NSS
Neutrosophic Sets and Systems, Vol. 78, 2025
University of New Mexico

Enhancing Neutrosophic Data Analysis: A Review of Neutrosophic Measures and Applications with Neutrostat

Zahid Khan[1,*] and Katrina Lane Krebs[2]
[1]Department of Quantitative Methods, University of Pannonia, Egyetem str. 10, Veszprem, H-8200, Hungary.; zahidkhan@hu.edu.pk
[2]SNM Social Sciences, Central Queensland University, 160 Ann St, Brisbane, QLD 4000, QLD, Australia; k.lane-krebs@cqu.edu.au
[*]Correspondence: zahidkhan@hu.edu.pk

Abstract. Neutrosophic statistical measures analyze data that is not fully determined, often due to imprecise observations. This type of data presents a major concern in neutrosophic statistics. The existing literature categorizes neutrosophic measures into two types: descriptive and inferential, aiming to broaden the categorization of statistics in these two major areas. Every statistical measure can be contextualized within the neutrosophic framework by acknowledging the inherent imprecision, vagueness, or not fully defined. In this study, the major focus is on reviewing existing neutrosophic measures rather than proposing new ones. The aim is to enhance current neutrosophic structures to make them more beneficial for end users in their analysis. Additionally, one of the contemporary challenges in neutrosophic data analysis is the dimension of data. We develop the R library neutrostat to efficiently describe complex and larger imprecise datasets. Finally, real-world examples are provided to review the effectiveness of the neutrostat package for analyzing neutrosophic data and evaluating existing neutrosophic measures.

Keywords: Descriptive measures; inferential measures; neutrosophic measures; neutrosophic probability; R package

1. Introduction

In data analysis, there are various statistical methods available for analyzing and interpreting data in specific situations [1]. It is important to understand the assumptions and conditions of these methods to choose the most appropriate option for data analysis [2]. The two main types of statistical methods used are descriptive statistics, which summarize data with measures including presentation of data using graphs, measuring central tendency and dispersion, and inferential statistics, which make conclusions based on sample data using tests like the z-test and t-test or constructing confidence intervals [3, 4]. Choosing the appropriate statistical method is crucial and depends on three key factors: the purpose and goals of the research, the type and distribution of data, and the characteristics of the observations being analyzed [5]. Choosing the right statistical method is crucial when analyzing real-world data [6]. Selecting an inappropriate method can lead to issues in interpreting findings and impact the overall validity and conclusion of the study [7]. It is essential to understand the assumptions and conditions of different statistical methods to select the most correct option for data analysis [8]. Many researchers make avoidable mistakes in statistical analysis by using the wrong approach, such as using a z-test instead of a t-test or chi-square test instead of correlation analysis, using a normal distribution test method for skewed data, applying methods for correlated data but they actually developed under independent identically distribution scenarios—these are but a few common examples [9]. Many statistical software packages, such as SPSS, R, Stata, SAS, S-Plus, MATLAB, etc., are available in the market, which makes analysis easier for high-dimensional data. While all computational tools help with analysis, they do not assist in selecting the appropriate method [10, 11]. Among research communities, R software appears as the most popular compared to other commercial statistical software because it is free and open source. The R software is flexible, easy to use, and allows users to create their own functions or packages for specific tasks [12]. Despite advancements in statistical analysis tools, choosing the right test remains challenging, especially for those without a statistical background. Aside from understanding statistical methods, it is crucial to consider the nature and type of data collected and the study's objective. The statistical methods chosen should align with the study's goal and be suitable for the data at hand.

When all data or some part of it involves imprecise, ambiguous, or incomplete portions, here the methods of neutrosophic statistics comes into play [13, 14, 15]. Neutrosophic statistics provides a new foundation for dealing with real-world scenarios that have indeterminate information. Fundamental concepts of neutrosophy were introduced by Smarandache in 1995 as a generalization for fuzzy logic and intuitionistic fuzzy logic [16]. Expanding traditional statistical measures with neutrosophic logic allows for parameters to have undetermined values, enabling better handling of situations involving imprecise or clear statistical data, particularly when working with vague and imprecise statistical information [17, 18]. According to the concept of neutrosophic logic, crisp values can be represented by a neutrosophic structure with determinate and indeterminate parts. When the indeterminate part of a number becomes zero, it converts to a classical structure [19, 20]. Thus, neutrosophic statistical methods have an advantage over classical methods due to their generalization with the extra benefit of analyzing neutrosophic data. Neutrosophic logic has been utilized to transform various statistical measures to accommodate imprecise datasets[21, 22, 23]. Some descriptive measures have been thoroughly explored in [24], with many methods still pending further examination. For inferential statistics in the domain of neutrosophic logic, significant contributions can be found in the research conducted by Smarandache [13]. Many statistical distributions that reflect the mathematical data generated models beyond any data source have also been converted into a neutrosophic framework [25, 26, 27].

