SEMARAK ILMU PUBLISHING

Semarak Engineering Journal

Journal homepage: https://semarakilmu.my/index.php/sej/index
ISSN: 3036-0145

Comparison Height-Level of Triangular Fuzzy Linear Programming using Simulation Data

Muhammad Ammar Shafi[1,*], Aliya Syaffa Zakaria[1,♠], Mohd Arif Mohd Zim[2], Zulfana Lidinillah[1], Nurnadiah Nordin[3], Siti Mukhaiyarah Mahtar[3], Aisya Natasya Musa[3], Chandrashekar Ramasamy[1]

[1] Department of Technology and Management, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, Malaysia
[2] Consultant Pulmonologist & Internal Medicine, Damansara Specialist Hospital 2, Bukit Lanjan, 60000 Kuala Lumpur, Malaysia
[3] Department of Medicine, Hospital Al-Sultan Abdullah, UiTM Bandar Puncak Alam,42300, Puncak Alam, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 4 March 2025
Received in revised form 13 April 2025
Accepted 21 May 2025
Available online 30 June 2025

Keywords:
Archimedes screw turbine; small-scale hydropower; renewable energy; screw design optimization; sustainable energy

ABSTRACT

The world's ever-increasing energy needs have necessitated the search for renewable means to satisfy such energy needs. One of those means is small-scale hydropower, and in particular, the Archimedes screw turbine (AST), which has been viewed as one such future option. Initially designed for water lifting, ASTs use incoming water kinetic energy to generate electricity today. This review paper reports on the advances made today in the design of ASTs, emphasizing the geometrical parameters of pitch, diameter, number of blades, and inclination as design parameters affecting efficiency. Improvement was made by either conducting experiments or by optimizing the simulation. Flow variation, sediment deposition, material degradation, and other challenges may be overcome. Still, innovations in composite materials, coatings for corrosion protection, and AI design improvements should be considered to enhance their durability and efficiency. Economic and policy barriers discourage collateral investments, such as high upfront capital costs and competition from other renewables. Modern integration with solar and innovative grid systems opens more expansive opportunities for sustainable energy. Future research and development advancements will improve AS output viability for decentralized eco-friendly power generation in remote areas.

1. Introduction

Regression analysis is a statistical method for estimating the relationship between variables that have a cause and an effect. Regression analysis is a potent method for comprehending (including forecasting and explaining) the causal influence on a population’s result [1]. Regression analysis's main goals are to quantify the relationship between variables, ascertain the effects of each additional independent variable, and forecast the dependent variable's value in relation to the independent

* *Corresponding author.*
E-mail address: ammar@uthm.edu.my
♠ *Corresponding author.*
E-mail address: syaffayila98@gmail.com

https://doi.org/10.37934/sej.9.1.7686

variables. The most widely used statistical analysis and modelling technique, such as in commercial and medical analysis, is regression analysis. This is due to the fact that regression analysis is user-friendly and applicable to a wide range of real-world scenarios. The study yielded the statistical equation that describes the two relationships between the dependent and independent variables. Its multidimensional nature contributes significantly to its explanatory power. It is simply interpretable and comes in computer packages. In the applied sciences, economics, engineering, computer science, social sciences, and other domains, it is also widely employed [2].

Many methods have been developed for performing regression analysis. The regression function in well-known techniques like linear, fuzzy, and ordinary least squares regression is described in terms of a limited number of unknown parameters that are estimated from the data. This makes these techniques parametric. Techniques known as "nonparametric regression" permit the regression function to fall inside a given set of functions, some of which may be infinite-dimensional.

Nevertheless, regression models are highly sensitive to outliers. A data point that dramatically deviates from most other observations is referred to as an outlier. Variability in measurement may be indicative of an experimental error, and an outlier in regression analysis might cause significant difficulty. Actual world data and problems are also oversimplified by regression analysis models, as data are rarely linearly separable. A researcher, Lotfi A. Zadeh, is the first person to develop the model that can handle the vagueness phenomenon, such as the fuzzy model and overcome the outliers [3].

