 + \frac{\varphi_3(\varphi_3 - 2)}{2} f_{rN} C_{xN}^2 + \varphi_3 f_{rN} \rho_{xyN} C_{xN} C_{yN},$$

$$E_3 = 1 + \left\{ \frac{g(g+1)}{2} \theta^2 - \varphi_3 g \theta + \frac{\varphi_3(\varphi_3 - 2)}{2} \right\} f_{rn_N} C_{x_N}^2 + (\varphi_3 - g \theta) f_{rn_N} \rho_{xy_N} C_{x_N} C_{y_N},$$

3 Proposed neutrosophic imputation approaches

In the presence of imperfect and ambiguous data, the suggested novel neutrosophic imputation approaches are crucial for elevating the accuracy and dependability of statistical studies. To handle the intricacies of real-life datasets, which typically contain unclear, uncertain, and vague information, the conventional imputation approaches usually fall short. Through simultaneous consideration of the truth, falsity, and indeterminacy of the data, neutrosophic imputation approaches efficiently manage these uncertainties by utilizing the concepts of neutrosophic logic. In domains like environment, marketing, healthcare, social sciences, and economics where data-driven choices carry significant consequences, this thorough treatment of data defects guarantees more reliable and accurate estimation. We propose some novel neutrosophic imputation approaches for estimating population mean under SRS as

The resultant proposed novel neutrosophic estimators under situations I-III are

$$T_{k_{1N}} = \left[\gamma_1 \bar{y}_{r_N} \left\{ \frac{1}{2} \left(\frac{\bar{X}_N}{\bar{x}_{n_N}} + \frac{\bar{x}_{n_N}}{\bar{X}_N} \right) \right\} + \delta_1 (\bar{X}_N - \bar{x}_{n_N}) \right] \left\{ 1 + \log \left(\frac{\bar{x}_{n_N}}{\bar{X}_N} \right) \right\}^{\varphi_1}, \quad (3.1)$$

$$T_{k_{2N}} = \left[\gamma_2 \bar{y}_{r_N} \left\{ \frac{1}{2} \left(\frac{\bar{X}_N}{\bar{x}_{r_N}} + \frac{\bar{x}_{r_N}}{\bar{X}_N} \right) \right\} + \delta_2 (\bar{X}_N - \bar{x}_{r_N}) \right] \left\{ 1 + \log \left(\frac{\bar{x}_{r_N}}{\bar{X}_N} \right) \right\}^{\varphi_2}, \quad (3.2)$$

$$\text{and } T_{k_{3N}} = \left[\gamma_3 \bar{y}_{r_N} \left\{ \frac{1}{2} \left(\frac{\bar{x}_{n_N}}{\bar{x}_{r_N}} + \frac{\bar{x}_{r_N}}{\bar{x}_{n_N}} \right) \right\} + \delta_3 (\bar{x}_{n_N} - \bar{x}_{r_N}) \right] \left\{ 1 + \log \left(\frac{\bar{x}_{r_N}}{\bar{x}_{n_N}} \right) \right\}^{\varphi_3}, \quad (3.3)$$

where γ_j , $j = 1, 2, 3$, δ_j , and φ_j are optimizing scalars.

Theorem 3.1 *The bias, MSE, and minimum MSE of the resultant novel neutrosophic estimators $T_{k_{jN}}$, $j = 1, 2, 3$ of the proposed novel neutrosophic imputation approaches are given by*

$$\text{Bias}(T_{k_{jN}}) = \bar{Y}_N (\gamma_j J_j + \delta_j K_j - 1), \quad (3.4)$$

Situation I

$$y_{i_{k_{1N}}} = \begin{cases} y_{i_N} & \text{for } i \in A_N \\ \frac{1}{n_N - r_N} \left(\left[n_N \left[\gamma_1 \bar{y}_{r_N} \left\{ \frac{1}{2} \left(\frac{\bar{X}_N}{\bar{x}_{n_N}} + \frac{\bar{x}_{n_N}}{\bar{X}_N} \right) \right\} + \delta_1 (\bar{X}_N - \bar{x}_{n_N}) \right] \times \left\{ 1 + \log \left(\frac{\bar{x}_{n_N}}{\bar{X}_N} \right) \right\}^{\varphi_1} \right] - r_N \bar{y}_{r_N} \right) & \text{for } i \in \bar{A}_N \end{cases}$$

Situation II

$$y_{i_{k_{2N}}} = \begin{cases} y_{i_N} & \text{for } i \in A_N \\ \frac{1}{n_N - r_N} \left(\left[n_N \left[\gamma_2 \bar{y}_{r_N} \left\{ \frac{1}{2} \left(\frac{\bar{X}_N}{\bar{x}_{r_N}} + \frac{\bar{x}_{r_N}}{\bar{X}_N} \right) \right\} + \delta_2 (\bar{X}_N - \bar{x}_{r_N}) \right] \times \left\{ 1 + \log \left(\frac{\bar{x}_{r_N}}{\bar{X}_N} \right) \right\}^{\varphi_2} \right] - r_N \bar{y}_{r_N} \right) & \text{for } i \in \bar{A}_N \end{cases}$$

Situation III

$$y_{i_{k_{3N}}} = \begin{cases} y_{i_N} & \text{for } i \in A_N \\ \frac{1}{n_N - r_N} \left(\left[n_N \left[\gamma_3 \bar{y}_{r_N} \left\{ \frac{1}{2} \left(\frac{\bar{x}_{n_N}}{\bar{x}_{r_N}} + \frac{\bar{x}_{r_N}}{\bar{x}_{n_N}} \right) \right\} + \delta_3 (\bar{x}_{n_N} - \bar{x}_{r_N}) \right] \times \left\{ 1 + \log \left(\frac{\bar{x}_{r_N}}{\bar{x}_{n_N}} \right) \right\}^{\varphi_3} \right] - r_N \bar{y}_{r_N} \right) & \text{for } i \in \bar{A}_N \end{cases}$$

$$MSE(T_{k_{jN}}) = \bar{Y}_N^2(1 + \gamma_j^2 G_j + \delta_j^2 H_j + 2\gamma_j \delta_j I_j - 2\gamma_j J_j - 2\delta_j K_j), \quad (3.5)$$

$$\text{and } \min. MSE(T_{k_{jN}}) = \bar{Y}_N^2 \left\{ 1 - \frac{(G_j K_j^2 + H_j J_j^2 - 2I_j J_j K_j)}{(G_j H_j - I_j^2)} \right\}. \quad (3.6)$$

Proof Consider the first proposed resultant novel neutrosophic estimator under situation I as

$$T_{k_{1N}} = \left[\gamma_1 \bar{y}_{rN} \left\{ \frac{1}{2} \left(\frac{\bar{X}_N}{\bar{x}_{nN}} + \frac{\bar{x}_{nN}}{\bar{X}_N} \right) \right\} + \delta_1 (\bar{X}_N - \bar{x}_{nN}) \right] \left\{ 1 + \log \left(\frac{\bar{x}_{nN}}{\bar{X}_N} \right) \right\}^{\varphi_1}.$$

Utilizing the notations defined in (1.1), we can express the above estimator as

$$T_{k_{1N}} = \left[\begin{array}{l} \gamma_1 \bar{Y}_N (1 + e_{0N}) \left\{ \frac{1}{2} ((1 + e_{2N})^{-1} + (1 + e_{2N})) \right\} \\ - \delta_1 \bar{X}_N e_{2N} \times \left\{ 1 + e_{2N} - \frac{e_{2N}^2}{2} \right\}^{\varphi_1} \end{array} \right], \quad (3.7)$$

$$T_{k_{1N}} - \bar{Y}_N = \bar{Y}_N \left[\begin{array}{l} \gamma_1 \left\{ 1 + e_{0N} + \frac{e_{2N}^2}{2} + \varphi_1 e_{2N} \right. \\ \left. + \varphi_1 e_{0N} e_{2N} \right. \\ \left. + \left(\frac{\varphi_1^2}{2} - \varphi_1 \right) e_{2N}^2 \right\} \\ \left. + \frac{\delta_1}{R} \left\{ -e_{2N} - \varphi_1 e_{2N}^2 \right\} - 1 \right]. \quad (3.8)$$

Taking expectation both side to (3.8), we get the bias of the resultant neutrosophic estimator $T_{k_{1N}}$ as

$$Bias(T_{k_{1N}}) = \bar{Y}_N \left[\begin{array}{l} \gamma_1 \left\{ 1 + \left(\frac{\varphi_1^2}{2} - \varphi_1 + \frac{1}{2} \right) f_{nN} C_{xN}^2 \right. \\ \left. + \varphi_1 f_{nN} \rho_{xyN} C_{xN} C_{yN} \right\} \\ \left. + \delta_1 \left(\frac{-\varphi_1}{R} f_{nN} C_{xN}^2 \right) - 1 \right]. \quad (3.9)$$

