
Estimasi Model Regresi Kuantil Spline Kuadratik pada Data Trombosit dan Hematokrit Pasien DBD

Bunga Aprilia^{1*}, Anna Islamiyati², Anisa³, Nirwan Ilyas⁴
^{1,2,3,4}Departemen Statistika, Fakultas MIPA,
Universitas Hasanuddin, Makassar, 90245, Indonesia

* Corresponding author, email: bungaprilialia021@gmail.com

Abstract

Nonparametric quantile 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 subtle 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 proposes to get the results of the estimation platelet count model to the hematocrit value of DHF. The optimal model obtained from the estimation of quadratic spline quantile regression is at quantile 0.5 with one knot and the GCV value is 41.5. The results of the estimation show that there is a decrease in platelet counts as the percentage of hematocrit increase.

Keywords: Hematocrit, Quantile, Quadratic, Spline, Platelet.

Abstrak

Regresi kuantil nonparametrik digunakan untuk memperkirakan fungsi regresi ketika asumsi tentang bentuk kurva regresi tidak diketahui dan hanya diasumsikan *smooth* dengan melibatkan nilai-nilai kuantil. Salah satu penaksir dalam regresi nonparametrik adalah spline. Sifat tersegmentasi dari spline memberikan lebih banyak fleksibilitas daripada polinomial biasa. Oleh karena itu, sifat spline memungkinkan untuk beradaptasi secara lebih efektif dengan karakteristik fungsi atau data lokal. Penelitian ini bertujuan untuk memperoleh estimasi model jumlah trombosit terhadap nilai hematokrit pasien DBD. Diperoleh model kuantil spline kuadratik yang optimal dengan nilai GCV minimum 41,5 pada kuantil 0,5 menggunakan satu titik knot. Hasil estimasi menunjukkan bahwa terjadi penurunan jumlah trombosit seiring dengan bertambahnya persentase hematokrit.

Kata Kunci: Hematokrit, Kuantil, Kuadratik, Spline, Trombosit.

1. Pendahuluan

Regresi merupakan analisis statistika yang digunakan untuk menyelidiki pola hubungan dan pengaruh beberapa variabel prediktor terhadap variabel respon [1]. Suatu model regresi dikatakan baik apabila memenuhi beberapa asumsi klasik, seperti asumsi normalitas dan homokedastisitas [2]. Namun asumsi ini seringkali tidak terpenuhi ketika data mengandung pencilan (*outlier*). Jika data mengandung pencilan, maka data tidak lagi berbentuk simetris sehingga nilai mean menjadi sangat peka dengan adanya data *outlier* dan kurang tepat digunakan [3]. Penanganan outlier dapat dilakukan diantaranya melalui regresi robust [4] dan regresi kuantil [5]. Regresi kuantil memungkinkan

menduga fungsi kuantil dari sebaran bersyarat respon pada berbagai nilai kuantil yang diinginkan. Setiap kuantil mencirikan titik tertentu (pusat atau ekor) dari sebaran bersyarat [6].

Pada kondisi tertentu, terdapat beberapa pola data yang tidak dapat dimodelkan dengan regresi kuantil parametrik karena akan menghasilkan *error* dan variansi yang besar sehingga disarankan untuk menggunakan pendekatan regresi nonparametrik. Regresi nonparametrik kuantil digunakan untuk mengestimasi fungsi regresi ketika asumsi tentang bentuk kurva regresi tidak diketahui dan hanya diasumsikan *smooth* dengan melibatkan nilai-nilai kuantil [7]. Untuk regresi nonparametrik, terdapat beberapa estimator diantaranya spline *truncated* [8,9,10], spline *smoothing* [11] dan spline *penalized* [12, 13, 14]. Spline terdiri atas beberapa potongan polinomial yang memiliki sifat tersegmen dan kontinu serta berorde tertentu yang saling bersambung pada titik-titik knot. Titik knot merupakan titik perpaduan bersama yang terjadi karena terdapat perubahan perilaku pola pada interval yang berlainan [15, 16]. Pendekatan spline dalam mengatasi pola data yang menunjukkan naik/turun yang tajam dapat dilakukan dengan bantuan titik-titik knot, serta kurva yang dihasilkan relatif mulus atau *smooth* [17].

