

# ESTIMASI PARAMETER REGRESI KUANTIL DENGAN FUNGSI SPLINE *TRUNCATED* PADA KASUS DEMAM BERDARAH *DENGUE* DI KOTA SURABAYA

A. M. Balami<sup>i</sup>, M. Y. Matdoan<sup>ii</sup>

<sup>i</sup> Universitas Pattimura, malik.balami@fmipa.unpatti.ac.id

<sup>ii</sup> Universitas Pattimura, yahya.matdoan@fmipa.unpatti.ac.id

**ABSTRAK**, Metode regresi kuantil merupakan perluasan model regresi pada kuantil bersyarat dimana distribusi kuantil bersyarat dari variabel respon dinyatakan sebagai fungsi dari kovariat yang diamati. Spline merupakan potongan polinomial yang kontinu, sehingga dapat menggambarkan karakteristik lokal pada data. Penelitian ini memodelkan Regresi Kuantil dengan Fungsi Spline *Truncated* pada kasus DBD di Kota Surabaya. Estimasi regresi kuantil dengan fungsi spline *truncated* diperoleh koefisien determinasi ( $R^2$ ) sebesar 72% pada kuantil 0,95 dengan menggunakan tiga titik knot. Lebih lanjut diperoleh hasil bahwa faktor utama penyebab penyebaran DBD di Kota Surabaya adalah faktor persentase rumah/bangunan bebas jentik nyamuk *Aedes Aegypti*, faktor rasio tenaga medis (dokter umum), faktor rasio sarana kesehatan Puskesmas dan faktor persentase rumah tangga yang memiliki tempat sampah sehat

**Kata Kunci:** Estimasi Parameter, Regresi Kuantil, Fungsi Spline *Truncated*, Demam Berdarah *Dengue*.

## 1. PENDAHULUAN

Analisis regresi digunakan untuk mengetahui pola hubungan dan pengaruh variabel prediktor terhadap variabel respon, dengan mengestimasi kurva regresinya. Terdapat tiga pendekatan untuk mengestimasi kurva regresi yaitu pendekatan regresi parametrik, regresi nonparametrik dan regresi semiparametrik [3]. Informasi mengenai hubungan fungsional antara variabel prediktor (independen) dengan variabel respon (dependen) dapat diperkirakan dengan melihat bentuk pola hubungan pada diagram pencar (*scatter plot*). Bentuk kurva regresi parametrik diasumsikan diketahui. Penggunaan pendekatan regresi parametrik memerlukan pengetahuan masa lalu tentang karakteristik data yang akan diselidiki. Berbeda dengan pendekatan regresi nonparametrik, dalam regresi nonparametrik bentuk kurva regresi di asumsikan tidak diketahui dalam ruang *sobolev* dan hanya di asumsikan *smooth* (mulus, halus, licin) [2].

Metode regresi kuantil pertama kali diperkenalkan oleh [10]. Metode regresi kuantil merupakan perluasan model regresi pada kuantil

bersyarat dimana distribusi kuantil bersyarat dari variabel respon dinyatakan sebagai fungsi dari kovariat yang diamati. Metode ini dapat digunakan mengukur efek *explanatory variable* tidak hanya di pusat sebaran data, tetapi juga pada bagian atas atau bawah ekor sebaran. Setiap kuantil mencirikan titik tertentu (pusat atau ekor) dari sebaran bersyarat. Kombinasi berbagai nilai kuantil akan menghasilkan deskripsi lengkap tentang sebaran bersyarat. Analisis ini berguna untuk sebaran bersyarat yang asimetris, padat di ekor sebarannya, atau sebarannya terpotong. Keuntungan utama dari regresi kuantil dibandingkan regresi OLS adalah fleksibilitas dalam pemodelan data dengan sebaran bersyarat yang heterogen.

Spline merupakan potongan polinomial yang kontinu, sehingga dapat menggambarkan karakteristik lokal pada data [6]. Salah satu teknik estimasi dalam regresi nonparametrik adalah estimator spline. Spline adalah salah satu jenis *piecewise* polinomial, yaitu polinomial yang memiliki sifat tersegmen. Kelebihan sifat tersegmen ini memberikan fleksibilitas lebih dari polinomial biasa, sehingga memungkinkan untuk menyesuaikan diri secara lebih efektif terhadap karakteristik lokal suatu fungsi atau data. Spline juga mempunyai keunggulan dalam mengatasi pola data yang cenderung naik/turun secara tajam, serta kurva yang dihasilkan relatif mulus. Spline *truncated* adalah basis fungsi dalam spline yang merupakan model polinomial yang tersegmen atau terbagi pada suatu titik fokus yang disebut knot.