The above expression can further be expressed as

$$Bias(T_{k_{1N}}) = \bar{Y}_N (\gamma_1 J_1 + \delta_1 K_1 - 1), \quad (3.10)$$

where

$$J_1 = 1 + \left(\frac{\varphi_1^2}{2} - \varphi_1 + \frac{1}{2} \right) f_{nN} C_{xN}^2 + \varphi_1 f_{nN} \rho_{xyN} C_{xN} C_{yN};$$

$$K_1 = \frac{-\varphi_1}{R} f_{nN} C_{xN}^2; \text{ and } R = \frac{\bar{Y}_N}{\bar{X}_N}.$$

Again, squaring and taking expectation both side to (3.8), we can get the proposed estimator's MSE up to the first order approximation as

$$MSE(T_{k_{1N}}) = \bar{Y}_N^2 E \left[\begin{array}{l} 1 + \gamma_1^2 \{ 1 + e_{0N}^2 + (2\varphi_1^2 - 2\varphi_1 + 1) e_{2N}^2 \\ + 4\varphi_1 e_{0N} e_{2N} \} + \frac{\delta_1^2}{R^2} e_{2N}^2 \\ + 2\gamma_1 \delta_1 \left(-2\varphi_1 e_{2N}^2 - e_{0N} e_{2N} \right) \\ - 2\gamma_1 \left\{ 1 + \left(\frac{\varphi_1^2}{2} - \varphi_1 + \frac{1}{2} \right) e_{2N}^2 \right. \\ \left. + \varphi_1 e_{0N} e_{2N} \right\} \\ \left. - 2\delta_1 \left(\frac{-\varphi_1}{R} e_{2N}^2 \right) \right] \quad (3.11)$$

$$MSE(T_{k_{1N}}) = \bar{Y}_N^2 \left[\begin{array}{l} 1 \\ + \gamma_1^2 \left\{ \begin{array}{l} 1 + f_{rN} C_{yN}^2 \\ + (2\varphi_1^2 - 2\varphi_1 + 1) f_{nN} C_{xN}^2 \\ + 4\varphi_1 f_{nN} \rho_{xyN} C_{xN} C_{yN} \end{array} \right\} \\ + \frac{\delta_1^2}{R^2} f_{nN} C_{xN}^2 + 2\gamma_1 \delta_1 (-2\varphi_1 f_{nN} C_{xN}^2 \\ - f_{nN} \rho_{xyN} C_{xN} C_{yN}) \\ - 2\gamma_1 \left\{ 1 + \left(\frac{\varphi_1^2}{2} - \varphi_1 + \frac{1}{2} \right) f_{nN} C_{xN}^2 \right. \\ \left. + \varphi_1 f_{nN} \rho_{xyN} C_{xN} C_{yN} \right\} - 2\delta_1 \left(\frac{-\varphi_1}{R} \right) f_{nN} C_{xN}^2 \end{array} \right]. \quad (3.12)$$

The above expression can further be expressed as

$$MSE(T_{k_{1N}}) = \bar{Y}_N^2 (1 + \gamma_1^2 G_1 + \delta_1^2 H_1 + 2\gamma_1 \delta_1 I_1 - 2\gamma_1 J_1 - 2\delta_1 K_1), \quad (3.13)$$

where,

$$G_1 = 1 + f_{rN} C_{yN}^2 + (2\varphi_1^2 - 2\varphi_1 + 1) f_{nN} C_{xN}^2 + 4\varphi_1 f_{nN} \rho_{xyN} C_{xN} C_{yN},$$

$$H_1 = \frac{1}{R^2} f_{nN} C_{xN}^2,$$

$$I_1 = \frac{1}{R} (-2\varphi_1 f_{nN} C_{xN}^2 - f_{nN} \rho_{xyN} C_{xN} C_{yN}),$$

$$J_1 = 1 + \left(\frac{\varphi_1^2}{2} - \varphi_1 + \frac{1}{2} \right) f_{nN} C_{xN}^2 + \varphi_1 f_{nN} \rho_{xyN} C_{xN} C_{yN},$$

$$K_1 = \frac{-\varphi_1}{R} f_{nN} C_{xN}^2.$$

The optimum values of γ_1 , δ_1 , and φ_1 are given by

$$\gamma_{1(opt)} = \frac{H_1 J_1 - I_1 K_1}{G_1 H_1 - I_1^2}, \quad \delta_{1(opt)} = \frac{G_1 K_1 - I_1 J_1}{G_1 H_1 - I_1^2}, \quad (3.14)$$

$$\text{and } \varphi_{1(opt)} = -\rho_{xyN} \frac{C_{yN}}{C_{xN}}.$$

The minimum MSE can be obtained by putting optimum values of γ_1 and δ_1 in (3.13) as

$$\min. MSE(T_{k_{1N}}) = \bar{Y}_N^2 \left(1 - \frac{G_1 K_1^2 + H_1 J_1^2 - 2I_1 J_1 K_1}{G_1 H_1 - I_1^2} \right). \quad (3.15)$$

Similarly, we can find out the bias and MSE of the resultant novel neutrosophic estimators under situations II and III. In general, we can write the bias and MSE of the resultant estimators under situations I-III as

$$Bias(T_{k_{jN}}) = \bar{Y}_N (\gamma_j J_j + \delta_j K_j - 1) \quad (3.16)$$

$$\text{and } MSE(T_{k_{jN}}) = \bar{Y}_N^2 (1 + \gamma_j^2 G_j + \delta_j^2 H_j + 2\gamma_j \delta_j I_j - 2\gamma_j J_j - 2\delta_j K_j). \quad (3.17)$$

The optimum values of the scalars γ_j , $i = 1, 2, 3$, δ_j , and φ_j are given by

$$\gamma_{j(opt)} = \frac{H_j J_j - I_j K_j}{G_j H_j - I_j^2}, \quad \delta_{j(opt)} = \frac{G_j K_j - J_j I_j}{G_j H_j - I_j^2}, \quad (3.18)$$

$$\text{and } \varphi_{j(opt)} = -\rho_{xyN} \frac{C_{yN}}{C_{xN}}.$$

In general, the minimum MSE can be obtained by putting optimum values of γ_j and δ_{jN} in (3.17) as

$$\min.MSE(T_{k_{jN}}) = \bar{Y}_N^2 \left(1 - \frac{G_j K_j^2 + H_j J_j^2 - 2I_j J_j K_j}{G_j H_j - I_j^2} \right), \quad (3.19)$$

where,

$$\begin{aligned} G_2 &= 1 + f_{rN} C_{yN}^2 + (2\varphi_1^2 - 2\varphi_1 + 1) f_{rN} C_{xN}^2 \\ &\quad + 4\varphi_2 f_{rN} \rho_{xyN} C_{xN} C_{yN}; \quad H_2 = \frac{1}{R^2} f_{rN} C_{xN}^2; \\ I_2 &= \frac{1}{R} (-2\varphi_1 f_{rN} C_{xN}^2 - f_{rN} \rho_{xyN} C_{xN} C_{yN}); \\ J_2 &= 1 + \left(\frac{\varphi_1^2}{2} - \varphi_1 + \frac{1}{2} \right) f_{rN} C_{xN}^2 + \varphi_1 f_{rN} \rho_{xyN} C_{xN} C_{yN}; \\ K_2 &= \frac{-\varphi_1}{R} f_{rN} C_{xN}^2; \quad G_3 = 1 + f_{rN} C_{yN}^2 \\ &\quad + (2\varphi_1^2 - 2\varphi_1 + 1) f_{rnN} C_{xN}^2 + 4\varphi_3 f_{rnN} \rho_{xyN} C_{xN} C_{yN}; \\ H_3 &= \frac{1}{R^2} f_{rnN} C_{xN}^2; \quad I_3 = \frac{1}{R} (-2\varphi_3 f_{rnN} C_{xN}^2 - f_{rnN} \rho_{xyN} C_{xN} C_{yN}); \\ J_3 &= 1 + \left(\frac{\varphi_1^2}{2} - \varphi_1 + \frac{1}{2} \right) f_{rnN} C_{xN}^2 + \varphi_3 f_{rnN} \rho_{xyN} C_{xN} C_{yN}; \\ \text{and } K_3 &= \frac{-\varphi_3}{R} f_{rnN} C_{xN}^2. \end{aligned}$$

□

4 Efficiency comparisons

The efficiency comparison of the proposed novel neutrosophic imputation approaches is crucial for determining their practical applicability and effectiveness in real-life applications. Deducing the efficiency involves assessing computational complexity, processing time, and resource utilization, which are critical components for extensive data analysis. By comparing the efficiency of different approaches, one can identify the most optimized approaches, balancing accuracy with computational demands. Moreover, understanding the efficiency of these approaches help in improving their scalability and adaptability, making them more accessible for numerous applications. Taking these advantages into consideration, we carry out an efficiency comparison by comparing the MSE of the proposed and adapted resultant neutrosophic estimators under situations I-III.

