NSS
Neutrosophic Sets and Systems {Special Issue: NeutroGeometry, NeutroAlgebra, and Their Applications}, Vol. 71, 2024
University of New Mexico

Uncertainty Analysis in Prediction Intervals Using Neutrosophic Numbers

Héctor Félix Cerna Maguiña[1], Domingo Guzmán Chumpitaz Ramos[2], Ada Lucía Gallegos Ruiz Conejo[3], Nancy Isabel Pool Painted[4], and Jesús Isabel Baltazar Ángeles [5]

[1] National University of San Marcos, Lima. Peru. E-mail: hcernam@unmsm.edu.pe
[2] Ricardo Palma University, Lima. Peru. E-mail: dchumpitazr@unmsm.edu.pe
[3] Ricardo Palma University, Lima. Peru. E-mail: agallegosr@unmsm.edu.pe
[4] Ricardo Palma University, Lima. Peru. E-mail: nalbercapintado@gmail.com
[5] Ricardo Palma University, Lima. Peru. E-mail: eb4060@gmail.com

Abstract. This study analyzes uncertainty in prediction intervals by applying neutrosophic numbers to quality of life forecasts based on characteristics such as compensation, job hours, years of experience, and academic level. Three examples were studied, with indeterminacy values ranging from 0.589 to 0.628. These statistics show that between 58.9% and 62.8% of the prediction intervals are affected by uncertainty. Higher indeterminacy levels reflect a stronger effect of external influences or unpredictability within the data. The analysis demonstrates that as variables like compensation and labor hours grow, so does the level of uncertainty in the projections. Neutrosophic numbers provide a valuable framework for quantifying this uncertainty, providing for a better understanding of the model's limitations and the unpredictable nature of the dataset.

Keywords: Neutrosophic numbers, uncertainty analysis, prediction intervals, indeterminacy, remuneration, quality of life predictions.

1 Introduction

The quality of life of teachers is an important factor in the academic sector, which directly influences their performance and professional happiness. This research focuses on examining how high salary and another factor effects the quality of life of professors at the Faculty of Accounting Sciences of the Universidad Nacional Mayor de San Marcos, using neutrosophic regression models to handle the uncertainty inherent to this issue. Historically, the influence of pay on quality of life has been the topic of various research. Early methods concentrated on direct relationships between wage and satisfaction, but over time, it has been recognized that the link is more complicated and multidimensional [1].

Currently, with the growth of sophisticated analytical models, such as neutrosophic models, new perspectives have been established to analyze the connection between economic conditions and well-being [2]. This change in emphasis illustrates the growth in the knowledge of how financial incentives impact teachers and academic workers more broadly.

Neutrosophic machine learning [3] is an emerging field that combines neutrosophic theory, developed by Florentin Smarandache [4, 5], with modern machine learning algorithms to address uncertainty, imprecision, and incompleteness in data. Traditional machine learning assumes data is accurate and complete, but data often contains ambiguities and gaps that can affect model performance. Neutrosophic methods offer a new approach by incorporating uncertainty directly into the learning process, enhancing model accuracy, flexibility, and interpretability. These models are particularly useful in fields like image analysis, natural language processing, medicine, and engineering, where data is often imperfect. However, implementing neutrosophic techniques can be complex, requiring significant customization of existing algorithms, and ongoing research is needed to fully validate their advantages over traditional methods. Despite these challenges, neutrosophic machine learning has the potential to significantly improve outcomes in various applications by better managing uncertainty and incomplete data. In essence, neutrosophic machine learning provides a novel progression in the way models manage uncertainty and imprecision. Although it currently faces hurdles and needs more validation, its ability to give a richer and more flexible representation of data shows tremendous promise to increase the accuracy and application of machine learning models in diverse scenarios.

The purpose of this research is, first, to utilize neutrosophic regression models factor that improves the quality of life of teachers, taking into consideration the uncertainty and variety in individual views. Secondly, it tries to make suggestions based on the results to enhance compensation policies in the academic setting. These aims are related with the study topic and will be elaborated throughout the paper, presenting a complete perspective on how remuneration effects the well-being of teachers and how this connection might be better handled in the academic sector [6].

2 Preliminaries

2.2. Neutrosophic Machine Learning

Machine learning (ML) [7] utilizes mathematical formulations to develop models that can learn from data to make predictions or judgments without being explicitly programmed to accomplish such tasks. Interval prediction in machine learning refers to the approach of predicting a range of probable outcomes for a given input rather than a single point estimate. By offering intervals, these approaches give not just forecasts but also insight into the dependability and uncertainty of the predictions, which is vital for decision making in uncertain contexts [8].

