BAREKENG: Jurnal Ilmu Matematika dan Terapan
March 2023 Volume 17 Issue 1 Page 0147–0154
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss1pp0147-0154

REGRESSION NONPARAMETRIC SPLINE ESTIMATION ON BLOOD GLUCOSE OF INPATIENTS DIABETES MELLITUS AT SAMARINDA HOSPITAL

Ar Ruum Mia Sari [1], **Sifriyani** [2*], **Moh. Nurul Huda** [3]

[1,2,3] *Statistic Study Program, Faculty of Math and Science, Mulawarman University, Barong Tongkok Street, Campus Gn. Kelua, Samarinda, 75123, Indonesia*

Corresponding author's e-mail: *sifriyani@fmipa.unmul.ac.id

Article History:

Received: 6th September 2022
Revised: 5th December 2022
Accepted: 18th January 2023

Keywords:

Diabetes Mellitus;
GCV;
Knots;
Nonparametric Regression;
Spline Truncated.

ABSTRACT

This study used a biresponse nonparametric regression method with truncated spline estimation that used two response variables. Nonparametric regression method is used when the regression curve is not known for its shape and pattern. One of the nonparametric regression model approaches that is often used is the spline. The truncated spline approach has a segmented polynomial function that provides flexibility. The data used in the study were blood glucose levels in patients with diabetes mellitus, cholesterol levels, and triglyceride levels in 2020. From the results of the study, the best nonparametric biresponse spline truncated regression model with three-knot points has been obtained where the minimum GCV value is 7,923.352 *and has the an* R^2 *value of* 98.77%..

How to cite this article:

A. R. M. Sari, Sifriyani, and M. N. Huda, "REGRESSION NONPARAMETRIC SPLINE ESTIMATION ON BLOOD GLUCOSE OF INPATIENTS DIABETES MELLITUS AT SAMARINDA HOSPITAL", *BAREKENG: J. Math. & App.*, vol. 17, iss. 1, pp. 0147-0154, March 2023.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article • **Open Access**

1. INTRODUCTION

Health cases that have a high risk of death are diabetes. Diabetes mellitus (DM) is a chronic disease characterized by blood glucose levels exceeding normal. An early sign that can be seen that a person suffers from DM can be seen directly from the effect of increasing blood glucose levels. The diagnosis of DM can be made if the results of the fasting blood glucose test reach the level of 126mg/dl or even more, and the blood glucose examination two hours after fasting reaches the level of 180mg/dl [4]. Measurement of fasting blood glucose levels can be done if you have previously fasted (not eating and drinking except water) for 8-12 hours [1]. There are various classifications or types of diabetes, but in Indonesia, the most commonly found is type 2 diabetes [10]. The poor quality of insulin, one of which is caused by the amount of fat content. In general, fat in the human body consists of total cholesterol, LDL cholesterol, HDL cholesterol, and triglycerides.

Diabetes not only causes premature death worldwide. It is also a leading cause of blindness, heart disease, and kidney failure. The International Diabetes Federation (IDF) organization estimates that at least 463 million people aged 20-79 years in the world suffer from diabetes in 2019 or equivalent to a prevalence rate of 9.3% of the total population at the same age. The prevalence of diabetes is estimated to increase as the population ages to 19.9% or 111,2 million people aged 65-79 years. The number is predicted to continue to increase to reach 578 million in 2030 and 700 million in 2045 [3].

Research in the field of nonparametric spline regression continues to be developed by experts in the field of statistics, including determining the estimation of the GWR model with a spline approach [8]. Further research is the development of a geographically weighted nonparametric regression model using a spline truncated approach [9]. Research using nonparametric spline truncated regression, among others, [15] conducted nonparametric spline truncated regression modeling on human development index data on the island of Kalimantan. [12] conducted nonparametric spline biresponse regression modeling on the percentage of poor people and the poverty depth index in East Kalimantan in 2015.