Penelitian mengenai regresi kuantil nonparametrik telah banyak dilakukan, diantaranya, Balami dan Matdoan (2019) mengestimasi parameter regresi kuantil menggunakan fungsi spline *truncated* [18]. Chen, Chen dan Zhang (2019) mengestimasi kuantil *small area* melalui regresi spline [19]. Jatmiko (2018) menggunakan analisis regresi kuantil *B-Splines* [20]. Islamiyati, Raupong dan Anisa (2019) menggunakan *penalized spline* untuk mengidentifikasi pola perubahan gula darah pasien diabetes [21]. Oleh sebab itu, penelitian ini menggunakan regresi kuantil dengan estimator spline kuadratik untuk memodelkan jumlah trombosit pasien Demam Berdarah *Dengue* (DBD). Serangan virus terjadi pada sel trombosit sehingga indikator utama seseorang terkena DBD dalam pengujian medis adalah pengukuran sel trombosit [22]. Namun selain jumlah trombosit, diperlukan pemeriksaan medis lain diantaranya pemeriksaan kandungan hematokrit karena keduanya merupakan sel darah.

2. Material dan Metode

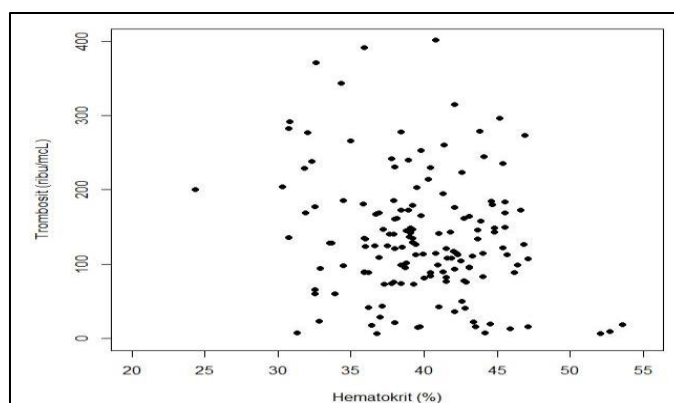
Data yang digunakan dalam penelitian ini adalah data sekunder dari pasien penderita Demam Berdarah *Dengue* (DBD) pada tahun 2013 – 2017 di Rumah Sakit Pendidikan Universitas Hasanuddin. Data terdiri angka trombosit pasien DBD sebagai respon (y) serta nilai hematokrit sebagai prediktor (x) yang diambil sejak hari pertama pasien dirawat di rumah sakit. Model yang digunakan adalah model regresi kuantil dengan estimator spline kuadratik sebagai berikut:

$$y(\theta) = \sum_{j=0}^2 \beta_j(\theta)x^j + \sum_{k=1}^r \beta_{2+k}(\theta)(x - K_k)_+^2 \quad (1)$$

Tahap pertama memperoleh model regresi kuantil spline linier adalah membuat scatter plot untuk melihat pola perubahan trombosit pasien berdasarkan hematokritnya. Selanjutnya pada tahap kedua menentukan titik knot optimal melalui nilai GCV spline kuadratik satu sampai tiga titik knot. Memodelkan regresi kuantil spline kuadratik $\theta = 0.25, 0.5,$ dan 0.75 berdasarkan titik knot optimal. Pada tahap ketiga dilakukan seleksi model untuk mendapatkan model kuantil spline kuadratik $\theta = 0.25, 0.5,$ dan 0.75 dengan titik knot optimal melalui nilai GCV minimum. Tahap keempat membandingkan nilai GCV model regresi kuantil spline kuadratik untuk mendapatkan model regresi kuantil spline optimal. Tahap terakhir adalah menginterpretasikan pola hubungan jumlah trombosit dengan hematokrit pada pasien DBD berdasarkan model optimal.