For a data set with independent variables $X = [x_1, x_2, \dots, x_n]$ and a dependent variable, the goal of regression analysis is to model the relationship between X and y accurately. This relationship is expressed mathematically as [9]:

$$y \approx f(X;\, \theta) \qquad (1)$$

Where:

$y,$ is the dependent variable or the objective to be predicted.

$X,$ rrepresents the independent or explanatory variables that are used to predict.

f is the regression function, which can vary in shape depending on the type of regression model used (linear, polynomial, logistic, etc.).

θ are the parameters or coefficients of the model, adjusted during the training process to minimize a loss function, typically the mean square error (MSE) in regression [10].

In regression analysis, representing predictions as prediction intervals provides a more complete view of the uncertainty associated with the predictions. A prediction interval provides a range within which we expect the true value of the dependent variable to fall with a certain probability, typically 95% or 99%. This is particularly useful because it takes into account variability in the data that might not be captured by prediction alone [2].

To calculate a prediction interval, both the uncertainty in the regression model estimate and the inherent variability of the data must be considered. The interval is built around the predicted value and is usually symmetrical, extending a certain amount above and below the predicted value. This range is determined based on the standard error of the prediction and the residual standard deviation, which reflects the dispersion of the residuals or errors of the model [11, 12].

For example, in a simple linear regression, the prediction interval for a new observation is given by [12]:

$$\hat{y}_0 \pm t_{\propto\,/2,n-2} \cdot SE \qquad (2)$$

Where $\hat{y}_0$ is the predicted value of y from the t distribution for a specific confidence level $\propto$and $n-2$ degrees of freedom, and?

Using prediction intervals in regression analysis is beneficial because they offer a realistic spectrum of possible outcomes, which aids in the decision-making process. This recognizes that a single predicted value is not absolute but rather a likely scenario within a range of potential outcomes. This forecasting method effectively incorporates the inherent uncertainties associated with future predictions, providing a more accurate description of what to expect. To further refine this model, neutrosophic statistics can be applied, which excel at managing ambiguity and indeterminacy in data. By converting the interval to a neutrosophic number, the traditional interval is enhanced to include a component of indeterminacy. This addition captures the uncertainty and imprecision that is typically present in real-world data, offering a more nuanced understanding of data variability. The neutrosophic treatment of the interval is as follows [13, 14]:

$$\hat{y}_0 - t_{\propto/\ 2,n-2} \cdot SE + (\hat{y}_0 + t_{\propto\ /2,n-2} \cdot SE)I \quad (3)$$

Here, $I_N$represents the indeterminacy factor associated with the prediction, where $I_N \in [I_l, I_u]$, this notation introduces the limits of indeterminacy. I_l (lower indeterminacy) and I_u (upper indeterminacy), which define the range of possible deviations due to uncertain elements that affect the forecast [15].

Neutrosophic techniques are particularly beneficial for fusing diverse predictive models because they allow for the inclusion of ambiguity, indeterminacy, and even contradicting information that may originate from numerous data sources or model outputs. This technique adds to increasing the robustness and reliability of prediction models by presenting a more complete framework that takes into account many types of uncertainty. The neutrosophic mean denoted as X_n, is calculated considering the neutrosophic inclusion $I_N$that belongs to the interval. $[I_l, I_u]$.This mean consists of two main elements: X_l, which is the mean of the bottom part of the neutrosophic samples , and X_u, which is the mean of the top part. The respective definitions are [16]:

$$X_l = \frac{\sum_{i=1}^{n_l} X_{il}1}{n_l} \quad (4)$$

$$X_u = \frac{\sum_{i=1}^{n_u} X_{iu}}{n_u} \quad (5)$$

where $n_l$and $n_u$represent the number of elements in the lower and upper parts of the neutrosophic samples, respectively. Therefore, the neutrosophic mean $X_n$is expressed as the sum of $X_l$and X_u, adjusted by the interval of indeterminacy I_n [17]:

$$X_N = X_l + X_u I_N;\ I_N \in [I_l, I_u] \quad (6)$$

$I_{l,}$=0, andI_u

$$I_u = \frac{X_u - X_l}{X_u} \quad (7)$$

3 Methods.

To better understand the process, it can be broken down into four essential steps, as illustrated in Figure 1.

