NSS
Neutrosophic Sets and Systems, Vol. 66, 2024
University of New Mexico

Analyzing the Income-Education Nexus in Ecuador: A Neutrosophic Statistical Approach

Guido Macas-Acosta[1], Fidel Márquez-Sánchez [2*], Arnaldo Vergara-Romero[3], Jesús Estupiñán Ricardo[4]

[1] Universidad ECOTEC; Guayaquil, Ecuador, gmacas@ecotec.edu.ec
[2] Universidad de Especialidades Espíritu Santo, Ecuador; fmarquez@uees.edu.ec
[3] Universidad ECOTEC; Guayaquil, Ecuador, avergarar@ecotec.edu.ec
[4] Universidad Regional Autónoma de Los Andes, Ambato. Ecuador, ua.jesusestupinan@uniandes.edu.ec

* Correspondence: fmarquez@uees.edu.ec

Abstract: This study examines the relationship between education and entrepreneur income in Ecuador, a prototypical small economy grappling with employment challenges. Utilizing quantile regression and neutrosophic statistical methods, the research uncovers varied impacts of education on earnings across different income levels. The findings reveal a decrease in indeterminacy as individuals ascend income quintiles, indicating that higher education levels yield more predictable and substantial economic returns at the upper end of the income spectrum. These insights are vital for policymakers to develop targeted educational interventions. The study's methodological novelty lies in its adoption of neutrosophic statistics, which embraces the indeterminacy intrinsic to economic data, offering a refined lens for understanding the income-education interplay. Future research directions include longitudinal studies to trace the temporal effects of education and cross-comparative analyses across economies. The potential integration of machine learning with neutrosophic statistics promises enhanced predictive models, contributing to data-driven economic policy formulation.

Keywords: Neutrosophic Statistics, Quantile Regression, Education, Entrepreneur Income, Economic Policy

1. Introduction

In the dynamic landscape of global economies, small economies face the pressing challenge of generating sufficient employment to sustain their populations. This scarcity of traditional job opportunities compels individuals, irrespective of age or gender, to seek alternative income sources to meet their basic survival needs. Amidst this backdrop, entrepreneurship and self-employment have emerged as significant avenues for employment, accounting for 33.3% of job occupation worldwide, trailing behind paid employment (53.4%), which includes positions in both the private and public sectors as reported by the International Labor Organization [1]. In Ecuador, a country representative of these small economies, understanding the dynamics that influence entrepreneur income is critical for economic development and policy formulation [2].

This study delves into the intricacies of the income-education nexus within Ecuador, employing a neutrosophic statistical approach [3] to uncover the principal determinants of entrepreneur income. By leveraging quantile regression, we aim to dissect the varying impacts of independent variables across different distribution points, providing insights into the heterogeneity of effects. This methodological choice is particularly pertinent in the presence of heteroscedasticity [4], a condition where error variances are not constant, thereby allowing for a nuanced analysis that traditional regression models might overlook.

Furthermore, the application of neutrosophic statistics introduces a novel perspective to our analysis. This theoretical framework operates on the premise that models and parameters should be defined over intervals rather than fixed numerical values, offering a unique advantage in terms of flexibility and adaptability [5]. By embracing the indeterminacy inherent in economic data [6] and modeling, we aspire to achieve a more accurate depiction of the income-education relationship in Ecuador, thereby contributing to a deeper understanding of economic phenomena in small economies.

The paper continues with a description of the methodologies employed in the investigation. Subsequently, the outcomes of the analysis are presented showing the impact of socioeconomic determinants such as educational attainment, marital status, gender, and geographical location on the income distribution among entrepreneurs. Following the presentation of results, the discussion segment interprets the findings. Finally, conclusions consolidate the principal discoveries and propose future areas of investigation.

2. Some notions on Neutrosophic Statistics

This section introduces foundational aspects of neutrosophic statistics, an extension of classical statistics that utilizes set values instead of precise numerical values. It encompasses Neutrosophic Descriptive Statistics, which summarize and delineate the characteristics of neutrosophic numerical data, and Neutrosophic Inferential Statistics, allowing for extrapolation from a neutrosophic sample to its broader population [7,8].