3. Hasil dan Diskusi

Sebelum dianalisis menggunakan analisis regresi kuantil spline terlebih dahulu akan dilakukan identifikasi pola hubungan antara variabel prediktor dengan variable jumlah trombosit pasien DBD dengan faktor yang mempengaruhi yaitu nilai hematokrit. Identifikasi dilakukan menggunakan scatter plot dan ditunjukkan pada Gambar 1.



Gambar 1. Plot Pengaruh Nilai Hematokrit (x) terhadap Jumlah trombosit (y)

Gambar 1 menunjukkan bahwa pola hubungan trombosit dengan hematokrit cenderung bervariasi. Banyak pasien DBD memiliki nilai hematokrit normal dengan jumlah trombosit yang abnormal. Ada juga pasien yang didiagnosis DBD memiliki trombosit normal dengan jumlah hematokrit di bawah normal yaitu $<38\%$. Sehingga hubungan kedua variabel tersebut cenderung tidak memiliki pola melalui sebaran data yang tidak beraturan, sehingga model estimasi regresi parametrik tidak dapat digunakan. Oleh karena itu, antara jumlah trombosit pasien DBD dengan nilai hematokrit diselesaikan dengan menggunakan pendekatan regresi nonparametrik melalui estimator spline kuadratik. Bentuk regresi yang digunakan adalah regresi kuantil. Ini terkait dengan distribusi data yang heterogen.

Pemodelan diawali dengan pemilihan titik knot optimal satu sampai tiga titik knot melalui nilai GCV spline kuadratik. Pemilihan titik knot dilakukan melalui *trial and error* dengan mengambil titik yang berada pada interval nilai variabel respon hingga didapatkan nilai GCV paling minimum. Hasil olahan ditunjukkan pada Tabel 1.

Tabel 1. Nilai GCV spline kuadratik satu sampai tiga titik knot

No.	Titik knot	GCV	Titik knot (K)		GCV	Titik knot			GCV
	K1		K1	K2		K1	K2	K3	
1	39.1	6264.6	43.4	46.9	6277.9	30.7	39.8	46.9	6318.1
2	40.8	6237.02	30.7	38	6258.8	30.5	44.2	47.1	6303.3
3	45.4	6192.4	31.8	42.6	6250.9	32.5	44.6	47.1	6301.6
4	44.8	6187.4	33.5	44.2	6247.1	40.4	45.2	47.1	6321.2
5	44.2	6188.5	30.7	45.2	6241.4	43.1	45.2	47.1	6306.9
6	44.6	6187.02*	30.3	44.7	6235.1	32.5	44.8	47.1	6298.6
7	44.7	6187.1	30.3	44.2	6233.5	30.7	44.8	47.1	6291.6
8	43.7	6192.9	30.7	44.2	6233.6	30.7	44.8	46.9	6292.9
9	45.2	6190.2	30.3	45.2	6241.3	30.3	44.8	47.1	6291.5
10	47.1	6207.6	30.7	44.6	6234.5	31.8	44.8	46.9	6296.7

Sumber: Data diolah 2020

Berdasarkan Tabel 1 nilai GCV minimum terbaik yang diperoleh dengan spline kuadratik adalah sebesar 6187.02 bersesuaian dengan titik knot optimal $K1 = 44.6$. Selanjutnya dilakukan estimasi model regresi kuantil spline kuadratik menggunakan titik knot optimal $K1 = 44.6$ pada kuantil 0.25, 0.5, 0.75