Based on this description, the researcher is interested in conducting a study entitled “Regression Nonparametric Spline Estimation on Blood Sugar of Inpatients Diabetes Mellitus at Samarinda Hospital”.

2. RESEARCH METHODS

2.1 Descriptive Statistics

Descriptive statistics is an arrangement of numbers that provides an overview of the data presented in tables, diagrams, histograms, and others [7].

2.2 Regression Analysis

Regression analysis is a study that explains and evaluates the relationship between one or more independent variables and one dependent variable with the aim of estimating the value of the dependent variable based on the known value of the independent variable [13]. The purpose of regression analysis is to investigate the pattern of functional relationships between one or more variables and can also be used to predict [2].

2.3 Nonparametric Regression

Nonparametric regression is a regression method approach where the shape of the curve of the regression function is unknown. In nonparametric regression, the regression curve is only assumed to be smooth in the sense that it is contained in a certain function space so that it has high flexibility [14]. In general, the regression function can be expressed as in the following equation:

$$y_i = f(x_i) + \varepsilon_i; i = 1,2, \dots, n \quad (1)$$

2.2 Nonparametric Regression with Truncated Spline Approach

One of the existing methods in nonparametric regression is the spline truncated nonparametric regression method. In general, a spline function of order m is any function that can be written in the form [15]:

$$f(x_i) = \sum_{k=1}^{m} \beta_k x_i^k + \sum_{h=1}^{r} \beta_{m+h} (x_i - K_h)_+^m \tag{2}$$

where $\sum_{k=1}^{m} \beta_k x_i^k$ is the component of the polynomial and $\sum_{h=1}^{r} \beta_{m+h} (x_i - K_h)_+^m$ is a truncated component.

2.3 Research Data

Using data on blood glucose levels of diabetes mellitus patients two hours after fasting, total cholesterol levels and triglyceride levels were taken from the Abdul Wahab Sjahranie Hospital, Samarinda City, East Kalimantan Province in 2020.

2.4 Data Analysis

This study uses several stages of analysis as follows:

1. Identify the response variables and predictor variables that are thought to influence.
2. Create descriptive statistics of response variables and predictor variables.
3. Identify the data pattern formed between the response variables and each predictor variable using a scatterplot.
4. Identifying multicollinearity on predictor variables with the following equation:

$$VIF_k = \frac{1}{1 - R_k^n}; k = 1, 2, \dots, p \tag{6}$$

5. Data modeling of diabetes mellitus blood sugar levels used multivariable spline truncated nonparametric regression at one-knot point, two-knot points, and three-knot points.
6. Selecting the optimal knot point using Generalized Cross Validation (GCV) in the following equation [11]:

$$\text{GCV}(k) = \frac{\text{MSE}(K)}{(n^{-1} trace[\mathbf{I} - \mathbf{A}(\mathbf{K})])^2} \tag{7}$$

7. Performing the parameter significance test simultaneously (simultaneously) in Equation (8) and followed by a partial test on Equation (9) if the decision to reject H_0 is obtained in the simultaneous test, and identify the coefficient of determination R^2 in Equation (10) [5].

$$F_{hitung} = \frac{MSR}{MSE} \tag{8}$$

$$t_{hitung} = \frac{\hat{\beta}_{j,(1+k)}}{SE(\hat{\beta}_{j,(1+k)})} \tag{9}$$

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} \tag{10}$$

8. Interpret the model obtained and draw conclusions.

3. RESULTS AND DISCUSSION

3.1. Descriptive Statistical Analysis

Descriptive statistical analysis used in this study are the average value, minimum value, maximum value, and standard deviation. The results of descriptive statistical analyses can be seen in Table 1.

Tabel 1. Descriptive Statistics

Variable	Average	*Minimum*	Maximum	Satuan
Y	213.4348	60	427	mg/dl
X_1	154.0096	54	255	mg/dl
X_2	153.0452	63	274	mg/dl

From **Table 1**, it can be seen that the average value of blood glucose levels in patients with diabetes mellitus is 213.4348 mg/dl, with the lowest value of 60 mg/dl and the highest value of 427 mg/dl.

