International Journal of Analysis and Applications

Neutrosophic Statistics with Outliers Data: Utilizing Neutrosophic Median Absolute Deviation to Estimate the Mean Parameter Using Neutrosophic Modified One Step M-Estimator

Nadia Hanim Abd Gahni, Nora Muda*

Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia

**Corresponding author:* noramuda@ukm.edu.my, nadhanim89@gmail.com

ABSTRACT. Classical statistical methods rely on precise data to estimate population means using auxiliary information but often face issues like bias and high mean squared error (MSE). Neutrosophic statistics extend classical approaches by incorporating vague, indeterminate, and uncertain data. This study introduces the Modified One-Step M-estimator (NMOM), which utilizes auxiliary information to improve estimation accuracy. The Neutrosophic Median Absolute Deviation (NMAD) is also employed to measure robustness against outliers and uncertainty. Empirical studies and simulations compare NMOM with the Neutrosophic Standard Mean (NSM) using metrics such as mean, median, standard deviation, covariance, NMAD, number of outliers, and NMSE. Results show that NMOM is more robust than NSM, particularly in managing outliers, reducing variance, and achieving lower MSE. The use of NMAD strengthens NMOM's ability to produce reliable estimates under uncertain data conditions. This highlights NMOM's effectiveness in fields like finance, engineering, and medicine, where data imprecision is a key concern.

1. Introduction

The presence of outliers in data can significantly impact the reliability and accuracy of parameter estimation. Outliers are data points that deviate markedly from the overall pattern of a dataset and may arise due to measurement errors, data entry mistakes, or inherent variability in the observed phenomena. While outliers can provide meaningful insights in some cases, they often distort the results of statistical analyses, leading to biased parameter estimates, reduced efficiency, and poor model performance. Handling outliers is, therefore, a critical step in ensuring the robustness and reliability of statistical inference.

Received May 2, 2025

2020 Mathematics Subject Classification. 62G35.

Key words and phrases. outliers; neutrosophic statistics; robust; mean estimator.

https://doi.org/10.28924/2291-8639-23-2025-208

ISSN: 2291-8639

In classical statistics, most standard methods for mean parameter estimation are highly sensitive to outliers. These methods assume that the data follow a well-defined distribution, often normal, and are free from extreme deviations. However, when outliers are present, the mean can shift considerably, and can become unreliable, leading to conclusions that misrepresent the underlying data structure. This sensitivity to outliers undermines the validity of the results, particularly in fields such as Iin finance, engineering, and medical research, where real-world datasets often exhibit irregularities and contamination.

Limited attention has been given to detecting and addressing outliers in unclear, ambiguous, or indeterminate, or when they are in the form of intervals. This challenge also affects parameter estimation, making it a critical area for further research. To address such situations, fuzzy logic is a valuable tool that handles data with imprecision. Fuzzy statistics are used to analyze such datasets introduced by [1] but does not account for the degree of indeterminacy inherent in the data. [2] further expanded this concept by introducing neutrosophy, a framework that incorporates both determinate and indeterminate aspects of uncertain or imprecise data. This distinctive feature makes neutrosophic approaches highly versatile and well-suited for handling uncertainty, indeterminacy, or vagueness.

Recent advancements in neutrosophic statistics have addressed key challenges related to uncertainty and ambiguity in robust parameter estimator. For example, [3] proposed a neutrosophic predictive estimator for the finite population mean using kernel regression, offering a robust alternative to classical methods. Similarly, [4] introduced a neutrosophic calibration approach to improve stratification weights and estimate the empirical cumulative distribution function (CDF) of finite populations, showcasing a novel application of neutrosophic techniques. Further developments include a neutrosophic estimator with minimum mean squared error (MSE) for improved population mean estimation [5] and a neutrosophic modified ratio-cum-product log-type estimator utilizing medians of auxiliary variables [6].

Furthermore, [7] advanced the field by proposing neutrosophic ratio-type estimators, which leverage auxiliary information for more effective finite population mean estimation. Additionally, three neutrosophic exponential ratio-type estimators were developed, utilizing auxiliary variables to enhance precision [8]. Addressing sampling gaps, a "neutrosophic median ranked set sampling" method was introduced to estimate population means in ambiguous datasets [9]. Collectively, these advancements underscore the growing relevance of neutrosophic statistics in tackling real-world data uncertainties and improving parameter estimation methodologies. It can be seen that all these approaches did not focus on identifying the outliers in the neutrosophic dataset. In the meantime, [10] and [11] have developed the Grubbs test and the Dixon test under neutrosophic environments in order to identify outliers. However, both studies did not extend to mean parameter estimation, highlighting the need for a robust technique to address this gap.

In the classical framework, the popular robust statistical technique for analysing classical datasets is the Median Absolute Deviation (MAD) as it is robust to outliers [12] and the Modified One-Step M-Estimator (MOM) is preferable by [13]. While both methods are robust and designed to handle non-normal or contaminated data, they serve different purposes and are used in distinct contexts. MAD is a measure of variability that is highly resistant to outliers because it relies on the median, which is less sensitive to extreme values compared to the mean. It calculates the median of the absolute deviations from the dataset's median, making it an effective tool for estimating scale in the presence of outliers or skewed distributions.

On the other hand, MOM is a robust method for estimating the central tendency of a dataset. It empirically determines how many observations should be trimmed, allowing for different amounts of trimming in the tails or even no trimming at all. Together, MAD and MOM provide robust solutions for analyzing data that deviates from normality or contains contamination, with MAD focusing on variability and MOM on central tendency.

Many studies have explored the use of robust methods. For instance, [14] applied MAD in conjunction with neural network training, demonstrating its utility in enhancing model robustness. [15] investigated the joint asymptotic normality of MAD with the sample median, providing theoretical insights into its statistical properties. Additionally, [16] highlighted MAD's effectiveness as a less sensitive alternative to traditional measures when dealing with extreme values, making it a preferred choice for analyzing datasets with outliers. On the other hand, [17] pioneered M-estimators by extending maximum likelihood estimation to handle outliers, laying the foundation for robust statistical techniques.