Neutrosophic Data embodies indeterminacy, with discrete neutrosophic data represented by distinct points and continuous neutrosophic data by one or multiple intervals. This data can be quantitative, such as an uncertain number within an interval, or qualitative, like color identification with uncertainty. Observations can be univariate, focusing on a single neutrosophic attribute, or multivariate, encompassing multiple attributes [9].

A Neutrosophic Statistical Number [10] consists of a determinate part (d) and an indeterminate part (I), expressed as N = d + I. The concept of Neutrosophic Frequency Distribution arises from categorizing frequencies and relative frequencies amidst indeterminacy, often stemming from imprecise or unknown data, thus affecting the precision of relative frequencies. Similarly, Neutrosophic Survey Results and Populations incorporate indeterminacy, impacting the certainty of membership within a population.

Sampling methods, like simple random neutrosophic sampling and stratified random neutrosophic sampling, adapt to this framework by accounting for indeterminacy in sample selection. Unlike interval statistics, which accumulate uncertainty from one operation to the next, neutrosophic statistics reduce or even eliminate uncertainty. The paper also delves operations between neutrosophic numbers, with operations defined as follows [11]:

Addition $(N_1 + N_2) = a_1 + a_2 + (b_1 + b_2)I$

Subtraction $(N_1 - N_2) = a_1 - a_2 + (b_1 - b_2)I$

Multiplication $(N_1 \times N_2) = a_1a_2 + (a_1b_2 + b_1a_2 + b_1b_2)I$

Division $\frac{N_1}{N_2} = \frac{a_1+b_1I}{a_2+b_2I} = \frac{a_1}{a_2} + \frac{a_2b_1-a_1b_2}{a_2(a_2+b_2)}I$

In machine learning, intervals play a crucial role in gauging the level of uncertainty tied to estimates and predictive analyses. Prediction intervals [12] set probabilistic boundaries, both upper and lower, for the anticipated outcome based on a predetermined confidence level, such as 95%. This suggests that, typically, 95 of every 100 future observations should fall within these boundaries. On the other hand, confidence intervals [13] aim to capture the uncertainty of an estimate to demonstrate a machine learning algorithm's effectiveness on unfamiliar data. Differing from prediction intervals that delineate the range for a single data point, confidence intervals define the expected range for a population metric, like the average.

2. Materials and Methods

The dataset in this research comes from the National Survey of Employment, Unemployment, and Underemployment (ENEMDU), carried out by the Ecuadorian Institute of Statistics and Census (INEC) in December 2021. The initial data was processed and filtered to obtain the most efficient and consistent estimates. The database was configured with a two-stage, stratified, and cluster sampling design to obtain population estimators. Then, individuals between the ages of 15 and 80 are selected, and their job occupation is classified as self-employed workers, which is considered a business or enterprise.

This research will focus on quantile regression [14] to measure this relationship. It involves fitting separate line segments, restricted by a quartile or quintile, that account for nonlinearity between the predictor and the outcome.

The next step is quantile regression to predict the earnings (lingemp) logarithm in the different quartiles. It will produce a model for various quintiles or quartiles ranging from 5 to 95. The following expression expresses the model:

$$lingem_i = \delta eduemp_i + \vartheta edaemp_i + \gamma exp_i + \theta sex_i + \varphi urbano_i + u_i \qquad (1)$$

lingem, the logarithm of income of entrepreneurs.
eduemp, years of entrepreneurship education.
edaemp, age of entrepreneurs.
exp, work experience in entrepreneurship.
sexo, gender of entrepreneurs.
urbano, urban location.
u, is the stochastic disturbance term or random error.

With quantile regression, it is possible to observe if there are differences in the effects of the independent variables depending on the point of the distribution that is analyzed. Furthermore, this type of regression is helpful in the presence of heteroscedasticity. It occurs when the variance of the errors is not constant.

