

Application of Fuzzy Linear Regression with Symmetric Parameter for Predicting Tumor Size of Colorectal Cancer

Muhammad Ammar Shafi^{1,*}, Mohd Saifullah Rusiman², Siti Nabilah Syuhada Abdullah³

¹Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia, Malaysia

²Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Malaysia

³Faculty of Science, Universiti Teknologi Malaysia, Malaysia

Received September 20, 2020; Revised November 1, 2020; Accepted January 28, 2021

Cite This Paper in the following Citation Styles

(a): [1] Muhammad Ammar Shafi, Mohd Saifullah Rusiman, Siti Nabilah Syuhada Abdullah, "Application of Fuzzy Linear Regression with Symmetric Parameter for Predicting Tumor Size of Colorectal Cancer," *Universal Journal of Educational Research*, Vol. 9, No. 1, pp. 36-40, 2021. DOI: 10.13189/ms.2021.090106.

(b): Muhammad Ammar Shafi, Mohd Saifullah Rusiman, Siti Nabilah Syuhada Abdullah (2021). *Application of Fuzzy Linear Regression with Symmetric Parameter for Predicting Tumor Size of Colorectal Cancer*. *Universal Journal of Educational Research*, 9(1), 36-40. DOI: 10.13189/ms.2021.090106.

Copyright©2021 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract The colon and rectum is the final portion of the digestive tube in the human body. Colorectal cancer (CRC) occurs due to bacteria produced from undigested food in the body. However, factors and symptoms needed to predict tumor size of colorectal cancer are still ambiguous. The problem of using linear regression arises with the use of uncertain and imprecise data. Since the fuzzy set theory's concept can deal with data not to a precise point value (uncertainty data), this study applied the latest fuzzy linear regression to predict tumor size of CRC. Other than that, the parameter, error and explanation for the both models were included. Furthermore, secondary data of 180 colorectal cancer patients who received treatment in general hospital with twenty five independent variables with different combination of variable types were considered to find the best models to predict the tumor size of CRC. Two models; fuzzy linear regression (FLR) and fuzzy linear regression with symmetric parameter (FLRWSP) were compared to get the best model in predicting tumor size of colorectal cancer using two measurement statistical errors. FLRWSP was found to be the best model with least value of mean square error (MSE) and root mean square error (RMSE) followed by the methodology stated.

Keywords Colorectal Cancer, Fuzzy Linear Regression with Symmetric Parameter, Mean Square

Error, Root Mean Square Error

1. Introduction

Recently, applied linear statistical models have been used in many fields such as medication, economy, social science and many more [1, 11, 12]; [1-12]. The main objective of linear statistical analysis is to predict the relationship of a respondent variable in terms of predictor variables in a linear function or multiple linear functions. From a linear regression approach, there are some assumptions that need to be fulfilled by researchers [2]. The assumptions are linear relationship, multivariate normality, no auto correlation and also homoscedascity. Statistical linear regression model can be applied only if the dependent variables are continuous and distributed according to a statistical model. For fuzzy data, fuzzy membership function must be in line with fuzzy set theory [3, 4].

In constructing a fuzzy and vagueness models, there are three key characteristics of every system model such as complexity, credibility and uncertainty (vagueness) attempt to maximize its useful-ness. The relationship is not fully covered by these three characteristics. On the upside,

vagueness tends to reduce complexity and increase credibility of the resulting model when vagueness is the main characteristic in modeling. The vagueness is allowable to solve and estimate by developing methods for each modeling problem [5].

Vagueness in modeling is generally acceptable in publication of a seminal paper by Lotfi A. Zadeh in 1965 [6, 14]. In his paper, Za-deh introduced how to deal with vagueness information using a theory of fuzzy. The significance of Zadeh’s paper was challenged not only in vagueness modeling theory but also the probability theory. After Zadeh’s paper was introduced, modeling in fuzzy area gained more interest especially in prediction of vagueness phenomenon [1]. Tanaka et al (1982) explains fuzziness of respondents or fuzzy uncertainty of dependent variables in fuzzy regression model. Hence, there are three categories data of fuzzy regression model;

- i). Non-fuzzy input and output.
- ii). Non-fuzzy input but fuzzy output.
- iii). Fuzzy input and output.