In 1964, Lotfi A. Zadeh, a student at the University of California, Berkeley, published the first article on fuzzy sets. The concept of grade membership, harsh criticism from the academic world, and government money waste are a few of the topics covered in the paper. Furthermore, Lotfi A. Zadeh kept expanding the groundwork for fuzzy set theory from 1965 until 1975. Fuzzy multistage decision making, fuzzy similarity relations, fuzzy constraints, and four linguistic hedges are all provided by the idea of fuzzy set theory. A broader interpretation of fuzzy logic is the notion of fuzzy sets. Fuzzy logic serves these two purposes by easing the burden that traditional mathematical methods have on constructing and analysing complicated systems and by showing how human thinking can make use of ideas and information that lack clear, demarcated limits. Examples include a tall person, a lamp that is lighter, and other objects [3].

Apart from that, in 1982, Hideo Tanaka became the first person to design fuzzy linear regression as a research method and statistical tool. He focused on applying fuzzy linear functions to regression analyses of nebulous phenomena in his study. Observation errors are typically attributed to variations in the regression model between the estimated and observed values. It was hypothesised that the ambiguity of the system structure was the cause of these system parameter variations. The information considered the input-output relationships, whose ambiguity the system's structure [4].

Fuzzy models have many advantages in analysis and can be applied without making any assumptions. The data is still usable even if its error is not regularly distributed. It differs from a different regression analysis in terms of statistics. A fundamental mathematical framework for handling vagueness is offered by fuzzy logic.

Fuzzy linear regression provides tools for studying the relationship between variables when certain assumptions of multiple linear regression fail. It also provides a fundamental mathematical and statistical framework for acknowledging the imprecision of data. In all statistical studies, researchers seek the most recent techniques for minimizing the statistical measurement error value [5,6]. A wide variety of fuzzy linear models can be used for approximating a linear dependence according to a set of observations in fuzzy regression analysis [7]. It can also aid in reducing the interference of unnecessary information, thereby improving the precision of the results [8].

A fuzzy regression model is applied to evaluate the functional relationship between dependent and independent variables in a fuzzy environment. Fuzzy linear regression analysis is also an effective alternative to the generally utilized statistics-based regression techniques. Numerous forms of fuzzy regression models are presented in the literature, along with a diversity of estimation techniques for fuzzy model parameters [9]. Previous research has explored and developed a variety of applications of fuzzy linear regression and its advantages. A recent study has demonstrated that a fuzzy approach, including fuzzy linear regression, is an appropriate framework for lentil yield management compared to multiple linear regression, which is ineffective as some of its preconditions are not met, as in the data, where some variables were fuzzy numbers [10]. Furthermore, it clearly shows its ability to deal with the perceptual uncertainties involved in strength prediction issues by providing a standard equation to estimate output values. This leads to accurate predictions of the predicted cement strength values, thereby enhancing the design and providing a valuable modelling tool for the engineering field [11].

2. Methodology

2.1 Materials

Statistical analysis is adaptable and applicable to various fields, particularly the linear regression technique. Fuzzy linear regression is a form of regression analysis in which fuzzy numbers represent certain model elements. Fuzzy linear functions were proven to be a helpful technique for ambiguous occurrences in linear regression models. The statistical software Microsoft Excel, SPSS and MATLAB were used to analyse the data.

2.2 Methods

2.2.1 Multiple linear regression

Linear regression is a statistical method for determining the value of a dependent variable in relation to the value of an independent variable. It is also known as a method utilised to establish the relationship between two variables. It is a technique for predicting a dependent variable based on one or more independent variables. Among all statistical methods, linear regression analysis is the most applied.

The assumptions underpinning the multiple linear regression model are as follows [12]:

i. The population mean of y within the level of the patient's population was defined by the *x*'s following a linear and additive pattern.
ii. The y observations were assumed to be statistically independent.
iii. The standard deviation of *y* within *x*-strata was constant over all values of *x*.
iv. The distribution of *y* within *x*-strata was normal.