A quantile regression is estimated that establishes the relationship between the variation in income and a series of explanatory variables for seven quintiles of the endogenous variable with the following specifications:

$$lingem_{i1} = \beta_{0.05} + \delta_{0.05} eduemp_{i1} + \alpha_{0.05} edaemp_{i1} + \gamma_{0.05} exp_{i1} + \theta_{0.05} sex_{i1} + \varphi_{0.05} urbano_{i1} + u_{i1} \quad (2)$$

$$lingem_{i2} = \beta_{0.10} + \delta_{0.10} eduemp_{i2} + \alpha_{0.10} edaemp_{i2} + \gamma_{0.10} exp_{i2} + \theta_{0.10} sex_{i2} + \varphi_{0.10} urbano_{i2} + u_{i2} \quad (3)$$

$$lingem_{i3} = \beta_{0.25} + \delta_{0.25} eduemp_{i3} + \alpha_{0.25} edaemp_{i3} + \gamma_{0.25} exp_{i3} + \theta_{0.25} sex_{i3} + \varphi_{0.25} urbano_{i3} + u_{i3} \quad (4)$$

$$lingem_{i4} = \beta_{0.50} + \delta_{0.50} eduemp_{i4} + \alpha_{0.50} edaemp_{i4} + \gamma_{0.50} exp_{i4} + \theta_{0.50} sex_{i4} + \varphi_{0.50} urbano_{i4} + u_{i4} \quad (5)$$

$$lingem_{i5} = \beta_{0.75} + \delta_{0.75} eduemp_{i5} + \alpha_{0.75} edaemp_{i5} + \gamma_{0.75} exp_{i5} + \theta_{0.75} sex_{i5} + \varphi_{0.75} urbano_{i5} + u_{i5} \quad (6)$$

$$lingem_{i6} = \beta_{0.90} + \delta_{0.90} eduemp_{i6} + \alpha_{0.90} edaemp_{i6} + \gamma_{0.90} exp_{i6} + \theta_{0.90} sex_{i6} + \varphi_{0.90} urbano_{i6} + u_{i6} \quad (7)$$

$$lingem_{i7} = \beta_{0.95} + \delta_{0.95} eduemp_{i7} + \alpha_{0.95} edaemp_{i7} + \gamma_{0.95} exp_{i7} + \theta_{0.95} sex_{i7} + \varphi_{0.95} urbano_{i7} + u_{i7} \quad (8)$$

In neutrosophic statistics, the confidence interval can be interpreted as an indicator of indeterminacy. In this context, Equation 9 defines the indeterminacy of an interval Im as the difference between the upper limit [15]:

$$\gamma(\text{Im}) = a2 - a1 \quad (9)$$

where Im = [a1, a2] be an interval and a2, a2 are lower and upper bounds of the confidence interval, respectively

The measure γ(Im) quantifies how much we do not know about the parameter we are estimating. The greater this value, the more indeterminacy there is. Graphically, a wider confidence band (a larger shaded area) would indicate more uncertainty in the estimation of the coefficients at that specific quantile [16].

3. Results

Quantile regression was developed (Table 1) to delve into how the returns on various factors such as education, marital status, sex, and location differ across the income distribution of entrepreneurs.

Table 11. Result of the Models with Logarithm of the Dependent Variable.

Variables	OLS	QR_05	QR_10	QR_25	QR_50	QR_75	QR_90	QR_95
eduemp	0.056***	0.049***	0.0477***	0.052***	0.054***	0.056***	0.062***	0.067***
	(0.001)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
edaemp	0.079***	0.102***	0.0951***	0.091***	0.077***	0.064***	0.058***	0.065***
	(0.002)	(0.005)	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
edaemp2	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
exp	0.003***	0.006***	0.006***	0.005***	0.003***	0.002***	0.003***	0.003***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
sex	0.547***	0.854***	0.848***	0.730***	0.513***	0.352***	0.265***	0.278***
	(0.009)	(0.021)	(0.016)	(0.013)	(0.010)	(0.009)	(0.012)	(0.017)
married	0.115***	0.059***	0.078***	0.103***	0.147***	0.151***	0.166***	0.180***
	(0.009)	(0.025)	(0.021)	(0.013)	(0.010)	(0.009)	(0.012)	(0.017)
urban	0.300***	0.460***	0.423***	0.378***	0.315***	0.222***	0.165***	0.163***
	(0.010)	(0.027)	(0.022)	(0.014)	(0.012)	(0.011)	(0.014)	(0.18)
_cons	2.295***	0.023***	0.632***	1.393***	2.447***	3.374***	3.938***	3.996***
	(0.043)	(0.1113)	(0.087)	(0.061)	(0.050)	(0.047)	(0.057)	(0.081)
Observations	55.857	55.857	55.857	55.857	55.857	55.857	55.857	55.857
R-squared	0.1982	0.1124	0.1172	0.1233	0.1106	0.0963	0.0927	0.0965