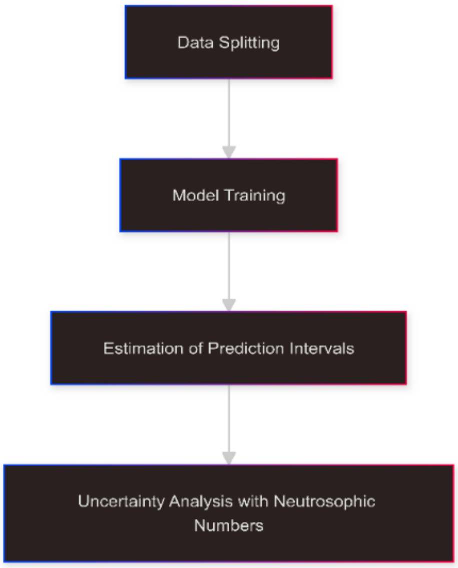

Figure 1. Modeling Process with Uncertainty Analysis using Neutrosophic Numbers

1. Splitting the data: The first step is to split the data into two sets: one for training and one for testing. This separation is critical since it enables you to evaluate the model with data you have not seen before, ensuring that its performance is not simply due to overfitting the training set. Generally, experts opt for a 70-30 or 80-20 split, with most of it utilized for training, and the remainder for testing.
2. Training the models: After separating the data, the next step is to train the models with the training set. This method entails modifying the model parameters so that they correspond as well as feasible with the data. Common regression models such as linear regression, ridge regression, and random forests are employed here. During training, we strive to optimize a loss function to properly capture the underlying pattern in the data.
3. Estimation of prediction intervals: Once the models are trained, the next step is to determine the prediction intervals for the data. Methods for estimating these intervals may vary based on the properties of the model and the nature of the data.
4. Uncertainty analysis using neutrosophic numbers: The final stage is to undertake a complete uncertainty analysis using neutrosophic numbers. This entails studying how the indeterminate component of these numbers varies based on the model, which might reveal insights about the complexity or unpredictability of the data set. For example, higher indeterminacy might suggest the effect of external influences or an unpredictable aspect to the data.

4 Results

This study aims to analyze the impact of high remuneration on the quality of life of teachers at the Faculty of Accounting Sciences of the Universidad Nacional Mayor de San Marcos by applying a neutrosophic logistic regression model. The main motivation for this analysis lies in the need to understand how salary factors influence essential aspects of quality of life, considering the uncertainty and indeterminacy in the data and results.

For this analysis, dataset of 55 events was used that includes information on teacher compensation and various metrics related to their quality of life. The dependent variable is quality of life is a continuous variable with values ranging from 0 to 10.

- **Independent Variables (X):**
  - **Monthly Remuneration (in soles)**
  - **Weekly Work Hours**
  - **Years of Experience**
  - **Academic Level**
- **Dependent Variable (Y):**
  - ✓ Quality of Life High (Values ranging from 0 to 10)

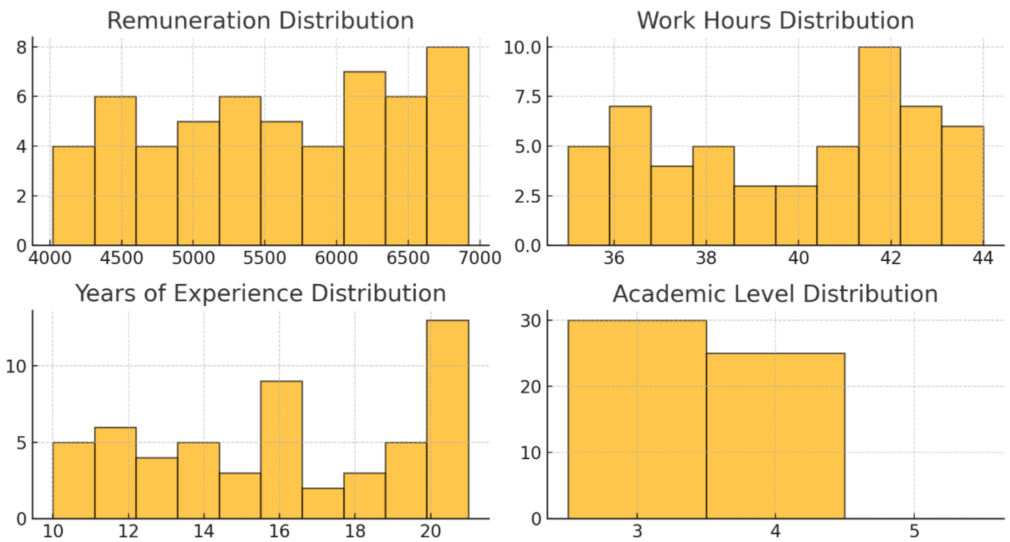

Figure 2. Distribution of Remuneration, Work Hours, Experience, and Academic Level in a Workforce Dataset

The data was divided into 70% for training and 30% for testing.
A linear regression model used is expressed as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

Where

Y is the dependent variable (Quality of Life)
X_1, X_2, X_3, X_4 , are the independent variables (predictors), which could represent Remuneration, Work Hours, Years of Experience, and Academic Level.