It is necessary to ensure that the multiple linear regression assumptions (i), (ii), (iii), and (iv) are satisfied before analysing the data. This study made three assumptions, which are constant variance, normality and multicollinearity [12]. If these assumptions meet, the results will be reliable.

There are several predictor variables in multiple linear regression, including the first order with two predictor variables and the first order with more than two variables. The model of multiple linear regression can be expressed as follows [13].

$$Y=\beta +\beta X +\beta X +\ldots+\beta X +\varepsilon$$

Where:
$\beta_o, \beta_{1,\ldots,} \beta_j$; are constants
$X_{i1,\ldots,} X_{ij}$; are unknown parameter/ independent variable

i = 1,…, *n* (1)

The equations in the table ANOVA represent the analysis of variance, in addition to mean square regression (MSR) and mean square error (MSE). MSE is a risk function whose value corresponds to the expected square error loss or quadratic loss. MSE estimates the mean square of "errors." Error is the difference between the estimator's implied value and the quantity being estimated. The difference is due to randomness or the estimator's failure to account for information that could lead to a more precise estimate. Table 1 provides an overview of the analysis of variance (ANOVA). The equations are as follows:

Table 1
ANOVA table formula

Source of Variation	SS	df	MS
Regression	SSR = b X'Y - (1/n)Y'JY	p-1	MSR = SSR / p-1
Error	SSE = (Y- Xb)' (Y- Xb)	*n*-p	MSR = SSR / p-1
Total	SST = Y'Y -(1/n) Y'JY.	*n*-1	MSE = SSE /n-p

2.2.2 Fuzzy linear regression

To formulate a fuzzy linear regression model, the following were assumed to hold:
1) The data can be represented by a fuzzy linear model:

$$Y_e* = A_1 * x_{e\ 1} + \ldots + A_g * x_{e\ g} \triangleq A^* x_e, \tag{2}$$

Where, Fuzzy parameter A_g
Variable of fuzzy parameter x_e
Equation of the fuzzy parameter Y_e^*

$$\mu_{Y_{e*}}(y) = 1 - \frac{|y_e - x_e^T \alpha|}{\varsigma^T |x_e|} \tag{3}$$

2) The degree of the fitting of the estimated fuzzy linear model $Y_e* = A * x_e$ to the given data $Y_e = (y_e, \varepsilon_e)$ was measured by the following index h_e, which maximizes h subject to $Y_e^h \subset Y_e^{*h}$ where:

$$Y_e^h = \{y | \mu_{Ye}(y) \geq h\}$$

$$Y_e^* = \left\{y \,|\mu_{Y_e^*}(y) \geq h\right\} \tag{4}$$

Which is h-level sets. This index h_e is illustrated in Figure 1. The degree of the fitting of the fuzzy linear model for all data $Y_1, \ldots, Y_N$ is defined by $\min_f [h_f]$.

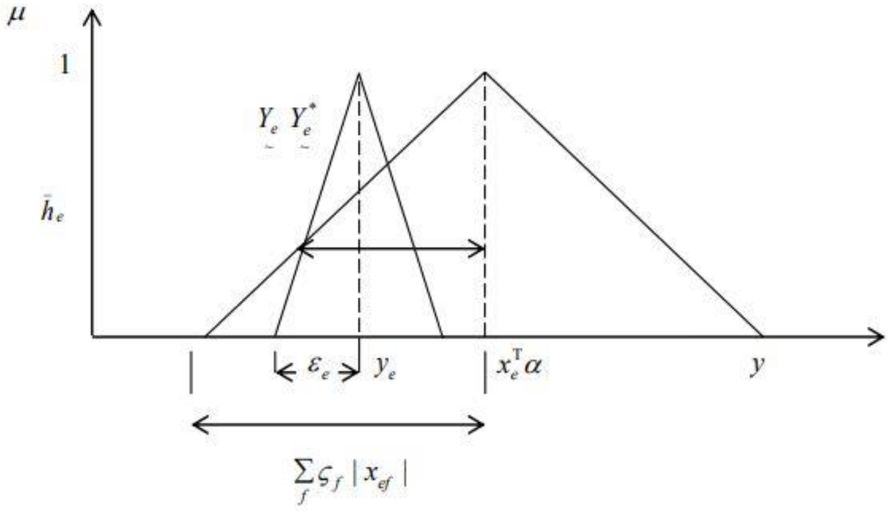

Fig. 1. Degree of fitting of a given fuzzy data

3) The vagueness of the fuzzy linear model is defined by:

$$JJ = \varsigma_1 + \ldots + \varsigma_g \tag{5}$$

The problem was elucidated by acquiring fuzzy parameters A*, which minimized *JJ* subject to $\bar{h}_e \geq H$ for all $_e$, where *H* was selected by the decision maker as the degree of fit of the fuzzy linear model. The $\bar{h}_e$ can be acquired by utilizing:

$$\bar{h}_e = 1 - \frac{|y_e - x_e^T \alpha|}{\sum_f \varsigma_f |x_{ef}| - \varepsilon_e} \tag{6}$$

Tanaka [5] model estimated the fuzzy parameter $A_e^* = (\alpha_e, \varsigma_e),$, which are the solutions of the following linear programming problem:

$$\min_{\alpha,\varsigma} = \varsigma_1 + \ldots + \varsigma_g$$

Subject to $\varsigma \geq 0$ and

$$\alpha^T x_e + (1 - H) \sum_f \varsigma_f |x_{ef}| \geq y_e + (1 - H)\varepsilon_e$$

$$-\alpha^T x_e + (1 - H) \sum_f \varsigma_f |x_{ef}| \geq -y_e + (1 - H)\varepsilon_e \tag{7}$$

The best-fitting model for the given data may be obtained by solving the conventional linear programming problem in (7). The number of constraints, 2 *N*, was generally substantially greater than the number of variables, *g*. As a result, solving the dual problem of (7) was easier than solving the primal problem of (7).

The fuzzy linear regression model (FLRM) can be stated as:

$$Y = A_0\,(\,\alpha_0\,,\varsigma_0\,) + x_g \tag{8}$$

2.3 Methodology

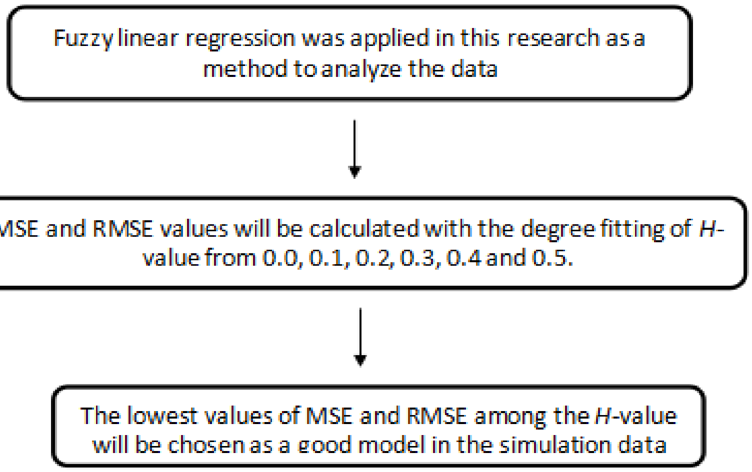

Fig. 2. Steps of analyzing the data using the fuzzy linear regression model

Based on Figure 2 above, fuzzy linear regression was applied in this paper as a method to evaluate the data. There were six degrees of fitting *H*-values from 0.0 to 0.5 that were applied to analyze the center (a_i) and width (c_i) of each variable using the fuzzy linear regression method. The estimated parameter for simulation data will be extracted from the width and center of each variable. The next step was calculating the errors of each *H*-value using the standard statistical measurement errors, which are mean square error (MSE) and root mean square error (RMSE). The lowest value of error will be chosen as the best prediction model for the fuzzy linear regression model in the research.