[18] emphasized that the MOM estimator specifically trims only the extreme values in a dataset, with the trimming process tailored to the data distribution, making it particularly effective for managing non-normal data. Furthermore, [19] endorsed the MOM estimator for its strong capability to detect and handle outliers in data distributions, further solidifying its reputation as a robust and reliable statistical tool. Together, these studies underscore the versatility and effectiveness of MAD and MOM in precise and crisp datasets. However, the primary goal of this study is to introduce an improved robust method above within a neutrosophic framework.

Therefore, this study contributes by introducing a neutrosophic approach that integrates robust statistical techniques for uncertainty, imprecision, and indeterminacy which are the Neutrosophic Median Absolute Deviation (NMAD), for robust outlier detection with a Neutrosophic Modified One-step M-estimator (NMOM) for accurate parameter estimation. This approach aims to enhance parameter estimation in the presence of outliers, providing a reliable and flexible framework for analysing uncertain and imprecise data. The study also evaluates the effectiveness of the proposed method using both simulation studies and real-world temperature datasets, highlighting its practical applicability in environments where data irregularities are prevalent.

Thus, this study is conducted to develop and propose a robust method that can effectively account for uncertainty, imprecision, and indeterminacy in data while accurately identifying outliers and estimating parameters. To validate the effectiveness of the proposed method, simulations are carried out under different scenarios that consider a wide range of data variations. These simulations allow for a systematic evaluation of the method's performance across multiple conditions, such as datasets with varying levels of variance and outliers. Following the simulations, the method is further validated by applying it to real-world datasets. This step ensures the practical applicability of the method and demonstrates its reliability and versatility in handling real-world challenges, such as environmental measurements, economic indicators, or social survey data, where uncertainty and indeterminacy are common.

2. Neutrosophic and its properties

The neutrosophic independent random variable is defined as $X_N = X_L + I_N X_U; I_N \in [I_L, I_U]$, where X_L is the determined and $I_N X_U$ is the indetermined part. Also, $I_N \in [I_L, I_U]$ is an undetermined interval. To identify the neutrosophic random variable, subscript N is used as shown above. Now consider a neutrosophic sample of size $n_N \in [n_L, n_U]$. The neutrosophic sample mean and standard deviation can be written as below

$$\overline{X_N} = \overline{X_L} + \overline{X_U} I_N;\ \text{where, } \overline{X_L} = \frac{1}{n_N}\sum_{i=1}^{n_N} X_L,\ \ \overline{X_U} = \frac{1}{n_N}\sum_{i=1}^{n_N} X_U \tag{1}$$

$$SD_N = \sqrt{\frac{\sum_{i=1}^{n_N}(X-\bar{X}_N)^2}{n_N}};\ \text{where } SD_N \in [SD_L, SD_U] \tag{2}$$

According to [20], $(X - \bar{X}_N)^2$ given by

$$(X-\bar{X}_N)^2 = \begin{bmatrix} \min\begin{pmatrix} (a_i b_i I_L)(\bar{a}+\bar{b}I_L), (a_i + b_i I_L)(\bar{a}+\bar{b}I_U) \\ (a_i b_i I_U)(\bar{a}+\bar{b}I_L), (a_i + b_i I_U)(\bar{a}+\bar{b}I_U) \end{pmatrix} \\ \max\begin{pmatrix} (a_i b_i I_L)(\bar{a}+\bar{b}I_L), (a_i + b_i I_L)(\bar{a}+\bar{b}I_U) \\ (a_i b_i I_U)(\bar{a}+\bar{b}I_L), (a_i + b_i I_U)(\bar{a}+\bar{b}I_U) \end{pmatrix} \end{bmatrix}, I_N \in [I_L, I_U]. \tag{3}$$

2.1 Neutrosophic Median & Median Absolute Deviation About Median

The neutrosophic median (NM) is determined from the neutrosophic sample data set by selecting the value that lies at the midpoint of the observations each at lower and upper data. Let $\widehat{M}_N$ denote the sample median of the neutrosophic data where $\widehat{M}_N \in [\widehat{M}_L, \widehat{M}_U]$.

Then, the neutrosophic median $\widehat{M}_N$ is expressed as follows:

If n is odd, $\widehat{M}_N = x_{N\left(\frac{n+1}{2}\right)}$. (4)

If n is even, $\widehat{M}_N = \frac{x_{N\left(\frac{n}{2}\right)} + x_{N\left(\frac{n}{2}+1\right)}}{2}$. (5)

Next, the neutrosophic median absolute deviation (NMAD) is found by first compute the absolute differences each lower and upper data and their median value. Then, find the neutrosophic median of the absolute difference for $|X_L - \widehat{M}_L|$ and $|X_U - \widehat{M}_U|$. We denote the NMAD as $NMAD_N$ where $NMAD_N \in [NMAD_L, NMAD_U]$.

The neutrosophic median absolute deviation about median ($NMADN$) is a robust measure of statistical dispersion, often used to detect outliers. It is derived from the neutrosophic Median Absolute Deviation (NMAD) but scaled to approximate the standard deviation under a normal distribution. This is to ensure that $NMADN$ provides as estimate that aligns with the standard deviation in a normal distribution [21].

$$NMADN_N = \frac{NMAD_N}{0.6745} \quad (6)$$

where $NMADN_N \in [NMADN_L, NMADN_U]$.
The value X_N is considered an outlier if it meets the criteria of the robust decision rule in Eq. 7-8:

$$\frac{|X_N - \widehat{M}_N|}{NMADN_N} > 1.4826 \quad (7)$$

Or when $\frac{|X_N - \widehat{M}_N|}{NMADN_N} < -1.4826$, (8)

To ensure that the Normalized Median Absolute Deviation (MADN) is comparable to the standard deviation for normally distributed data, it is multiplied by a common scaling factor of 1.4826, which is derived based on the assumption of normality.