β_0 is the intercept of the model
$\beta_1, \beta_2, \beta_3, \beta_4$ are the coefficients associated with each independent variable.

Table 1 shows the predicted quality of life for three different people with varying combinations of pay, hours worked, years of experience, and educational attainment.

Table 1. Prediction intervals and neutrosophic forms

Remuneration	Work Hours	Years of Experience	Academic Level	Prediction Interval [Lower Bound, Upper Bound]	Neutrosophic form
4500	**36**	**10**	**3**	**[3.7, 9]**	**3.7+9**I; $I \in [0, 0.589]$
5500	**40**	**16**	**4**	**[3.4, 8.7]**	**3.4+8.7**I; $I \in [0, 0.609]$
6500	**44**	**20**	**5**	**[3.2, 8.6]**	**3.2+8.6**I; $I \in [0, 0.0.628]$

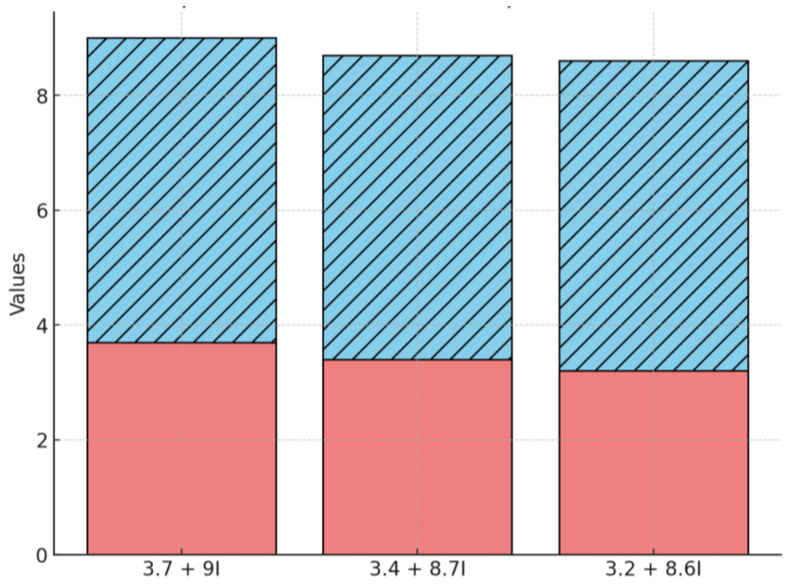

Figure 3. Representation of Neutrosophic Numbers

The analysis reveals varying levels of indeterminacy across the three cases. As the remuneration, work hours, and academic level increase, the indeterminacy slightly increases, indicating more unpredictability. This might suggest the impact of external factors or complexity in the data that the model cannot fully capture. The use of neutrosophic numbers helps quantify and visualize this uncertainty, making it clearer how much of the prediction interval is influenced by indeterminate factors.

The findings point to the possibility of enhancing academic compensation schemes by taking into account the effect that pay has on teachers' general well-being. According to the data, there is more uncertainty in predicting quality of life at higher compensation levels, even if these factors are similarly connected with longer work hours and more experience. This implies that improving well-being might need more than just paying more. Rather, policies ought to strive for a well-rounded strategy that takes into account variables including professional assistance, career advancement, and workload. Academic institutions should better promote the entire well-being of their workers and create a more content and productive workforce by addressing both the financial and non-financial parts of remuneration. This viewpoint will be broadened throughout the study to offer a thorough grasp of the relationship between compensation and well-being in the academic sector.

Conclusion

The findings of this study highlight the importance of understanding the influence of remuneration on the quality of life of instructors, particularly in the academic sector. By using neutrosophic numbers, the analysis captures the uncertainty and indeterminacy present in the relationship between remuneration and well-being. The results suggest that while remuneration plays a significant role in influencing quality of life, its impact is not uniform and varies depending on other factors like work hours, years of experience, and academic level. The inclusion of neutrosophic analysis offers a clearer understanding of this complexity, showing that financial incentives alone are not sufficient. Academic institutions can benefit from this approach by designing compensation policies that consider both financial and non-financial factors, ensuring a more balanced and supportive work environment for their staff.