Next, make a scatter plot between the variable blood glucose levels of patients with diabetes mellitus and predictor variables that are thought to have an effect. It can be seen in **Figure 1** that the data pattern between the response variable and the predictor variable does not form a certain pattern, so the right solution to use in this case is to use a nonparametric approach.

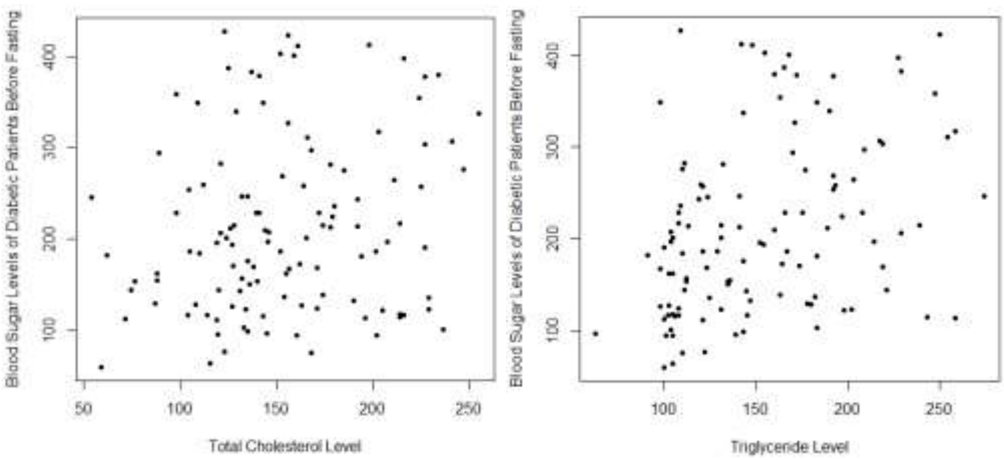

Figure 1. Scatter Plot Between Response Variable and Predictor Variable

It can be seen in **Figure 1** that the data pattern between the response variable and the predictor variable does not form a certain pattern, so the right solution to be used in this case is to use a nonparametric approach.

3.2. Multivariable Nonparametric Regression Model with Truncated Spline

The selection of the best model from the spline biresponse regression can be seen from the minimum GCV value and the maximum R^2 value. **Table 2** shows the GCV value and R^2 value for each knot point.

Table 2. Minimum GCV Value and R^2 Value of Each Knot Point

Knot Points	GCV Value	R^2
1 Knot Point	8,027.968	97.15%
2 Knot Point	7,938.818	97.75%
3 Knot Point	7,923.352	**98.77%**

Table 2 shows that the modeling that produces the smallest GCV value and the largest R^2 value is at three-knot points with an R^2 value of 98.77%. The nonparametric spline regression model in this study has the smallest GCV value at three-knot points of 7,923.352. Therefore, it was decided that the best model to be chosen was a nonparametric spline truncated regression model using three-knot points. So that the results of the knot point position for each variable and the parameter values of the three-knot point linear model are obtained in **Table 3** and **Table 4** below:

Table 3. Optimal Knot Point Position with Three-Knot Points

Predictor Variable	Y
x_1	137.1725
	157.9655
	164.8965
x_2	150.3103
	172.1379
	179.4137

Table 4. Estimation of Model Parameters with Three-Knot Points

Predictor Variable	Y	
	$\hat{\theta}$	$\hat{\emptyset}$
x_1	0.0495	2.8082
		−3.5956
		1.7773
x_2	0.0474	−0.0671
		0.2605
		−0.2406

The three-knot points spline model can be written in the form of **Equation (4)** as follows:

$$\hat{Y} = 98.0029 + 0.0495x_1 + 2.8082(x_1 - 137.1724)_+ - 3.5956(x_1 - 157.9655)_+ + 1.7773(x_1 - 164.8965)_+ + 0.0474x_2 - 0.0671(x_2 - 150.3103)_+ + 0.2605(x_2 - 172.1379)_+ - 0.2406(x_2 - 179.4137)_+$$