Lemma 4.1 *The proposed novel neutrosophic imputation approaches $y_{i_{k_{jN}}}$, $j = 1, 2, 3$ outperform the neutrosophic mean imputation approach $y_{i_{mN}}$ under situations I-III, if*

$$\begin{aligned} \min.MSE(T_{k_{jN}}) &< V(T_{mN}) \\ \Rightarrow \frac{G_j K_j^2 + H_j J_j^2 - 2I_j J_j K_j}{G_j H_j - I_j^2} &> 1 - f_{rN} C_{yN}^2. \end{aligned} \quad (4.1)$$

Lemma 4.2 (i). *The proposed novel neutrosophic imputation approach $y_{i_{k_{1N}}}$ outperforms the neutrosophic ratio imputation approach $y_{i_{r_{1N}}}$ under situation I, if*

$$\begin{aligned} \min.MSE(T_{k_{1N}}) &< MSE(T_{r_{1N}}) \\ \Rightarrow \frac{G_1 K_1^2 + H_1 J_1^2 - 2I_1 J_1 K_1}{G_1 H_1 - I_1^2} &> 1 - f_{rN} C_{yN}^2 - f_{nN} (C_{xN}^2 - 2\rho_{xyN} C_{xN} C_{yN}). \end{aligned} \quad (4.2)$$

(ii). *The proposed novel neutrosophic imputation approach $y_{i_{k_{2N}}}$ outperforms the neutrosophic ratio imputation approach $y_{i_{r_{2N}}}$ under situation II, if*

$$\begin{aligned} MSE(T_{k_{2N}}) &< MSE(T_{r_{2N}}) \\ \Rightarrow \frac{G_2 K_2^2 + H_2 J_2^2 - 2I_2 J_2 K_2}{G_2 H_2 - I_2^2} &> 1 - f_{rN} (C_{yN}^2 + C_{xN}^2 - 2\rho_{xyN} C_{xN} C_{yN}). \end{aligned} \quad (4.3)$$

(iii). *The proposed novel neutrosophic imputation approach $y_{i_{k_{3N}}}$ outperforms the neutrosophic ratio imputation approach $y_{i_{r_{3N}}}$ under situation III, if*

$$\begin{aligned} \min.MSE(T_{k_{3N}}) &< MSE(T_{r_{3N}}) \\ \Rightarrow \frac{G_3 K_3^2 + H_3 J_3^2 - 2I_3 J_3 K_3}{G_3 H_3 - I_3^2} &> 1 - f_{nN} C_{yN}^2 - f_{rnN} (C_{yN}^2 + C_{xN}^2 \\ &\quad - 2\rho_{xyN} C_{xN} C_{yN}). \end{aligned} \quad (4.4)$$

Lemma 4.3 (i). *The proposed novel neutrosophic imputation approach $y_{i_{k_{1N}}}$ outperforms the neutrosophic regression imputation approach $y_{i_{lr_{1N}}}$ and neutrosophic logarithmic imputation approach $y_{i_{l_{1N}}}$ under situation I, if*

$$\begin{aligned} \min.MSE(T_{k_{1N}}) &< \min.MSE(t_{1N}^*), \text{ where } t_{1N}^* = T_{lr_{1N}} \text{ and } T_{l_{1N}} \\ \Rightarrow \frac{G_1 K_1^2 + H_1 J_1^2 - 2I_1 J_1 K_1}{G_1 H_1 - I_1^2} &> 1 - f_{rN} C_{yN}^2 + f_{nN} C_{yN}^2 \rho_{xyN}^2. \end{aligned} \quad (4.5)$$

(ii). *The proposed novel neutrosophic imputation approach $y_{i_{k_{2N}}}$ outperforms the neutrosophic regression imputation approach $y_{i_{lr_{2N}}}$ and neutrosophic logarithmic imputation approach under situation II, if*

$$\begin{aligned} \min.MSE(T_{k_{2N}}) &< \min.MSE(t_{2N}^*), \text{ where } t_{2N}^* = T_{lr_{2N}} \text{ and } T_{l_{2N}} \\ \Rightarrow \frac{G_2 K_2^2 + H_2 J_2^2 - 2I_2 J_2 K_2}{G_2 H_2 - I_2^2} &> 1 - f_{rN} C_{yN}^2 (1 - \rho_{xyN}^2). \end{aligned} \quad (4.6)$$

(iii). *The proposed novel neutrosophic imputation approach $y_{i_{k_{3N}}}$ outperforms the neutrosophic regression imputation approach $y_{i_{lr_{3N}}}$ and neutrosophic logarithmic imputation approach $y_{i_{l_{3N}}}$ under situation III, if*

$$\begin{aligned}
\min.MSE(T_{k_{3N}}) &< MSE(t_{3N}^*), \text{ where } t_{3N}^* = T_{lr_{3N}} \text{ and } T_{l_{3N}} \\
&\Rightarrow \frac{G_3 K_3^2 + H_3 J_3^2 - 2I_3 J_3 K_3}{G_3 H_3 - I_3^2} \\
&> 1 - f_{n_N} C_{y_N}^2 - f_{rn_N} C_{y_N}^2 (1 - \rho_{xy_N}^2).
\end{aligned} \quad (4.7)$$

Lemma 4.4 (i). The proposed novel neutrosophic imputation approach $y_{i_{k_{1N}}}$ outperforms Kumar and Kumar [34] neutrosophic imputation approach $y_{i_{e_{1N}}}$ under situation I, if

$$\begin{aligned}
\min.MSE(T_{k_{1N}}) &< \min.MSE(T_{e_{1N}}) \\
&\Rightarrow \frac{G_1 K_1^2 + H_1 J_1^2 - 2I_1 J_1 K_1}{G_1 H_1 - I_1^2} \\
&> \frac{A_1 E_1^2 + B_1 D_1^2 - 2C_1 D_1 E_1}{A_1 B_1 - C_1^2}.
\end{aligned} \quad (4.8)$$

(ii). The proposed novel neutrosophic imputation approach $y_{i_{k_{2N}}}$ outperforms Kumar and Kumar [34] neutrosophic imputation approach $y_{i_{e_{2N}}}$ under situation II, if

$$\begin{aligned}
MSE(T_{k_{2N}}) &< \min.MSE(T_{e_{2N}}) \\
&\Rightarrow \frac{G_2 K_2^2 + H_2 J_2^2 - 2I_2 J_2 K_2}{G_2 H_2 - I_2^2} \\
&> \frac{A_2 E_2^2 + B_2 D_2^2 - 2C_2 D_2 E_2}{A_2 B_2 - C_2^2}.
\end{aligned} \quad (4.9)$$

(iii). The proposed novel neutrosophic imputation approach $y_{i_{k_{3N}}}$ outperforms Kumar and Kumar [34] neutrosophic imputation approach $y_{i_{e_{3N}}}$ under situation III, if

$$\begin{aligned}
\min.MSE(T_{k_{3N}}) &< \min.MSE(T_{e_{3N}}) \\
&\Rightarrow \frac{G_3 K_3^2 + H_3 J_3^2 - 2I_3 J_3 K_3}{G_3 H_3 - I_3^2} \\
&> \frac{A_3 E_3^2 + B_3 D_3^2 - 2C_3 D_3 E_3}{A_3 B_3 - C_3^2}.
\end{aligned} \quad (4.10)$$

5 Simulation study

The simulation study has crucial ramifications for confirming the robustness and practicality of the suggested approaches. The simulation study offers empirical support for the idea that neutrosophic imputation is a more effective approach for managing the complexity of uncertain and vague data. Moreover, the simulation study helps practitioners apply the neutrosophic technique as effectively as possible by providing insights into the circumstances in which it works best. In domains where data quality is critical, this empirical validation is essential for fostering method adoption and boosting trust in the approach. In light of this, we conduct a simulation study using Singh and Horn [49] as an outline and create a few symmetric and asymmetric populations utilizing the neutrosophic normal (NN) and neutrosophic gamma (NG) distributions. Below are the models that were used to produce the neutrosophic data.