3. Results and Discussion

The fuzzy linear regression machine model provided 124 rows of data as simulation data. This model was used to study and analyze six predictor variables, such as A_1,A_2, A_3,A_4,A_5,A_6. Microsoft Excel, SPSS and MATLAB were applied to obtain the results. The common measurement errors of the cross-validation technique, which are MSE (mean square error) and RMSE (root mean square error), will then be used to acquire the errors of multiple linear regression and fuzzy linear regression models. Multiple linear regression will be analyzed by the assessment for the significance of variables and ANOVA. The degree of fitting *H*-values, which are 0.0, 0.1, 0.2, 0.3, 0.4 and 0.5, will be compared for the fuzzy linear regression model by calculating the center, a_i and width c_i of six variables for each *H*-value. a_i is the center of the fuzzy parameter, while c_i is the fuzziness of the parameter (width). The center and the width will then be used to calculate the MSE and RMSE. The error of multiple linear regression and degree of fitting *H*-values will be analyzed to obtain the smallest error value of MSE and RMSE, then the model will be the best value of this simulation data.

3.1 Analysis of Multiple Linear Regression
3.1.1 Assessment for significance of individual predictor variable

Multiple linear regression is a prevalent statistical model referred to as the basic linear regression. This model is used to examine and assess six predictor variables. According to Table 2, statistically significant variables have significance values less than 0.05. The three assumptions, such as constant variance, normality and multicollinearity fulfilled by the test and analysis.

Table 2
Parameter estimation of a multiple linear regression model

Variables	Beta (β)	Sig. Value
(Constant)	10.701	0.005
A_1	3.740	0.402
A_2	1.872	0.636
A_3	0.135	0.342
A_4	1.006	0.847
A_5	3.364	0.411
A_6	8.660	0.242

*Significance at 0.05

The following are the predicted outcomes of a model for multiple linear regression on simulation data:

$$\hat{Y} = 10.701 - 3.740\, A_1 + 1.872\, A_2 + 0.135\, A_3 + 1.006\, A_4 + 3.364\, A_5 + 8.660\, A_6$$

3.1.2 Analysis of Variance (ANOVA)

The purpose of analysis of variance (ANOVA) is to assess significant values and obtain information about the mean within a regression model. The value 306.257 is the mean square error term. With a *P*-value for the *F* test statistic less than 0.05, the null hypothesis is strongly refuted. The results of the ANOVA for multiple linear regression are displayed in Table 3.

Table 3
ANOVA for multiple linear regression

Source	Sum of squares	df	Mean square	F-value	P-value
Regression	1270.905	10	1276.091	0.773	0.005
Residual	34607.704	113	306.257		
Total	47367.980	123			

3.2 Analysis of Fuzzy Linear Regression

Fuzzy linear regression (FLR) was used for predicting manufacturing income. This model was compared by six degrees of fitting *H*-values simultaneously. The center, a_i and width, c_i of six variables for each *H*-value have been portrayed in Tables 4 to 8 below.

Table 4
H-value of 0

Variables	Center (a_i)	Width (c_i)
A_1	1.078	0
A_2	1.078	0
A_3	180.850	18.403
A_4	126.003	0
A_5	-1.803	0
A_6	-187.292	0

The estimated model parameter for simulation data is stated as below:

$\hat{Y}$ = (1.078, 0) A_1 + (1.078, 0) A_2 + (180.850, 18.403) A_3 + (126.003, 0) A_4 + (-1.803, 0) A_5 + (-187.292, 0) A_6

Table 5
H-value of 0.1

Variables	Center (a_i)	Width (c_i)
A_1	942.098	0
A_2	1.108	0
A_3	178.834	19.529
A_4	126.182	0
A_5	-1.769	0
A_6	-176.339	0

The estimated model parameter for simulation data is indicated as follows:

$\hat{Y}$ = (942.098, 0) A_1 + (1.108, 0) A_2 + (178.834, 19.529) A_3 + (126.182, 0) A_4 + (-1.769, 0) A_5 + (-176.339, 0) A_6