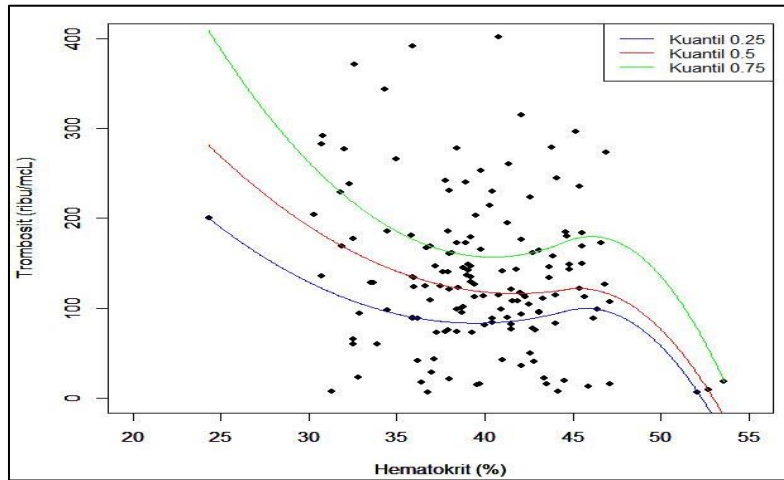
$$\hat{y}(0,25) = 872 - 39.9x + 0.5x^2 - 2.8(x - 44.6)_+^2$$

$$\hat{y}(0,5) = 1060.1 - 45.2x + 0.5x^2 - 2.6(x - 44.6)_+^2$$

$$\hat{y}(0,75) = 1741 - 78.4x + 0.9x^2 - 3.8(x - 44.6)_+^2$$

Model regresi untuk setiap kuantil yang diperoleh berbeda, tetapi kecenderungan perubahan trombosit tampak tidak berbeda secara signifikan. Selanjutnya kurva estimasi model kuantil spline kuadratik dengan satu titik knot yang menunjukkan pola perubahan di setiap kuantil disajikan pada Gambar 2. Pola perubahan yang terlihat cenderung sama pada jumlah trombosit pasien DBD ketika nilai hematokrit pasien dibawah 44.6% dan ketika nilai hematokrit meningkat di atas 44.6%. Pola perubahan trombosit berdasarkan jumlah hematokrit menunjukkan bahwa terjadi penurunan jumlah trombosit secara kuadratik seiring dengan meningkatnya nilai hematokrit hingga melewati batas normal, yaitu >45%. Hasil yang diperoleh telah sesuai dengan penelitian lain bahwa semakin tinggi persentase hematokrit jumlah trombosit akan semakin menurun. Hal ini dapat terjadi karena kebocoran plasma di pembuluh darah kapiler menyebabkan peningkatan nilai hematokrit dan trombosit yang berperan dalam pembekuan darah berusaha

menutupi kebocoran tersebut, sehingga kadar trombosit pun menjadi rendah [23]. Hasil ini diharapkan dapat menjadi informasi bagi tim medis untuk menyiapkan langkah-langkah awal perawatan sebelum nilai hematokrit pasien meningkat di atas 45%.



Gambar 2. Estimasi kurva kuantil Spline kuadratik Satu Knot antara Jumlah Trombosit Pasien DBD dengan Nilai Hematokrit

Selanjutnya, nilai-nilai GCV dari masing-masing model regresi untuk tiga kuantil dibandingkan untuk menunjukkan model yang paling mampu menjelaskan kondisi distribusi data dengan baik. Model terbaik adalah model yang menghasilkan nilai GCV minimum. Hasil olahan ditunjukkan pada Tabel 2.

Tabel 1. Nilai GCV untuk setiap model regresi kuantil spline kuadratik

Kuantil	0.25	0,5	0,75
GCV	56.2	41,5	51,9

Sumber: Data diolah 2020

Berdasarkan Tabel 2, dapat dilihat bahwa model kuantil spline kuadratik yang memiliki nilai GCV terkecil adalah 41,5 pada kuantil 0,5. Ini berarti bahwa median pembagian model lebih mampu menjelaskan keragaman data. Selain itu, apabila dibandingkan nilai GCV yang dihasilkan oleh model kuantil spline kuadratik jauh lebih kecil dari nilai GCV spline kuadratik. Artinya model regresi kuantil spline kuadratik jauh lebih baik digunakan untuk melihat pengaruh hematokrit terhadap perubahan jumlah trombosit pasien DBD. Berikut estimasi model terbaik yaitu model spline kuadratik dengan satu titik knot ($K = 44,6$) pada kuantil 0,5 sebagai berikut:

$$\hat{y}(0,5) = 1060.1 - 45.2x + 0.5x^2 - 2.6(x - 44.6)_+^2$$

$$= \begin{cases} 1060.1 - 45.2x + 0.5x^2 & , \quad x < 44.6 \\ -4165.4 + 189.1x - 2.1x^2 & , \quad x \geq 44.6 \end{cases} \quad (2)$$

4. Kesimpulan

Pola perubahan trombosit pada makalah ini diinterpretasikan melalui model regresi kuantil spline orde dua. Hal tersebut berdasarkan kriteria GCV minimum yang dihasilkan dibandingkan dengan model spline biasa. Terdapat kecenderungan nilai trombosit pasien semakin menurun ketika persentase nilai hematokrit pasien meningkat. Semakin tinggi nilai hematokrit terlihat trombosit semakin turun secara kuadratik. Titik knot optimal pada titik 44,6 yang menunjukkan bahwa setelah nilai hematokrit mencapai 44,6 % ada perubahan pola trombosit yang terjadi pada pasien DBD.

Makalah ini masih mempertimbangkan satu faktor yaitu hematokrit dalam mengevaluasi perubahan trombosit pada pasien DBD. Oleh sebab itu, kajian ini dapat dikembangkan ke penggunaan prediktor yang lebih banyak dengan menggunakan estimator spline yang lainnya, seperti spline smoothing, spline penalized, atau pun bentuk spline lainnya.

Daftar Pustaka

- [1] Islamiyati, A., Fatmawati & Chamidah, N. Estimation of Covariance Matrix on Bi-Response Longitudinal Data Analysis with Penalized Spline Regression. *Journal of Physics: Conference Series*, 979 (1) 012093, 2018.
- [2] Yanuar, F. Penerapan Metode Regresi Kuantil pada Kasus Pelanggaran Asumsi Kenormalan Sisaan. Skripsi. Universitas Andalas. 2016.
- [3] Balami, A.M. Estimasi Parameter Regresi Kuantil Pada Kasus Demam Berdarah Dengue di Kota Surabaya. Departemen Statistika FMIPA ITS, 2017.
- [4] Nurdin, N., Raupong & Islamiyati, A. Penggunaan Regresi Robust pada Data yang Mengandung Pencilan dengan Metode Momen. *Jurnal Matematika, Statistika & Komputasi*, 10 (2) : 114-123, 2014.
- [5] Aprilia, B., Islamiyati, A. & Anisa. Platelet Modeling Based On Hematocrit in DHF Patients with Spline Quantile Regression. *International Journal of Academic and Applied Research*, 3 (12) : 51-54, 2019.
- [6] Furno, M. Prediction on Quantile Regression. *Open Journal of Statistics*, 4 : 504-517, 2014.
- [7] Putri, W.N.A., Islamiyati, A. & Anisa. Penggunaan Regresi Multivariat pada Perubahan Trombosit Pasien Demam Berdarah Dengue. *ESTIMASI: Journal of Statistics and Its Application*, 1 (1) : 1-9, 2020.
- [8] Islamiyati, A. Taksiran Kurva Regresi Spline pada Data Longitudinal dengan Kuadrat Terkecil. *Jurnal Matematika, Statistika & Komputasi*, 11(1) : 97-102, 2014.
- [9] Ramdhani, Z.A., Islamiyati, A. & Raupong. Hubungan Faktor Kolesterol Terhadap Gula Darah Diabetes dengan Spline Kubik Terbobot. *ESTIMASI: Journal of Statistics and Its Application*, 1 (1) : 32-39, 2020.
- [10] Arifin, S., Islamiyati, A. & Raupong. Kemampuan Estimator Spline Linear dalam Analisis Komponen Utama. *ESTIMASI: Journal of Statistics and Its Application*, 1 (1) : 40-47, 2020.

