Contemporary Mathematics
https://ojs.wiserpub.com/index.php/CM/

Research Article

Analyzing the Impact of Uncertainty on Multiple Regression Metrics and Model Performance

Muhammad Aslam[1*], Turki Alhuzali[2], Mohammed Alsubaie[2], Bayan Adel Shukr[3], Afnan Ali Almazmomi[3]

[1] Department of Statistics, Faculty of Science, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[2] Department of Mathematics and Statistics, AlKhurmah University College, Taif University, P.O. Box 11099, Taif, 21944, Saudi Arabia
[3] Department of Management Information Systems, College of Business Administration, Taibah University, Tayba, Madinah, 42353, Saudi Arabia
E-mail: magmuhammad@kau.edu.sa

Received: 4 August 2025; **Revised:** 7 November 2025; **Accepted:** 21 November 2025

Abstract: Regression models based on classical statistics have been extensively applied across various fields. However, these traditional models fail to account for the degree of indeterminacy when dealing with uncertainty. To address these limitations, this manuscript presents multiple regression models within the neutrosophic statistics framework, emphasizing key metrics relevant to the proposed regression approach. The concept of the neutrosophic random variable is introduced, with its essential properties and the formulation of the neutrosophic regression line, along with metrics such as multiple R, the coefficient of determination, and neutrosophic analysis of variance. Extended simulation studies examine how varying degrees of uncertainty influence multiple R, standard error, and F-test values. Analysis of a real-world example shows that uncertainty levels significantly affect both predicted and residual values generated by the regression model. The results indicate that variations in the degree of indeterminacy lead to significant differences in the multiple R, coefficient of determination, and F-test values. The findings further suggest that regression modeling outcomes may diverge from traditional regression analysis under neutrosophic statistics, making the proposed multiple regression method suitable for situations where data uncertainty plays a critical role.

Keywords: regression, correlation, classical statistics, simulation, residual

MSC: 62A86

1. Introduction

Regression analysis is employed across diverse domains to evaluate the connection between a dependent variable and independent variables. This analysis serves to gauge the strength of relationships between variables, aiding in estimation and prediction tasks. While simple linear regression entails consideration of a single independent variable, multiple regressions encompasses the incorporation of multiple independent variables. Following the application of multiple regression models, predicted values and residual values are computed for future reference. Given its versatility, regression analysis finds widespread application across various fields. Eyduran et al. [1] applied the multiple regression for analyzing

DOI: https://doi.org/10.37256/cm.7220268132

body weight data. Kang and Zhao [2] used the multiple regression analysis for denoising data. Ruiz and Freyne [3] applied the multiple regression analysis using the concrete data. Trunfio et al. [4] applied the regression model using the healthcare data. Dorta-González [5] applied the multiple regression for the analysis of journal's contribution to the social attention. The detail about the methodology and application of the multiple regression can be seen in Sun et al. [6]. More applications of the regression analysis can be seen in Moxley et al. [7]. Neutrosophic statistics, a branch of mathematical science, is highly valuable for handling, analyzing, presenting, and interpreting imprecise data. The neutrosophic methods offer several advantages over fuzzy sets, which use only a single membership degree, and intuitionistic fuzzy approaches, which consider both membership and non-membership degrees. In contrast, neutrosophic methods generalize these fuzzy-based approaches by introducing three components—truth, falsity, and indeterminacy. These components make neutrosophic approaches more informative and effective for dealing with incomplete and interval data [8]. The relationship between the neutrosophic approach and fuzzy-based methods is further discussed in Rezaei et al. [9]. Smarandache [10] demonstrated through several examples that neutrosophic statistics outperform interval statistics. Moreover, neutrosophic statistics extend fuzzy statistics [11], which do not account for the degree of indeterminacy. Neutrosophic statistics extends classical statistics by incorporating degrees of uncertainty often overlooked in traditional approaches. Neutrosophic statistics offers enhanced flexibility for analyzing imprecise data, as demonstrated by its introduction by Smarandache [12] and subsequent studies. Recent research, including work by Smarandache [13], Chen et al. [14], Chen et al. [15], and Duan et al. [16], highlighted the efficiency of neutrosophic statistical analysis, particularly when dealing with complex or e-commerce data. Nagarajan et al. [17] delved into neutrosophic multiple regression analysis, while AlAita et al. [18] focused on split-plot design for neutrosophic data analysis. AlAita and Aslam [19] explored analysis of covariance for imprecise data, while Shahzadi [20] investigated neutrosophic statistical analysis in the context of temperature variations across cities. The proposal of neutrosophic kernel regression for mean estimation further expands the application of neutrosophic statistical techniques. More information to deal the measurement error can be seen in Mushtaq and Butt [21] and Khali et al. [22].

Extensive literature exists on the multiple regression model within classical statistics, which is applicable solely when dealing with certain or precise data. However, in practical scenarios, complete certainty among all observations is unattainable. When faced with imprecise or interval data, the conventional multiple regression analysis within classical statistics fails to suffice for prediction and estimation purposes. To bridge this gap, the introduction proposed neutrosophic multiple regressions, leveraging the concept of the neutrosophic random variable. In this paper, the neutrosophic random variable introduces along with its defining properties. Subsequently, the neutrosophic regression line presents alongside associated measures such as multiple R, coefficient of determination, and neutrosophic analysis of variance. This study extends further to include simulation studies, examining the impact of uncertainty levels on the aforementioned regression measures. Additionally, data analysis conducts utilizing both the proposed neutrosophic regression model and conventional regression analysis within classical statistics. Through simulation and real data analysis, it is anticipated that the proposed model will prove more suitable for analyzing imprecise data, offering enhanced adequacy compared to existing methodologies.

2. Neutrosophic random variables

Consider two neutrosophic random variables, $X_N = X_L + X_L I_N$ and $Y_N = Y_L + Y_L I_N$, each comprised of two components. The initial parts, X_L and Y_L, represent the determinate aspects akin to classical statistics. Meanwhile, the additional terms $X_L I_N$ and $Y_L I_N$ signify the indeterminate facets, where $I_N \in [I_L, I_U]$ denotes the degree of indeterminacy or uncertainty. Notably, when $I_L = 0$, these neutrosophic random variables reduce to classical statistical variables. Assuming X_L and Y_L follow normal distribution μ_x and μ_y, and variances σ_x^2 and σ_y^2 respectively, as stated by Granados [23], neutrosophic logic extends fuzzy logic, with $I_N^2 = I_N, \ldots, I_N^n = I_N$ for $n \in N$ and when $I_N \in [0, 1]$. Given this information, the properties of the proposed neutrosophic random variables are outlined in Eq. (1) to Eq. (5), following the approach of Aslam [24].