2.2 Neutrosophic Modified One Step M-estimator & MSE

The neutrosophic modified one step M estimator is defined by

$$\widetilde{\mu_m} = \frac{\sum_{i=i_1+1}^{n-i_2} X_{iN}}{n-i_1-i_2}, \widetilde{\mu_m} \in [\widetilde{\mu_L}, \widetilde{\mu_U}]. \quad (9)$$

The MSE is given as

$$MSE(\widetilde{\mu_m}) = Var(\widetilde{\mu_m}) + Bias\,(\widetilde{\mu_m})^2 \quad (10)$$

where

$Var(\widetilde{\mu_m})$ is the variance of the estimator and $Bias(\widetilde{\mu_m}) = E(\widetilde{\mu_m}) - \widetilde{\mu_m}$ is the bias of the estimator.

Let $\widetilde{\mu_m} = \frac{1}{n'}\sum_{i=i_1+1}^{n-i_2} X_{iN}$ where $n' = n - i_1 - i_2$

The expectation of NMOM is

$$E(\widetilde{\mu_m}) = E\left[\frac{1}{n'}\sum_{i=i_1+1}^{n-i_2} X_{iN}\right] = \frac{1}{n'}\sum_{i=i_1+1}^{n-i_2} E[X_{iN}] \tag{11}$$

Assuming that $E[X_{iN}] = \mu$, we get

$$E(\widetilde{\mu_m}) = \mu. \tag{12}$$

which means the estimator is unbiased.

Next, to evaluate the performance of NMOM, we utilized variance neutrosophic given by:

$$Var(\widetilde{\mu_m}) = E\left(\frac{\sum_{i=i_1+1}^{n-i_2} X_{iN}}{n'}\right)^2 - \left[E\left(\frac{\sum_{i=i_1+1}^{n-i_2} X_{iN}}{n'}\right)\right]^2. \tag{13}$$

Expectation of the squared $\widetilde{\mu_m}$ is

$$E\left(\frac{\sum_{i=i_1+1}^{n-i_2} X_{iN}}{n-i_1-i_2}\right)^2 = \frac{1}{n'^2} E\left[\sum_{i=i_1+1}^{n-i_2} X_{iN}^2 + 2\sum_{i<j} X_{iN}X_{jN}\right]. \tag{14}$$

Since $E\left[X_{iN}^2\right] = Var\,(X_{iN}) + (E[X_{iN}])^2 = \sigma^2 + \mu^2$, and assuming independence:

$$E\left(\frac{\sum_{i=i_1+1}^{n-i_2} X_{iN}}{n-i_1-i_2}\right)^2 = \frac{1}{n'^2}\left(n'(\sigma^2 + \mu^2) + 2\sum_{i<j} \mu^2\right). \tag{15}$$

Since there are $\binom{n'}{2} = \frac{n'(n'-1)}{2}$ terms in the second sum, therefore

$$E\left(\frac{\sum_{i=i_1+1}^{n-i_2} X_{iN}}{n-i_1-i_2}\right)^2 = \frac{n'(\sigma^2+\mu^2)+n'(n'-1)\mu^2}{n'^2} = \frac{n'\sigma^2+n'\mu^2+n'^2\mu^2-n'\mu^2}{n'^2} = \frac{n'\sigma^2+n'^2\mu^2}{n'^2}. \tag{16}$$

Hence, compute variance of NMOM is

$$E\left(\frac{\sum_{i=i_1+1}^{n-i_2} X_{iN}}{n-i_1-i_2}\right)^2 - \left[E\left(\frac{\sum_{i=i_1+1}^{n-i_2} X_{iN}}{n-i_1-i_2}\right)\right]^2 = \frac{n'\sigma^2+n'^2\mu^2}{n'^2} - \mu^2 = \frac{\sigma^2}{n'} = \frac{\sigma^2}{n-i_1-i_2}. \tag{17}$$

where $Var(\widetilde{\mu_m}) \in [Var(\widetilde{\mu_L}),\ Var(\widetilde{\mu_U})]$.

Furthermore, the NMSE of NMOM is

$$MSE(\widetilde{\mu_m}) = E[(\widetilde{\mu_m} - \mu)^2] = Var(\widetilde{\mu_m}) + [E(\widetilde{\mu_m}) - \mu]^2. \quad (18)$$

If $\widetilde{\mu_m}$ is unbiased, then:

$$MSE(\widetilde{\mu_m}) = Var(\widetilde{\mu_m}) = \frac{\sigma^2}{n-i_1-i_2}. \quad (19)$$

This result shows that the NMSE of NMOM will be smaller than MSE of the standard mean. We can express this as a ratio where

$$\frac{MSE(\bar{X})}{MSE\ (\widetilde{\mu_m})} = \frac{n}{n-i_1-i_2} \geq 1.$$

3. Simulation Procedure

This section presents a Monte Carlo simulation study conducted to evaluate the performance of the proposed NMOM compared to the neutrosophic standard mean under varying percentages of outliers and variances. The study examines the robustness and efficiency of these estimators in handling contaminated datasets, providing insights into their relative performance in the presence of outliers. Neutrosophic random samples of sizes 50 and 300 were generated from a neutrosophic normal distribution with a mean interval of [5,6]. Specified percentages of outliers (e.g., 0%, 5%, 10%, and 15%) were introduced into the contaminated datasets to simulate diverse conditions outlined in Tables 4 and 5. The simulation study was repeated 10,000 times to ensure reliable and stable results, offering a comprehensive analysis of the estimators' behavior under different conditions.

The simulation study will be discussed in the following manner. The process begins by generating data from a neutrosophic normal distribution with a mean interval of [5,6] and varying levels of variance and contamination percentages (0%, 5%, 10%, 15%). Next, several neutrosophic statistical measures are computed: the standard mean, median, standard deviation, median absolute deviation, and median absolute deviation about the median. The process then identifies outliers using Equation (7-8). Based on whether outliers are present, the analysis branches into two different approaches for estimating the mean parameter:

1. If no outliers are detected, the mean parameter is estimated using the neutrosophic standard approach.
2. If outliers are present, the mean parameter is estimated using a neutrosophic modified one-step M-estimator.

