International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 3 Issue 12, December – 2019, Pages: 51-54

Platelet Modeling Based On Hematocrit in DHF Patients with Spline Quantile Regression

Bunga Aprilia[1] Anna Islamiyati[2] and Anisa[3]
[1,2,3]Departement of Statistics, Hasanuddin University, Makassar, 90245, Indonesia
bungaprilia021@gmail.com
nkalondeng@gmail.com
correspondence author: annaislamiyati@unhas.ac.id

***Abstract**: Quantile nonparametric regression is used to estimate the regression function when assumptions about the shape of the regression curve are unknown. It is only assumed to be smooth by involving quantile values. One estimator in nonparametric regression is spline. The segmented properties of the spline provide more flexibility than ordinary polynomials. Therefore, the nature of the spline makes it possible to adapt more effectively to the local characteristics of a function or data. This study makes a model for platelet counts based on hematocrit of Dengue Hemorrhagic Fever patients with linear spline quantile regression. The optimal model obtained from the estimation of linear spline quantile regression is at quantile 0.5 with two knots and the GCV value is 40,799. Based on the model, there are three segments of platelet change based on hematocrit. There is a decreased platelet segment with increasing percentage of hematocrit. However, there is a hematocrit segment of 36.9-46.9 which should receive attention where platelets increase in that interval.*

Keywords— hematocrit; linear spline; platelet; quantile regression; segment

1. Introduction

Testing assumptions in the regression model under certain conditions must be done because it is related to the accuracy of the estimation of our model. Assumptions include homoscedastic [1], autocorrelation [2], and multicollinearity [3]. However, these assumptions are often not met when the data contain outliers. It on the data will interfere with the accuracy of the estimation results so it needs to be overcome [4]. One method studied by statistical researchers to model data containing outliers is quantile regression. The quantile regression method divides the data into specific quantile and predicts conditional quantile functions of a data distribution [5]. The model obtained from quantile regression is a complete result of data behavior, both in the middle and tail distribution. Furno has compared the use of OLS with quantile regression [6]. Alhamzawi et al. used the Bayesian approach in quantile regression [7]. Sauzet et al. used a quantile regression in a patient satisfaction survey [8].

Under certain conditions, there are some data patterns that cannot be modeled by parametric quantile regression because it will produce large errors and variances. If the shape of the regression curve is unknown then it is recommended to use a nonparametric regression approach [9]. There are several spline estimator that can be used to estimate the regression function. Ahmad et al. [10] have used cubic splines. Chamidah et al. developed spline with reproducing kernel hilbert space [11]. Islamiyati et al. developed the penalized spline on longitudinal data [12]. Lestari et al. developed spline smoothing [13]. Furthermore, nonparametric regression is used for responses with quantile conditional functions. That is used when assumptions about the shape of the regression curve are unknown and are only assumed to be smooth by involving quantile values. Yuan developed the quantile smoothing spline [14]. Marrie developed quantile regression and restricted cubic splines [15]. Wang and Yang stated that the spline consists of several pieces of polynomial that have segmented and continuous properties and certain sequences that are joined together at knot points [16].

Segmentation of the spline provides more flexible properties than ordinary polynomials, making it possible to adapt more effectively to the local characteristics of a function or data. Therefore, this article uses linear quantile spline regression in modeling the platelet counts of Dengue Hemorrhagic Fever (DHF) patients. Attack of the virus occurs in platelet cells so that the main indicator of a person affected by DHF in medical testing is the measurement of platelet cells. But in addition to platelet counts, other medical examinations are needed, including examination of hematocrit content because both are blood cells. Laboratory tests that can support the diagnosis of DHF are examination of hematocrit and platelet count [17].

2. Material and Methods

The data in this study are the data of patients suffering from Dengue Hemorrhagic Fever in 2013-2017 at the Hasanuddin University Teaching Hospital. Data consisted of response and predictor variables, namely the number of DHF patient platelets (thousand/mcL) as response variables (y) and hematocrit values (%) as predictor variables (x). The platelet count and hematocrit of the patient were taken from the first day the patient was hospitalized.