The expected value of the random variable X_N is given by

$$E(X_N) = E(X_L + X_L I_N) = (1 + I_N)\mu_x. \quad (1)$$

The expected value of the random variable Y_N is given by

$$E(Y_N) = E(Y_L + Y_L I_N) = (1 + I_N)\mu_y. \quad (2)$$

The variance of the random variable X_N is given by

$$\mathrm{Var}(X_N) = \mathrm{Var}(X_L + X_L I_N) = (1 + I_N)^2 \sigma_x^2. \quad (3)$$

The variance of the random variable Y_N is given by

$$\mathrm{Var}(Y_N) = \mathrm{Var}(Y_L + Y_L I_N) = (1 + I_N)^2 \sigma_y^2. \quad (4)$$

The additive properties of variance is given by

$$\mathrm{Var}(X_N + Y_N) = (1 + I_N)^2 \sigma_x^2 + (1 + I_N)^2 \sigma_y^2. \quad (5)$$

3. Methodology

Suppose that $Y_N = Y_L + Y_L I_N$ be dependent neutrosophic random variable and $X_{N1} = X_{L1} + X_{L1} I_N$, $X_{N2} = X_{L2} + X_{L2} I_N$, ..., $X_{Nn} = X_{Ln} + X_{Ln} I_N$ be independent neutrosophic random variables of sample size n. Note here that the first values of dependent and the independent random variable shows the determinate parts and presents the random variable under classical statistics. Using these variables, the neutrosophic multiple regression line is defined in Eq. (6).

$$(1 + I_N) Y_L = a + (1 + I_N) b_1 X_{L1} + (1 + I_N) b_2 X_{L2} + (1 + I_N) b_3 X_{L2} + \cdots + (1 + I_N) b_n X_{Ln} + e. \quad (6)$$

The neutrosophic regression model is expressed in Eq. (7).

$$Y_N = a + b_1 X_{N1} + b_2 X_{N2} + b_3 X_{N3} + \cdots + b_n X_{Nn} + e. \quad (7)$$

Note here that in the proposed neutrosophic multiple regression model, a presents the intercept of the model, b_1, b_2, b_3, ..., b_n denote the slope of the neutrosophic regression model and e denotes the random error. The estimated neutrosophic regression model is presented in Eq. (8) and defined as follows:

$$\hat{Y}_N = a + b_1 X_{N1} + b_2 X_{N2} + b_3 X_{N3} + \cdots + b_n X_{Nn}. \quad (8)$$

The neutrosophic multiple regression model for two neutrosophic independent random variable is presented in Eq. (9) and expressed as follows

$$\hat{Y}_N = a + b_1 X_{N1} + b_2 X_{N2}. \tag{9}$$

Note here the proposed neutrosophic regression model is the generalization of the multiple regression models under classical statistics. The proposed neutrosophic regression model reduces to the classical multiple regression model under classical statistics when $I_L = 0$. For the neutrosophic multiple regression line, the intercept a is provided in Eq. (10) and can be calculated as follows:

$$a = (\bar{Y}_L + \bar{Y}_L I_N) - b_1 (\bar{X}_{L1} + \bar{X}_{L1} I_N) - b_2 (\bar{X}_{L2} + \bar{X}_{L2} I_N); \quad I_N \in [I_L, I_U]. \tag{10}$$

The first slope b_1, given in Eq. (11) can be calculated as follows

$$b_1 = \frac{\sum(1+N)^2 X_{L2}^2 \sum (X_{L1} + X_{L1} I_N)(Y_L + Y_L I_N) - \sum (X_{L1} + X_{L1} I_N)(X_{L2} + X_{L2} I_N) \sum (X_{L2} + X_{L2} I_N)(Y_L + Y_L I_N)}{\sum(1+N)^2 X_{L1}^2 \sum(1+N)^2 X_{L2}^2 - \left(\sum (X_{L1} + X_{L1} I_N)(X_{L2} + X_{L2} I_N)\right)^2}. \tag{11}$$

The first slope b_2, given in Eq. (12) can be calculated as follows

$$b_2 = \frac{\sum(1+N)^2 X_{L1}^2 \sum (X_{L2} + X_{L2} I_N)(Y_L + Y_L I_N) - \sum (X_{L1} + X_{L1} I_N)(X_{L2} + X_{L2} I_N) \sum (X_{L1} + X_{L1} I_N)(Y_L + Y_L I_N)}{\sum(1+N)^2 X_{L1}^2 \sum(1+N)^2 X_{L2}^2 - \left(\sum (X_{L1} + X_{L1} I_N)(X_{L2} + X_{L2} I_N)\right)^2}. \tag{12}$$

3.1 *Multiple regression coefficient*

The multiple regression coefficient, known as multiple R, quantifies the highest level of association between independent variables and a single dependent variable. However, traditional multiple R faces limitations when dealing with imprecise or uncertain data observations. To address this, it is aimed to enhance multiple R by integrating the principles of neutrosophic statistics. By doing so, it is seek to develop a revised multiple R that can effectively accommodate imprecise data while accounting for the inherent uncertainty within the dataset.

The multiple regression coefficients (multiple R), say $R_{N1.23}$ can be derived as follows

The neutrosophic regression line with two independent variables is defined in Eq. (13) as follows:

$$X_{N1} = a + b_{12.3} X_{N2} + b_{13.2} X_{N3}. \tag{13}$$

It is supposed that the three neutrosophic random variable x_{N1}, x_{N2}, and x_{N3} can be measured from their respective means. The neutrosophic regression equation of x_{N1} dependent on x_{N2} and x_{N3} is presented in Eq. (14) as follows:

$$x_{N1} = a + b_{12.3} x_{N2} + b_{13.2} x_{N3}. \tag{14}$$

where $\left(X_{N1} - \overline{X}_{N1}\right) = (1 + I_N) x_{L1}$, $\left(X_{N2} - \overline{X}_{N2}\right) = (1 + I_N) x_{L2}$, and $\left(X_{N3} - \overline{X}_{N3}\right) = (1 + I_N) x_{L3}$.

Note here that $(1 + I_N) \sum x_{L1} = (1 + I_N) \sum x_{L2} = (1 + I_N) \sum x_{L3} = 0$.