Recently, there is one model of fuzzy which is commonly used by professional researchers such as FLR. This study aims to provide prediction improvements among fuzzy linear regression and FLRWSP models. The FLR and FLRWSP model have been applied to a colorectal cancer data. The comparison among both models was done based on measurement error values such as MSE and RMSE [7, 8, 9, 10]; [7-10].

2. Materials and Methods

2.1. Fuzzy Linear Regression Model (Tanaka, 1982)

Statistical analysis is versatile and can be used in any of fields especially with regards of the method of linear regression. Fuzzy linear regression is a fuzzy type of regression analysis in which some elements of the model are represented by fuzzy number. FLRM was an approach explored by Hideo Tanaka in 1982. In his research, the main objective was to obtain fuzzy sets which represent the fuzziness of the system structure from estimated values, meanwhile the conventional confidential interval is related to the observation errors. No assumptions are compulsory in fuzzy model.

The data input and output data vagueness is derived from the existence of fuzzy parameters. In the model, the deviations among data are explained as the vagueness of the system structure expressed by fuzzy parameters [14].

Fuzzy output denoted as $Y_i = (y_i, e_i)$, where y_i is a center and e_i is a width of fuzzy triangular diagram. The linear function of fuzzy linear regression as;

$$Y = A_0(\alpha_0, \varsigma_0) + A_1(\alpha_1, \varsigma_1) x_1 + \dots + A_g(\alpha_g, \varsigma_g) x_g \quad (1)$$

Where $X = [\alpha_i, c_i]$ is a vector of independent variables and $A = [A_0, A_1 \dots A_g]$ is a vector of fuzzy coefficient presented in form of triangular fuzzy number. In FLR, there are fitting model can be fine by the data given and solving the linear programming problem. Other than that, the fuzzy parameter can be fined by following linear programming problem:

$$\begin{aligned} \alpha'xi + (1-H)\sum_j c_j |x_{ij}| \geq y_i + (1-H)e_i \\ -\alpha'xi + (1-H)\sum_j c_j |x_{ij}| \geq -y_i + (1-H)e_i \end{aligned} \quad (2)$$

2.2. Fuzzy Linear Regression with Symmetric Parameter (Zolfaghari, 2014)

Fuzzy linear regression with symmetric parameter (FLRWSP) is one of the most used models by professional researcher in fuzzy phenomena. Fuzzy linear regression with symmetric parameter represent some conditions in vagueness and unclear. In the study, the researcher applied fuzzy linear regression to determine the quality of food products especially fried donut. From the science and engineering point of view, the theory of model is useful for conceptual framework and results that can be directly applied in models of systems using fuzzy approach and recent development in fuzzy logic [15].

If $\hat{A}_i (i = 0, 1 \dots n)$ is symmetrical fuzzy number and x_i is crisp real number, and will be a triangular fuzzy number and defined as, $\hat{Y} = (f^c(x), f^s(x))$ when $f^c(x)$ is the mode and $f^s(x)$ is spread of triangular fuzzy number.

The model of FLRWSP can be writing as follow:

$$\begin{aligned} f^s(x) &= s_0 + s_1x_1 + \dots + s_nx_n \\ f^c(x) &= a_0 + a_1x_1 + \dots + a_nx_n \end{aligned} \quad (3)$$

Target function is defined in a symmetric condition of triangular fuzzy number as follows:

$$\begin{aligned} (1-h)s_0^L + (1-h)\sum_{i=1}^n (s_i^L |x_{ji}|) - a_0 - \sum_{i=1}^n (a_i x_{ji}) \geq -y_j \\ (1-h)s_0 + (1-h)\sum_{i=1}^n (k_i s_i^L |x_{ji}|) + a_0 + \sum_{i=1}^n (a_i x_{ji}) \geq -y_j \end{aligned} \quad (4)$$

2.3. Structure and Procedure of a Hybrid Model

The structure framework is based on fuzzy linear regression with symmetric parameter model proposed by Zolfaghari (2014). Steps to produce the model are illustrated in Figure 1:

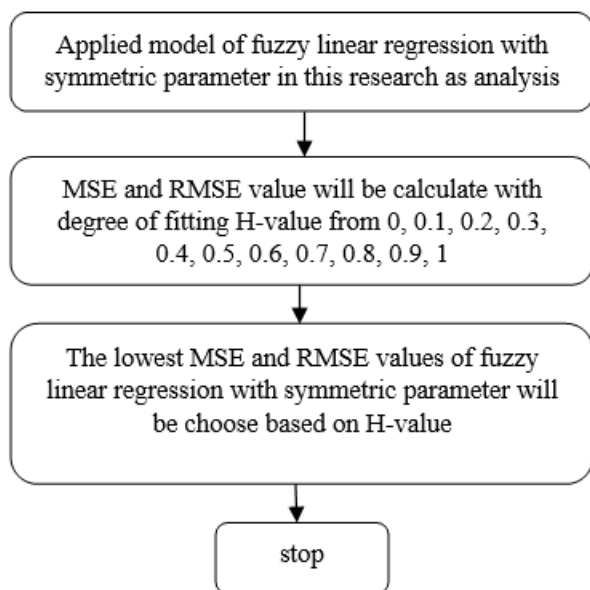


Figure 1. Steps of FLRWSP model

3. Results

This study used secondary data consisting of 180 respondents for the estimated models. The dependent variable or outcome of model is tumor size. In actuality, there were twenty five independent variables in binary and continuous values. The software used to harness results were Microsoft excel, matlab and social science package (SPSS). The comparison among both models was done using cross validation statistical technique also known as MSE and RMSE. This comparison was conducted to find the best model in predicting outcome based on its MSE and RMSE values.

3.1. Fuzzy Linear Regression

FLR model proposed by Tanaka (1982) was used to predict manufacturing income. This model was evaluated by using two performance measures such as MSE and RMSE. Performance of the two methods could also be evaluated by degree of fitting (H-Value = 0.5) in Table 1 and statistical measurement error stated in Table 2. The smallest error values indicate the best model in FLR.

Table 1. Parameter for FLR

Fuzzy Parameter	Center α_i	Width c_i
(Constant)	1.915	0
Gender	1.492	0
Age at Diagnosis	-0.196	0.9575
Ethnic Group	5.319	0
ICD 10 Site	5.372	0
TNM Staging	1.229	0
Family History	-6.083	0
Diabetes Mellitus	5.612	0
Crohn's Disease	14.086	0
Ulcerative colitis	0.687	0
Polyp	-0.919	0
History of cancer	6.332	0
Endometrial	5.201	0
Gastric	-4.508	0
Small bowel	5.038	0
Hepatobiliary	1.893	0
Urinary tract	2.947	0
Ovarian	13.725	0
Other cancer	6.357	0
Intestinal Obstruction	-3.093	0
Colorectal	4.263	0
weight loss	7.747	0
Diarrhoea	3.369	0
blood stool	5.108	0
Anaemia	-0.890	0
Abdominal	4.266	0

Table 2. Value of errors

Methods	Value
Mean square error	277.952
Root mean square error	16.671
Mean absolute error	13.752
Mean absolute percentage error	29.30

The FLR model with H-value = 0.5 as below:

$$\hat{Y} = 1.915 + (1.492, 0) \text{ gender} + (-0.196, 0) \text{ age} + (5.319, 0.957) \text{ ethnic} + (5.372, 0) \text{ icd10} + (1.229, 0) \text{ TNM Staging} + (-6.083, 0) \text{ family history} + (5.612, 0) \text{ diabetes mellitus} + (14.086, 0) \text{ Crohn's disease} + (0.687, 0) \text{ ulcerative colitis} + (-0.919, 0) \text{ polyp} + (6.332, 0) \text{ history of cancer} + (5.201, 0) \text{ endometrial} + (-4.508, 0) \text{ gastric} + (5.038, 0) \text{ small bowel} + (1.893, 0) \text{ hepatobiliary} + (2.947, 0) \text{ urinary tract} + (13.725, 0) \text{ ovarian} + (6.357, 0) \text{ other cancer} + (-3.093, 0) \text{ intestinal obstruction} + (4.263, 0) \text{ colorectal} + (7.747, 0) \text{ weight loss} + (3.369, 0) \text{ diarrhea} + (5.108, 0) \text{ blood stool} + (-0.890, 0) \text{ anemia} + (4.266, 0) \text{ abdominal}.$$