Penyakit Demam Berdarah *Dengue* (DBD) merupakan penyakit yang disebabkan virus *dengue* yang dibawa oleh nyamuk *Aedes aegypti*. Penyakit ini mulanya ditemukan sekitar tahun 1950-an di Filipina berdasarkan laporan Quintos mengenai epidemi penyakit dengan gejala panas, pendarahan akut dan *shock*. Hal ini menandai permulaan penyebaran DBD di kawasan Asia

Tenggara. Berdasarkan laporan *World Health Organization* (WHO), bila dibandingkan dengan negara-negara lain di kawasan Asia Tenggara, maka kasus DBD di Indonesia adalah yang paling tinggi. Bahkan Filipina sebagai negara asal kemunculan penyakit ini memiliki jumlah kasus yang lebih rendah.

Di Indonesia demam berdarah pertama kali ditemukan di kota Surabaya pada tahun 1968, dimana sebanyak 58 orang terinfeksi dan 24 orang diantaranya meninggal dunia [9]. Berdasarkan Data Profil Kesehatan Jawa Timur Tahun 2013 menunjukkan bahwa Surabaya merupakan kota dengan persentase angka tertinggi, yang diikuti dengan Kab. Bojonegoro, Jombang, Kediri dan Bangkalan. Perkembangan kasus DBD per bulan di kota Surabaya pada lima tahun terakhir menunjukkan bahwa telah terjadi peningkatan. Perkembangan kasus DBD per bulan selalu mulai mengalami peningkatan pada bulan Februari setiap tahunnya, puncak kasusnya adalah pada bulan Maret [5]. Hal ini menunjukkan bahwa penanganan DBD di Kota Surabaya membutuhkan perhatian yang sangat serius.

Penelitian yang berkaitan dengan DBD di Indonesia sudah banyak dilakukan. Penelitian mengenai DBD yang menerapkan analisis regresi parametrik pernah dilakukan oleh Rahmawati *et al.* [13], di mana penelitiannya bertujuan untuk mengetahui pengaruh karakteristik wilayah terhadap jumlah kasus DBD. Yussanti (2012) melakukan pemodelan DBD di Jawa Timur berdasarkan faktor iklim dan sosial ekonomi dengan pendekatan regresi panel semiparametrik. Sedangkan penelitian yang menggunakan regresi nonparametrik sebelumnya dilaksanakan oleh Mubarak [12] yang memanfaatkan pendekatan spline *truncated* untuk memodelkan kematian penderita DBD di Jawa Timur. Penelitian-penelitian terdahulu yang sudah disebutkan sebelumnya, tidak ada satupun yang menggunakan pendekatan analisis regresi kuantil. Hal ini menunjukkan bahwa metode analisis ini belum digunakan khususnya oleh peneliti dalam negeri. Beberapa penelitian yang pernah memanfaatkan regresi kuantil antara lain adalah Abrevaya [1] dalam penelitiannya di bidang demografi, Koenker dan Billias [10] untuk data durasi dan Machado dan Mata [11]

yang menerapkan regresi kuantil fungsi pendapatan di Portugal.

Bentuk hubungan fungsional antara variabel respon dengan variabel prediktor pada regresi kuantil dapat dikembangkan dalam bentuk nonparametrik. Salah satu bentuk hubungan fungsional nonparametrik adalah spline

## 2. TINJAUAN PUSTAKA

### ANALISIS REGRESI

Analisis regresi digunakan untuk melihat hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Misalkan terdapat data berpasangan  $(x_i, y_i)$  untuk  $n$  pengamatan, maka hubungan antara variabel  $x_i$  dan variabel  $y_i$  dapat dinyatakan sebagai berikut:

$$y_i = f(x_i) + \varepsilon_i ; i = 1, 2, \dots, n \quad (1.1)$$

dimana :

$y_i$  : Variabel respon pada pengamatan ke- $i$

$f(x_i)$  : Kurva regresi pada pengamatan ke- $i$

$\varepsilon_i$  : Residual (error) pada pengamatan ke- $i$

### REGRESI KUANTIL

Regresi Kuantil merupakan suatu pendekatan dalam analisis regresi yang dikenalkan oleh Koenker dan Bassett (1978). Pendekatan ini menduga berbagai fungsi kuantil dari suatu distribusi  $Y$  sebagai fungsi dari  $X$ . Regresi Kuantil sangat berguna jika distribusi data tidak homogen (*heterogenous*). Misalkan diberikan data  $\{X_{i1}, X_{i2}, \dots, X_{in}, Y_i\}$ ,  $i = 1, 2, \dots, n$  merupakan himpunan berpasangan dari variabel random yang berdistribusi secara independen dan tidak identik dengan kuantil  $\theta \in (0, 1)$ .