The right hand side of Eq. (1) is expected value of x_{L1} that depends of x_{L2} and x_{L3} is given in Eq. (15) and can be expressed as follows

$$x_{N1.23} = b_{12.3}x_{N2} + b_{13.2}x_{N3}. \tag{15}$$

The residual $e_{N1.23}$ is given in Eq. (16) and can be expressed by

$$e_{N1.23} = x_{N1} - b_{12.3}x_{N2} - b_{13.2}x_{N3} = x_{N1} - x_{N1.23}. \tag{16}$$

From Eq. (13), the resulting expression is obtained and presented in Eq. (17).

$$x_{L1.23} = x_{N1} - e_{N1.23}. \tag{17}$$

The neutrosophic multiple correlation is presented in Eq. (18) and defined as follows:

$$R_{N1.23} = \frac{\operatorname{Cov}(x_{N1}, x_{N1.23})}{\sqrt{\operatorname{Var}(x_{N1}) \cdot \operatorname{Var}(x_{N1.23})}}. \tag{18}$$

Thus, the covariance is presented in Eq. (19) and defined as follows:

$$\operatorname{Cov}(x_{N1}, x_{N1.23}) = \frac{1}{N}\sum x_{N1}x_{N1.23}. \tag{19}$$

After some simplification,

$$\operatorname{Cov}(x_{N1}, x_{N1.23}) = \left(\sigma_{N1}^2 - \sigma_{N1.23}^2\right)$$

and $\operatorname{Var}(x_{N1}) = \sigma_{N1}^2$ and $\operatorname{Var}(x_{N1.23}) = \left(\sigma_{N1}^2 - \sigma_{N1.23}^2\right)$.

The $R_{N1.23}$ in Eq. (15) can be written as in Eq. (20)

$$R_{N1.23} = \frac{\left(\sigma_{N1}^2 - \sigma_{N1.23}^2\right)}{\sqrt{\sigma_{N1}^2\left(\sigma_{N1}^2 - \sigma_{N1.23}^2\right)}}; \quad I_N \in [I_L, I_U]. \tag{20}$$

The neutrosophic form of $R_{N1.23}$ is presented in Eq. (21) and can be expressed as follows:

$$R_{N1.23} = \frac{\left(\sigma_{L1}^2 - \sigma_{L1.23}^2\right)}{\sqrt{\sigma_{L1}^2\left(\sigma_{L1}^2 - \sigma_{L1.23}^2\right)}} + \frac{\left(\sigma_{U1}^2 - \sigma_{U1.23}^2\right)}{\sqrt{\sigma_{U1}^2\left(\sigma_{U1}^2 - \sigma_{U1.23}^2\right)}} I_{RN1.23}; \quad I_{RN1.23} \in [I_{RL1.23}, I_{RU1.23}]. \tag{21}$$

Note that $I_{RN1.23} \in [I_{RL1.23}, I_{RU1.23}]$ be the degree of uncertainty in $R_{N1.23}$.

3.2 *Coefficient of determination*

The coefficient of determination $R^2_{N1.23}$ is used to explain the proportion of variation in dependent variable due to the prediction of the independent variables. It is expressed the coefficient of determination $R^2_{N1.23}$ for the neutrosophic statistics as follows:

$R^2_{N1.23}$ can be expressed as shown in Eq. (22)-(24).

$$R^2_{N1.23} = 1 - \frac{\sigma^2_{N1.23}}{\sigma^2_{N1}} \tag{22}$$

and

$$R^2_{N1.23} = 1 - \frac{r^2_{N12} + r^2_{N13} - 2r_{N12}r_{N13}r_{N23}}{1 - r^2_{N23}}$$

or

$$R^2_{N1.23} = \frac{r^2_{N12} + r^2_{N13} - 2r_{N12}r_{N13}r_{N23}}{1 - r^2_{N23}}$$

and

$$R_{N1.23} = \sqrt{\frac{r^2_{N12} + r^2_{N13} - 2r_{N12}r_{N13}r_{N23}}{1 - r^2_{N23}}} \tag{23}$$

where $\sigma^2_{N1.23}$ is variance of the error and can be expressed as

$$\sigma^2_{N1.23} = \frac{\sigma^2_{N1}}{1 - r^2_{N23}}(1 - r^2_{N23} - r^2_{N12} - r^2_{N13} + 2r_{N12}r_{N13}r_{N23})$$

The neutrosophic expression of $R^2_{N1.23}$ is given by

$$R^2_{N1.23} = \frac{r^2_{L12} + r^2_{L13} - 2r_{L12}r_{L13}r_{L23}}{1 - r^2_{L23}} + \frac{r^2_{U12} + r^2_{u13} - 2r_{U12}r_{U13}r_{U23}}{1 - r^2_{U23}} I_{R^2N1.23}; \quad I_{R^2N1.23} \in [I_{R^2L1.23}, I_{R^2U1.23}] \tag{24}$$

Note that $I_{R^2N1.23} \in [I_{R^2L1.23}, I_{R^2U1.23}]$ be the degree of uncertainty in $R^2_{N1.23}$.

3.3 *Neutrosophic analysis of variance*

In this section, the purpose is to discuss the Neutrosophic Analysis of Variance (NANOVA) for the multiple regression analysis. The neutrosophic sum of residuals is presented in Eq. (25) and defined as follows:

$$RSS_N = \sum \left(y_N - \widehat{Y}_N \right)^2. \tag{25}$$

The neutrosophic total sum of squares is presented in Eq. (26) and defined as follows:

$$RSS_N = \sum \left(y_N - \bar{y}_N \right)^2. \tag{26}$$

The NANOVA is given in Table 1.

Table 1. NANOVA

	Degrees of freedom (df)	Sum-of-Squares (SS)	Mean Squares (MS)	F_N
Regression	k	$\Sigma \left(\widehat{Y}_N - \bar{y}_N \right)^2$	$\dfrac{\Sigma \left(\widehat{Y}_N - \bar{y}_N \right)^2}{k}$	$\dfrac{\Sigma \left(\hat{Y}_N - \bar{y}_N \right)^2}{k} / \dfrac{\Sigma \left(y_N - \hat{Y}_N \right)^2}{n-k-1}$
Residual	$n-k-1$	$\Sigma \left(y_N - \widehat{Y}_N \right)^2$	$\dfrac{\Sigma \left(y_N - \widehat{Y}_N \right)^2}{n-k-1}$	
Total	$N-1$	$\Sigma \left(y_N - \bar{y}_N \right)^2$		

The neutrosophic expression of F-test is given by

$$F_N = \frac{\Sigma \left(\hat{Y}_L - \bar{y}_L \right)^2}{k} / \frac{\Sigma \left(y_L - \hat{Y}_L \right)^2}{n-k-1} + \frac{\Sigma \left(\hat{Y}_U - \bar{y}_U \right)^2}{k} / \frac{\Sigma \left(y_U - \hat{Y}_U \right)^2}{n-k-1} I_{FN}; \quad I_{FN} \in [I_{FL}, I_{FU}].$$

Note that $I_{FN1.23} \in [I_{FL1.23}, I_{FU1.23}]$ be the degree of uncertainty in F_N. The neutrosophic F-test reduces to F-test under classical statistics when $I_{FL} = 0$.