Finally, the process concludes by comparing the Mean Squared Error (MSE) of the neutrosophic approaches, likely to evaluate which method performs better under different contamination scenarios. This methodology appears to be designed to handle uncertainty and imprecision in data through neutrosophic statistics, with special consideration for robust estimation when outliers are present. The flowchart of the study using neutrosophic data is illustrated in Figure 1.

Table 1: Descriptive statistics under neutrosophic statistics for simulation study when $n_N = [300, 300]$

	Neutrosophic Standard Mean estimator with variance [1,1]			Neutrosophic Modified one-step estimator with variance [1,1]		
Outliers' percentage / Statistics	0%	10%	15%	0%	10%	15%
$\bar{X}_N$	[5.5352, 6.8868]	[6.5039, 8.143]	[6.9806, 8.7063]	[5.4971, 6.8478]	[5.6806, 7.1289]	[5.8269, 7.5021]
$\widehat{M}_N$	[5.5347, 6.8863]	[6.5003, 8.1375]	[6.9798, 8.7044]	[5.4971, 6.8481]	[5.6812, 7.1251]	[5.825, 7.4903]
MAD_N	[0.0541, 0.0798]	[0.1844, 0.238]	[0.2136, 0.2579]	[0.0594, 0.086]	[0.0697, 0.1161]	[0.0808, 0.1575]
SD_N	[0.0544, 0.0798]	[0.1876, 0.2416]	[0.2145, 0.2541]	[0.0589, 0.0864]	[0.073, 0.1157]	[0.0847, 0.1567]

As shown in Table 1, the comparison between the Neutrosophic Standard Mean Estimator (NSM) and the Neutrosophic Modified One-Step Estimator (NMOM) indicates that both estimators yield similar performance when no outliers are present. However, when 15% outliers are introduced, notable differences emerge. The $\bar{X}_N$ and median $\widehat{M}_N$ of the estimator increase significantly, indicating that it is more sensitive to the presence of outliers. In conclusion, while both estimators perform similarly in the absence of outliers, the NMOM estimator proves to be more robust and reliable in datasets contaminated with outliers, as it shows less variability and distortion across increasing levels of contamination.

Table 2: Neutrosophic MSE and Bias for NSM and NMOM for $n_n = [50, 50]$.

		Estimator			
		Neutrosophic Standard Mean		Neutrosophic Modified One Step M-estimator	
Outliers' percentage	Variance	Bias	MSE	Bias	MSE
0%	[1, 1]	[−0.002779, 0.1345]	[0.01956, 0.03878]	[−0.002191, 0.1345]	[0.02132, 0.04059]
	[4, 8]	[0.0004884, 0.13782]	[0.1160, 0.1410]	[0.00073, 0.1373]	[0.1272, 0.1525]
	[8,8]	[−0.007861,0.1292]	[0.1565,0.1811]	[−0.006198, 0.13036]	[0.1705, 0.1961]
5%	[1, 1]	[0.1772, 0.31901]	[0.05096, 0.12246]	[0.0975, 0.2369]	[0.03203, 0.07992]
	[4, 8]	[0.4392, 0.5875]	[0.3103,0.4687]	[0.2438, 0.3867]	[0.1938, 0.2908]
	[8,8]	[0.5013, 0.6511]	[0.4077, 0.5884]	[0.2758, 0.4195]	[0.2562, 0.3652]
10%	[1,1]	[0.2972, 0.4420]	[0.10789, 0.21605]	[0.1738, 0.3154]	[0.0538, 0.1244]
	[4, 8]	[0.7316, 0.8873]	[0.6558, 0.9140]	[0.4312, 0.5789]	[0.3264, 0.4828]
	[8,8]	[0.8407,0.9989]	[0.8631, 1.1623]	[0.4917, 0.6409]	[0.4305,0.6091]
15%	[1, 1]	[0.4772,0.6265]	[0.2473, 0.4132]	[0.3095,0.4547]	[0.1211, 0.2335]
	[4, 8]	[1.170, 1.337]	[1.4978, 1.922]	[0.761, 0.917]	[0.7324,1.0020]
	[8,8]	[1.3498, 1.5208]	[1.9783, 2.4773]	[0.8755, 1.0342]	[0.9686, 1.2822]

Table 3: Neutrosophic MSE and Bias for NSM and NMOM for $n_n = [300, 300]$.

		Estimator			
		Neutrosophic Standard Mean		Neutrosophic Modified One Step M-estimator	
Outliers' percentage	Variance	Bias	MSE	Bias	MSE
0%	[1, 1]	$[4.085e-4, 0.1378]$	$[3.389e-3, 0.02262]$	$[5.479e-4, 0.1372]$	$[3.667e-3, 0.02274]$
	[4, 8]	[0.0020, 0.1392]	[0.0590, 0.0815]	[0.0032, 0.1396]	[0.0643, 0.0872]
	[8,8]	[−0.0046, 0.1328]	[0.0782, 0.0999]	[−0.0045, 0.1323]	[0.0847, 0.1066]
5%	[1, 1]	[0.1504, 0.2915]	[0.02601,0.08864]	[0.08309, 0.2222]	[0.01075,0.05345]
	[4, 8]	[0.3673, 0.5136]	[0.1954, 0.3275]	[0.2047, 0.3466]	[0.1093, 0.1910]
	[8,8]	[0.4195, 0.5677]	[0.2542, 0.4047]	[0.2282, 0.3710]	[0.1400, 0.2300]
10%	[1,1]	[0.3004, 0.4453]	[0.0936,0.2019]	[0.1786, 0.3203]	[0.0359, 0.1069]
	[4, 8]	[0.7326, 0.8880]	[0.6013, 0.8567]	[0.4377, 0.5853]	[0.2635, 0.4183]
	[8,8]	[0.8438,1.0026]	[0.7903, 1.0877]	[0.4985, 0.6480]	[0.3405, 0.5166]
15%	[1, 1]	[0.4504, 0.5991]	[0.2062, 0.3625]	[0.2910,0.4357]	[0.0889, 0.1944]
	[4, 8]	[1.0979, 1.2625]	[1.2766, 1.6690]	[0.7089, 0.8631]	[0.5815, 0.8284]
	[8,8]	[1.2681, 1.4375]	[1.6863, 2.1489]	[0.8135, 0.9713]	[0.7592, 1.0459]