In this paper, we propose enhancements to the descriptive aspects of neutrosophic structures. We aim to rectify inconsistencies encountered in published literature, where the principles of neutrosophic structures are inaccurately utilized. Our objective is not to critique existing studies but to point out notable discrepancies that require consideration for the successful application of neutrosophic measures in practical situations. Additionally, an orientation of some R functions, which are part of our ongoing project neutrostat package, is also presented. In real-world scenarios, dimensions can become quite large, and this R tool can simplify the calculations. It is not feasible to discuss every neutrosophic measure due to the limited source of this work; however, we suggest that many existing works could benefit from refinement and proper implementation of the neutrosophic framework discussed in this work.

This paper is structured as follows: In Section 2, some neutrosophic data and their algebraic operations are presented. In Section 3, some neutrosophic descriptive measures with major discrepancies are discussed. In Section 4, the neutrostat package with some key functions is described. Finally, the major findings of the work are concluded in Section 5.

2. Neutrosophic Data

Neutrosophic data is a collection of information where some or all of the data is uncertain to a certain extent. This neutrosophic data can be in the form of discrete and continuous. For example, the following data represent the neutrosophic form of data: $2, 3, 5, 6, 3 + I_1, 8, 7 + I_2$ where I_1 and I_2 represent the indeterminate parts of two numbers 3 and 7 with any numerical values say $I_1 = [4, 6]$ and $I_2 = [3, 5]$. Due to these indeterminate parts, we do not know the exact value of some data observations. It could be 7 and 9 if we consider I_1 or it can be 10 and 12 if we consider I_2. So, in a general structure, neutrosophic number can be expressed as:

$$N = D + I \tag{1}$$

where D represents the determinate part of the number and I represents the indeterminate part of the number. It is obvious from (1) that when indeterminate becomes zero, the number turns to a crisp number. In this way, the neutrosophic number has a more generated structure as compared to the crisp numbers. When the data is comprised of neutrosophic numbers, the operations of arithmetic operations have specific operations for interval analysis. To understand this interval analysis let us assume that $Z_1 = [\underline{Z}_1, \overline{Z}_1]$ and $Z_2 = [\underline{Z}_2, \overline{Z}_2]$ be the two closed bounded neutrosophic numbers with operation of addition, subtraction, multiplication and division denoted by star (*) then:

$$\left[\underline{Z}_1, \overline{Z}_1\right] * \left[\underline{Z}_2, \overline{Z}_2\right] = [\alpha, \beta] \tag{2}$$

where $[\alpha, \beta] = \left\{Z_1 * Z_2 \mid \underline{Z}_1 < Z_1 < \overline{Z}_1, \underline{Z}_2 < Z_2 < \overline{Z}_2\right\}$.
Equation (2) can be further simplified as:

$$\left[\underline{Z}_1, \overline{Z}_1\right] + \left[\underline{Z}_2, \overline{Z}_2\right] = \left[\underline{Z}_1 + \underline{Z}_2, \overline{Z}_1 + \overline{Z}_2\right] \tag{3}$$

$$\left[\underline{Z}_1, \bar{Z}_1\right] - \left[\underline{Z}_2, \bar{Z}_2\right] = \left[\underline{Z}_1 - \bar{Z}_2, \overline{Z}_1 - \underline{Z}_2\right] \tag{4}$$

$$\frac{\left[\underline{Z}_1, \overline{Z}_1\right]}{\left[\underline{Z}_2, \overline{Z}_2\right]} = \left[\underline{Z}_1, \overline{Z}_1\right] * \left[\frac{1}{\underline{Z}_2}, \frac{1}{\overline{Z}_2}\right] \tag{5}$$