Persamaan umum regresi kuantil linier yaitu sebagai berikut:

$$Y_i = \beta_0(\theta) + \beta_1(\theta)X_{i1} + \dots + \beta_k(\theta)X_{ik} + \varepsilon_i(\theta) \quad (1.2)$$

Selanjutnya persamaan (2), dapat ditulis dalam bentuk model linier berikut:

$$Y = X\beta(\theta) + \varepsilon(\theta) \quad (1.3)$$

### FUNGSI SPLINE TRUNCATED

Spline mempunyai keunggulan dalam mengatasi pola data yang menunjukkan naik atau turun yang

tajam dengan bantuan titik-titik knot, serta kurva yang dihasilkan relatif mulus. Titik knot merupakan perpaduan bersama yang menunjukkan pola perilaku fungsi spline pada selang yang berbeda [8].

Fungsi spline memiliki knot yang merupakan titik perpaduan bersama yang menunjukkan perubahan kurva. Spline orde  $m$  dengan titik knot  $K_1, K_2, \dots, K_k$  dapat ditulis sebagai berikut:

$$f(x_i) = \sum_{j=1}^m \beta_j x_i^j + \sum_{k=1}^N \beta_{m+k} (x_i - K_k)_+^m \quad (1.4)$$

dengan  $\beta_j$  merupakan parameter-parameter model dan

$$(x_i - K_k)_+^m = \begin{cases} (x_i - K_k)^m, & \text{jika } x \geq K_k \\ 0, & \text{jika } x < K_k \end{cases} \quad (1.5)$$

Berdasarkan bentuk matematis fungsi spline, dapat dikatakan bahwa spline merupakan model polinomial yang sepotong-sepotong (*piecewise polynomial*) dan spline masih bersifat kontinu pada knot-knotnya. Knot diartikan sebagai suatu titik fokus dalam fungsi spline, sehingga kurva yang dibentuk tersegmen pada titik tersebut dan untuk setiap fungsi  $m$ , titik knot dapat dinyatakan dengan kombinasi linier. Fungsi spline merupakan suatu gabungan fungsi polinomial dimana penggabungan beberapa polinomial tersebut pada knot-knot dengan suatu cara yang menjamin sifat kontinuitas. Spline adalah potongan polinomial mulus yang masih memungkinkan memiliki sifat tersegmen [6].

Error random  $\varepsilon_i$  diasumsikan berdistribusi normal independen dengan mean nol dan variansi  $\sigma^2$ . Sebagai salah satu ilustrasi sederhana diberikan spline linier *truncated* dengan tiga knot pada  $x_1 = K_1 \leq x_2 = K_2 \leq x_3 = K_3$ .

### ESTIMASI PARAMETER DALAM REGRESI SPLINE LINIER

Diberikan model regresi  $Y_i = f(x_i) + \varepsilon_i, i = 1, 2, \dots, n$ .

Jika didekati dengan fungsi spline linier dengan tiga knot pada  $x = K_1, x = K_2$  dan  $x = K_3$  maka  $f$  dapat ditulis menjadi:

$$f(x) = \beta_1 x + \beta_2 (x - K_1)_+^1 + \beta_3 (x - K_2)_+^1 + \beta_4 (x - K_3)_+^1 \quad (1.6)$$

Berdasarkan persamaan (14), diperoleh model regresi spline linear:

$$Y_i = \beta_1 x + \beta_2 (x - K_1)_+^1 + \beta_3 (x - K_2)_+^1 + \beta_4 (x - K_3)_+^1 + \varepsilon_i, i = 1, 2, \dots, n \quad (1.7)$$