Table 1 $MSE_N \in [MSE_L, MSE_U]$ of resultant neutrosophic estimators using neutrosophic normal population

Estimators	ρ_{xy_N}				
	[0.05, 0.1]	[0.25, 0.3]	[0.45, 0.5]	[0.65, 0.7]	[0.85, 0.9]
T_{mN}	[1.26, 1.48]	[1.26, 1.47]	[1.25, 1.47]	[1.25, 1.46]	[1.25, 1.47]
Situation I					
$T_{r_{1N}}$	[2.38, 2.69]	[2.03, 2.27]	[1.62, 1.78]	[1.16, 1.23]	[0.66, 0.68]
$T_{lr_{1N}}$	[1.26, 1.47]	[1.19, 1.37]	[1.04, 1.17]	[0.81, 0.88]	[0.51, 0.55]
$T_{l_{1N}}$	[1.26, 1.47]	[1.19, 1.37]	[1.04, 1.17]	[0.81, 0.88]	[0.51, 0.55]
$T_{e_{1N}}$	[1.26, 1.46]	[1.19, 1.36]	[1.04, 1.16]	[0.81, 0.87]	[0.50, 0.52]
$T_{k_{1N}}$	[1.25, 1.46]	[1.18, 1.35]	[1.03, 1.15]	[0.80, 0.86]	[0.50, 0.51]
Situation II					
$T_{r_{2N}}$	[2.63, 2.96]	[2.20, 2.44]	[1.70, 1.84]	[1.14, 1.17]	[0.53, 0.56]
$T_{lr_{2N}}$	[1.26, 1.47]	[1.18, 1.34]	[1.00, 1.10]	[0.72, 0.74]	[0.34, 0.37]
$T_{l_{2N}}$	[1.26, 1.47]	[1.18, 1.34]	[1.00, 1.10]	[0.72, 0.74]	[0.34, 0.37]
$T_{e_{2N}}$	[1.26, 1.46]	[1.17, 1.33]	[0.99, 1.09]	[0.71, 0.74]	[0.34, 0.36]
$T_{k_{2N}}$	[1.25, 1.45]	[1.16, 1.32]	[0.98, 1.08]	[0.70, 0.73]	[0.33, 0.35]
Situation III					
$T_{r_{3N}}$	[1.51, 1.75]	[1.43, 1.65]	[1.33, 1.54]	[1.23, 1.41]	[1.12, 1.28]
$T_{lr_{3N}}$	[1.26, 1.48]	[1.24, 1.45]	[1.20, 1.40]	[1.15, 1.33]	[1.08, 1.25]
$T_{l_{3N}}$	[1.26, 1.48]	[1.24, 1.45]	[1.20, 1.40]	[1.15, 1.33]	[1.08, 1.25]
$T_{e_{3N}}$	[1.03, 1.21]	[1.01, 1.18]	[0.97, 1.13]	[0.92, 1.06]	[0.85, 0.98]
$T_{k_{3N}}$	[1.03, 1.20]	[1.01, 1.18]	[0.97, 1.13]	[0.92, 1.06]	[0.85, 0.98]

Table 2 $MSE_N \in [MSE_L, MSE_U]$ of resultant neutrosophic estimators using neutrosophic gamma population

Estimators	ρ_{xyN}				
	[0.05, 0.1]	[0.25, 0.3]	[0.45, 0.5]	[0.65, 0.7]	[0.85, 0.9]
T_{mN}	[2.54, 2.63]	[2.53, 2.68]	[2.52, 2.72]	[2.52, 2.73]	[2.52, 2.71]
Situation I					
$T_{r_{1N}}$	[4.30, 4.32]	[3.60, 3.63]	[2.81, 2.83]	[1.90, 1.93]	[1.14, 1.19]
$T_{lr_{1N}}$	[2.53, 2.61]	[2.40, 2.48]	[2.10, 2.16]	[1.64, 1.64]	[1.03, 1.09]
$T_{l_{1N}}$	[2.53, 2.61]	[2.40, 2.48]	[2.10, 2.16]	[1.64, 1.64]	[1.03, 1.09]
$T_{e_{1N}}$	[2.51, 2.59]	[2.37, 2.46]	[2.08, 2.14]	[1.61, 1.62]	[1.01, 1.05]
$T_{k_{1N}}$	[2.49, 2.57]	[2.35, 2.43]	[2.06, 2.11]	[1.59, 1.60]	[1.00, 1.01]
Situation II					
$T_{r_{2N}}$	[4.67, 4.70]	[3.80, 3.84]	[2.81, 2.85]	[1.71, 1.75]	[0.84, 0.87]
$T_{lr_{2N}}$	[2.53, 2.61]	[2.37, 2.44]	[2.01, 2.04]	[1.36, 1.39]	[0.70, 0.72]
$T_{l_{2N}}$	[2.53, 2.61]	[2.37, 2.44]	[2.01, 2.04]	[1.36, 1.39]	[0.70, 0.72]
$T_{e_{2N}}$	[2.51, 2.58]	[2.34, 2.41]	[1.98, 2.01]	[1.33, 1.37]	[0.68, 0.70]
$T_{k_{2N}}$	[2.48, 2.56]	[2.32, 2.38]	[1.96, 1.98]	[1.31, 1.35]	[0.66, 0.68]
Situation III					
$T_{r_{3N}}$	[2.94, 3.01]	[2.78, 2.89]	[2.60, 2.74]	[2.41, 2.56]	[2.21, 2.32]
$T_{lr_{3N}}$	[2.54, 2.63]	[2.50, 2.64]	[2.43, 2.59]	[2.32, 2.49]	[2.19, 2.31]
$T_{l_{3N}}$	[2.54, 2.63]	[2.50, 2.64]	[2.43, 2.59]	[2.32, 2.49]	[2.19, 2.31]
$T_{e_{3N}}$	[2.06, 2.13]	[2.03, 2.13]	[1.96, 2.09]	[1.85, 1.98]	[1.72, 1.80]
$T_{k_{3N}}$	[2.06, 2.13]	[2.02, 2.13]	[1.95, 2.08]	[1.85, 1.98]	[1.71, 1.80]

Table 3 $PRE_N \in [PRE_L, PRE_U]$ of resultant neutrosophic estimators using neutrosophic normal population

Estimators	ρ_{xyN}				
	[0.05, 0.1]	[0.25, 0.3]	[0.45, 0.5]	[0.65, 0.7]	[0.85, 0.9]
T_{mN}	[100.00, 100.00]	[100.00, 100.00]	[100.00, 100.00]	[100.00, 100.00]	[100.00, 100.00]
Situation I					
$T_{r_{1N}}$	[53.15, 55.21]	[62.11, 65.12]	[77.42, 82.68]	[107.46, 119.36]	[188.05, 216.18]
$T_{lr_{1N}}$	[100.20, 100.82]	[105.38, 107.9]	[119.85, 125.71]	[152.83, 166.91]	[244.58, 267.28]
$T_{l_{1N}}$	[100.20, 100.82]	[105.38, 107.9]	[119.85, 125.71]	[152.83, 166.91]	[244.58, 267.28]
$T_{e_{1N}}$	[100.62, 101.28]	[105.90, 108.51]	[120.50, 126.43]	[153.73, 167.97]	[246.27, 282.24]
$T_{k_{1N}}$	[101.10, 101.80]	[106.49, 109.16]	[121.30, 127.34]	[155.00, 169.52]	[249.18, 288.60]
Situation II					
$T_{r_{2N}}$	[48.14, 50.22]	[57.29, 60.44]	[73.72, 79.62]	[109.27, 124.73]	[233.80, 262.51]
$T_{lr_{2N}}$	[100.25, 101.01]	[106.66, 109.89]	[125.39, 133.33]	[173.16, 196.07]	[360.36, 397.30]
$T_{l_{2N}}$	[100.25, 101.01]	[106.66, 109.89]	[125.39, 133.33]	[173.16, 196.07]	[360.36, 397.30]
$T_{e_{2N}}$	[100.67, 101.48]	[107.22, 110.51]	[126.15, 134.20]	[174.39, 197.60]	[364.00, 408.34]
$T_{k_{2N}}$	[101.29, 102.16]	[108.01, 111.40]	[127.30, 135.55]	[176.52, 200.40]	[372.24, 420.00]
Situation III					
$T_{r_{3N}}$	[83.62, 84.72]	[88.06, 89.36]	[93.91, 95.55]	[101.56, 103.74]	[111.61, 114.61]
$T_{lr_{3N}}$	[100.04, 100.18]	[101.14, 101.66]	[103.82, 104.76]	[108.32, 109.78]	[115.12, 117.27]
$T_{l_{3N}}$	[100.04, 100.18]	[101.14, 101.66]	[103.82, 104.76]	[108.32, 109.78]	[115.12, 117.27]
$T_{e_{3N}}$	[122.68, 122.90]	[124.32, 125.11]	[128.36, 129.81]	[135.27, 137.56]	[146.01, 149.49]
$T_{k_{3N}}$	[122.79, 123.01]	[124.44, 125.23]	[128.49, 129.94]	[135.41, 137.71]	[146.16, 149.65]

Table 4 $PRE_N \in [PRE_L, PRE_U]$ of resultant neutrosophic estimators using neutrosophic gamma population