Equation (5) assumes that zero does not belong to $\left[\bar{Z}_2, \bar{Z}_2\right]$.

$$[\underline{Z}_1, \overline{Z}_1] \times [\underline{Z}_2, \overline{Z}_2] = [\alpha, \beta] \tag{6}$$

where $\alpha = \min\left\{\underline{Z}_1\underline{Z}_2, \underline{Z}_1\overline{Z}_2, \overline{Z}_1\underline{Z}_2, \overline{Z}_1\overline{Z}_2\right\}$ and $\beta = \max\left\{\underline{Z}_1\underline{Z}_2, \underline{Z}_1\overline{Z}_2, \overline{Z}_1\underline{Z}_2, \overline{Z}_1\overline{Z}_2\right\}$. Multiplication and division expressions can be simplified further if we know that $\underline{Z}_1 > 0, \quad \overline{Z}_2 < 0, \quad \underline{Z}_1 < 0 \quad \text{or} \quad \overline{Z}_2 < 0$.

To understand the neutrosophic numbers and their computation procedure, let us take a real-world example. A metrologist recorded the temperature in Veszprem, Hungary, over 10 days. The data may have some variability due to different measuring instruments and missing data. Some temperatures are reported as ranges, either due to estimates or less precise measurements, while others are precise readings from reliable instruments. Here is the recorded temperature data for the 10 days:

22, [18, 20], 25, [21, 23], 19, 26, [24, 27], 23, [22, 24], [20, 22]

Using the definitions of neutrosophic mean, median, and standard deviation [14], we can find the summary measures of temperature as:

Neutrosophic mean= [22, 23.1]

Neutrosophic median= [22, 23]

Neutrosophic standard deviation= [1.73, 3.36]

Note that we have correctly followed the principles defined for neutrosophic numbers and can compute these basic measures using some defined functions of our neutrostat R package in just a few seconds. This package can handle large and complex neutrosophic data efficiently. To ensure accuracy in statistical measures developed using neutrosophic theory, it is essential to verify the fundamental operations and their extensions defined in neutrosophic theory. Unfortunately, a critical review of various papers in this field has revealed violations of these principles in computation. Consequently, the research community in applied sciences is unable to benefit from the newly proposed neutrosophic statistical measures. In the following section, we have addressed some discrepancies found in neutrosophic measures to make them more user-friendly and practical for end users.

3. Neutrosophic Measures

In this section, we discuss some previously published studies related to neutrosophic statistical measures, highlighting that their computations do not adhere to the properly defined algebraic operations for neutrosophic data. In the study [28], authors analyzed the neutrosophic coefficient variation for utilizing the temperature data on different cities of the southeastern Anatolia region of Turkey. The neutrosophic coefficient of variation is defined as:

$$\mathrm{CV}_n = \frac{\text{Neutrosophic mean}}{\text{Neutrosophic standard deviation}} \tag{7}$$

The computed neutrosophic coefficient of variation reported in the source [28] is given in Table 1.

Table 1. Coefficient of variation reported for temperature data reported for different cities of Turkey

City	Neutrosophic mean	Neutrosophic standard deviation	CV_n
Adiyaman	[0.66, 7.47]	[2.04, 3.06]	[40.96, 309.06]
Batman	[1.33, 8.23]	[1.54, 2.56]	[31.10, 115.79]
Gaziantep	[0.80, 8.57]	[3.13, 7.08]	[82.61, 391.25]
Killis	[3.80, 10.57]	[2.14, 4.49]	[42.48, 56.31]
Mardin	[3.61, 11.23]	[1.83, 3.77]	[33.57, 50.69]

Now computed values of CV_n using the neutrostat R package for the same data are shown in Table 2.

Table 2. Neutrosophic coefficient of variation using neutrostat R package

City	Adiyaman	Batman	Gaziantep	Killis	Mardin
CV_n	[27.29, 463.65]	[18.71, 192.2]	[36.51, 885]	[20.24, 118.19]	[16.28, 104.42]

The results in Table 2 show significant discrepancies between the accurate results and the presented results from the original source [28]. Note that the neutrosophic coefficient of variation has been calculated using the neutrostat package, assuming that the neutrosophic mean and standard deviation were accurately reported. Incorrect computation of the neutrosophic coefficient of variation can have serious implications for decision-making and statistical analysis. This can mislead researchers or stakeholders into drawing incorrect conclusions about the consistency or volatility of the data. A question may arise is how can we trust that the neutrostat package provides accurate information? To address this, we implement some relevant functions of the neutrostat package on the example data provided by Smarandache [29]. The data consists of the values: 30, 6, [2, 5], [18, 24].