Apabila model regresi spline ini disajikan dalam matrik didapat:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 & (x_1 - K_1)_+^1 & (x_1 - K_2)_+^1 & (x_1 - K_3)_+^1 \\ x_2 & (x_2 - K_1)_+^1 & (x_2 - K_2)_+^1 & (x_2 - K_3)_+^1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & (x_n - K_1)_+^1 & (x_n - K_2)_+^1 & (x_n - K_3)_+^1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

dengan  $K_1 < K_2 < K_3$

$$(1.8)$$

Estimasi parameter  $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)^T$  diperoleh dengan menggunakan metode kuadrat terkecil. Khususnya estimator  $\hat{\beta}$  didapat dengan menyelesaikan optimasi:

$$\arg \min_{\beta} \{ \varepsilon^T \varepsilon \} = \arg \min_{\beta} \{ (y - \gamma[K_1, K_2, K_3] \beta)^T (y - \gamma[K_1, K_2, K_3] \beta) \} \quad (1.9)$$

Dengan  $y = [y_1, y_2, \dots, y_n]^T$  dan  $\gamma[K_1, K_2, K_3]$  matriks berukuran  $n \times 4$  yang diberikan oleh:

$$\gamma[K_1, K_2, K_3] = \begin{bmatrix} x_1 & (x_1 - K_1)_+^1 & (x_1 - K_2)_+^1 & (x_1 - K_3)_+^1 \\ x_2 & (x_2 - K_1)_+^1 & (x_2 - K_2)_+^1 & (x_2 - K_3)_+^1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & (x_n - K_1)_+^1 & (x_n - K_2)_+^1 & (x_n - K_3)_+^1 \end{bmatrix} \quad (1.10)$$

Digunakan derivatif parsial dan menyelesaikan persamaan di atas diperoleh estimator  $\hat{\beta}$ :

$$\hat{\beta} = (\gamma^T [K_1, K_2, K_3] \gamma [K_1, K_2, K_3])^{-1} \gamma^T [K_1, K_2, K_3] Y$$

Akibatnya estimasi untuk kurva regresi  $f(x)$  diberikan oleh:

$$\begin{aligned} \hat{f}(x) &= \gamma[K_1, K_2, K_3] \hat{\beta} \\ &= \gamma[K_1, K_2, K_3] (\gamma^T [K_1, K_2, K_3] \gamma [K_1, K_2, K_3])^{-1} \\ &\quad \gamma^T [K_1, K_2, K_3] Y \\ &= H[K_1, K_2, K_3] Y \end{aligned} \quad (1.11)$$

dengan

$$H[K_1, K_2, K_3] = \gamma [K_1, K_2, K_3] (\gamma^T [K_1, K_2, K_3] \gamma [K_1, K_2, K_3])^{-1} \gamma^T [K_1, K_2, K_3]$$

Terlihat bahwa  $\hat{f}(x)$  merupakan estimator linear dalam observasi  $Y = [Y_1, Y_2, \dots, Y_n]^T$

### PEMILIHAN TITIK KNOT OPTIMAL

Salah satu metode pemilihan titik knot optimal adalah *Generalized Cross Validation* (GCV).

Model spline yang terbaik dengan titik knot optimal didapat dari nilai GCV yang terkecil (Budiantara, 2000). Fungsi GCV seperti pada persamaan berikut

$$GCV(\lambda) = \frac{MSE(\lambda)}{[n^{-1}trace(\mathbf{I} - \mathbf{A}(\lambda))]^2} \quad (1.11)$$

dengan :  $MSE(\lambda) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$

### 3. METODOLOGI

#### Sumber Data

Penelitian ini menggunakan data sekunder melalui publikasi data profil kesehatan di Dinas Kesehatan Kota Surabaya, dan Badan Pusat Statistik (BPS) tahun 2017 dengan unit observasi adalah 31 kecamatan di kota Surabaya [5].

#### Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini terdiri dari variabel respon ( $Y$ ) yaitu persentase kasus DBD di Kota Surabaya tahun 2013. Sedangkan variabel prediktor yang digunakan yaitu Persentase rumah/bangunan bebas jentik nyamuk *Aedes Aegypti* ( $X_1$ ), persentase rumah tangga miskin ( $X_2$ ), rasio tenaga medis (dokter umum) ( $X_3$ ), rasio sarana kesehatan (Puskesmas) ( $X_4$ ), persentase rumah tangga yang memiliki tempat sampah sehat ( $X_5$ )

#### Metode Analisis

Metode penelitian yang digunakan sebagai langkah-langkah untuk mencapai tujuan penelitian yaitu melakukan identifikasi hubungan antara variabel banyaknya kasus DBD dengan faktor-faktor yang memengaruhinya menggunakan *scatter plot* serta korelasi. Melakukan pengujian multikolinieritas antar variabel prediktor terhadap kasus DBD di kota Surabaya. Jika tidak terdapat kasus multikolinieritas pada data, maka dapat dilanjutkan pada langkah berikutnya sebagai berikut:

1. Memodelkan regresi kuantil pada data banyaknya kasus DBD di kota Surabaya . langkah-langkah sebagai berikut.

- a. Mengestimasi parameter model regresi kuantil, yaitu dengan mencari nilai estimasi  $\hat{\beta}$  dengan meminimumkan fungsi

$$\min_{\beta} \sum_{i=1}^n |y_i - \mathbf{x}_i^T \beta|$$

- b. Mengoptimasi simpleks dengan proses sebagai berikut:

- Menghitung nilai  $s_1 = [y - \mathbf{x}\beta]_+$  dan nilai  $s_2 = [\mathbf{x}\beta - y]_+$  sehingga diperoleh permasalahan linier.
- Menghitung nilai  $\min_{\phi} \mathbf{d}^T \phi$  dan memperoleh nilai parameter dual dari  $\max_{\mathbf{d}} \mathbf{y}^T \mathbf{z}$  dimana  $\mathbf{B}^T \mathbf{z} \leq \mathbf{d}$
- Menentukan permasalahan minimalisasi yaitu dengan formulasi  $\max_{\mathbf{z}} \{ \mathbf{y}^T \mathbf{z} \mid \mathbf{X}^T \mathbf{z} = \mathbf{0}, \mathbf{z} \in [-1, +1]^n \}$

- Untuk  $\mathbf{X}^T \mathbf{z} = \mathbf{0}$  dapat ditransformasikan menjadi

$$\mathbf{X}^T \left( \frac{1}{2} \mathbf{z} + \frac{1}{2} \right)$$

- Jika  $\frac{1}{2} \mathbf{z} + \frac{1}{2} = \eta$  dan  $\frac{1}{2} \mathbf{X}^T \mathbf{1} = \mathbf{b}$  maka diperoleh  $\max \{ \mathbf{y}^T \mathbf{J} \mid \mathbf{X}^T \mathbf{J} = \mathbf{b}, \mathbf{J} \in [0, 1]^n \}$

- c. Mendapatkan estimator  $\hat{\beta}$  yaitu sebagai berikut:

$$\hat{\beta}(\theta) = \min_{\beta} \theta \sum_{j=1}^n |y_j - \mathbf{x}_j^T \beta| + (1 - \theta) \sum_{j=1}^n |y_j - \mathbf{x}_j^T \beta|$$

2. Memodelkan kasus DBD pada masing-masing kuantil dengan menggunakan fungsi spline truncated. Langkah analisisnya sebagai berikut:

- a. Memodelkan variabel dependen dan variabel independen dengan menggunakan spline linear ( $m = 1$ ) dan dengan menggunakan satu titik knot, dua titik knot dan tiga titik knot.
- b. Memilih model spline terbaik dengan memilih titik knot optimum berdasarkan GCV minimum.

3. Mendapatkan model regresi spline dengan titik knot optimal..

**4. PEMBAHASAN**

Pada bagian ini akan dilakukan analisis dan pembahasan dari data yang telah diolah dengan menggunakan beberapa metode analisis yang sudah dibahas pada bagian sebelumnya. Sebelum dilakukan pemodelan banyaknya kasus DBD di Surabaya, terlebih dahulu akan dilakukan identifikasi pola hubungan antara variabel persentase kasus DBD di Surabaya dengan faktor-faktor yang mempengaruhinya. Identifikasi dilakukan melalui pembuatan *scatter plot* dan korelasi antara variabel respon dengan beberapa prediktor yang mempengaruhinya disajikan pada Tabel 1.

**Tabel 1.** Hasil Korelasi antara variabel

Korelasi	Y	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>
X <sub>1</sub>	[0,02] (0,89)				
X <sub>2</sub>	[-0,16] (0,39)	[0,07] (0,70)			
X <sub>3</sub>	[0,26] (0,15)	[0,69] (0,00)	[-0,21] (0,24)		
X <sub>4</sub>	[0,16] (0,39)	[-0,16] (0,39)	[-0,04] (0,82)	[-0,05] (0,77)	
X <sub>5</sub>	[-0,04] (0,81)	[0,05] (0,78)	[-0,08] (0,68)	[0,16] (0,38)	[-0,00] (0,98)