3.3. Model Significance Test

There are two parameter estimation tests carried out, namely simultaneous parameter testing and partial testing. The following is the form of the simultaneous test hypothesis:

H_0: $\beta_0 = \beta_{1.1} = \dots = \gamma_{1.4} \dots = \beta_{4.1} = \dots = \gamma_{4.4} = 0$

H_1: there is at least one $\beta_{k(1+h)} \neq 0$; where $k = 1, 2, \dots, 4$ and $h = 1, 2, 3$

The test statistic value is 1,065.848 with a p-value of 1.4198×10^{-97}. If the p-value is compared with the significance level used, which is 0.01, a decision to reject H_0 can be taken. So, it can be concluded that there is at least one variable that has a significant influence on the model. The occurrence of H_0 rejection indicates that it is necessary to do a partial test to find out which variables have a significant influence on the model. Partial test results are presented in **Table 5**.

Table 5. Partial Regression Model Parameter Testing

Parameter Variable		Parameter Estimation	*P-value*	Desc.
Constant	β_0	98.0029	12.9782	
Total Cholesterol Level (x_1)	$\beta_{1.1}$	0.0495	0.8551	Sig
	$\gamma_{1.2}$	2.8082	9.3866	
	$\gamma_{1.3}$	−3.5956	−3.0821*	
	$\gamma_{1.4}$	1.7773	1.9832	
Triglyceride Level (x_2)	$\beta_{2.1}$	0.0474	2.0052	Sig
	$\gamma_{2.2}$	−0.0671	−0.2592*	
	$\gamma_{2.3}$	0.2605	0.2782*	
	$\gamma_{2.4}$	−0.2406	−0.2901*	

Table 5 shows that of the nine parameters in the nonparametric spline truncated regression model, there are three parameters that are not significant. Although there are five parameters that are not significant, overall, all variables are suspected to have an effect.

3.4. Model Fit Test

The comparison between data on blood glucose levels of patients with diabetes mellitus and data on blood glucose levels of patients with diabetes mellitus can be seen in the graph in **Figure 2**.

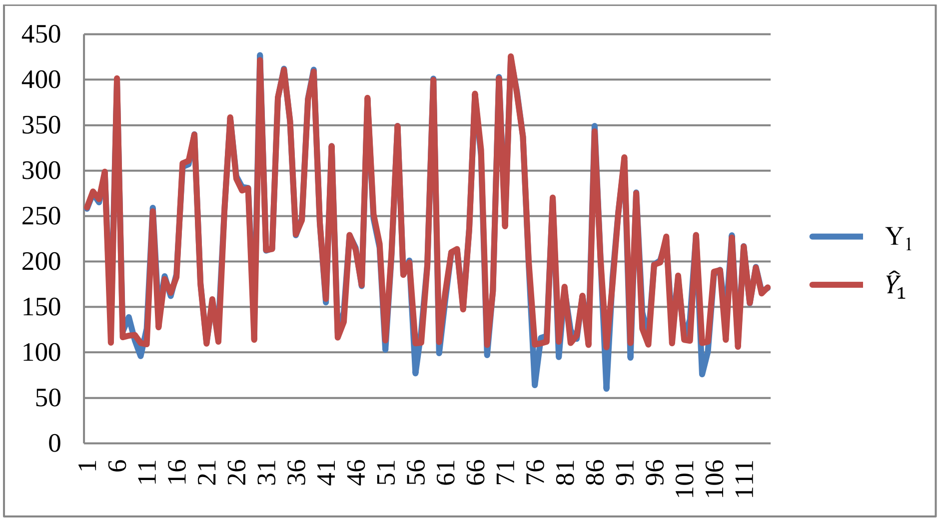

Figure 2. Graph of Actual and Predicted Response Variables for Truncated Spline Estimators

The pattern formed from the two lines in **Figure 2** does not show a significant difference.