The main findings are summarized below:

- Education: The study found that the benefit of an additional year of education on earnings varies across the income distribution but remains significant across all quintiles. Notably, higher returns on education are evident at the higher end of the income spectrum. In the 5th quintile, an additional year of education is associated with an increase in income by $1.84, while in the 95th quintile, the increase is $52.45.

- Marital Status: Being married or in a common-law union positively influences earnings across the distribution. The effect is more pronounced at the upper end, where partnered entrepreneurs earn 12.1% more than their single peers in the 95th quintile.
- Gender: The coefficient analysis reveals that male entrepreneurs consistently out-earn their female counterparts across all educational quintiles, with a substantial 80% higher earnings at the lower conditional quintiles.
- Urban Location: Entrepreneurs located in urban areas have a clear income advantage over those in rural areas, with earnings between 10% to 45% higher, showcasing the importance of geographical location in economic success.
- Age: Age shows a positive correlation with earnings, but its impact diminishes in the higher income quintiles, indicating that younger entrepreneurs tend to earn more in the lower quintiles of the distribution.
- Experience: While experience contributes to income, its impact is relatively minor, not exceeding 1%.
- Coefficients and Confidence Intervals: The study also highlights that the constants and slopes of these relationships significantly differ across quintiles, with a steeper slope observed from the 75th quintile upwards. This suggests an increasing return on these variables, especially education, in higher income brackets.

The quantile regression plot (Figure 1) depicts the impact of years of entrepreneurship education, labeled as "eduemp," on the dependent variable across different points of its distribution. The Y-axis measures the size of the "eduemp" coefficient, indicating the strength and direction of its association with the outcome variable at various quantiles, displayed on the X-axis, ranging from the lowest (0) to the highest (1).

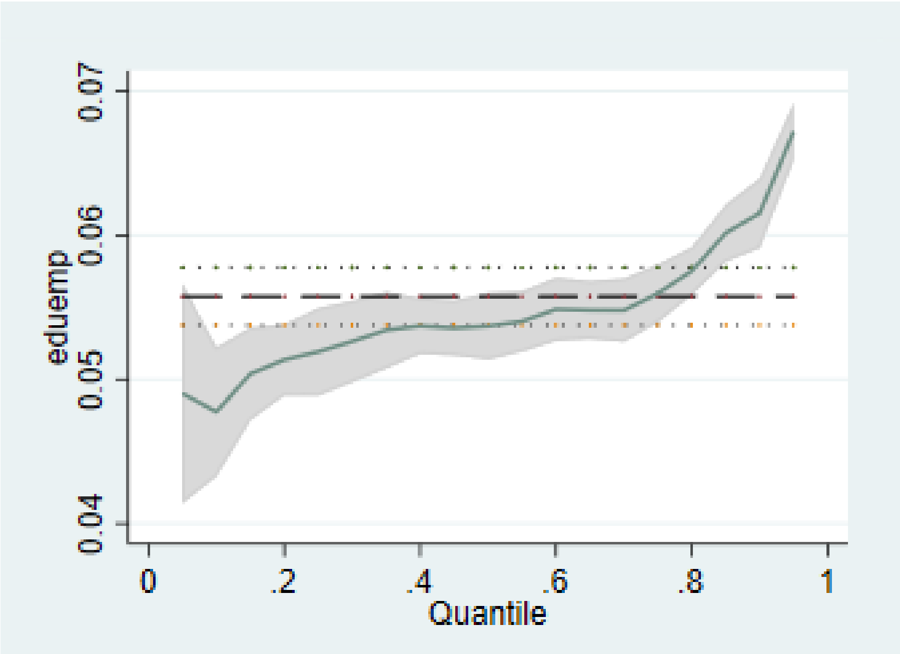

Figure 1. Quantile Regression of Entrepreneurship Education Years ('eduemp') on Outcome Variable