Now, let us see how we can input this data into the "nmean" and "nstd" functions of the R library "neutrostat":

Input:

$x < -list(30, 6, c(2, 5), c(18, 24))$

We can run the following lines in the R workspace to obtain the output:

"nmean(x)" and "nstd(x)"

Output:

Neutro mean = [14, 16.25] and Neutro standard deviation = [9.18, 12.88]. The output matches the values reported by Smarandache [29] Additionally, the neutrostat package offers the advantage of providing the summary function "nsummary(data)", which performs a more comprehensive analysis of the underlying data. Another study conducted by Shahzadi [30] on the computation of neutrosophic coefficient of determination for temperature data to measure the consistency of temperature across different cities of Pakistan.

TABLE 3. The CV_n values for temperature data for different cities of PaKistan

City	Gujranwala	Lahore	Karachi	Islamabad	Sialkot
μ_n	[29.68, 37.78]	[30.96, 39.07]	[28.54, 32.00]	[27.43, 35.42]	[27.75, 35.32]

However, we believe that there are major discrepancies in the results reported for neutrosophic standard deviation and reported values are not according to rules designed for interval arithmetic. These discrepancies result in neutrosophic standard deviations, leading to wrong statistics reported as evident in Table 3. The correct reported using the R package can be easily obtained using different built-in functions such as "nmean", "nstd" and "ncv" for reporting neutrosophic mean, standard deviation, and coefficient of variation, respectively, for temperature data. Correct statistics using the neutrsostat package are reported in Tables 4-6.

TABLE 4. Neutrosophic mean of temperature data across different cities in Pakistan

City	Gujranwala	Lahore	Karachi	Islamabad	Sialkot
μ_n	[29.68, 37.78]	[30.96, 39.07]	[28.54, 32.00]	[27.43, 35.42]	[27.75, 35.32]

Table 5. Neutrosophic variability of temperature data across different cities in Pakistan

City	Gujranwala	Lahore	Karachi	Islamabad	Sialkot
σ_n	[7.89, 9.83]	[7.82, 10.11]	[3.41, 4.38]	[7.91, 9.24]	[7.23, 9.61]

Table 6. Neutrosophic coefficient of variation of temperature data across different cities in Pakistan

City	Gujranwala	Lahore	Karachi	Islamabad	Sialkot
CV_n	[20.89, 33.14]	[20.00, 32.65]	[10.66, 15.37]	[22.33, 33.67]	[20.48, 34.61]

Table 5 displays notable differences between the precise findings and the reported data in the source material [30]. These discrepancies are common in various studies on neutrosophic measures, highlighting the importance of further research to ensure that end users can effectively utilize the innovative approach to neutrosophic statistics. Results obtained using the neutrostat package are more profound and reflective.

4. **Overview of neutrostat**

In this section, a brief description of neutrostat package with its basic functionalities is discussed. The neutrostat is an R software package with functions developed following the principles of neutrosophic statistics originated by Smarandache [13]. The neutrostat package works as a computational tool to analyze data based on vague and imprecise information. The latest release of version 0.0.1 of the package is available on CRAN at https://CRAN.R-project.org/package=neutrostat and the developmental version is hosted on GitHub at https://github.com/kzst/neutrostat. The initial version of the package is equipped with various descriptive functions for neutrosophic numbers, where indeterminacy, inconsistency, and uncertainty are inherent parts of the data. Along with a basic functions package, it also includes some real-world neutrosophic data such as gold prices across Indian cities (goldprice), dioxin ingestion in Japan (dioxin) and temperature data from different cities of Pakistan (citytemp). Additionally, some interval operators, such as addition, subtraction, multiplication, and division for interval arithmetic commonly used in interval calculations of neutrosophic numbers, are built into the package. These operators not only allow the computational procedures of many functions but also facilitate the addition features, such as the conversion of data to neutrosophic form or sorting of data in ascending or descending order. The summary of some key functions of the neutrostat package is outlined below:

Table 7. Key functions of the neutrostat package with short descriptions

Function	Description	Usage
interval_df	Converting data into neutrosophic numbers	interval_df(data)
interval_add	Return addition of neutrosophic numbers	interval_add(data)
interval_sub	Return subtraction of neutrosophic numbers	interval_sub(data)
interval_mul	Return multiplication of neutrosophic numbers	interval_mul(data)
interval_div	Return division of neutrosophic numbers	interval_div(data)
sort_interval	Return sorting of neutrosophic numbers	sort_interval(data)
nmean	Return mean of neutrosophic numbers	nmean(data)
nvar	Return variance of neutrosophic numbers	nvar(data)
nstd	Return standard deviation of neutrosophic numbers	nstd(data)
nmedian	Return median of neutrosophic numbers	nmedian(data)
nquant	Return quantiles of neutrosophic numbers	nquant(data)
nsk	Return Pearson coefficient of skewness of neutrosophic numbers	nsk(data)
nkur	Return kurtosis coefficient of neutrosophic numbers	nkur(data)
nsummary	Return basic summary of neutrosophic numbers	nsummary(data)

Table 7 indicates that the first version (0.0.1) of the neutrostat package is equipped with the necessary fundamental functions required for an initial glance at data analysis. Each function listed returns an interval value, indicating imprecision in data, making it ideal for real-world scenarios where data uncertainty commonly exist. Additionally, the functions listed in Table 7 are user-friendly. A comprehensive description of each function, and applied examples, can be found in the reference manual of the neutrostat package, which is available on CRAN for public use.

5. Conclusions

This research thoroughly examined various neutrosophic statistical measures and classified them into descriptive and inferential categories. Our findings emphasize the significance of integrating the neutrosophic approach in statistical analyses to accommodate imprecise, vague, or undefined data. By utilizing this framework, we can better represent and analyze such complex data commonly encountered in practical situations. Through this review, we have pinpointed both the advantages and drawbacks of existing neutrosophic metrics and proposed methods to improve their functionality for users. Additionally, we have underscored the importance of utilizing R software for managing intricate datasets with imprecise values. We have developed the neutrostat R package to make the neutrosophic data analysis easier for public use with its key functions outlined in this work. With neutrostat library, individuals can effectively conduct neutrosophic data analysis, even in situations where uncertainties and indeterminacies are present in high-dimension neutrosophic information.

As neutrosophic statistical methods continue to evolve, extending the neutrostat library to incorporate these advancements will be a key focus of future work.

Acknowledgments: This work has been implemented by the TKP2021-NVA-10 project with support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the 2021 Thematic Excellence Programme funding scheme.

Conflicts of Interest: The authors declare no conflict of interest.

Data availbility

The data used in this work is available within the manuscript.

References

1. Burr, I.W. Applied Statistical Methods; Elsevier, 2014.
2. Bastardoz, N.; Matthews, M.J.; Sajons, G.B.; Ransom, T.; Kelemen, T.K.; Matthews, S.H. Instrumental variables estimation: Assumptions, pitfalls, and guidelines. *Leadersh. Q.* **2023**, *34*, 101673.
3. Shatz, I. Assumption-checking rather than (just) testing: The importance of visualization and effect size in statistical diagnostics. *Behav. Res. Methods* **2024**, *56*, 826–845.
4. Casella, G.; Berger, R. Statistical Inference; CRC Press, 2024.
5. Choi, D.-A.; Tagore, P.; Siddiq, F.; Park, K.; Ewing, R. Descriptive statistics and visualizing data. In *Basic Quantitative Research Methods for Urban Planners*; Routledge, 2020; pp. 107–132.
6. Ravid, R. Practical Statistics for Educators; Rowman & Littlefield, 2024.
7. Di Leo, G.; Sardanelli, F. Statistical significance: p-value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach. *Eur. Radiol. Exp.* **2020**, *4*, 1–8.
8. McShane, B.B.; Bradlow, E.T.; Lynch Jr, J.G.; Meyer, R.J. "Statistical significance" and statistical reporting: Moving beyond binary. *J. Mark.* **2024**, *88*, 1–19.
9. Greenland, S.; Senn, S.J.; Rothman, K.J.; Carlin, J.B.; Poole, C.; Goodman, S.N.; Altman, D.G. Statistical tests, p-values, confidence intervals, and power: A guide to misinterpretations. *Eur. J. Epidemiol.* **2016**, *31*, 337–350.
10. Kopyt, M. Basic operations in the R software. In *Applied Spatial Statistics and Econometrics*; Routledge, 2020; pp. 1–35.
11. Siebert, J. Applications of statistical causal inference in software engineering. *Inf. Softw. Technol.* **2023**, *159*, 107198.
12. James, G.; Witten, D.; Hastie, T.; Tibshirani, R.; Taylor, J. An Introduction to Statistical Learning: With Applications in Python; Springer, 2023.
13. Smarandache, F. Neutrosophic logic—a generalization of the intuitionistic fuzzy logic. *Multispace Multistruct. Neutrosophic Transdisciplinarity (100 Collected Papers Sci.)* **2010**, *4*, 396.
14. Broumi, S.; Bakali, A.; Bahnasse, A. Neutrosophic sets: An overview; Infinite Study, 2018.
15. Garg, H. A new exponential-logarithm-based single-valued neutrosophic set and their applications. *Expert Syst. Appl.* **2024**, *238*, 121854.
16. Smarandache, F. A unifying field in logics: Neutrosophic logic. In *Philosophy*; American Research Press, 1999; pp. 1–141.