3.2. Fuzzy Linear Regression with Symmetric Parameter

FLRWSP model was proposed by Zolfaghari (2014). This model evaluated by two performance measures; MSE and RMSE. The performance of the two methods could also be evaluated by measuring degree of fitting (H-Value = 0.5). The model with smallest error would become the best model in predicting tumor size of colorectal cancer. This model was applied for the study in colorectal cancer data. The results for performance measurement of parameter and error are in Table 3 and Table 4.

Table 3. Parameter for FLRWSP

Fuzzy Parameter	Center α_i	Width c_i
(Constant)	1.915	0
Gender	1.492	0
Age at Diagnosis	-0.196	0.9575
Ethnic Group	5.319	0
ICD 10 Site	5.372	0
TNM Staging	1.229	0
Family History	-6.083	0
Diabetes Mellitus	5.612	0
Crohn's Disease	14.086	0
Ulcerative colitis	0.687	0
Polyp	-0.919	0
History of cancer	6.332	0
Endometrial	5.201	0
Gastric	-4.508	0
Small bowel	5.038	0
Hepatobiliary	1.893	0
Urinary tract	2.947	0
Ovarian	13.725	0
Other cancer	6.357	0
Intestinal Obstruction	-3.093	0
Colorectal	4.263	0
weight loss	7.747	0
Diarrhoea	3.369	0
blood stool	5.108	0
Anaemia	-0.890	0
Abdominal	4.266	0

Table 4. Value of errors

Methods	Value
Mean square error	277.952
Root mean square error	16.671
Mean absolute error	13.752
Mean absolute percentage error	29.30

There are 25 determinants which are significant in

predicting tumor size of colorectal cancer. The estimated fuzzy linear regression with symmetric parameter model for colorectal cancer of patients is as follows with H = 0.5 is stated below.

$$\hat{Y} = 1.915 + (1.492, 0) \text{ gender} + (-0.196, 0) \text{ age} + (5.319, 0.957) \text{ ethnic} + (5.372, 0) \text{ icd10} + (1.229, 0) \text{ TNM Staging} + (-6.083, 0) \text{ family history} + (5.612, 0) \text{ diabetes mellitus} + (14.086, 0) \text{ Crohn's disease} + (0.687, 0) \text{ ulcerative colitis} + (-0.919, 0) \text{ polyp} + (6.332, 0) \text{ history of cancer} + (5.201, 0) \text{ endometrial} + (-4.508, 0) \text{ gastric} + (5.038, 0) \text{ small bowel} + (1.893, 0) \text{ hepatobiliary} + (2.947, 0) \text{ urinary tract} + (13.725, 0) \text{ ovarian} + (6.357, 0) \text{ other cancer} + (-3.093, 0) \text{ intestinal obstruction} + (4.263, 0) \text{ colorectal} + (7.747, 0) \text{ weight loss} + (3.369, 0) \text{ diarrhea} + (5.108, 0) \text{ blood stool} + (-0.890, 0) \text{ anemia} + (4.266, 0) \text{ abdominal. (5.2)}$$

4. Conclusions

In statistics, MSE and RMSE of an estimator is one of many ways to quantify the difference between values implied by an estimator and the true values of the estimated quantity. The results for statistical error of measurement showed in Table 5.

Table 5. MSE and RMSE values in each model

Models	MSE	RMSE
FLR	277.952	16.671
FLRWSP	275.071	16.585

The results of MSE and RMSE models have been shown in Table 5. The observed Y is taken from total respondent of 180 patients. The value of MSE for fuzzy linear regression (Tanaka) is 277.952 and fuzzy linear regression with symmetric parameter is 275.071. Since, the value of MSE for fuzzy linear regression with symmetric parameter model is the smallest compared to fuzzy linear regression, which is 275.071, it is found that the FLRWSP model is the most appropriate and efficient model in predicting the size of tumour of colorectal cancer faced by patients in 2012.