Indeterminacy is calculated (Figure 2) and normalized to observe the relative variability and trends more clearly in the data across various quintiles.

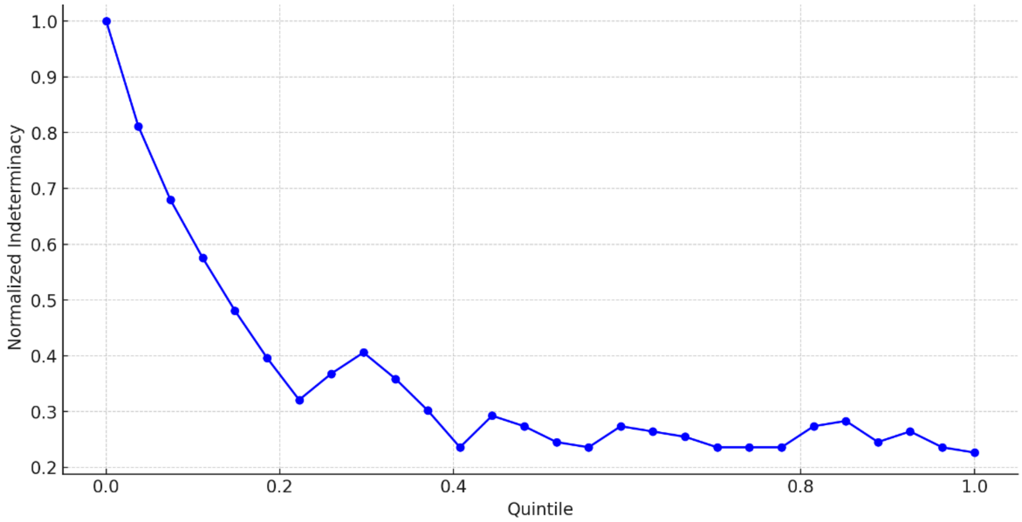

Figure 2. Normalized Indeterminacy Across Quintiles

The pattern depicted in the figure reveals a decline in indeterminacy, hinting at a progressively uniform influence of educational attainment on earnings as one ascends the income quintiles. This visual evidence underscores a pivotal consideration: the impact of education on income is not uniform across the spectrum. Consequently, this necessitates a nuanced approach in policy-making that acknowledges and addresses the variable effects of educational investment across different income segments.

4. Discussion

A few key observations can be drawn and related to the findings on education's impact on earnings and the calculation of indeterminacy [17, 18, 19]:

1. Decreasing Indeterminacy with Income: The graph shows a steep initial decline in indeterminacy, which then stabilizes as it progresses through the quintiles. This could suggest that at lower income levels, the impact of additional education on earnings is more variable. As we move up the income distribution, the effect of education becomes more predictable, with higher quintiles showing less indeterminacy and hence a more consistent return on investment in education.
2. Higher Returns at Higher Income Levels: The research indicated that additional education has higher returns at the upper end of the income spectrum, which corresponds to the lower indeterminacy at higher quintiles seen in the graph. This relationship suggests a positive correlation between the certainty of returns on education and the income level, reinforcing the finding that additional education is a more significant predictor of income increases for those already at the higher end of the income distribution.
3. Implications for Neutrosophic Statistics: In the framework of neutrosophic statistics, the varying levels of indeterminacy across quintiles can be seen as an opportunity to explore the elements of truth, falsity, and indeterminacy in the data. Neutrosophic statistics could be used to model these elements more explicitly, potentially offering richer insights into the nuanced way education impacts earnings.