From Table 1 - 3, the Neutrosophic Standard Mean (NSM) estimator and the Neutrosophic Modified One-Step (NMOM) estimator reveals that the $NMOM$ estimator exhibits greater robustness, particularly in the presence of outliers. As the percentage of outliers increases, both estimators experience an increase in bias and MSE. However, $NMOM$ consistently maintains a lower bias than NSM, indicating that it is less influenced by the presence of extreme values. Additionally, $NMOM$ shows lower MSE across all variance levels, further demonstrating its stability. When variance increases from [1,1] to [8,8], both estimators become less efficient, but NSM deteriorates more significantly compared to $NMOM$. This trend suggests that $NMOM$ remains more reliable across different variance conditions. At higher outlier percentages, such as 15%, the difference becomes even more pronounced, with NSM showing substantially higher MSE than $NMOM$, highlighting its susceptibility to contamination. Furthermore, for larger sample sizes, as shown in the Table 5 for $n_N = [300, 300]$, the $NMOM$ estimator maintains its advantage in both bias and MSE, reinforcing its robustness. At 15% outliers with high variance ([8,8]), NSM has a significantly higher bias and MSE compared to $NMOM$, demonstrating that NSM is more prone to error when the data is contaminated. This further confirms $NMOM's$ superior performance, especially when dealing with large-scale datasets where outliers may be present. Overall, these findings suggest that the $NMOM$ estimator is the superior choice for datasets containing outliers, as it offers greater accuracy and resilience against data contamination.

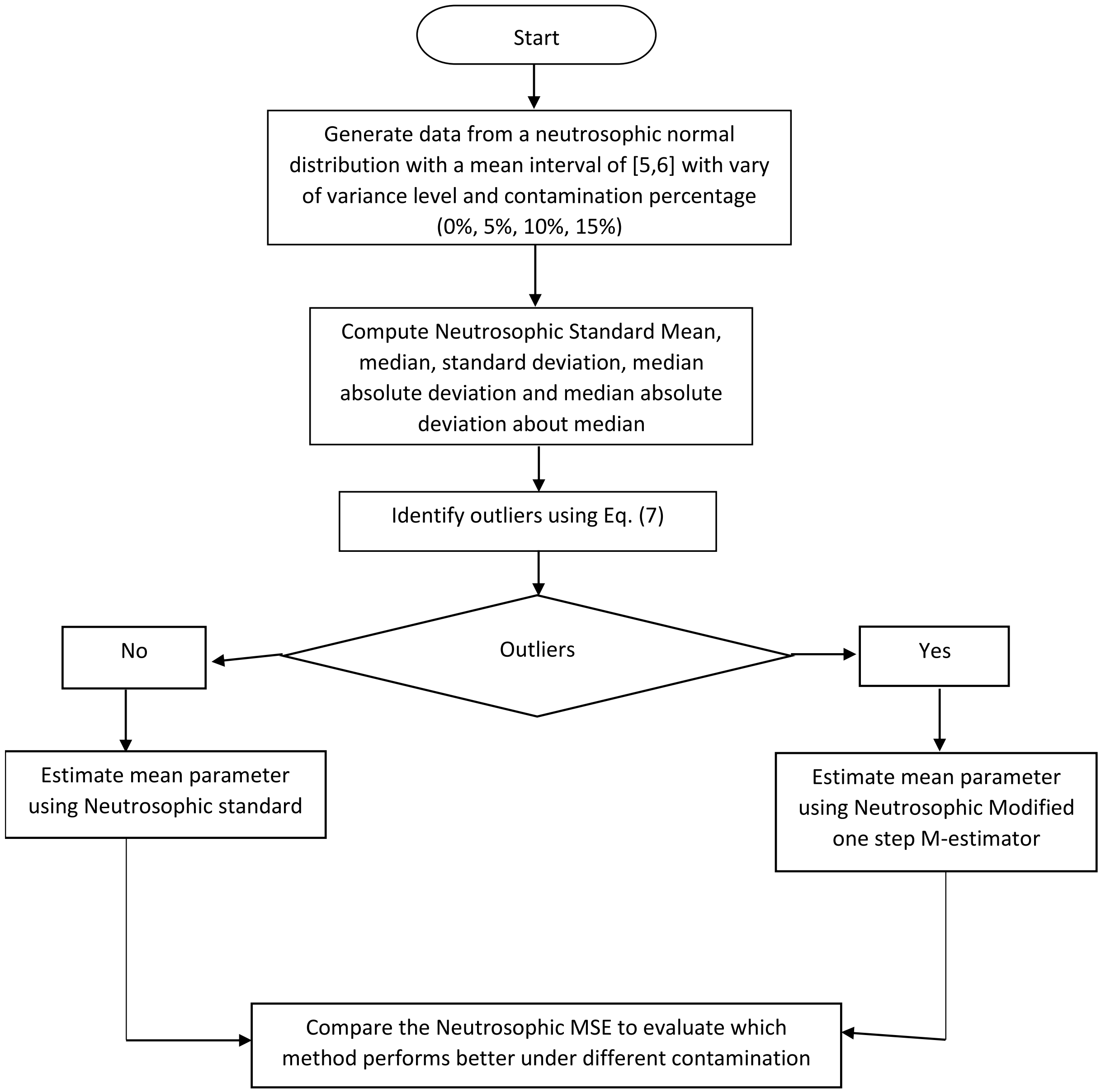

Figure 1. Process of the Neutrosophic Modified One-Step M Estimator Under Contamination

4. Real-world Data

The application of the Neutrosophic Modified One-step M-estimator (NMOM) is demonstrated through a comparative analysis of two distinct datasets, each presenting unique characteristics that make them valuable for this statistical approach. The first dataset, originally documented by [22], consists of COVID-19 mortality rates from the Netherlands spanning 30 days (March 31 to April 30, 2020), characterized by rough or imprecise measurements that inherently contain degrees of uncertainty. The second dataset, sourced from [23], comprises 50 observations of thickness measurements with holes and sheets, representing manufacturing quality control data with its own form of indeterminacy. These datasets are particularly important for this application because they represent different domains (public health and manufacturing), contain different forms of uncertainty, feature varying sample sizes (30 versus 50 observations), and have been previously analyzed in published literature—enabling meaningful comparisons with established results. By applying NMOM to these diverse real-world datasets rather than simulated data, researchers can effectively demonstrate the method's versatility and robustness in handling outliers and indeterminacy across different fields, thereby establishing its practical value as a statistical tool for analyzing neutrosophic data in situations where traditional methods might be inadequate due to data imprecision or contamination. The two data sets in neutrosophic form is given in Table 4 below.