3.5. Interpretation of Nonparametric Spline Truncated Regression Model

The best data modeling of blood glucose levels in diabetes mellitus patients using a nonparametric spline regression approach is a nonparametric spline regression model with three-knot points. The interpretation of the spline model is described as follows:

1. If the variable of triglyceride levels in patients with diabetes mellitus is considered constant, then the effect of total cholesterol levels in patients with diabetes mellitus is as follows:

$$\hat{y} = 0.0495x_1 + 2.8082(x_1 - 137.1724)_+ - 3.5956(x_1 - 157.9655)_+ + 1.7773(x_1 - 164.8965)_+$$

$$\begin{cases} 0.0495x_1 & ;\ x_1 < 137.1724 \\ 2.8577x_1 - 385.2075 & ;\ 137.1724 \leq x_1 < 157.9655 \\ -0.7379x_1 + 182.7732 & ;\ 157.9655 \leq x_1 < 164.8965 \\ 1.0394x_1 - 110.2973 & ;\ x_1 \geq 164.8965 \end{cases}$$

From the equation above, if the total cholesterol level of a diabetic patient is less than 137.1724mg/dl, then every 1mg/dl addition of the total cholesterol level in a diabetic patient will increase the blood glucose level of a diabetic patient by 0.0495mg/dl. If the total cholesterol level of a diabetic patient is 137.1724mg/dl to less than 157.9655mg/dl, then each additional 1mg/dl of total cholesterol level in a diabetic patient will increase the blood glucose level of a diabetic patient by 2.8577mg/dl. If the total cholesterol level in diabetic patients is 157.9655mg/dl to less than 164.8965mg/dl, then every 1mg/dl addition of total cholesterol levels in diabetic patients will reduce blood glucose levels in diabetic patients by 0.7379mg/dl. If the total cholesterol level of a diabetic patient is 164.8965mg/dl or more, then every 1mg/dl addition of the total cholesterol level in a diabetic patient will increase the blood glucose level of a diabetes mellitus patient by 1.0394mg/dl.

2. If the variable total cholesterol levels in patients with diabetes mellitus are considered constant, then the effect of triglyceride levels in patients with diabetes mellitus is as follows:

$$\hat{y} = 0.0474x_1 - 0.0671(x_1 - 150.3103)_+ + 0.2605(x_1 - 172.1379)_+ - 0.2406(x_1 - 179.4137)_+$$

$$\begin{cases} 0.0474x_2 & ; x_2 < 150.3103 \\ -0.0197x_2 + 10.0858 & ; 150.3103 \leq x_2 < 172.1379 \\ 0.2408x_2 - 34.7561 & ; 172.1379 \leq x_2 < 179.4137 \\ 0.0002x_2 + 8.4108 & ; x_2 \geq 179.4137 \end{cases}$$

From the equation above, if the triglyceride level of a diabetic patient is less than 150.3103mg/dl, then every 1mg/dl addition of a triglyceride level in a diabetic patient will increase the blood glucose level of a diabetic patient by 0.0474mg/dl. If the triglyceride level of a diabetic patient is 150.3103mg/dl to less than 172.1379mg/dl, then every 1mg/dl addition of triglyceride level in a diabetic patient will lower the blood glucose level of a diabetic patient by 0.0197mg/dl. If triglyceride levels in diabetic patients are 172.1379mg/dl to less than 179.4137mg/dl, then every 1mg/dl addition of triglyceride levels in diabetic patients will increase blood glucose levels in diabetic patients by 0.2408mg/dl. If triglyceride levels in diabetic patients are 179.4137mg/dl or more, then every 1mg/dl addition of total cholesterol levels in diabetic patients will increase the blood glucose levels of diabetes mellitus patients by 0.0002mg/dl.