For instance, where indeterminacy is high, the neutrosophic approach could help to uncover the underlying factors contributing to this uncertainty, such as differing quality of education, the varying economic value of different fields of study, or the impact of network effects and social capital that might not be captured in a traditional statistical model[20]. Conversely, at higher quintiles, where there is less indeterminacy, neutrosophic statistics could help validate the robustness of the observed trends, confirming the strong positive impact of additional education on earnings for individuals in these segments.

Incorporating neutrosophic statistics into this analysis could thus enhance the understanding of educational returns [21] across income levels, enabling more targeted and effective educational policies and interventions that are responsive to the observed heterogeneity in the data [22].

5. Conclusions

The research presented in this paper reveals a pattern of decreasing indeterminacy with ascending income quintiles, indicating that the influence of educational attainment on earnings stabilizes at higher income levels. This suggests that the economic returns on additional education are more predictable and potentially greater for individuals in the upper levels of the income distribution. These findings are instrumental for policymakers, as they highlight the necessity of crafting nuanced educational policies that consider the heterogeneous effects of education across the income spectrum. By focusing on the areas where educational investment yields substantial benefits, policymakers can better allocate resources to enhance economic outcomes.

Looking ahead, the application of neutrosophic statistics in this study introduces an innovative methodology that more accurately reflects the uncertainties present in economic data. Prospective research could build on this foundation by applying a neutrosophic framework to analyze other socio-economic factors, thereby unraveling the intricate factors influencing income. Future initiatives could include longitudinal analyses to track the long-term effects of education on earnings or cross-country comparisons to understand these dynamics in various economic contexts. Moreover, integrating machine learning with neutrosophic statistical methods could lead to the development of more sophisticated and flexible predictive models, thus equipping policymakers with advanced tools for fostering economic growth and development.

Funding: This research received no external funding.

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

References

[1]. ILO, International Labour Organization. Social and Employment Perspectives in the World. Geneva: International Labour Organization, 2022. Accessed on January 17, 2024. Available at: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_848464.pdf.

[2]. Pérez-Campdesuñer, R., García-Vidal, G., Sánchez-Rodríguez, A., Martínez-Vivar, R., de Miguel-Guzmán, M., & Guilarte-Barinaga, E. Influence of the socio-economic environment on the entrepreneurs behavior. Cases of Cuba and Ecuador. International Journal of Engineering Business Management 2021, 13, 1847979021994509.

[3]. Smarandache, F. Plithogenic Probability & Statistics are generalizations of MultiVariate Probability & Statistics. Neutrosophic Sets and Systems 2021, 43(1), 20.

[4]. Guan, H.; Guan, S.; Zhao, A. Forecasting model based on neutrosophic logical relationship and Jaccard similarity. Symmetry 2017, 9(9), 191.

[5]. Kandemir, H. Ş.; Aral, N. D.; Karakaş, M.; Et, M. Neutrosophic Statistical Analysis of Temperatures of Cities in the Southeastern Anatolia Region of Turkey. Neutrosophic Systems with Applications 2024, 14, 50-59.

[6]. Moscoso-Paucarchuco, K.; Raiser Vásquez-Ramírez, M.; Yupanqui-Villanueva, H. R.; Yupanqui-Villanueva, W. F.; Vivanco-Nuñez, O. A.; Beraún-Espíritu, M. M.; Fernández-Jaime, R. J.; Gutiérrez-Gómez, E. Study of the relationship among economic variables in cattle production in a region of Peru, based on Plithogenic Statistics. Neutrosophic Sets and Systems 2024, 65, 1-10. DOI: 10.5281/zenodo.10729952.