Table 6
H-value of 0.2

Variables	Center (a_i)	Width (c_i)
A_1	806.618	0
A_2	1.137	0
A_3	176.816	20.936
A_4	126.361	0
A_5	-1.736	0
A_6	-165.388	0

The estimated model parameter for simulation data is stated as follows:

$\hat{Y}$ = (806.618, 0) A_1 + (1.137, 0) A_2 + (176.816, 20.936) A_3 + (126.361, 0) A_4 + (-1.736, 0) A_5 + (-165.388, 0) A_6

Table 7
H-value of 0.3

Variables	Center (a_i)	Width (c_i)
A_1	671.138	0
A_2	1.167	0
A_3	174.799	22.746
A_4	126.539	0
A_5	-1.702	0
A_6	-154.437	0

The estimated model parameter for simulation data is stated as follows:

$\hat{Y}$ = (671.138, 0) A_1 + (1.167, 0) A_2 + (174.799, 22.746) A_3 + (126.539, 0) A_4 + (-1.702, 0) A_5 + (-154.437, 0) A_6

Table 8
H-value of 0.4

Variables	Center (a_i)	Width (c_i)
A_1	535.659	0
A_2	1.196	0
A_3	172.781	25.159
A_4	126.718	0
A_5	-1.668	0
A_6	-143.485	0

The estimated model parameter for simulation data is indicated as follows:

$\hat{Y}$ = (535.659, 0) A_1 + (1.196, 0) A_2 + (172.781, 25.159) A_3 + (126.718, 0) A_4 + (-1.668, 0) A_5 + (-143.485, 0) A_6

Table 9
H-value of 0.5

Variables	Center (a_i)	Width (c_i)
A_1	400.179	0
A_2	1.226	0
A_3	170.764	28.536
A_4	126.897	0
A_5	-1.634	0
A_6	-132.533	0

The estimated model parameter for simulation data is indicated as below:

$\hat{Y}$ = (400.179, 0) A_1 + (1.226, 0) A_2 + (170.764, 28.536) A_3 + (126.897, 0) A_4 + (-1.634, 0) A_5 + (-132.533, 0) A_6

Fuzzy linear regression model analyses performance using two statistical measurement errors, including MSE and RMSE. The performance of the two methods will also be assessed using the degree of fit (*H*-value) in Table 9. The lowest error values determine the best model in both multiple and fuzzy linear regression.

Table 10
MSE and RMSE value

Model	*H*	MSE	RMSE
MLR	-	306.257	17.500
FLR	0.0	9.993	3.161
	0.1	78.861	8.869
	0.2	78.651	8.869
	0.3	68.125	8.254
	0.4	57.600	7.589
	0.5	47.07	6.860

Based on Table 10 above, *H*-value of 0.0 has the lowest measurement error, which are mean square error value of 9.993 and a root mean square error value of 3.161 compared to the multiple linear regression model. The value of the fuzzy parameter is very low, and has a value even though the value is 0. The obtained values in fuzzy parameters are shown in Table 3, which includes six predictor variables. The fuzzy mean value of the simulation data can be explained by A_3 with the highest fuzzy parameter = 180.850.

4. Conclusion

The purpose of this study is to determine the best prediction model with the lowest measurement error between fuzzy linear regression and multiple linear regression models. The results of the fuzzy parameter show that an *H*-value of 0.0 is a good prediction model for simulation data in the fuzzy linear regression model, as it has the lowest measurement error among another model. Table 9 displays the summary evaluation of multiple linear regression and fuzzy linear regression with mean square error and root mean square error. The mean square error for *H*-value of 0.0 was 9.993, and the root mean square error was 3.161. Fuzzy linear regression can be found in various domains in future applications, particularly for inaccurate data. Although only fuzzy linear regression is presented in this paper, another model can be applied by the same approach.

In future studies, other researchers can compare other models with the fuzzy linear regression model to prove that it is the most accurate model with the least measurement errors. The researchers can also add more variables and rows of data in the data analysis.