The model used is quantile regression with the linear spline estimator.

$$y(\theta) = \beta_0(\theta) + \beta_1(\theta)x + \beta_2(\theta)(x - K_1)_+^1 + \cdots + \beta_p(\theta)(x - K_k)_+^1$$

The first step in this research is the making of scatter plots to see the pattern of changes in patient platelets based on hematocrit. The second step is modeling the quantile regression functions of 0.25, 0.5 and 0.75 with linear spline. The knots that are tried are one to three knots. In the third stage, we selected the model to obtain optimal quantile regression models at knots and quantile points that varied based on the minimum GCV value. Next, the final step is to explain the pattern of changes in the number of platelets with hematocrit in DHF patients based on the optimal model.

3. Result and Discussion

Identification of the pattern of the relationship between hematocrit and platelet count of DHF patients is shown through scatter plots as in Figure 1.

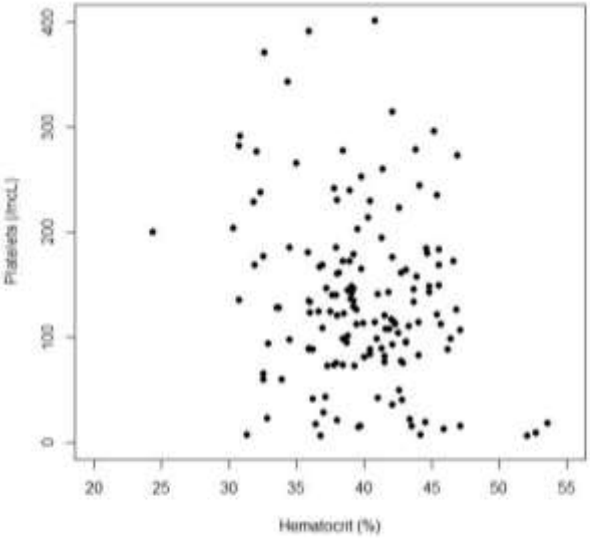

Figure 1. Scatter plot of the effect of hematocrit (x) on platelet count (y)

In Figure 1, we can see the pattern of the relationship of platelets with hematocrit tends to vary. Many DHF patients have normal hematocrit values with abnormal platelet counts. There are also patients diagnosed with DHF has normal platelets with a hematocrit number below normal at <38%. The pattern in Figure 1 does not follow the parametric pattern so the parametric regression estimation model cannot be used. Therefore, the data were solved using a nonparametric regression approach through a linear spline estimator. The form of regression that we use is quantile regression. This is related to the distribution of heterogeneous data. The distribution of quantile and the estimation of the data distribution function of each quantile through the spline estimator can provide a model that represents the data.

3.1 Modeling Hematocrit BAsed On Platelet in Quantile Regression using Linear Spline Estimator

Modeling begins with the selection of optimal knot points from one to three knots. The choice of the knot point is done through trial and error by taking the point which is in the interval of the hematocrit value until the minimum GCV value is obtained. Minimum GCV values for one, two and three knot points are presented in Table 1.

Table 1. GCV values at one, two and three knot points

Knot Number	Knot Point			GCV
	K1	K2	K3	
1	39,2	-	-	6322,557
	44,2	-	-	6276,181
	45,4	-	-	6213,514
	46,2	-	-	6207,157
	46,9	-	-	6195,423
	47,1	-	-	6199,968
2	39,6	46,9	-	61912,481
	37,8	46,9	-	6163,614
	36,6	47,1	-	6150,799
	36,8	**46,9**	**-**	**6141,853**
	36	46,4	-	6159,594
	36,6	46,9	-	6142,721
3	39	44	46,9	6247,654
	36,8	46,9	47,1	6148,378
	35,5	46,9	47,1	6182,586
	36,8	38,7	46,9	6211,385
	36,8	46,9	48	6159,293
	35	36,8	46,9	6202,296

Based on Table 1, we get the most optimal knot point (K) with a minimum GCV value of 6141,853 located at 36.8% and 46.9% of the hematocrit value. The optimal knot point is the same for each try quantile. These results indicate that there are changes in platelets when the patient's hematocrit values are abnormal. The linear spline quantile regression model for the two knots oft the quantile θ is as follows.

$$\hat{y}(\theta) = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)x + \hat{\beta}_2(\theta)(x - K_1)_+^1 + \hat{\beta}_3(\theta)(x - K_2)_+^1 + \hat{\varepsilon}_i$$

The results of parameter estimation and linear spline quantile regression curves are respectively shown in Table 2 and Figure 2.