4. Simulation study

In this section, the aim is to present a simulation study examining how the degree of uncertainty impacts the values of multiple R, standard error, and F-test. It is explored for various combinations of I_N and sample size (n) to investigate these effects. For $n = 10$, it is displayed the values of multiple R, standard error, and F-test in Tables 2-4 respectively. Furthermore, it is extended the current analysis to $n = 15$, showcasing the values of multiple R, standard error, and F-test in Tables 5-7 correspondingly. Examining Tables 2 and 5, it is observed a general increasing trend in the values of multiple R as I_N varies from 0.1 to 1. For instance, in Table 2, the second row indicates a multiple R value of 0.2786 for $I_N = 0.1$, while it rises to 0.3285 for $I_N = 1$. Table 3 and Table 6 present the standard error values for multiple regressions across different I_N values and n. The simulated results demonstrate a significant impact on standard error as I_N varies. Overall, there is a noticeable decreasing trend in standard error values across various combinations of I_N and n. Analyzing Tables 3 and 6, it becomes evident that changes in the degree of uncertainty directly influence standard error values. For instance, in Table 3, the first row indicates a standard error value of 3.2437 for $I_N = 0.1$, whereas it decreases to 3.1276 for $I_N = 0.9$. Table 4 and Table 7 present the values of F-test across different I_N values and n. The simulated results demonstrate a significant impact on the values of F-test as I_N varies. Overall, there is a noticeable decreasing trend in

F-test values across various combinations of I_N and n. Analyzing Tables 4 and 7, it becomes evident that changes in the degree of uncertainty directly influence F-test values. For instance, in Table 4, the first row indicates a F-test value of 0.6133 for $I_N = 0.1$, whereas it decreases to 0.5932 for $I_N = 0.9$. The simulation studies illustrate the significant impact of uncertainty on various aspects of multiple regression analysis, including the generation of multiple R values, standard error values, and F-test values. Thus, decision-makers must exercise caution when implementing multiple regression analysis in uncertain conditions.

Table 2. Effect on multiple R when $n = 10$

$I_N = 0$	$I_N = 0.1$	$I_N = 0.2$	$I_N = 0.3$	$I_N = 0.4$	$I_N = 0.5$	$I_N = 0.6$	$I_N = 0.7$	$I_N = 0.8$	$I_N = 0.9$	$I_N = 1$
0.3854	0.3861	0.3865	0.3866	0.3864	0.3859	0.3851	0.3840	0.3825	0.3807	0.3785
0.2728	0.2786	0.2844	0.2902	0.2959	0.3015	0.3071	0.3126	0.3180	0.3233	0.3285
0.6374	0.6433	0.6493	0.6553	0.6613	0.6674	0.6734	0.6795	0.6855	0.6915	0.6974
0.2512	0.2515	0.2519	0.2523	0.2527	0.2531	0.2535	0.2539	0.2543	0.2547	0.2552
0.6114	0.6126	0.6138	0.6150	0.6161	0.6173	0.6184	0.6195	0.6206	0.6217	0.6228
0.4506	0.4400	0.4300	0.4209	0.4127	0.4055	0.3992	0.3939	0.3895	0.3859	0.3831
0.6154	0.5812	0.5502	0.5228	0.4989	0.4784	0.4612	0.4468	0.4350	0.4255	0.4179
0.0820	0.0854	0.0902	0.0963	0.1033	0.1112	0.1197	0.1287	0.1382	0.1480	0.1581
0.2499	0.2210	0.1939	0.1697	0.1494	0.1340	0.1242	0.1200	0.1208	0.1254	0.1326
0.3568	0.3717	0.3839	0.3939	0.4022	0.4093	0.4153	0.4204	0.4249	0.4288	0.4322

Table 3. Effect on standard error when $n = 10$

$I_N = 0$	$I_N = 0.1$	$I_N = 0.2$	$I_N = 0.3$	$I_N = 0.4$	$I_N = 0.5$	$I_N = 0.6$	$I_N = 0.7$	$I_N = 0.8$	$I_N = 0.9$	$I_N = 1$
3.2734	3.2437	3.2172	3.1940	3.1742	3.1578	3.1449	3.1356	3.1298	3.1276	3.1290
4.9000	4.8619	4.8257	4.7914	4.7592	4.7292	4.7013	4.6755	4.6520	4.6307	4.6117
4.2200	4.1517	4.0844	4.0184	3.9536	3.8901	3.8279	3.7672	3.7080	3.6503	3.5943
9.3271	9.2793	9.2325	9.1866	9.1416	9.0976	9.0545	9.0125	8.9714	8.9314	8.8924
16.05	16.0105	15.9651	15.9203	15.8760	15.8323	15.7892	15.7467	15.7048	15.6635	15.6227
17.6792	18.1159	18.6097	19.1563	19.7513	20.3904	21.0697	21.7853	22.5338	23.3121	24.1172
18.5615	20.0335	21.5385	23.0702	24.6234	26.1945	27.7803	29.3785	30.9872	32.6048	34.2300
22.1627	22.0737	21.9884	21.9070	21.8293	21.7555	21.6857	21.6198	21.5578	21.4999	21.4460
9.3224	9.5756	9.8761	10.2197	10.6022	11.0196	11.4681	11.9442	12.4446	12.9667	13.5078
20.7196	22.7285	24.7746	26.8496	28.9472	31.0627	33.1929	35.3350	37.4870	39.6472	41.8145