Table 4: Two set of neutrosophic data

Data Set 1	Data Set 2
(14.918,15.66390), (10.056,11.18880), (12.274,12.88770), (10.289,10.80345), (10.832,11.37360), (7.099,7.45395), (5.928, 6.22440), (13.211,13.87155), (7.968,8.36640), (7.584,7.96320), (5.555,5.83275), (6.027,6.32835), (4.097,4.30185), (3.611,3.79155), (4.960,5.20800), (7.498,7.87290), (6.940,7.28700), (5.307,5.57235), (5.048,5.30040), (2.857,2.99985), (2.254,2.36670), (5.431,5.70255), (4.462,4.68510), (3.883,4.07715), (3.461,3.63405), (3.647,3.82935), (1.974,2.07270), (1.273,1.33665), (1.416,1.48680), (4.235,4.44675).	(0.03,0.05), (0.01,0.03), (0.05,0.07), (0.11,0.13), (0.13,0.15), (0.09,0.085), (0.21,0.223), (0.11,0.121), (0.07,0.082), (0.25,0.262), (0.23,0.243), (0.03,0.043), (0.13,0.142), (0.15,0.162), (0.07,0.083), (0.25,0.264), (0.322,0.34), (0.27,0.284), (0.13,0.145), (0.15,0.164), (0.23,0.243), (0.21,0.224), (0.10,0.122), (0.10,0.184), (0.20,0.242), (0.30,0.323), (0.12,0.164), (0.12,0.142), (0.05,0.082), (0.11,0.164), (0.20,0.245), (0.10,0.163), (0.11,0.324), (0.10,0.182), (0.20,0.243), (0.20,0.222), (0.10,0.163), (0.11,0.123), (0.20,0.242), (0.05,0.063), (0.01,0.024), (0.17,0.1840), (0.21,0.225), (0.11,0.143), (0.05,0.062), (0.03,0.042), (0.12,0.143), (0.25,0.263), (0.15,0.182), (0.10,0.162).

5. Results and Discussion

From Table 4, the Neutrosophic Modified One Step M-estimator demonstrates greater robustness and accuracy compared to the Neutrosophic Standard Mean. It yields a lower mean and median, indicating that it is less influenced by extreme values. Additionally, its lower standard deviation and covariance suggest that it is more stable and exhibits less variability. A significant difference is observed in the detection of outliers, where the modified estimator identifies seven outliers at lower and upper data, while the standard mean detects none.

Furthermore, the Mean Squared Error (MSE) for the modified estimator is significantly lower, highlighting its improved accuracy. Overall, the Neutrosophic Modified One Step M-estimator is a more reliable approach, as it effectively minimizes the impact of outliers and provides more precise estimates.

Table 5: The descriptive summary of data set 1 with sample size [30,30]

	Neutrosophic Standard Mean	Neutrosophic Modified One Step M-estimator
Mean	[6.14, 6.46]	[4.84, 5.08]
Median	[5.37, 5.64]	[4.96, 5.21]
Standard deviation	[3.45, 7.10]	[1.81, 3.70]
Covariance	[1.94, 7.79]	[0.37, 0.73]
No of outliers	[0,0]	[7,7]
MSE	[11.9, 13.7]	[3.26, 4.09]

From table 5, while both methods yield similar mean and median values, the NMOM exhibits a lower standard deviation [0.0670,0.134] compared to the NSM [0.0766,0.156] indicating greater stability. Additionally, the covariance is slightly lower for the NMOM, suggesting reduced variability in relationships between variables. A significant distinction is in the detection of outliers, where the NSM detects none, whereas the NMOM estimator identifies three at lower data, making it more sensitive to extreme values. In terms of accuracy, the NMOM has a slightly higher lower bound for MSE [0.00678,0.00794] compared to the NSM [0.00586,0.00967] but it remains within a similar range.

Table 6: The descriptive summary of data set 2 with sample size [50,50]

	Neutrosophic Standard Mean	Neutrosophic Modified One Step M-estimator
Mean	[0.138, 0.167]	[0.128, 0.167]
Median	[0.12, 0.163]	[0.11, 0.163]
Standard deviation	[0.0766, 0.156]	[0.0670, 0.134]
Covariance	[0.555, 0.935]	[0.524, 0.801]
No of outliers	[0,0]	[3,0]
MSE	[0.00586, 0.00967]	[0.00678, 0.00794]

In addition, by comparing the Neutrosophic Standard Mean and the Neutrosophic Modified One Step M-estimator across different sample sizes highlights the robustness of the modified estimator. In both cases, the NMOM estimator shows lower variability, as indicated by the reduced standard deviation and covariance values. This suggests that it provides more stable estimates compared to the standard mean. Additionally, the NMOM estimator detects more outliers, reinforcing its ability to account for anomalies in the dataset. Although the Mean Squared Error for the NMOM estimator is slightly higher in the larger sample size, it remains relatively low, ensuring reliable performance. Overall, the findings confirm that the Neutrosophic Modified One Step M-estimator is a more effective approach for handling uncertainty and outliers, leading to more precise and robust statistical estimates.