For future research, it is recommended to explore the impact of various non-economic fac-tors on teachers' quality of life, such as work-life balance, job security, and institutional support. Additionally, extending the sample size and applying neutrosophic models to other academic sectors or areas could provide new insights into how multiple variables combine to promote well-being. Future research should also focus on strengthening neu-trosophic models to better handle multidimensional and dynamic datasets, enabling for even more detailed examination of the factors affecting teachers' quality of life in varied educational environments.

References

[1] Szromek, A. R., & Wolniak, R. (2020). Job satisfaction and problems among academic staff in higher education. Sustainability, 12(12), 4865.

[2] Vázquez, M. L., & Torres, L. C. (2024). Integrating Neutrosophic Numbers in Regression Analysis for Enhancing Predictive Modelling through Uncertainty Representation. Plithogenic Logic and Computation, 1, 120-126.

[3] Liu, J. (2021). Exploring teacher attrition in urban China through interplay of wages and well-being. Education and Urban Society, 53(7), 807-830.

[4] KERVANCI, I. S. (2023). A Review Hybrid Structure of Neutrosophy and Machine Learning Algorithms for Different Types of Problems. 2023 Neutrosophic SuperHyperAlgebra And New Types of Topologies, 133.

[5] [Smarandache, F. (1999). A unifying field in Logics: Neutrosophic Logic. In Philosophy (pp. 1-141). American Research Press.] KERVANCI, I. S. (2023). A Review Hybrid Structure of Neutrosophy and Machine Learning Algorithms for Different Types of Problems. 2023 Neutrosophic SuperHyperAlgebra And New Types of Topologies, 133.

[6] Leyva Vázquez, M., Estupiñan, J., & Smarandache, F. (2020). Neutrosofía en Latinoamérica, avances y perspectivas. Neutrosophic Computing and Machine Learning, 14, 1-7.

[7] Shinde, P. P., & Shah, S. (2018, August). A review of machine learning and deep learning applications. In 2018 Fourth international conference on computing communication control and automation (ICCUBEA) (pp. 1-6). IEEE.

[8] Alanazi, B. and Alrashdi, I. (2023) "Anomaly Detection in Smart Agriculture Systems on Network Edge Using Deep Learning Technique", Sustainable Machine Intelligence Journal, 3, pp. (4):1–31.

[9] Bilal, M., Ali, M. A., Nichol, J. E., Qiu, Z., Mhawish, A., & Khedher, K. M. (2023). Reduced major axis

regression. In Encyclopedia of Mathematical Geosciences (pp. 1199-1203). Cham: Springer International Publishing.
[10] Anwar, M. B., Hanif, M., Shahzad, U., Emam, W., Anas, M. M., Ali, N., & Shahzadi, S. (2024). Incorporating the neutrosophic framework into kernel regression for predictive mean estimation. Heliyon, 10(3).
[11] Shrestha, D. L., & Solomatine, D. P. (2006). Machine learning approaches for estimation of prediction interval for the model output. Neural networks, 19(2), 225-235.
[12] Doksum, K., & Koo, J. Y. (2000). On spline estimators and prediction intervals in nonparametric regression. Computational statistics & data analysis, 35(1), 67-82.
[13] Smarandache, F., & Aslam, M. (Eds.). (2023). Cognitive Intelligence with Neutrosophic Statistics in Bioinformatics. Elsevier
[14] Leyva Vázquez, M., & Smarandache, F. (2021). Resolución de sistemas de ecuaciones lineales neutrosóficas mediante computación simbólica. Neutrosophic Computing and Machine Learning, 15, 41-46.
[15] Delgado Estrada, S. M., Rocio Ruiz Molina, K. D., Ponce Orella-na, F. E., & Chabusa Vargas, J. L. (2024). Neutrosophic Statistics for Enhanced Time Series Analysis of Unemployment Trends in Ecuador. Neutrosophic Sets and Systems, 67(1), 14.
[16] Castro Sánchez, F., Almeida Blacio, J. H., Flores Bracho, M. G., Andrade Santamaria, D. R., & Sánchez Casanova, R. (2021).
[17] Neutrosophic and Plithogenic Statistical Analysis in Educational Development. Neutrosophic Sets and Systems, 44(1), 26.

Received: June 23, 2024. Accepted: August 15, 2024