Table 4. Effect on F-value when $n = 10$

$I_N = 0$	$I_N = 0.1$	$I_N = 0.2$	$I_N = 0.3$	$I_N = 0.4$	$I_N = 0.5$	$I_N = 0.6$	$I_N = 0.7$	$I_N = 0.8$	$I_N = 0.9$	$I_N = 1$
0.5694	0.6133	0.6148	0.5673	0.6145	0.6127	0.6096	0.6053	0.5999	0.5932	0.5855
0.2815	0.2946	0.3081	0.3219	0.3359	0.3502	0.3646	0.3792	0.394	0.4088	0.4236
2.3959	2.4721	2.5517	2.6381	2.7213	2.8115	2.9052	3.0020	3.1033	0.1026	3.5153
0.2358	0.2365	0.2372	0.2379	0.2387	0.2395	0.2403	0.2412	0.2420	0.2429	0.2439
2.0903	2.1033	2.1162	2.1292	2.1421	2.1550	2.1678	2.1806	2.1934	2.2061	2.2188
0.8921	0.8405	0.7943	0.7538	0.7187	0.6888	0.6638	0.6430	0.6262	0.6127	0.6022
2.1339	1.7855	1.5199	1.3165	1.1600	1.0392	0.9457	0.8733	0.8172	0.7739	0.7406
0.0237	0.0257	0.0287	0.0327	0.0377	0.0438	0.0509	0.0590	0.0681	0.0784	0.0897
0.2332	0.1797	0.1367	0.1038	0.0799	0.0640	0.0549	0.0512	0.0519	0.0559	0.0626
0.5106	0.5614	0.6051	0.6429	0.6757	0.7044	0.7294	0.7516	0.7712	0.7887	0.8044

Table 5. Effect on multiple R when $n = 15$

$I_N = 0$	$I_N = 0.1$	$I_N = 0.2$	$I_N = 0.3$	$I_N = 0.4$	$I_N = 0.5$	$I_N = 0.6$	$I_N = 0.7$	$I_N = 0.8$	$I_N = 0.9$	$I_N = 1$
0.3699	0.3596	0.3512	0.3435	0.3390	0.3346	0.3310	0.3281	0.3257	0.3238	0.3221
0.2335	0.2248	0.2164	0.2085	0.2010	0.1942	0.1881	0.1828	0.1783	0.1748	0.1721
0.2235	0.2143	0.2049	0.1953	0.1856	0.1758	0.1660	0.1564	0.1469	0.1376	0.1287
0.4567	0.4583	0.4592	0.4596	0.4592	0.4583	0.4566	0.4542	0.4511	0.4474	0.4430
0.3191	0.3175	0.3159	0.3144	0.3128	0.3112	0.3097	0.3081	0.3066	0.3050	0.3035
0.2227	0.2432	0.2637	0.2838	0.3032	0.3216	0.3390	0.3552	0.3701	0.3838	0.3962
0.4811	0.4806	0.4800	0.4795	0.4790	0.4784	0.4779	0.4773	0.4768	0.4762	0.4757
0.0525	0.0530	0.0537	0.0545	0.0554	0.0565	0.0578	0.0591	0.0606	0.0623	0.0640
0.6870	0.6868	0.6865	0.6862	0.6859	0.6856	0.6853	0.6850	0.6846	0.6843	0.6839
0.4749	0.4748	0.4746	0.4744	0.4742	0.4740	0.4738	0.4736	0.4734	0.4732	0.4730

Table 6. Effect on standard error when $n = 15$

$I_N = 0$	$I_N = 0.1$	$I_N = 0.2$	$I_N = 0.3$	$I_N = 0.4$	$I_N = 0.5$	$I_N = 0.6$	$I_N = 0.7$	$I_N = 0.8$	$I_N = 0.9$	$I_N = 1$
17.39	18.92	20.51	22.14	23.81	25.50	27.22	28.96	30.72	32.49	34.27
17.72	17.77	17.83	17.91	18.01	18.13	18.27	18.42	18.60	18.79	18.99
15.92	15.93	15.98	16.04	16.13	16.24	16.38	16.53	16.71	16.91	17.13
7.30	7.22	7.16	7.10	7.06	7.04	7.0296	7.0265	7.0396	7.0655	7.1044
24.93	24.89	24.86	24.8338	24.8061	24.7815	24.7602	24.7422	24.7273	24.7157	24.7074
5.2448	5.2324	5.2340	5.2497	5.2794	5.3228	5.3796	5.4493	5.5316	5.6257	5.7313
14.6642	14.6581	14.6522	14.6465	14.6410	14.6357	14.6306	14.6257	14.6210	14.6165	14.6122
5.6264	5.6115	5.5971	5.5832	5.5698	5.5568	5.5443	5.5323	5.5208	5.5098	5.4992
10.8529	10.8483	10.8440	10.8400	10.8362	10.8328	10.8295	10.8266	10.8239	10.8215	10.8194
13.0138	13.0042	12.9948	12.9855	12.9765	12.9677	12.9591	12.9508	12.9426	12.9346	12.9269

Table 7. Effect on F-value when $n = 15$

$I_N = 0$	$I_N = 0.1$	$I_N = 0.2$	$I_N = 0.3$	$I_N = 0.4$	$I_N = 0.5$	$I_N = 0.6$	$I_N = 0.7$	$I_N = 0.8$	$I_N = 0.9$	$I_N = 1$
0.9513	0.8912	0.8445	0.8080	0.7793	0.7566	0.7385	0.7240	0.7123	0.7027	0.6949
0.3462	0.3195	0.2949	0.2727	0.2528	0.2353	0.2202	0.2075	0.1971	0.1891	0.1833
0.3154	0.2888	0.2629	0.2379	0.2140	0.1914	0.1702	0.1505	0.1324	0.1159	0.1011
1.5817	1.5953	1.6038	1.6068	1.6040	1.5953	1.5805	1.5598	1.5335	1.5018	1.4653
0.6803	0.6729	0.6655	0.6582	0.6509	0.6437	0.6366	0.6295	0.6225	0.6157	0.6089
0.3131	0.3772	0.4485	0.5257	0.6075	0.6925	0.7793	0.8663	0.9525	1.0366	1.1176
1.8072	1.8022	1.7971	1.7920	1.7858	1.7816	1.7763	1.7710	1.7657	1.7603	1.7548
0.0166	0.0169	0.0173	0.0178	0.0185	0.0192	0.0201	0.0211	0.0221	0.0234	0.0247
5.3652	5.3574	5.3453	5.3409	5.3323	5.3234	5.3142	5.3048	5.2951	5.2852	5.2750
1.7479	1.7462	1.7446	1.7428	1.7411	1.7393	1.7374	1.7355	1.7336	1.7316	1.9296