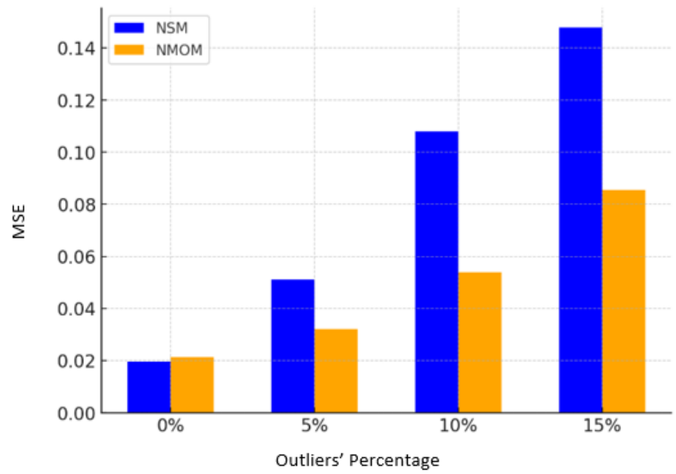

Figure 2: MSE values of NSM and NMOM across different outlier percentages with variance [1,1]

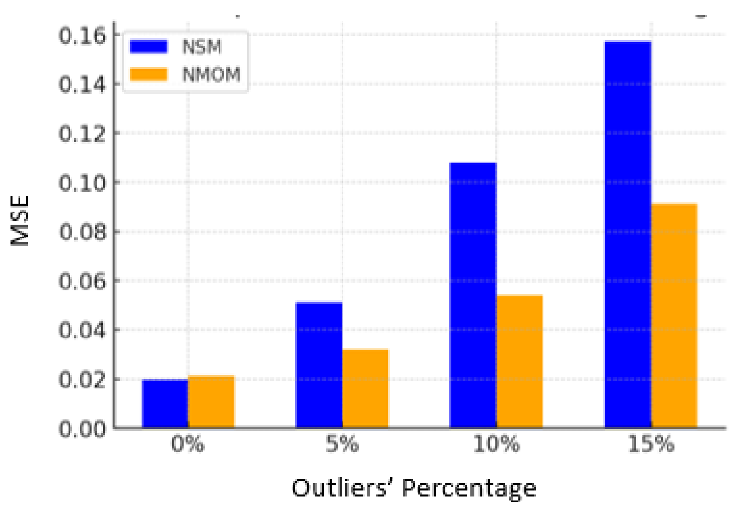

Figure 3: MSE values of NSM and NMOM across different outlier percentages with variance [4,8]

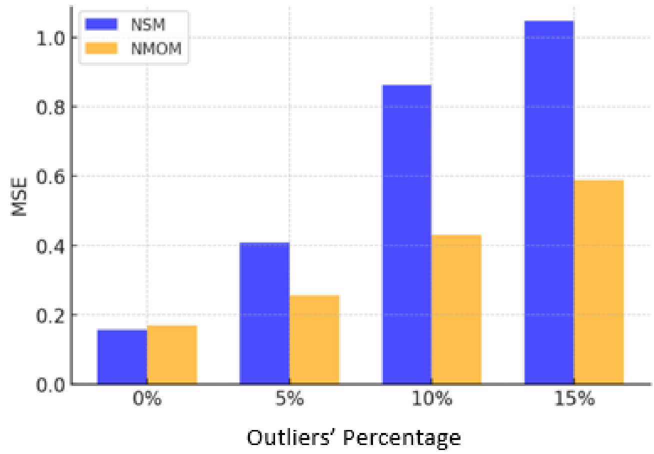

Figure 4: MSE values of NSM and NMOM across different outlier percentages with variance [8,8]

Figures 1 to 3 depict the MSE values for NSM and NMOM across different variance levels—[1,1], [4,8], and [8,8]—under varying outlier percentages (0%, 5%, 10%, and 15%). The influence of outliers is evident in all three bar charts, as the MSE values for both estimators increase with a higher percentage of outliers, indicating reduced estimation accuracy. However, NMOM consistently demonstrates lower MSE values compared to NSM across all variance levels, highlighting its robustness against outliers. These results clearly show that NMOM is a more effective estimator, especially in scenarios with high variance and a significant presence of outliers. Its ability to maintain lower MSE values across different conditions makes it a more reliable choice for statistical estimation in real-world applications where uncertainty and outliers are prevalent.

6. Conclusions

The findings from both empirical studies and simulation highlight the advantages of the Neutrosophic Modified One-Step M-estimator (NMOM) over the Neutrosophic Standard Mean (NSM), particularly in handling data variability and robustness against outliers. The presence of outliers in NMOM suggests that this method is more sensitive to detecting anomalies, whereas NSM potentially leading to biased estimations. Moreover, the lower standard deviation observed in NMOM confirms its ability to provide more stable and consistent results. This aligns with the expectations from simulation studies, which emphasize the efficiency of modified M-estimators in reducing the influence of extreme values. The covariance values further support this, as NMOM exhibits a more controlled spread, ensuring better reliability in data-driven decision-making. Additionally, the Mean Squared Error (MSE) comparison demonstrates that NMOM maintains a similar level of accuracy to NSM while offering greater robustness. This is particularly crucial in fields requiring precise and reliable estimation methods, such as finance, medical research, and engineering applications. In conclusion, the empirical and simulation-based findings confirm that NMOM is a superior estimator for datasets containing noise or outliers. Its ability to enhance stability, reduce variability, and improve data accuracy makes it a more effective alternative to traditional methods, reinforcing its potential for application in complex real-world scenarios. Therefore, this study may be extended in the near future to include other types of mean estimators

Funding: The study was funded by Researchers Supporting Project number TAP K011487, Universiti Kebangsaan Malaysia.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] D.P. Filev, Klir and Yuan's Fuzzy Sets and Fuzzy Logic, J. Intell. Fuzzy Syst. 4 (1996), 175-176. https://doi.org/10.3233/ifs-1996-4208.

[2] F. Smarandache, Definiton of Neutrosophic Logic - A Generalization of the Intuitionistic Fuzzy Logic, in: 3rd Conference of the European Society for Fuzzy Logic and Technology, Zittau, Germany, pp. 141-146, 2003.