17. Smarandache, F. A Unifying Field in Logics: Neutrosophic Logic. Neutrosophy, Neutrosophic Set, Neutrosophic Probability; Infinite Study, 2005.
18. Alhabib, R.; Ranna, M.M.; Farah, H.; Salama, A. Some neutrosophic probability distributions. *Neutrosophic Sets Syst.* **2018**, *22*, 30–38.
19. Khan, Z.; Al-Bossly, A.; Almazah, M.M.; Alduais, F.S. On statistical development of neutrosophic gamma distribution with applications to complex data analysis. *Complexity* **2021**, *2021*, 3701236.
20. Duan, W.-Q.; Khan, Z.; Gulistan, M.; Khurshid, A. Neutrosophic exponential distribution: Modeling and applications for complex data analysis. *Complexity* **2021**, *2021*, 5970613.
21. Sherwani, R.A.K.; Shakeel, H.; Saleem, M.; Awan, W.B.; Aslam, M.; Farooq, M. A new neutrosophic sign test: An application to COVID-19 data. *PLoS ONE* **2021**, *16*, 0255671.
22. Khan, Z.; Almazah, M.M.; Odhah, H.H.; Alshanbari, H.M. Generalized Pareto model: Properties and applications in neutrosophic data modeling. *Math. Probl. Eng.* **2022**, *2022*, 3686968.
23. Ibrahim, A.M.M.; Khan, Z. Neutrosophic Laplace distribution with properties and applications in decision making. *Int. J. Neutrosophic Sci.* **2024**, *1*, 73–3.
24. Smarandache, F.; Abdel-Basset, M. Neutrosophic Operational Research: Methods and Applications; Springer, 2021.
25. Khan, Z.; Gulistan, M.; Kausar, N.; Park, C. Neutrosophic Rayleigh model with some basic characteristics and engineering applications. *IEEE Access* **2021**, *9*, 71277–71283.
26. Ibrahim, A.M.M.; Alduais, F.S.; Khan, Z.; Amin, A.; LaneKrebs, K. Neutrosophic structure of the geometric model with applications. *Neutrosophic Sets Syst.* **2023**, *61*, 382–394.
27. Al-Duais, F.S. Neutrosophic log-gamma distribution and its applications to industrial growth. *Neutrosophic Sets Syst.* **2024**, *72*, 22.
28. Kandemir, H.S.; Aral, N.D.; Karakas, M.; Et, M. Neutrosophic statistical analysis of temperatures of cities in the southeastern Anatolia region of Turkey. *Neutrosophic Syst. Appl.* **2024**, *14*, 50–59.
29. Smarandache, F. Introduction to Neutrosophic Statistics; Infinite Study, 2013.
30. Shahzadi, I. Neutrosophic statistical analysis of temperature of different cities of Pakistan; Infinite Study, 2023.

Received: Aug 29, 2024. Accepted: Nov 20, 2024