Estimators	ρ_{xy_N}				
	[0.05, 0.1]	[0.25, 0.3]	[0.45, 0.5]	[0.65, 0.7]	[0.85, 0.9]
T_{mN}	[100.00, 100.00]	[100.00, 100.00]	[100.00, 100.00]	[100.00, 100.00]	[100.00, 100.00]
Situation I					
$T_{r_{1N}}$	[59.07, 60.93]	[70.27, 73.84]	[89.69, 96.09]	[132.68, 141.53]	[219.51, 227.74]
$T_{lr_{1N}}$	[100.20, 100.82]	[105.38, 107.94]	[119.85, 125.71]	[152.83, 166.91]	[244.58, 248.62]
$T_{l_{1N}}$	[100.20, 100.82]	[105.38, 107.94]	[119.85, 125.71]	[152.83, 166.91]	[244.58, 248.62]
$T_{e_{1N}}$	[101.10, 101.76]	[106.48, 109.09]	[121.20, 127.15]	[156.52, 168.90]	[247.69, 258.09]
$T_{k_{1N}}$	[101.92, 102.58]	[107.47, 110.11]	[122.51, 128.52]	[158.49, 171.15]	[252.20, 268.32]
Situation II					
$T_{r_{2N}}$	[54.38, 56.06]	[66.58, 69.78]	[89.67, 95.26]	[147.36, 155.93]	[298.89, 311.49]
$T_{lr_{2N}}$	[100.25, 101.01]	[106.66, 109.89]	[125.39, 133.33]	[185.29, 196.07]	[360.36, 376.38]
$T_{l_{2N}}$	[100.25, 101.01]	[106.66, 109.89]	[125.39, 133.33]	[185.29, 196.07]	[360.36, 376.38]
$T_{e_{2N}}$	[101.16, 101.97]	[107.84, 111.14]	[126.95, 135.03]	[189.47, 198.83]	[366.72, 387.15]
$T_{k_{2N}}$	[102.21, 103.04]	[109.16, 112.51]	[128.81, 137.04]	[192.36, 202.83]	[379.25, 398.56]
Situation III					
$T_{r_{3N}}$	[86.27, 87.53]	[90.96, 92.70]	[96.91, 99.10]	[104.39, 106.97]	[113.76, 116.63]
$T_{lr_{3N}}$	[100.04, 100.18]	[101.14, 101.66]	[103.82, 104.76]	[108.32, 109.78]	[115.12, 117.27]
$T_{l_{3N}}$	[100.04, 100.18]	[101.14, 101.66]	[103.82, 104.76]	[108.32, 109.78]	[115.12, 117.27]
$T_{e_{3N}}$	[123.13, 123.33]	[124.78, 125.56]	[128.82, 130.27]	[135.74, 138.03]	[146.49, 149.97]
$T_{k_{3N}}$	[123.32, 123.51]	[124.98, 125.75]	[129.04, 130.48]	[135.98, 138.27]	[146.76, 150.23]

$$y_N = 13.9 + \sqrt{(1 - \rho_{xy_N}^2)} y_N^* + \rho_{xy_N} \left(\frac{S_{y_N}}{S_{x_N}} \right) x_N^* \quad (5.1)$$

and $x_N = 13.5 + x_N^*$.

The neutrosophic variables x_N^* and y_N^* are independent and have distinct neutrosophic distributions. The following describes the simulation process.

- (i) Making use of (5.1), we produce the subsequent populations: **(a)**. Following Saleem and Aslam [41], create a symmetric population from a neutrosophic normal (NN) distribution of size $N = 1200$ such that $x_N^* \sim NN([6, 7], [14, 15])$ and $y_N^* \sim NN([4, 5], [12, 13])$ with varying neutrosophic correlation coefficient $\rho_{xy_N} \in ([0.05, 0.1], [0.25, 0.3], [0.45, 0.5], [0.65, 0.7], [0.85, 0.9])$. **(b)**. Following Saleem and Aslam [41], create an asymmetric population from a neutrosophic gamma (NG) distribution of size $N = 1200$ such that $x_N^* \sim NG([0.06, 0.07], [0.014, 0.015])$ and $y_N^* \sim NG([0.04, 0.05], [0.012, 0.013])$ with varying neutrosophic correlation coefficients $\rho_{xy_N} \in ([0.05, 0.1], [0.25, 0.3], [0.45, 0.5], [0.65, 0.7], [0.85, 0.9])$.
- (ii) Choose a sample of size $n_N = [120, 120]$ from the neutrosophic normal and neutrosophic gamma populations produced in the above step, and assume that the selected sample n_N provides a response as $r_N = [100, 100]$ units.

- (iii) Take the chosen sample n_N and drop down $n_N - r_N$ units.

- (iv) Taking into account 15,000 replications and utilizing the corresponding formulas provided in (5.2), calculate the $MSE_N \in [MSE_L, MSE_U]$ and $PRE_N \in [PRE_L, PRE_U]$ of the resultant neutrosophic estimators for the samples and responding units chosen in the preceding step.

$$MSE(T^*) = \frac{1}{15,000} \sum_{i=1}^{15,000} (T_i^* - \bar{Y}_N)^2 \quad (5.2)$$

and $PRE = \frac{V(T_{mN})}{MSE(T^*)} \times 100,$

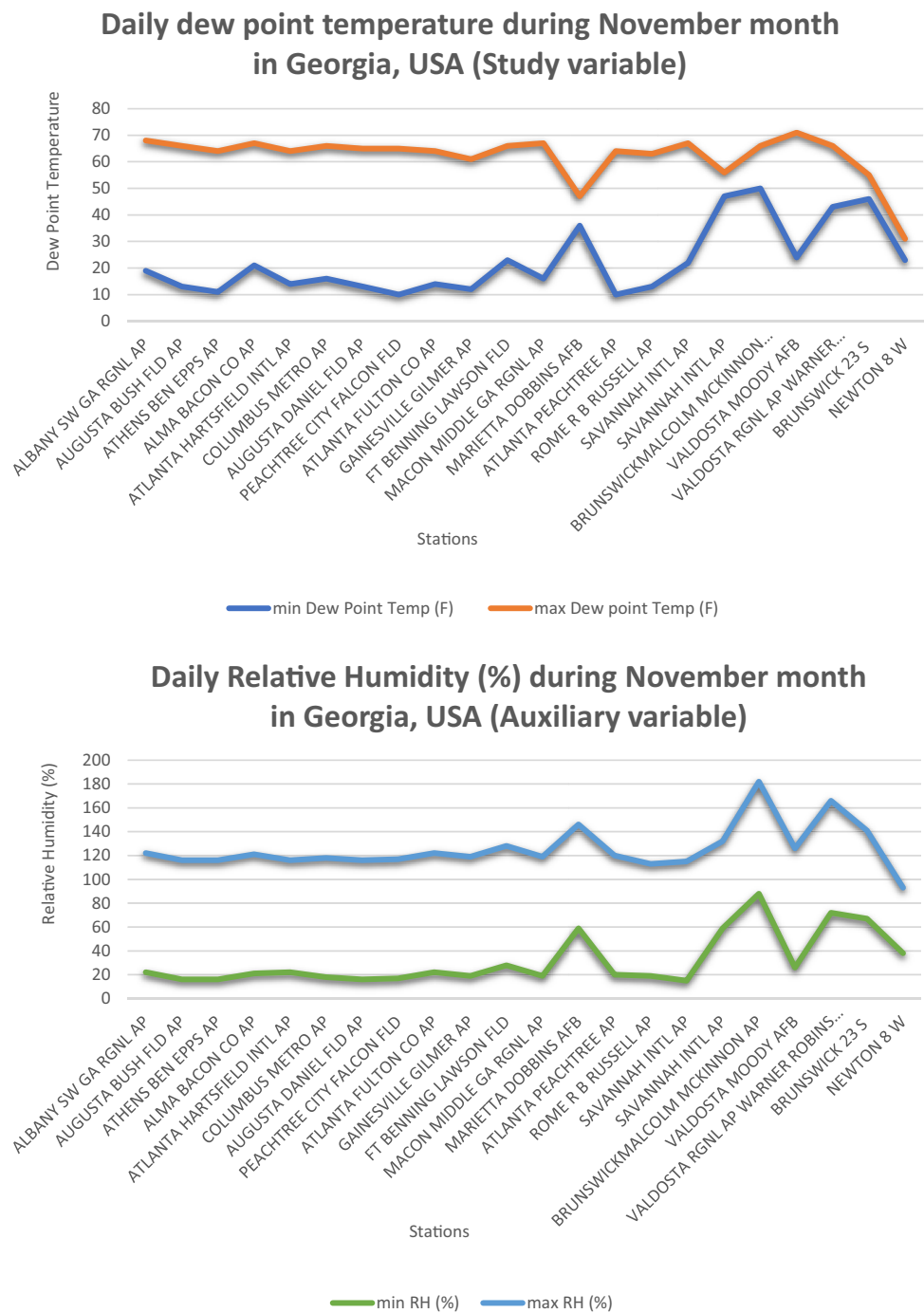
where

- $T^* = T_{mN}, T_{r_{jN}}, j = 1, 2, 3, T_{lr_{jN}}, T_{l_{jN}}, T_{e_{jN}}, T_{k_{jN}}$
- (v) For neutrosophic normal and neutrosophic gamma populations, respectively, the simulated MSE is shown in Tables 1-2 and the simulated PRE is shown in Tables 3-4.