5. Comparative studies

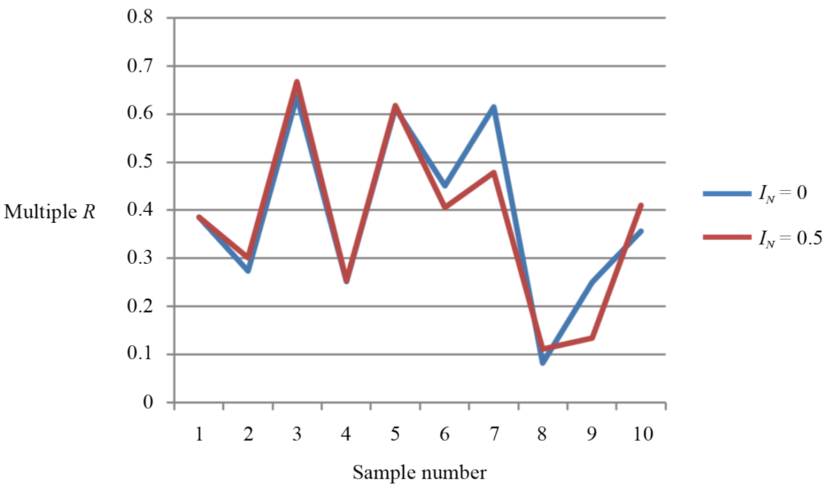

Figure 1. Multiple R curves from two regressions when $n = 10$

In this section, the aim is to compare the results from classical statistics and neutrosophic statistics, focusing on multiple R, standard error, and F-test values obtained from both traditional and proposed multiple regression analyses. The proposed method is essentially an expansion of classical regression analysis, designed to handle data with uncertain or imprecise observations. In cases without such data issues, the proposed method aligns with traditional classical statistical methods. For clarity and fairness in comparison, it has presented the results of the traditional regression analysis in the first column of Tables 2-7. These tables highlight differences in the values of multiple R, standard error, and F-test between the classical and the proposed neutrosophic statistical approaches. For instance, in Table 2 (second last row), the multiple R value is 0.2499 when $I_N = 0$ and decreases to 0.1939 when $I_N = 0.2$. Similarly, in Table 3 (fifth row), the standard error is 16.05 for $I_N = 0$ and slightly lower at 15.9651 for $I_N = 0.2$. Likewise, in Table 4 (last row), the F-test value is 0.5106 for $I_N = 0$ and increases to 0.6051 for $I_N = 0.2$. Additionally, Figures 1-3 illustrate the trends in multiple R, standard error, and F-test values. These figures distinctly showcase differences between the outcomes derived from classical statistical analysis and those obtained through neutrosophic statistical analysis. This study suggests that utilizing multiple regressions in both certain and uncertain environments can yield varied interpretations of the model. Employing multiple regression analysis in scenarios with imprecise or uncertain observations might misguide decision-makers in their predictions and forecasts. Consequently, it is advised that decision-makers exercise caution when employing regression models in the presence of uncertainty.

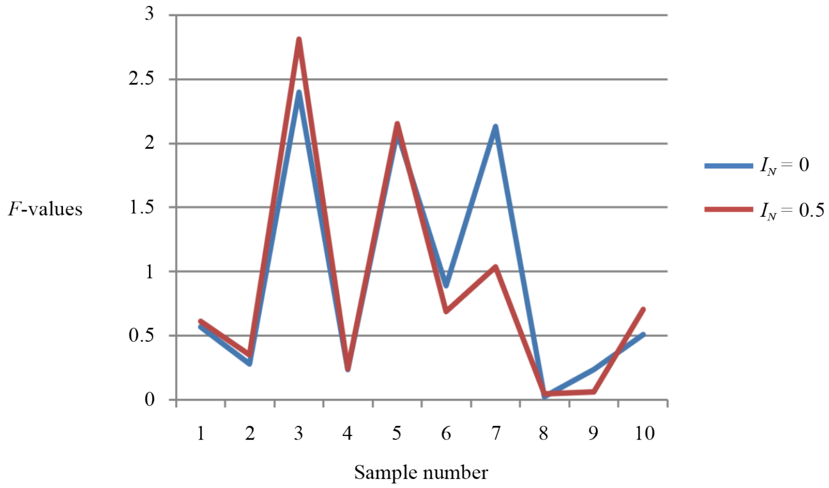

Figure 2. Standard error curves from two regressions when $n = 10$

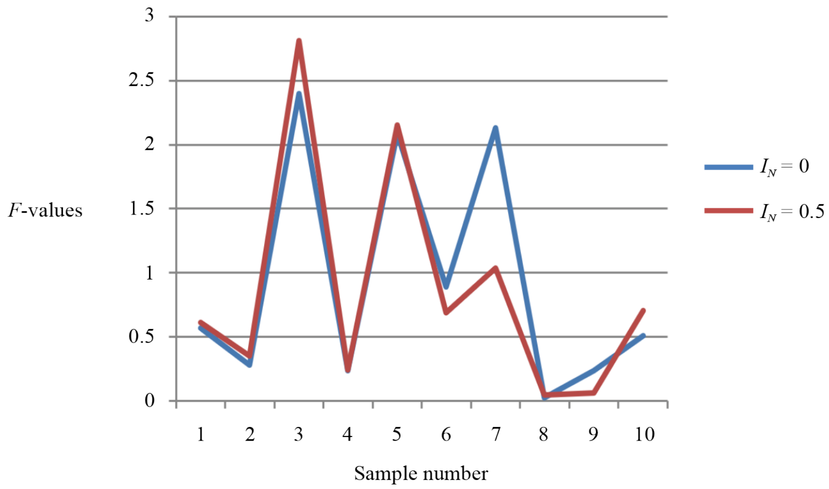

Figure 3. F-values curves from two regressions when $n = 10$

6. Application

In this segment, the aim is to delve into the practical application of the proposed multiple regression model using authentic data sourced from the realm of education. The neutrosophic dataset it is utilized to obtained from the work of [17]. This dataset comprises student interviews focusing on aptitude and personality tests, as well as the measurement of conscientiousness levels among students. For clarity, it is designated the neutrosophic variable representing conscientiousness as Y_N, the aptitude test as X_{N1}, and the personality test as X_{N2}. The data is meticulously outlined in Table 8, with determinate values sourced directly from [17] and indeterminate values computed assuming

$I_N = 0.2$. Subsequently, a comprehensive analysis of the dataset is conducted, encapsulating that the current findings in Table 9, and perform NANOVA as detailed in Table 10. Additionally, the residual output is presented in Table 11. To visually depict the regression models, showcase the predicted values curve under classical statistics alongside the proposed regression model is shown in Figure 4, followed by a comparison of residual plots in Figure 5. Upon scrutiny of Table 9, slight fluctuations are observed in the multiple R values (from 0.9322 to 0.9317), R square values (from 0.8692 to 0.8691), and standard error values (from 1.7542 to 2.1231). Similarly, Table 10 illustrates marginal variations in F-test values (from 23.25 to 23.02). Notably, Table 11 and Figures 4 and 5 underscore disparities between predicted and residual values in both regression models. This investigation underscores a crucial insight: relying solely on deterministic analysis in the presence of underlying uncertainty can misguide decision-makers in their predictions and forecasts. Hence, decision-makers must exercise caution when applying regression models, acknowledging and accounting for uncertainty to mitigate its impact on forecasting and estimation accuracy.