[7]. Patro, S. K.; Smarandache, F. The Neutrosophic Statistical Distribution - More Problems, More Solutions. Neutrosophic Sets and Systems 2016, 12, 73-79. doi:10.5281/zenodo.571153
[8]. Reales Chacón, L.; Campos Moposita, A.; Marquina Amón, D.; Garcia Camacho, V.; Bonilla Ayala, J.; Caiza Lema, J.; Aguilar Zapata, Y.; Cantuña Vallejo, P. Study of the impact of cardiovascular exercises and their functional responses on diabetic peripheral neuropathy using neutrosophic statistics. Neutrosophic Sets and Systems 2024, 64, 77-83. DOI: 10.5281/zenodo.10699498
[9]. Smarandache, F. Neutrosophic statistics is an extension of interval statistics, while plithogenic statistics is the most general form of statistics (third version). Bulletin of Pure & Applied Sciences-Mathematics and Statistics 2022, 41(2), 172-183.
[10]. Hassan, M. K.; Aslam, M. Birnbaum Saunders distribution for imprecise data: statistical properties, estimation methods, and real life applications. Scientific Reports 2024, 14(1), 6955.
[11]. Smarandache, F. Neutrosophic Statistics is an extension of Interval Statistics, while Plithogenic Statistics is the most general form of statistics (second version). International Journal of Neutrosophic Science (IJNS) 2022, 19(1).
[12]. Nourani, V.; Paknezhad, N. J.; Tanaka, H. Review of Prediction Interval Estimation Methods in Artificial Neural Network (ANN)-based Hydro-climatic Modeling. Sustainability 2021, 13(4), 1633.
[13]. Bayle, P.; Bayle, A.; Janson, L.; Mackey, L. Establishing Confidence Intervals for Test Error Through Cross-validation. Advances in Neural Information Processing Systems 2020, 33, 16339-16350.
[14]. Koenker, R. Quantile Regression: 40 Years On. Annual Review of Economics 2017, 9, 155-176.
[15]. Huera Castro, D. E.; Dávila Castillo, M. R.; Escobar Gonzáles, E. J.; Hidalgo Ruiz, M. R. Applying Neutrosophic Statistics to Analyze Tax Evasion Causes. Neutrosophic Sets and Systems 2021, 44, 427-435. DOI: 10.5281/zenodo.5163700.
[16]. Shah, F.; Aslam, M.; Khan, Z. New Control Chart Based on Neutrosophic Maxwell Distribution with Decision Making Applications. Neutrosophic Sets and Systems 2023, 53(1), 18.
[17]. Hernández, N. B.; Vázquez, M. Y. L.; Caballero, E. G.; Cruzaty, L. E. V.; Chávez, W. O.; Smarandache, F. A New Method to Assess Entrepreneurship Competence in University Students Using Plithogenic Numbers and SWOT Analysis. International Journal of Fuzzy Logic and Intelligent Systems 2021, 21(3), 280-292.
[18]. Nagarajan, D.; Broumi, S.; Smarandache, F.; Kavikumar, J. Analysis of Neutrosophic Multiple Regression. Neutrosophic Sets and Systems 2021, 43, 44-54.
[19]. Pupo, I. P.; Pérez, P. Y. P.; Vacacela, R. G.; Bello, R.; Santos, O.; Vázquez, M. Y. L. Extensions to Linguistic Summaries Indicators Based on Neutrosophic Theory, Applications in Project Management Decisions. Neutrosophic Sets & Systems 2018, 22.
[20]. Caballero, E. G.; Smarandache, F.; Leyva Vázquez, M. On Neutrosophic Offuninorms. Symmetry 2019, 11(9), 1136.
[21]. Barrientos Gutiérrez, P.; Flores-Ledesma, K. N.; Guanilo Gómez, S. L.; López Bulnes, J. L.; Jave Nakayo, J. L.; Cabrera Carranza, C. F.; Neri Ayala, A. C.; Ramos Yovera, S. E. Plithogenic Iadov Model to Study University Teaching Practices in the Complexity of the Educational Process of Comprehensive Training by Competencies. Neutrosophic Sets and Systems 2023, 62, 78-85. DOI: 10.5281/zenodo.10436629.
[22]. Sneyers, E., & De Witte, K. (2018). Interventions in Higher Education and Their Effect on Student Success: A Meta-Analysis. Educational Review, 70(2), 208-228.

Received: Jan 14, 2024. Accepted: Mar 31, 2024