5.1 Discussion of simulation results

An essential component of any research is the thorough discussion of the results and their consequences. We interpret the simulation results in this part by elucidating their importance in light of the goals of the study and the body

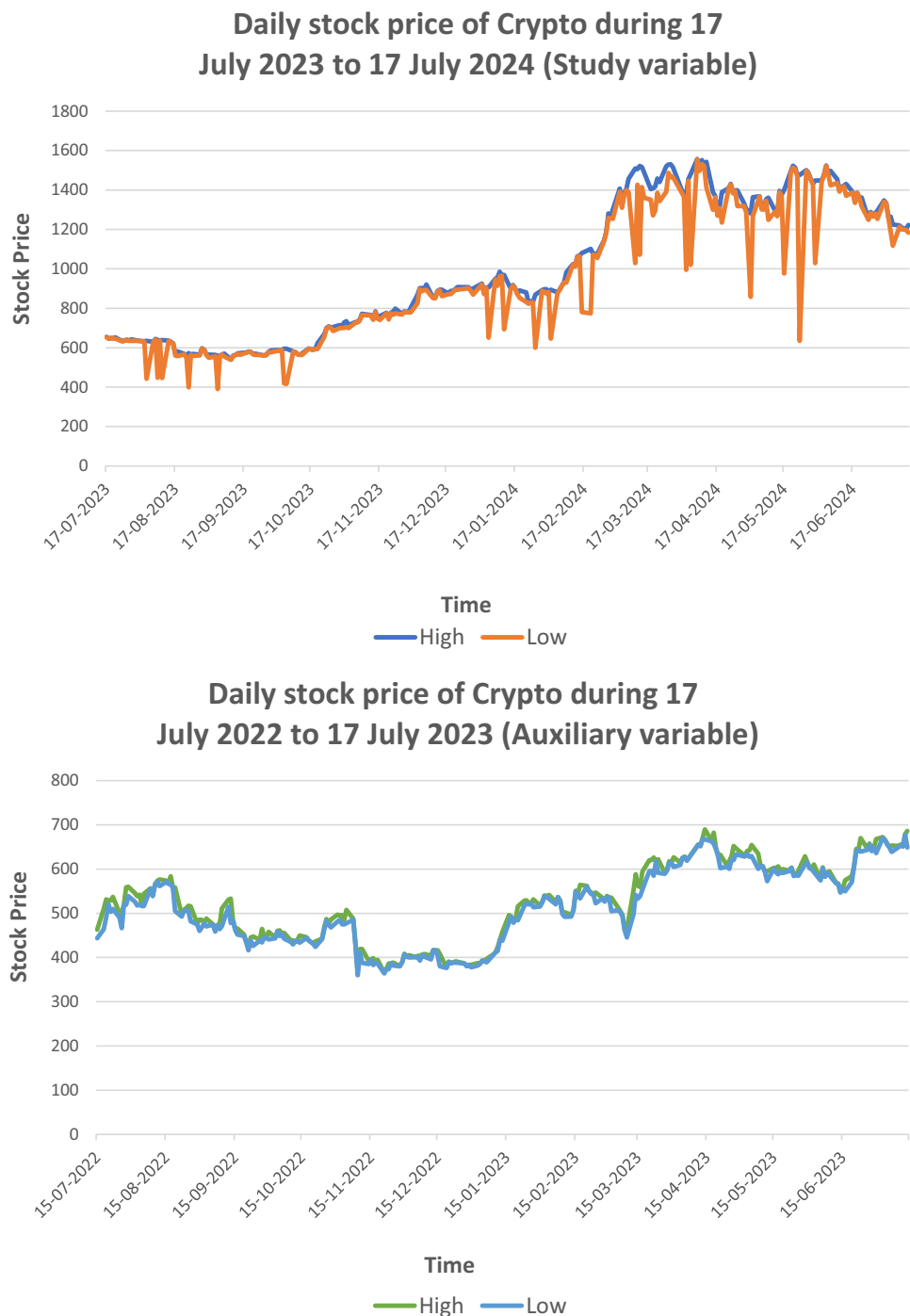
Fig. 2 Line chart to show dew point temperature (F) and relative humidity (%) at different stations in Georgia, USA



of current information. Additionally, it gives a chance to address any irregularities or unexpected patterns seen in the data, proposing theories or explanations that may be pursued further. A comprehensive discussion can also help practitioners adapt the study findings to real-life situations, provide actionable recommendations, and pinpoint domains that warrant further investigation. The simulation results shown in Tables 1, 2, 3, 4 are discussed as follows:

- The MSE_N values derived from the neutrosophic normal and neutrosophic gamma populations are reported in Tables 1, 2, respectively, demonstrate that, for every value of ρ_{xyN} , the proposed novel neutrosophic estimators $T_{k_{jN}}$, $j = 1, 2, 3$ obtain least MSE_N than the existing neutrosophic mean estimator T_{mN} , neutrosophic ratio estimators $T_{r_{jN}}$, neutrosophic regression estimators $T_{l_{r_{jN}}}$, neutrosophic logarithmic estimators $T_{l_{jN}}$, and Kumar and Kumar [34] estimators $T_{e_{jN}}$ under situations I-III. This shows improved accuracy and superior

Fig. 3 Line chart to show changes in daily stock price of Crypto over time



performance while handling with indeterminate and uncertain data.

- The PRE_N values obtained from the neutrosophic normal and neutrosophic gamma populations are reported in Tables 3, 4, respectively, show that, for every value of ρ_{xy_N} , the proposed novel neutrosophic estimators $T_{k_{jN}}$, $j = 1, 2, 3$ obtain highest PRE_N than the adapted neutrosophic mean estimator T_{mN} , neutrosophic ratio estimators $T_{r_{jN}}$, neutrosophic regression estimators

$T_{lr_{jN}}$, neutrosophic logarithmic estimators $T_{l_{jN}}$, and Kumar and Kumar [34] estimators $T_{e_{jN}}$ under situations I-III. This demonstrates enhanced performance and accuracy of the proposed approaches while working with indeterminate and uncertain data.

- It is clear from Tables 1, 2 that, for both neutrosophic normal and neutrosophic gamma populations, the MSE_N of the proposed novel neutrosophic estimators

Table 5 Descriptive statistics of real data sets

Descriptive statistics	Data set 1	Data set 2
N_N	[22, 22]	[252, 252]
n_N	[8, 8]	[30, 30]
r_N	[6, 6]	[20, 20]
$\bar{Y}_N$	[31.77, 93.86]	[950.4153, 992.4203]
$\bar{X}_N$	[22.55, 62.23]	[510.7419, 520.3006]
C_{yN}	[0.126, 0.689]	[0.3503, 0.3463]
C_{xN}	[0.139, 0.579]	[0.1646, 0.1643]
ρ_{xyN}	[0.885, 0.948]	[0.647, 0.650]

Table 6 $MSE_N \in [MSE_L, MSE_U]$ of neutrosophic estimators for real data sets

Estimators	Data set 1	Data set 2
T_{mN}	[1.9423, 506.9272]	[5104.5470, 5436.9920]
Situation I		
$T_{r_{1N}}$	[1.0046, 211.8097]	[3843.4310, 4077.2790]
$T_{lr_{1N}}$	[0.9439, 207.9545]	[3741.0670, 3970.1260]
$T_{l_{1N}}$	[0.9439, 207.9545]	[3741.0670, 3970.1260]
$T_{e_{1N}}$	[0.9411, 196.3449]	[3722.9890, 3951.3640]
$T_{k_{1N}}$	[0.9379, 186.7844]	[3719.6390, 3947.8090]
Situation II		
$T_{r_{2N}}$	[0.5135, 57.2244]	[3127.6620, 3305.5500]
$T_{lr_{2N}}$	[0.4210, 51.3497]	[2967.2010, 3137.5800]
$T_{l_{2N}}$	[0.4210, 51.3497]	[2967.2010, 3137.5800]
$T_{e_{2N}}$	[0.4186, 45.5098]	[2954.0630, 3123.9770]
$T_{k_{2N}}$	[0.4136, 33.2314]	[2949.4250, 3119.0570]
Situation III		
$T_{r_{3N}}$	[1.4511, 352.3419]	[4388.7780, 4665.2630]
$T_{lr_{3N}}$	[1.4193, 350.3224]	[4330.6800, 4604.4470]
$T_{l_{3N}}$	[1.4193, 350.3224]	[4330.6800, 4604.4470]
$T_{e_{3N}}$	[0.7504, 169.8358]	[2474.6850, 2627.7860]
$T_{k_{3N}}$	[0.7493, 166.2713]	[2473.4430, 2626.4680]

$T_{k_{jN}}$, $j = 1, 2, 3$ under situations I-III decreases as ρ_{xyN} increases.