[3] M.B. Anwar, M. Hanif, U. Shahzad, W. Emam, M.M. Anas, N. Ali, S. Shahzadi, Incorporating the Neutrosophic Framework into Kernel Regression for Predictive Mean Estimation, Heliyon 10 (2024), e25471. https://doi.org/10.1016/j.heliyon.2024.e25471.

[4] H. Abbasi, U. Shahzad, W. Emam, M. Hanif, N. Ali, M. Mukhtar, Calibrated Empirical Neutrosophic Cumulative Distribution Function Estimation for Both Symmetric and Asymmetric Data, Symmetry 16 (2024), 633. https://doi.org/10.3390/sym16050633.

[5] H. Abbasi, U. Shahzad, W. Emam, M. Hanif, N. Ali, M. Mukhtar, Calibrated Empirical Neutrosophic Cumulative Distribution Function Estimation for Both Symmetric and Asymmetric Data, Symmetry 16 (2024), 633. https://doi.org/10.3390/sym16050633.

[6] R. Singh, A. Kumari, Neutrosophic Ranked Set Sampling Scheme for Estimating Population Mean: An Application to Demographic Data, Neutrosophic Sets Syst. 68 (2024), 246–270. https://doi.org/10.5281/zenodo.11479519.

[7] R. Singh, S.N. Tiwari, Improved Estimator for Population Mean Utilizing Known Medians of Two Auxiliary Variables Under Neutrosophic Framework, Neutrosophic Syst. Appl. 25 (2025), 38-52. https://doi.org/10.61356/j.nswa.2025.25443.

[8] Z. Tahir, H. Khan, M. Aslam, J. Shabbir, Y. Mahmood, F. Smarandache, Neutrosophic Ratio-Type Estimators for Estimating the Population Mean, Complex Intell. Syst. 7 (2021), 2991-3001. https://doi.org/10.1007/s40747-021-00439-1.

[9] P. Poonam, P. Singh, P. Sharma, B. Aloraini, Estimation of Population Mean Using Neutrosophic Exponential Estimators with Application to Real Data, Int. J. Neutrosophic Sci. 25 (2025), 322-338. https://doi.org/10.54216/ijns.250329.

[10] A. Kumari, R. Singh, F. Smarandache, New Modification of Ranked Set Sampling for Estimating Population Mean: Neutrosophic Median Ranked Set Sampling with an Application to Demographic Data, Int. J. Comput. Intell. Syst. 17 (2024), 210. https://doi.org/10.1007/s44196-024-00548-y.

[11] M. Aslam, Introducing Grubbs's Test for Detecting Outliers Under Neutrosophic Statistics – An Application to Medical Data, J. King Saud Univ. - Sci. 32 (2020), 2696-2700. https://doi.org/10.1016/j.jksus.2020.06.003.

[12] C. Leys, C. Ley, O. Klein, P. Bernard, L. Licata, Detecting Outliers: Do Not Use Standard Deviation Around the Mean, Use Absolute Deviation Around the Median, J. Exp. Soc. Psychol. 49 (2013), 764-766. https://doi.org/10.1016/j.jesp.2013.03.013.

[13] R.R. Wilcox, H.J. Keselman, Modern Robust Data Analysis Methods: Measures of Central Tendency, Psychol. Methods 8 (2003), 254-274. https://doi.org/10.1037/1082-989x.8.3.254.

[14] O. Owolabi, D. Okoh, B. Rabiu, A. Obafaye, K. Dauda, A Median Absolute Deviation-Neural Network (MAD-NN) Method for Atmospheric Temperature Data Cleaning, MethodsX 8 (2021), 101533. https://doi.org/10.1016/j.mex.2021.101533.

[15] S. Mazumder, R. Serfling, Bahadur Representations for the Median Absolute Deviation and Its Modifications, Stat. Probab. Lett. 79 (2009), 1774-1783. https://doi.org/10.1016/j.spl.2009.05.006.

[16] D.C. Howell, Median Absolute Deviation, Encycl. Stat. Behav. Sci. (2005). https://doi.org/10.1002/0470013192.bsa384.

[17] P.J. Huber, Minimax Aspects of Bounded-Influence Regression, J. Am. Stat. Assoc. 78 (1983), 66-72. https://doi.org/10.1080/01621459.1983.10477928.

[18] A.R. Othman, H. Keselman, A.R. Padmanabhan, R.R. Wilcox, K. Fradette, Comparing Measures of the 'Typical' Score Across Treatment Groups, Br. J. Math. Stat. Psychol. 57 (2004), 215-234. https://doi.org/10.1348/0007110042307159.

[19] S.S.S. Abd Mutalib, S.Z. Satari, W.N.S. Wan Yusoff, A New Robust Estimator to Detect Outliers for Multivariate Data, J. Phys.: Conf. Ser. 1366 (2019), 012104. https://doi.org/10.1088/1742-6596/1366/1/012104.

[20] J. Chen, J. Ye, S. Du, Scale Effect and Anisotropy Analyzed for Neutrosophic Numbers of Rock Joint Roughness Coefficient Based on Neutrosophic Statistics, Symmetry 9 (2017), 208. https://doi.org/10.3390/sym9100208.

[21] R.R. Wilcox, H.J. Keselman, Modern Robust Data Analysis Methods: Measures of Central Tendency., Psychol. Methods 8 (2003), 254-274. https://doi.org/10.1037/1082-989x.8.3.254.

[22] H.M. Almongy, E.M. Almetwally, H.M. Aljohani, A.S. Alghamdi, E. Hafez, A New Extended Rayleigh Distribution with Applications of Covid-19 Data, Results Phys. 23 (2021), 104012. https://doi.org/10.1016/j.rinp.2021.104012.

[23] Z. Mi, S. Hussain, C. Chesneau, On a Special Weighted Version of the Odd Weibull-Generated Class of Distributions, Math. Comput. Appl. 26 (2021), 62. https://doi.org/10.3390/mca26030062.