Table 8. The neutrosophic data

Y_N	X_{N1}	X_{N2}
[1, 1.2]	[3, 3.6]	[2, 2.4]
[2, 2.4]	[2, 2.4]	[2, 2.4]
[2, 2.4]	[1, 1.2]	[3, 3.6]
[4, 4.8]	[2, 2.4]	[4, 4.8]
[1, 1.2]	[2, 2.4]	[4, 4.8]
[6, 7.2]	[2, 2.4]	[4, 4.8]
[2, 2.4]	[2, 2.4]	[1, 1.2]
[10, 12]	[5, 6]	[6, 7.2]
[14, 16.8]	[7, 8.4]	[8, 9.6]
[5, 6]	[7, 8.4]	[3, 3.6]

Table 9. Summary output

Regression statistics	
Multiple R	[0.9322, 0.9317]
R square	[0.8692, 0.8681]
Adjusted R square	[0.8318, 0.8304]
Standard error	[1.7542, 2.1231]
Observations	10

Table 10. NANOVA

	df	SS	MS	F	Significance F
Regression	2	[143.11, 207.62]	[71.55, 103.8]	[23.25, 23.02]	[0.00080, 0.00083]
Residual	7	[21.54, 31.55]	[3.07, 4.51]		
Total	9	[164.65, 239.18]			

Table 11. Residual output

Observation	Predicted YN	Residuals
1	[2.0485, 2.4112]	[−0.85, −1.21]
2	[1.4146, 1.6568]	[0.58, 0.74]
3	[2.2402, 2.6685]	[−0.24, −0.27]
4	[4.3337, 5.1891]	[−0.34, −0.39]
5	[4.3337, 5.1891]	[−3.33, −3.99]
6	[4.3337, 5.1891]	[1.67, 2.01]
7	[−0.0449, −0.1093]	[2.04, 2.51]
8	[9.1546, 10.9846]	[0.85, 1.02]
9	[13.3416, 16.0257]	[0.66, 0.77]
10	[6.0438, 7.1950]	[−1.04, −1.20]

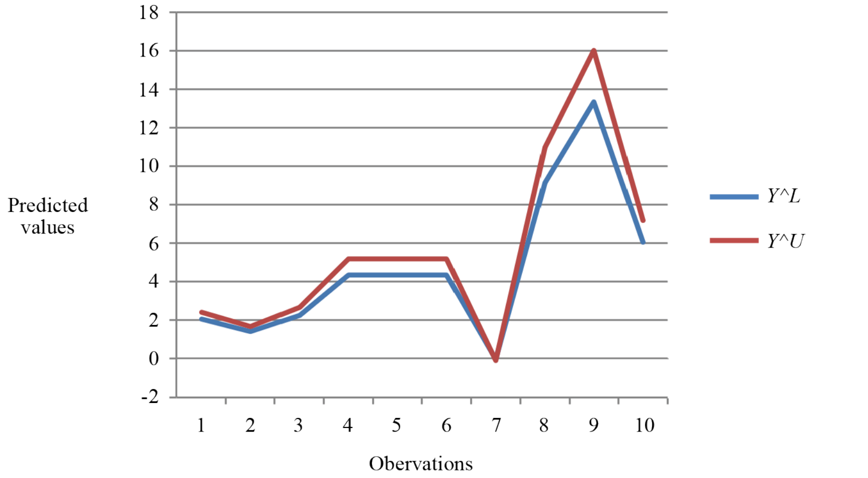

Figure 4. Predicted values from two regression

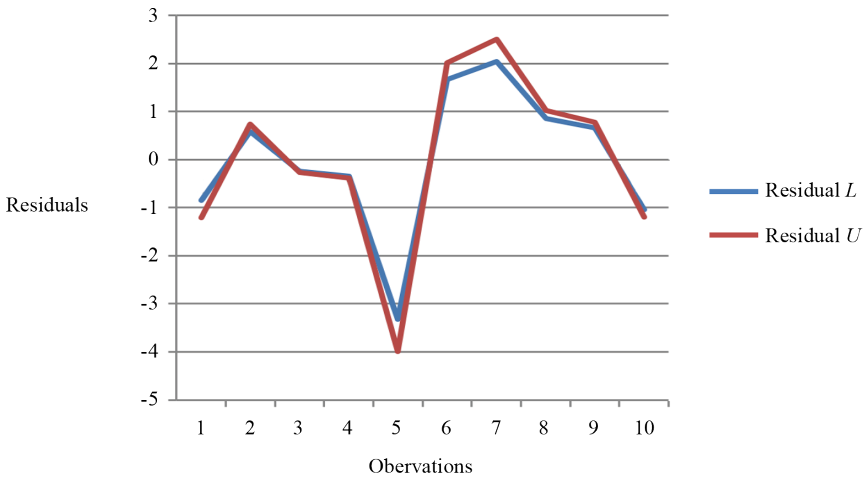

Figure 5. Residuals values of two regression model

7. Concluding remarks

In this manuscript, multiple regression models within the framework of neutrosophic statistics were introduced. The key metrics pertinent to the proposed regression model within this statistical paradigm were highlighted. Through extension simulation studies, it was demonstrated how varying degrees of uncertainty impact metrics such as multiple R, standard error, and F-test values. Additionally, by examining a real-world example, it is ascertained that uncertainty levels significantly influence the predicted values derived from the regression model, as well as the residual values. The analysis reveals that the outcomes of regression modeling can diverge under specific conditions when compared to traditional regression analysis conducted within neutrosophic statistics. Consequently, it is advocated for the application of the proposed multiple regression method in scenarios where data uncertainty is prevalent. This methodology holds promise across diverse domains including business, political science, banking, medical science, neural networks, and artificial intelligence. Furthermore, it is suggested to explore the potential application of the proposed method in logistic regression as a promising avenue for future research.