Table 2. Estimation results of linear spline quantile regression parameters using two point knots

Parameters	Quantile		
	0,25	0,5	0,75
$\hat{\boldsymbol{\beta}}_0$	413,481	576,105	737,889
$\hat{\boldsymbol{\beta}}_1$	-9,038	-12,281	-15,556
$\hat{\boldsymbol{\beta}}_2$	10,927	12,028	16,326
$\hat{\boldsymbol{\beta}}_3$	-19,763	-19,164	-23,791

Based on Table 2, the linear spline quantile regression model with two knots, we can write it as follows.

$$\hat{y}_{0,25} = 413{,}481 - 9{,}038x + 10{,}927(x-36{,}8)_+^1 - 19{,}763(x-46{,}9)_+^1$$

$$\hat{y}_{0,5} = 576{,}105 - 12{,}281x + 12{,}028(x-36{,}8)_+^1 - 19{,}164(x-46{,}9)_+^1$$

$$\hat{y}_{0,75} = 737{,}889 - 15{,}556x + 16{,}326(x-36{,}8)_+^1 - 23{,}791(x-46{,}9)_+^1$$

The regression model for each quantile obtained is different, but the tendency for changes in platelets does not appear to be significantly different. The pattern of changes in each quantile is shown in Figure 2.

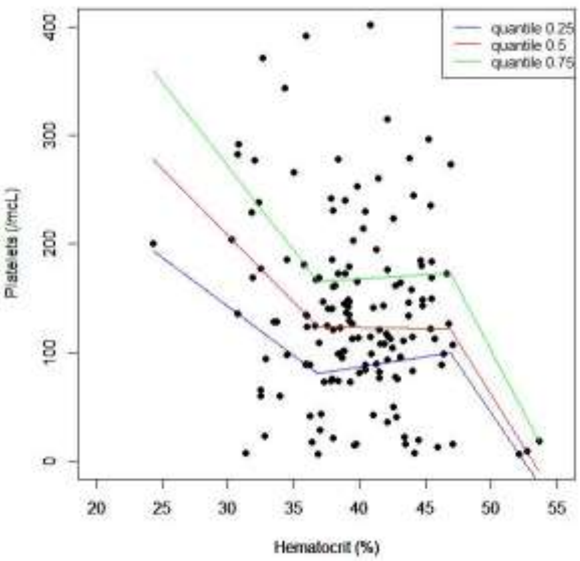

Figure 2. Estimation curves for linear spline quantile regression using two point knots

From the estimation results of quantile regression, there are three segments about the relationship between the number of platelets and the percentage of the hematocrit value of DHF patients in different quantile. The three models as shown in Figure 2 cover the ups and downs of platelets based on hematocrit. At low hematocrit up to 36.8, platelets tend to fall dramatically in each quantile. Furthermore, in hematocrit 36.8-46.9 have platelets, which tend to increase sharply. The last segment, platelets dropped dramatically after a hematocrit rise above 46.9. These results indicate that the second segment needs to be studied in depth with the medical team taking into account many factors.

Next, the GCV values from each model for the three quantile are compared to show the model that is best able to explain the conditions of data distribution as well. The processed results are shown in Table 3.

Table 3. GCV values for each linear spline quantile model

$\boldsymbol{\theta}$	0,25	0,5	0,75
GCV	54,236	40,799	52,469

Based on Table 3, it can be seen that the linear spline quantile model which has the smallest GCV value is 40.799 at the quantile 0.5. That means that the median division of models is better able to explain the diversity of data and the influence of hematocrit on changes in platelet counts of DHF patients.

3.2 Effect Hematocrit on Platelet

The number of normal platelets in a person's blood is around 150 to 400 thousand/mcL and the range of normal hematocrit value ranges between 38 - 45%. Patterns of platelet change based on hematocrit count indicate that there is a very sharp decrease in platelet count with the addition of hematocrit to 38%, and after exceeding 38%, platelet changes are very small. However, when the hematocrit continues to increase until it exceeds the normal limit, the platelet count returns to a significant decrease. These results certainly become information for the medical team to prepare treatment measures for each DHF patient.