4. CONCLUSIONS

Based on the parameter significance test, it is known that all variables have a significant effect on blood glucose levels in patients with diabetes mellitus. The two variables include total cholesterol levels and triglyceride levels. The model has an R-Squared value of 99.86%. The graph formed by the two lines in **Figure 3** does not show a significant difference. That is, the prediction results show a pattern similar to the actual data on blood glucose levels of patients with diabetes mellitus. This shows that the best model obtained is the result of selecting the most optimal three-knot points so as to produce a minimum GCV.

REFERENCES

[1] American Diabetes Association, "Diagnosis and Classification of Diabetes Mellitus," *Diabetes Care,* vol. 33, pp. S62-9, 2010.

[2] B. I. Nyoman, "Spline dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang," Surabaya: Institut Teknologi Sepuluh Nopember, 2009.

[3] Kementrian Kesehatan RI, "INFODATIN Pusat Data dan Informasi Kementrian Kesehatan RI Tetap Produktif, Cegah, dan Atasi Diabetes Melitus". Jakarta, 2020.

[4] Khomsah, "Penyakt Diabetes Melitus," 2008, [Online]. Available: http://www.infopenyakit.com/2008/03/penyakit-diabetes-melitus-dm.html [Accessed on March 21, 2021].

[5] N. Wianovita, Sifriyani, P. Ika, "Pemodelan Regresi Nonparametrik Spline Linier Persentase Penduduk Miskin di Kalimantan," *Jurnal Siger Matematika*, vol. 1, no. 2, pp. 37, 2020.

[6] O. Dhina, Regresi Spline Birespon untuk Memodelkan Kadar Gula Darah Penderita Diabetes Melitus, Surabaya: Institut Teknologi Sepuluh Nopember, 2011.

[7] S. Amirotun, "Statistik Deskriptif dalam Penelitian Kualitatif," *KOMUNIKA*, ISSN, vol. 10, no. 2, pp. 1978-1261, 2016.

[8] Sifriyani, Haryatmi, B.I. Nyoman, and Gunardi, "Geographically Weighted Regression with Spline Approach," *Far East Journal of Mathematical Sciences*, vol. 101, no. 6, pp. 1183-1196, 2017.

[9] Sifriyani, S.H. Kartiko, B.I. Nyoman, and Gunardi, "Development of Nonparametric Geographically Weighted Regression Using Truncated Spline Approach," *Songklanakarin Journal of Science and Technology*, vol. 40, no. 4, pp. 909-920, 2018.

[10] Subekti. *Apa Itu Diabetes: Patofisiologi, Gejala dan Tanda. Materi Penyuluhan Pasien pada Penatalaksanaan Diabetes Melitus Terpadu Edisi Kedua*. Jakarta: Balai Penerbit FKUI, 2009.

[11] Suparti and P. Alan, "Pemodelan Regresi Nonparametrik Menggunakan Pendekatan Polinomial Lokal pada Beban Listrik di Kota Semarang," *Media Statistika*, vol. 9, no. 2, pp. 85-93, 2016.

[12] T. Ronald, Y. Desi, and H. Memi, "Regresi Nonparametrik Spine Birespon untuk Memodelkan Persentase Penduduk Miskin dan Indeks Kedalaman Kemiskinan di Kalimantan Timur Tahun 2015," *Jurnal Eksponensial*, vol. 10, pp. 2085-7829, 2019.

[13] W. Agus. *Ekonometrika Teori dan Aplikasi*. Yogyakarta: Ekonisia FE UII, 2007.

[14] Winarti and Sony, "Pendekatan Regresi Semiparametrik Spline (Pada data nilai Ujian Nasional siswa SMKN 1 Nguling Pasuruan)," *Jurnal Sains dan Seni Pomits*, vol. 3, no. 2, pp. 194-199, 2010.

[15] Y. Izzatul, Sifriyani, and Wasono, "Pemodelan Regresi Nonparametrik *Spline Truncated* dan Aplikasinya pada Indeks Pembangunan Manusia di Pulau Kalimantan," *Proceedings of the National Seminar on Mathematics, Statistics and Applications*, vol. 5. pp. 2657-232X, 2019.