Acknowledgments

The research work is supported by Ministry of Higher Education, Malaysia. Fundamental Research Grant Scheme (FRGS) grant (Vot K297), reference number FRGS/1/2020/STG06/UTHM/02/4.

REFERENCES

[1] M.A.Shafi, M.S.Rusiman. The Use of Fuzzy Linear

- Regression Models For Tumor Size in Colorectal Cancer in Hospital Of Malaysia, *Applied Mathematical Sciences*, Vol. 9, 2749-2759, 2019.
- [2] M. H. Kutner, C. J. Nachtsheim, J. Neter, W. Li. *Applied Linear Statistical Models*. Fifth edition MC GRAW HILL, 2004.
- [3] T. Hideo, U.Satoru and A. Kiyoji. *Linear Regression Analysis with Fuzzy Model*. IEEE Transactions on Systems, Man and Cybernetics, Vol. 12, 903-907, 1982.
- [4] T. Hideo. *Fuzzy Data Analysis by Possibilities Linear Models*. Fuzzy Set Syst, Vol. 24, 363-375, 1987.
- [5] G. J. Klir, B. Yuan. *Fuzzy Sets and Fuzzy Logic*. Prentice hall of india, new delhi, 110001.
- [6] L.A. Zadeh. *Fuzzy Sets*. Inform Control, Vol. 8, 338-358, 1965.
- [7] B. M.A. Shafi, B. M.S. Rusiman, Y.N.S.H. Che. Determinants status of patient after receiving treatment at intensive care unit: a case study in Johor Bahru. *Computer, Communications, and Control Technology (I4CT)*, 2014 International Conference on, 80 – 82, 2014.
- [8] M. A. Shafi, M. S. Rusiman, N. S. A. Hamzah, S. F. Sufahani, A. Khamis, N. A. H. M. Azmi. Fuzzification of Quantitative Data to Predict Tumour Size of Colorectal Cancer. *Far East Journal of Mathematical Sciences*, Vol. 103, 951-959, 2018.
- [9] M. A.Shafi, M. S. Rusiman, S .N. A. M. Razali, M. Mohamad. The Effective Model of Linear Regressions for Colorectal Cancer Stages in General Hospital: A Case Study in Kuala Lumpur. *Australian Journal of Basic and Applied Sciences*, Vol. 9, 7-10, 2015.
- [10] M. A. Shafi, M. S. Rusiman, N. S. A. Hamzah, E. Nasibov, N. A. H. M. Azmi. The Use of Linear Statistical Model to Predict Tumour Size of Colorectal Cancer. *Journal of Science and Technology* 8.
- [11] M. A. Shafi, M. S. R., M. E. Nor, A. Khamis, S.N.S. Abdullah, M. S.Azmi, M. S. Z. Abidin, M. Ali. The Factors That Influence Job Satisfaction among Royal Malaysian Customs Department Employee. *Journal of Physics: Conference Series*, Vol. 995, 12-16, 2018.
- [12] M. A. Shafi, M.S. Rusiman, N. S. A. Hamzah, M. E. Nor, N. A. H. M. Azmi, M. F. A. Latip, A. H. Azman. The Analysis of Morphometric Data on Rocky Mountain Wolves and Artic Wolves Using Statistical Method. *Journal of Physics: Conference Series*, Vol. 995, 12-15, 2018.
- [13] N. Yongshen. *Fuzzy Correlation and Regression Analysis*. PhD Thesis of University of Oklahoma Graduate College, UMI number: 3163014, 2005.
- [14] L.A. Zadeh. The Concept of A Linguistic Variable And Its Application To Approximate Reasoning. *Information Sciences*, Part I: No. 8, pp. 199-249; Part II : No. 8, pp. 301-357 ; Part III : No. 9, pp. 43-80, 1975.
- [15] Z. Z. Sadat, M. Mohebbat, N. Marzieh. Application of Fuzzy Linear Regression Method for Sensory Evaluation Of Fried Donut. *Applied Soft Computing*, Vol. 22, pp. 417-423, 2014.
- [16] M. A. Shafi, M. S. Rusiman, S. Ismail, M. G. 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(IJACSA)*, Vol. 10, pp. 323-328, 2019.