Acknowledgement

The authors are deeply thankful to the editor and reviewers for their valuable suggestions to improve the quality and presentation of the paper. The authors would like to acknowledge the Deanship of Graduate Studies and Scientific Research, Taif University for funding this work. We acknowledge the use of ChatGPT to improve the clarity and grammar of the manuscript's English language.

Conflict of interest

The authors declare no competing financial interest.

References

[1] Eyduran E, Topal M, Sonmez AY. Use of factor scores in multiple regression analysis for estimation of body weight by several body measurements in brown trouts (Salmo trutta fario). *International Journal of Agriculture and Biology*. 2010; 12(4): 611-615.

[2] Kang H, Zhao H. Description and application research of multiple regression model optimization algorithm based on data set denoising. *Journal of Physics: Conference Series*. 2020; 1631: 012063. Availavle from: https://doi.org/10.1088/1742-6596/1631/1/012063.

[3] Ruiz E, Freyne S. Multiple regression model for load rating of reinforced concrete bridges. *Transportation Research Record*. 2020; 2674(7): 361-372.

[4] Trunfio TA, Scala A, Giglio C, Rossi G, Borrelli A, Romano M, et al. Multiple regression model to analyze the total LOS for patients undergoing laparoscopic appendectomy. *BMC Medical Informatics and Decision Making*. 2022; 22(1): 141. Availavle from: https://doi.org/10.1186/s12911-022-01884-9.

[5] Dorta-González P. A multiple linear regression analysis to measure the journal contribution to the social attention of research. *Axioms*. 2023; 12(4): 337. Availavle from: https://doi.org/10.3390/axioms12040337.

[6] Sun Y, Wang X, Zhang C, Zuo M. Multiple regression: methodology and applications. *Highlights in Science, Engineering and Technology*. 2023; 49: 542-548. Availavle from: https://doi.org/10.54097/hset.v49i.8611.

[7] Moxley TA, Johnson-Leung J, Seamon E, Williams C, Ridenhour BJ. Application of elastic net regression for modeling COVID-19 sociodemographic risk factors. *PLoS One*. 2024; 19(1): e0297065. Availavle from: https://doi.org/10.1371/journal.pone.0297065.

[8] Smarandache F. Neutrosophic set is a generalization of intuitionistic fuzzy set, inconsistent intuitionistic fuzzy set (picture fuzzy set, ternary fuzzy set), pythagorean fuzzy set, spherical fuzzy set, and q-rung orthopair fuzzy set,

while neutrosophication is a generalization of regret theory, grey system theory, and three-ways decision (revisited). *Journal of New Theory*. 2019; 29: 1-31.

[9] Rezaei A, Oner T, Katican T, Smarandache F, Gandotra N. A short history of fuzzy, intuitionistic fuzzy, neutrosophic and plithogenic sets. *International Journal of Neutrosophic Science*. 2022; 18(1): 99-116. Available from: https://doi.org/10.54216/IJNS.180109.

[10] Smarandache F. Neutrosophic statistics is an extension of interval statistics, while plithogenic statistics is the most general form of statistics. *International Journal of Neutrosophic Science*. 2022; 19(1): 148-165. Available from: https://doi.org/10.54216/IJNS.190111.

[11] Mohammad SI, Yogeesh N, Raja N, Chetana R, William P, Vasudevan A, et al. The synergy of simplicity and vagueness: exploring simple statistics in fuzzy mathematical frameworks. *Applied Mathematics*. 2025; 19(2): 457-465. Available from: http://dx.doi.org/10.18576/amis/190219.

[12] Smarandache F. *Introduction to Neutrosophic Statistics*. Romania: Educational Publisher; 2014.

[13] Smarandache F. Neutrosophic statistics is an extension of interval statistics, while plithogenic statistics is the most general form of statistics. *Neutrosophic Computing and Machine Learning*. 2022; 23: 21-38.

[14] Chen J, Ye J, Du S. Scale effect and anisotropy analyzed for neutrosophic numbers of rock joint roughness coefficient based on neutrosophic statistics. *Symmetry*. 2017; 9(10): 208. Available from: https://doi.org/10.3390/sym9100208.

[15] Chen J, Ye J, Du S, Yong R. Expressions of rock joint roughness coefficient using neutrosophic interval statistical numbers. *Symmetry*. 2017; 9(7): 123. Available from: https://doi.org/10.3390/sym9070123.

[16] Duan WQ, Khan Z, Gulistan M, Khurshid A. Neutrosophic exponential distribution: modeling and applications for complex data analysis. *Complexity*. 2021; 2021: 1-8. Available from: https://doi.org/10.1155/2021/5970613.

[17] Nagarajan D, Broumi S, Smarandache F, Kavikumar J. Analysis of neutrosophic multiple regression. *Neutrosophic Sets and Systems*. 2021; 43: 44-53.

[18] AlAita A, Talebi H, Albassam M, Sultan KA. Neutrosophic statistical analysis of split-plot designs. *Soft Computing*. 2023; 27(12): 7801-7811. Available from: https://doi.org/10.1007/s00500-023-08025-y.

[19] AlAita A, Aslam M. Analysis of covariance under neutrosophic statistics. *Journal of Statistical Computation and Simulation*. 2023; 93(3): 397-415. Available from: https://doi.org/10.1080/00949655.2022.2108423.

[20] Shahzadi I. Neutrosophic statistical analysis of temperature of different cities of Pakistan. *Neutrosophic Sets and Systems*. 2023; 53: 10.

[21] Mushtaq N, Butt AS. Joint parameter estimation in measurement errors and non-response for sensitive variables. *SCOPUA Journal of Applied Statistical Research*. 2025; 1(2). Available from: https://doi.org/10.64060/JASR.v1.i2.1.

[22] Khali R, AlAita A, Zai YA. The exact analysis of augmented incomplete Latin square design with one missing observation. *SCOPUA Journal of Applied Statistical Research*. 2025; 1(2). Available from: https://doi.org/10.64060/JASR.v1.i2.4.

[23] Granados C. Some discrete neutrosophic distributions with neutrosophic parameters based on neutrosophic random variables. *Hacettepe Journal of Mathematics and Statistics*. 2022; 51(5): 1442-1457. Available from: https://doi.org/10.15672/hujms.1099081.

[24] Aslam M. The neutrosophic negative binomial distribution: algorithms and practical application. *REVSTAT-Statistical Journal*. 2024. Available from: https://doi.org/10.57805/revstat.vi.781.