- It is clear from Tables 3, 4 that, for both neutrosophic normal and neutrosophic gamma populations, the PRE_N of the suggested novel neutrosophic estimators $T_{k_{jN}}$, $j = 1, 2, 3$ under situations I-III grows as ρ_{xyN} increases.
- Furthermore, in both populations, for high correlation, situation II yields superior results, while for low correlation, situation III provides superior results, for the suggested and adapted resultant estimators.

Table 7 $PRE_N \in [PRE_L, PRE_U]$ of neutrosophic estimators for real data sets

Estimators	Data set 1	Data set 2
T_{mN}	[100.0000, 100.0000]	[100.0000, 100.0000]
Situation I		
$T_{r_{1N}}$	[193.3315, 239.3314]	[132.8122, 133.3486]
$T_{lr_{1N}}$	[205.7577, 243.7684]	[136.4463, 136.9476]
$T_{l_{1N}}$	[205.7577, 243.7684]	[136.4463, 136.9476]
$T_{e_{1N}}$	[206.3752, 258.1820]	[137.1089, 137.5979]
$T_{k_{1N}}$	[207.0742, 271.3970]	[137.2323, 137.7218]
Situation II		
$T_{r_{2N}}$	[378.2504, 885.8582]	[163.2065, 164.4807]
$T_{lr_{2N}}$	[461.3078, 987.2058]	[172.0324, 173.2862]
$T_{l_{2N}}$	[461.3078, 987.2058]	[172.0324, 173.2862]
$T_{e_{2N}}$	[463.9526, 1113.8850]	[172.7975, 174.0407]
$T_{k_{2N}}$	[469.5658, 1525.4460]	[173.0692, 174.3153]
Situation III		
$T_{r_{3N}}$	[133.8457, 143.8737]	[116.3091, 116.5420]
$T_{lr_{3N}}$	[136.8426, 144.7030]	[117.8694, 118.0813]
$T_{l_{3N}}$	[136.8426, 144.7030]	[117.8694, 118.0813]
$T_{e_{3N}}$	[258.8316, 298.4808]	[206.2705, 206.9039]
$T_{k_{3N}}$	[259.2046, 304.8796]	[206.3741, 207.0078]

6 Real data applications

A real data application is essential to demonstrate the practical utilization and robustness of the proposed neutrosophic imputation approaches in addressing the practical difficulties. Simulated datasets frequently fail to properly portray several kinds of uncertainty and indeterminacy information that are present in real data. Through the use of real datasets, we may verify the accuracy and dependability of the neutrosophic imputation approaches in realistic situations. Furthermore, real-life data applications highlight the usefulness of the neutrosophic imputation in various domains, including the environmental sciences, stock marketing, social sciences, medical, and economics, where precise data analysis is essential for making decisions. For these reasons, we have used two distinct real data sets to conduct a real data application of the suggested neutrosophic imputation approaches.

The data set 1 is based on the humidity and temperature of the Georgia state, USA during November month. This dataset is taken from a publicly accessible website <http://mrcc.purdue.edu>. In this dataset, the neutrosophic study variable $y_N \in [y_L, y_U]$ is relative humidity, while the neutrosophic auxiliary variable $x_N \in [x_L, x_U]$ is the dew point temperature (F).

The data set 2 is related to the daily price of stock 'CMC Crypto 200 Index by Solacti (CMC200)' which can be

obtained from publicly available website <https://finance.yahoo.com/quote/%5ECMC200/history/>. The neutrosophic survey variable $y_N \in [y_L, y_U]$ is the varying price of the stock on each day from 17 July, 2023 to 17 July, 2024, while the neutrosophic auxiliary variable $x_N \in [x_L, x_U]$ is the varying price of the stock on each day from 17 July, 2022 to 17 July, 2023.

The line charts of dataset 1 and dataset 2 are provided in Fig. 2 and Fig. 3, respectively.

Using the basic arithmetic operators defined in notation section, we compute the required descriptive statistics of both datasets and reported in Table 5. The following equation is utilized to calculate the PRE_N .

$$PRE_N = \frac{V(T_{mN})}{MSE(T^*)} \times 100. \quad (6.1)$$

The real data results are shown by $MSE_N \in [MSE_L, MSE_U]$ in Table 6 and by $PRE_N \in [PRE_L, PRE_U]$ in Table 7. The real data findings presented in Table 6 demonstrate that the MSE_N of the proposed neutrosophic estimators T_{k_jN} , $j = 1, 2, 3$ is minimum when compared to the MSE_N of the neutrosophic mean estimator T_{mN} , neutrosophic ratio estimators T_{r_jN} , neutrosophic regression estimators T_{lr_jN} , neutrosophic logarithmic estimators T_{l_jN} , and Kumar and Kumar [34] estimators T_{e_jN} under situations I-III. Furthermore, the PRE_N of the suggested neutrosophic estimators T_{k_jN} is highest against the PRE_N of the adapted neutrosophic estimators, as shown by the real data results of Table 7. Moreover, among the three situations taken into consideration, situation II yields better results in dataset 1 for the suggested and adapted neutrosophic imputation approaches, while situation III provide better results in dataset 2 for the suggested neutrosophic imputation approaches.

7 Conclusions

This paper proposed some novel neutrosophic imputation approaches and the corresponding resultant estimators to deal with the problems caused by missing data in survey sampling. The proposed approaches provided a strong foundation for managing the inherent uncertainty and indeterminacy in real-life data by utilizing the ideas of neutrosophic logic. The proposed novel neutrosophic imputation approaches enhanced the accuracy and reliability of the population mean estimates by incorporating the degrees of truth, falsity, and indeterminacy simultaneously, thereby provided a more extensive solution than the conventional imputation approaches.

The superiority of the proposed approaches was established through rigorous mathematical derivations, extensive simulations using both symmetric and asymmetric neutrosophic populations, and real data applications. These findings collectively highlight not only the methodological novelty of the proposed approaches but also its practical robustness in dealing with incomplete and ambiguous information. Further, the real data analyses confirmed that the suggested imputations yield more reliable and valid statistical inferences compared to the existing and adapted neutrosophic imputations.

The broader significance of this work lies in its applicability across different fields such as environmental studies, economics, healthcare, and social sciences, where decision-making often depends on incomplete and uncertain data. By improving the reliability of the population mean estimation under such conditions, this study provided both a theoretical contribution and a practical tool for researchers and survey practitioners.

In forthcoming studies, we would concentrate our focus on developing enhanced neutrosophic approaches, investigating their potential in alternative fields, and developing approachable instruments to promote their extensive implementation.

Acknowledgements The authors are extremely grateful to the panel of reviewers and the Editor-in-Chief for their careful insights and constructive suggestions which led to considerable improvement of the paper.

Author contributions V.K. was involved in the writing original draft, formal analysis, investigation, and simulation. A.K. was involved in the conceptualization, methodology, validation, writing-review and editing, and supervision.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

Ethics declarations Not applicable.

References

1. Alhabib, R., Ranna, M.M., Farah, H., Salama, A.A.: Some neutrosophic probability distributions. *Neutrosophic Sets Syst* **22**, 30–38 (2018)
2. Alomair, A.M., Ahmad, S.: New comprehensive mean estimation using regression-cum-exponential type estimator: application with neutrosophic data. *Kuwait Journal of Science* **52**(1), 100346 (2025)