4. Conclusion

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference in the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

The second segment review can be developed in the second order quantile spline regression as has been developed by Islamiyati [18] in longitudinal studies. In addition, research still needs to be developed into a higher variable dimension.

5. References

[1] Chamidah, N., & Lestari, B. (2016). Spline estimator in homoscedastic multi-response nonparametric regression model in case of unbalanced number of observations. Far East Journal of Mathematical Sciences (FJMS), 100(9) : 1433-1453.

[2] Hamel, S., Yoccos, N.G., & Gaillard, J.M. (2012). Statistical evaluation of parameters estimating autocorrelation and individual heterogeneity in longitudinal study. Methods in Ecology and Evaluation, 3 : 731-742.

[3] Islamiyati, A. (2015). Estimasi parameter model regresi logistik biner komponen utama non linear dengan maksium likelihood. Jurnal Matematika, Statistika dan Komputasi, 11 (2) : 122-128.

[4] Nurdin, N., Raupong, & Islamiyati, A. (2018). Penggunaan regresi pada data yang mengandung pencilan dengan metode momen. Jurnal matematika, Statistika dan Komputasi, 10 (2).

[5] Buhai, S. (2005). Quantile Regression: Overview and Selected Application.

[6] Furno, M. (2014). Prediction on quantile regression. Open Journal of Statistics, 4 : 504-517.

[7] Alhamzawi, R., Yu, K., & Mallick, H. (2019). Quantile regression and beyond in statistical analysis of data. Hindawi Journal of Probability and Statistics, Vol. 2019.

[8] Sauzet, O., Razum, O., Widera, T., & Brzoska, P. (2019). Front Public Health, 11 June 2019.

[9] Islamiyati, A., Fatmawati & Chamidah, N. (2018). Estimation of covariance matrix on bi-response longitudinal data analysis with penalized spline regression. J. Phys.: Conf. Ser. **979** 012093.

[10] Ahmad, R.R., Ghazali, N., Rambeli, A.S., Din, U. K. S., & Hassan, N. (2012). Application of cubic spline in the implementation of braces for the case of a child. Journal of Mathematics and Statistics, 8 (1) : 144-149.

[11] Chamidah, N., Lestari, B., & Saifuddin, T. (2019). Predicting blood presures and heart rate associated with stress level using spline estimator : a theoretically discussion. International Journal of Academic and Applied Research (IJAAR), 3(10) : 5-12.

[12] Islamiyati, A., Fatmawati & Chamidah, I.N. (2019). Penalized spline estimator with multi smoothing parameters in biresponse multipredictor regression model for longitudinal data. Songklanakarin Journal of Science and Technology, In Press SJST-2018-0423.R2.

[13] Lestari, B., Fatmawati & Budiantara, I.N. (2019). Smoothing spline estimator in multiresponse nonparametric regression for predicting blood pressure and heart rate. International Journal of Academic and Applied Research (IJAAR), 3 (9) : 1-8.

[14] Yuan, M. (2006). GACV for quantile smoothing spline. Computational statistics & data analysis, 50 (3) : 813-829.

[15] Marrie, R.A et.al. (2009). Quantile regression and restricted cubic splines are usefull for exploring relationships between continuous variables. Journal of clinical Epidemiology, 62 : 511 – 517.

[16] Islamiyati, A., Fatmawati & Chamidah, N. (2017). Fungsi goodness of fit dalam kriteria penalized spline pada estimasi regresi nonparametrik birespon untuk data longitudinal. *Proseding Seminar Nasional Matematika dan Aplikasinya*. UNAIR Surabaya.

[17] Islamiyati, A. (2019). Regresi spline polynomial truncated biprediktor untuk identifikasi perubahan jumlah trombosit pasien demam berdarah dengue. Al khwarizmi, 7 (2) : 97-110.

[18] Islamiyati, A., Fatmawati and Chamidah, N. (2019) Ability of covariance matrix in bi-response multi-prredictor penalized spline model through longitudinal data simulation. International Journal of Academic and Applied Research (IJAAR), 3(3) : 8-11.