Acknowledgement
This research was supported by the Ministry of Higher Education (MOHE) through the Fundamental Research Grant Scheme (FRGS/1/2021/STG06/UTHM/03/1).

References
[1] Fahrmeir, Ludwig, Thomas Kneib, Stefan Lang, and Brian D. Marx. "Regression models." In *Regression: Models, methods and applications*, pp. 23-84. Berlin, Heidelberg: Springer Berlin Heidelberg, 2022. https://doi.org/10.1007/978-3-662-63882-8_2
[2] Iqbal, Muhammad Ahmad. "Application of regression techniques with their advantages and disadvantages." *Elektron Mag* 4 (2021): 11-17.
[3] Zadeh, Lotfi Asker. "Fuzzy sets." *Information and control* 8, no. 3 (1965): 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X
[4] Mundt, Philipp. "The formation of input–output architecture: Evidence from the European Union." *Journal of Economic Behavior & Organization* 183 (2021): 89-104. https://doi.org/10.1016/j.jebo.2020.12.031

[5] Asai, H. T. S. U. K., S. Tanaka, and K. Uegima. "Linear regression analysis with fuzzy model." *IEEE Trans. Systems Man Cybern* 12 (1982): 903-907. https://doi.org/10.1109/TSMC.1982.4308925
[6] Zolfaghari, Zahra Sadat, Mohebbat Mohebbi, and Marzieh Najariyan. "Application of fuzzy linear regression method for sensory evaluation of fried donut." *Applied Soft Computing* 22 (2014): 417-423. https://doi.org/10.1016/j.asoc.2014.03.010
[7] Škrabánek, P., and J. Marek. "Models used in fuzzy linear regression." In *Proceedings of the 17th Conference on Applied Mathematics—APLIMAT*, pp. 955-964. 2018.
[8] Kang, Hao, Ting Ai, Qiuyan Tian, and Qiaoling Zhang. "The Advantage of Fuzzy Regression Analysis and the Establishment of Uml Model." In *2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, pp. 2218-2220. IEEE, 2018. https://doi.org/10.1109/IMCEC.2018.8469757
[9] Pérez, Lisset Denoda, Gladys Casas Cardoso, J. L. Martínez, Laureano Rodríguez Corvea, and Emilio González Rodríguez. "Fuzzy Linear Regression Models: A Medical Application." *Departamento de Computación, Centro de Estudios de Informática. Facultad Matemática, Física y Computación, Universidad, All content following this page was uploaded by Lisset Denoda on* 9 (2015).
[10] Sorkheh, Karim, Ahmad Kazemifard, and Shakiba Rajabpoor. "A comparative study of fuzzy linear regression and multiple linear regression inagricultural studies: a case study of lentil yield management." *Turkish Journal of Agriculture and Forestry* 42, no. 6 (2018): 402-411. https://doi.org/10.3906/tar-1709-57
[11] Gkountakou, Fani, and Basil Papadopoulos. "The use of fuzzy linear regression and ANFIS methods to predict the compressive strength of cement." *Symmetry* 12, no. 8 (2020): 1295. https://doi.org/10.3390/sym12081295
[12] Wallisch, Christine, Paul Bach, Lorena Hafermann, Nadja Klein, Willi Sauerbrei, Ewout W. Steyerberg, Georg Heinze, Geraldine Rauch, and Topic Group 2 of the STRATOS Initiative. "Review of guidance papers on regression modeling in statistical series of medical journals." *PloS one* 17, no. 1 (2022): e0262918. https://doi.org/10.1371/journal.pone.0262918
[13] Shafi, Muhammad Ammar, Mohd Saifullah Rusiman, Shuhaida Ismail, and Muhamad Ghazali Kamardan. "A hybrid of multiple linear regression clustering model with support vector machine for colorectal cancer tumor size prediction." *International Journal of Advanced Computer Science and Applications* 10, no. 4 (2019). https://doi.org/10.14569/IJACSA.2019.0100439