3. Al-Marzouki, S., Ahmad, S.: Developing a generalized class of estimators for estimation of population mean using neutrosophic approach. *Eur. J. Pure Appl. Math.* **18**(3), 6477–6477 (2025)
4. Atanassov, K.: Intuitionistic fuzzy sets. *Fuzzy Sets and Systems* **20**(1), 87–96 (1986)
5. Ahmed, M.S., Al-Titi, O., Al-Rawi, Z., Abu-Dayyeh, W.: Estimation of a population mean using different imputation methods. *Stat. Transit.* **7**(6), 1247–1264 (2006)
6. Al-Omari, A.I., Bouza, C.N., Herrera, C.: Imputation methods of missing data for estimating the population mean using simple random sampling with known correlation coefficient. *Qual. Quant.* **47**, 353–365 (2013)
7. Al-Omari, A.I., Bouza, C.: Ratio estimators of the population mean with missing values using ranked set sampling. *Environmetrics* **26**(2), 67–76 (2015)
8. Alqudah, M.A., Zayed, M., Subzar, M., Wani, S.A.: Neutrosophic robust ratio type estimator for estimating finite population mean. *Heliyon* **10**(8), 1–12 (2024)
9. Almarashi, A.M., Aslam, M.: Correlated proportions test under indeterminacy. *J. Math.* **2021**(1), 6564006 (2021)
10. Anas, M.M., Huang, Z., Shahzad, U., Zaman, T., Shahzadi, S.: Compromised imputation based mean estimators using robust quantile regression. *Communications in Statistics* **53**(5), 1700–1715 (2022)
11. Aslam, M.: Neutrosophic analysis of variance: application to university students. *Complex Intell. Syst.* **5**(4), 403–407 (2019)
12. Aslam, M.: Monitoring the road traffic crashes using NEWMA chart and repetitive sampling. *International Journal of Injury Control and Safety Promotion* **28**(1), 39–45 (2020)
13. Aslam, M.: Testing average wind speed using sampling plan for Weibull distribution under indeterminacy. *Sci. Rep.* **11**(1), 7532 (2021)
14. Aslam, M., Arif, O.H., Sherwani, R.A.K.: New diagnosis test under the neutrosophic statistics: an application to diabetic patients. *BioMed Research International* **2020**(1), 2086185 (2020)
15. Aslam, M., Algarni, A.: Analyzing the solar energy data using a new Anderson-Darling test under indeterminacy. *Int. J. Photoenergy* **2020**(1), 6662389 (2020)
16. Arif, O.H., Aslam, M.: A new sudden death chart for the weibull distribution under complexity. *Complex Intell. Syst.* **7**(4), 2093–2101 (2021)
17. Bhushan, S., Pandey, S.: A cost-effective computational approach with non response on two occasions. *Communications in Statistics* **49**(20), 4951–4973 (2020)
18. Bhushan, S., Pandey, S.: Optimal random non response framework for mean estimation on current occasion. *Commun. Stat. Theory Methods*, 1–27 (2024)
19. Bhushan, S., Kumar, A., Pandey, A.P., Singh, S.: Estimation of population mean in presence of missing data under simple random sampling. *Communications in Statistics* **52**(12), 6048–6069 (2023)
20. Bhushan, S., Kumar, A., Zaman, T., Al Mutairi, A.: Efficient difference and ratio-type imputation methods under ranked set sampling. *Axioms* **12**(6), 558 (2023)
21. Bhushan, S., Kumar, A.: Imputation of missing data using multi auxiliary information under ranked set sampling. *Commun. Stat. Simul. Comput.*, 1–22 (2023)
22. Bhushan, S., Kumar, A.: Ranked set sampling imputation methods in presence of correlated measurement errors. *Commun. Stat. Theory Methods*, 1–22 (2024)
23. Bouza, C.N., Al-Omari, A.I.: Ranked set estimation with imputation of the missing observations: the median estimator. *Investig. Oper.* **32**(1), 30–37 (2011)
24. Chakraborty, A., Mondal, S.P., Alam, S., Dey, A.: Classification of trapezoidal bipolar neutrosophic number, de-bipolarization technique and its execution in cloud service-based MCGDM problem. *Complex Intell. Syst.* **7**(1), 145–162 (2021)
25. Chen, C.T.: Extensions of the TOPSIS for group decision-making under fuzzy environment. *Fuzzy Sets Syst.* **114**(1), 109 (2000)
26. Chen, J., Ye, J., Du, S.: Scale effect and anisotropy analyzed for neutrosophic numbers of rock joint roughness coefficient based on neutrosophic statistics. *Symmetry* **9**(10), 208 (2017)
27. Chhoker, P.K., Nagar, A.: Mathematical modeling and fuzzy availability analysis of stainless steel utensil manufacturing unit in steady state: a case study. *Int. J. Syst. Assur. Eng. Manag.* **6**(3), 304–318 (2015)
28. Diana, G., Perri, P.F.: Improved estimators of the population mean for missing data. *Communications in Statistics* **39**, 3245–3251 (2010)
29. Heitjan, D.F., Basu, S.: Distinguishing ‘Missing at Random’ and ‘Missing Completely at Random. *The American Statistician* **50**, 207–213 (1996)
30. Kumar, P., Dudeja, C.: Shadowed type 2 fuzzy-based Markov model to predict shortest path with optimized waiting time. *Soft. Comput.* **25**(2), 995–1005 (2021)
31. Kumar, A., Priya, Tripathi, V.: Optimal neutrosophic framework for population mean estimation under simple random sampling. *Neutrosophic Sets Syst.* **86**, 689–709 (2025)
32. Kumar, A., Bhushan, S., Mustafa, M.S., Aldallal, R., Aljohani, H.M., Almulhim, F.A.: Novel imputation methods under stratified simple random sampling. *Alexandria Engineering Journal* **95**, 236–246 (2024)
33. Kumar, A., Bhushan, S., Pokhrel, R., Emam, W.: Small area estimation using design based direct and synthetic logarithmic estimators. *Ain Shams Engineering Journal* (2024). <https://doi.org/10.1016/j.asej.2024.102836>
34. Kumar, A., Kumar, V.: Uncertainty based efficient neutrosophic imputation methods for population mean. *Natl. Acad. Sci. Lett.*, 1–11 (2025)
35. Lee, H., Rancourt, E., Sarndal, C.E.: Experiments with variance estimation from survey data with imputed values. *J. Off. Stat.* **10**, 231–243 (1994)
36. Mahdizadeh, M., Zamanzade, E.: Smooth estimation of a reliability function in ranked set sampling. *Statistics* **52**(4), 750–768 (2018)
37. Masood, S., Ibrar, B., Shabbir, J., Shokri, A., Movaheedi, Z.: Estimating neutrosophic finite median employing robust measures of the auxiliary variable. *Sci. Rep.* **14**(1), 10255 (2024)
38. Nabeeh, N.A., Smarandache, F., Abdel-Basset, M., El-Ghareeb, H.A., Aboelfetouh, A.: An integrated neutrosophic-TOPSIS approach and its application to personnel selection: a new trend in brain processing and analysis. *IEEE Access* **7**, 29734–29744 (2019)
39. Rao, G.S., Aslam, M.: Inspection plan for COVID-19 patients for weibull distribution using repetitive sampling under indeterminacy. *BMC Med. Res. Methodol.* **21**(1), 229 (2021)
40. Rubin, R.B.: Inference and missing data. *Biometrika* **63**(3), 581–592 (1976)
41. Saleem, M., Aslam, M.: Algorithms for generating neutrosophic gamma distributed data. *J. Reliab. Stat. Stud.* **18**(2), 289–312 (2025)
42. Smarandache, F.: A unifying field in logics: neutrosophic logic. *Philos. (London, England)*, American Research Press, 1–141 (1999)
43. Smarandache, F.: Neutrosophic set a generalization of the intuitionistic fuzzy set. *Int. J. Pure Appl. Math.* **24**(3), 287 (2005)
44. Smarandache, F.: Neutrosophic logic - a generalization of the intuitionistic fuzzy logic. *Multispace & Multistructure. Neutrosophic Transdisciplinarity (100 Collected Papers of Science)*, **4**, 396 (2010)

45. Smarandache, F.: Introduction to neutrosophic measure, neutrosophic integral, and neutrosophic probability. *Infinite Study* (2013)
46. Smarandache, F.: Introduction to neutrosophic statistics. Sitech & Education Publishing (2014)
47. Priya, Kumar, A.: Robust neutrosophic exponential estimators of population mean in the presence of uncertainty. *Qual. Quant.*, 1–24 (2025). <https://doi.org/10.1007/s11135-025-02150-6>
48. Punia, P., Raj, A., Kumar, P.: Enhanced zebra optimization algorithm for reliability redundancy allocation and engineering optimization problems. *Cluster Comput* **28**(4), 267 (2025)
49. Singh, H.P., Horn, S.: An alternative estimator for multi-character surveys. *Metrika* **48**, 99–107 (1998)
50. Singh, S., Horn, S.: Compromised imputation in survey sampling. *Metrika* **51**, 267–276 (2000)
51. Tahir, Z., Khan, H., Aslam, M., Shabbir, J., Mahmood, Smarandache, F., et al.: Neutrosophic ratio-type estimators for estimating the population mean. *Complex Intell. Syst.* **7**(6), 2991–3001 (2021)
52. Yadav, S.K., Smarandache, F.: Generalized neutrosophic sampling strategy for elevated estimation of population mean. *Neutrosophic Sets Syst.* **53**, 1–20 (2023)
53. Yadav, V.K., Prasad, S.: Neutrosophic estimators for estimating the population mean in survey sampling. *Measurement: Interdisciplinary Research and Perspectives* **22**(4), 373–397 (2024)
54. Zadeh, L.A.: Fuzzy sets. *Information and Control* **8**(3), 338–353 (1965)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.