- [11] Lestari, B., Fatmawati & Budiantara, I.N. Smoothing Spline Estimator in Multiresponse Nonparametric Regression for Predicting Blood Pressure and Heart Rate. *International Journal of Academic and Applied Research*, 3 (9) : 1-8, 2019.
- [12] Islamiyati, A., Fatmawati & Chamidah, N. Fungsi Goodness of Fit dalam Kriteria Penalized Spline pada Estimasi Regresi Nonparametrik Birespon untuk Data Longitudinal. *Proseding Seminar Nasional Matematika dan Aplikasinya*. UNAIR Surabaya, 2017.
- [13] Islamiyati, A., Fatmawati & Chamidah, N. Penalized Spline Estimator With Multi Smoothing Parameters in Biresponse Multipredictor Regression Model for Longitudinal Data. *Songklanakarinn Journal of Science and Technology*, 42 (4) : 897-909, 2020.
- [14] Islamiyati, A., Sunusi, N., Kalondeng, A., Wati, F. & Chamidah, N. Use of Two Smoothing Parameters in Penalized Spline Estimator for Bi-Variate Predictor Non-Parametric Regression Model. *Journal of Sciences, Islamic Republic of Iran*, 31 (2) : 175-183, 2020.
- [15] Ahmad, R.R., Ghazali, N., Rambeli, A.S., Din, U. K. S. & Hassan, N. Application of Cubic Spline in the Implementation of Braces for the Case of a Child. *Journal of Mathematics and Statistics*, 8 (1) : 144-149, 2012.
- [16] Islamiyati, A., Fatmawati & Chamidah, N. Changes in Blood Glucose 2 Hours After Meals in Type 2 Diabetes Patients Based On Length of Treatment at Hasanuddin University Hospital, Indonesia. *Rawal Medical Journal*, 45 (1) : 31-34, 2020.
- [17] Islamiyati, A. Spline Polynomial Truncated dalam Regresi Nonparametrik. *Jurnal Matematika, Statistika & Komputasi*, 14 (1) : 54-60, 2017.
- [18] Balami, A. M. & Matdoan, M.Y. Estimasi Parameter Regresi Kuantil dengan Fungsi Spline Truncated pada Kasus Demam Berdarah Dengue di Kota Surabaya. *JURNAL MSA*, 7 (1), 2019.
- [19] Chen, Z., Chen, J. & Zhang, Q. Small Area Quantile Estimation Via Spline Regression and Empirical Likelihood. *Statistics Canada, Catalogue No. 12-001-X*, 2019.
- [20] Jatmiko, Y.A. Analisis Regresi Kuantil B-Splines Monoton Naik pada Hubungan Rata-Rata Lama Sekolah dan Pengeluaran Rumah Tangga Per Kapita di Provinsi Yogyakarta. *JURNAL MATEMATIKA "MANTIK"*, 4 (2), 2018.
- [21] Islamiyati, A., Raupong & Anisa. Use of Penalized Spline Linear to Identify Change in Pattern of Blood Sugar Based On the Weight of Diabetes Patients. *International Journal of Academic and Applied Research*, 3(12) : 75-78, 2019.
- [22] Islamiyati, A. Regresi Spline Polynomial Truncated Biprediktor Untuk Identifikasi Perubahan Jumlah Trombosit Pasien Demam Berdarah Dengue. *Al khwarizmi*, 7 (2) : 97-110, 2019.
- [23] Syumarta, Y., Akmal, M. & Hanif, E. R. Hubungan Jumlah Trombosit, Hematokrit dan Hemoglobin dengan Derajat Klinik Demam Berdarah Dengue pada Pasien Dewasa di RSUP . *M.Djamil Padang*. 38, 492–498, 2013.