Ket : [ ] Korelasi Pearson, ( ) *P-value*, Berdasarkan tabel 1 dapat diketahui bahwa terdapat tiga variabel yang berkorelasi positif terhadap persentase DBD, yaitu persentase rumah/bangunan bebas jentik nyamuk *Aedes Aegypti* (X<sub>1</sub>), rasio tenaga medis (dokter umum) (X<sub>3</sub>) dan rasio sarana kesehatan puskesmas (X<sub>4</sub>). Hal ini menunjukkan bahwa jika terjadi peningkatan terhadap setiap variabel prediktor akan mengakibatkan semakin tingginya persentase kasus DBD. Sementara itu, variabel persentase rumah tangga miskin (X<sub>2</sub>) dan persentase rumah tangga yang memiliki tempat sampah sehat (X<sub>5</sub>) memiliki korelasi negatif, artinya jika terjadi peningkatan pada variabel tersebut maka akan mengakibatkan penurunan pada kasus DBD. Bila dilihat dari koefisien korelasi antar variabel prediktor dapat disimpulkan bahwa tidak terjadi multikolinieritas karena nilai koefisien korelasinya tidak lebih dari 0,95. Untuk lebih memastikan ada atau tidaknya multikolinieritas, dapat dilihat dari nilai *Variance Inflation Factors* (VIF) dari tiap

variabel prediktor. Berikut adalah nilai *Variance Inflation Factors* (VIF) dari masing-masing variabel prediktor yang digunakan dalam penelitian ini:

**Tabel 2.** Nilai Variance Inflation Factor (VIF)

Prediktor	Nilai VIF
X <sub>1</sub>	2,21
X <sub>2</sub>	1,17
X <sub>3</sub>	2,31
X <sub>4</sub>	1,03
X <sub>5</sub>	1,03

Tabel diatas menunjukkan bahwa nilai VIF untuk semua variabel prediktor adalah lebih kecil dari 10 yang artinya tidak terjadi multikolinieritas diantara variabel prediktor.

**Estimasi Parameter Regresi Kuantil**

Dalam penelitian ini metode regresi kuantil digunakan untuk mendapatkan model pada tiap kuantil yang menggambarkan seberapa besar pengaruh variabel prediktor terhadap persentase kasus DBD pada tiap kuantil sehingga diperoleh besarnya koefisien untuk masing-masing variabel prediktor. Selain itu, dapat dilihat seberapa besar efek variabel prediktor mempengaruhi persentase kasus DBD pada tiap kuantil. Hasil estimasi parameter disajikan pada Tabel 3.

**Tabel 3.** Hasil estimasi parameter regresi kuantil

Paramet er	Kuantil				
	0,05	0,25	0,5	0,75	0,95
$\beta_0$	-	-	-	2,67	-
$\beta_1$	0,71046	1,23	1,31303	-	0,39052
$\beta_2$	-	-	-	0,00	-
$\beta_3$	0,00992	0,00 6	0,00806	0,05	0,10957
$\beta_4$	-	-	-	0,00	-
$\beta_5$	0,02293	0,02 7	0,01215	2	0,01298
$\beta_6$	0,02487	0,02 6	0,01442	0,04	0,13020
$\beta_7$	0,00371	0,00 8	0,01003	0,01 3	0,04335
$\beta_8$	-	-	-	-	-
$\beta_9$	0,44361	0,78	1,96369	2,77	5,25682

Berdasarkan hasil estimasi parameter pada tabel 3, maka model regresi kuantil adalah sebagai berikut:

**Kuantil 0,05**

$$Q_{0,05}(y|x) = -0,71046 - 0,00992x_1 - 0,02293x_2 + 0,02487x_3 + 0,00704x_4 - 0,44361x_5$$

**Kuantil 0,50**

$$Q_{0,50}(y|x) = -1,31303 + 0,00806x_1 - 0,01215x_2 + 0,01442x_3 + 0,01003x_4 - 1,96369x_5$$

**Kuantil 0,95**

$$Q_{0,95}(y|x) = -0,39052 - 0,10957x_1 - 0,01298x_2 + 0,13020x_3 + 0,04335x_4 - 5,25682x_5$$

Model diatas adalah model kuantil 0,05; 0,50 dan 0,95. Model lain dapat dituliskan mengacu pada Tabel 3.

Dari beberapa kuantil yang ditampilkan terlihat bahwa dalam mempengaruhi DBD, beberapa variabel prediktor mempunyai pengaruh positif dan negatif. Selanjutnya akan dilihat variabel yang signifikan mempengaruhi DBD pada setiap kuantil. Hasil signifikansi parameter disajikan pada Tabel 4.

**Tabel 4.** *P-value* dari Hasil estimasi parameter

Parameter	Kuantil				
	0,05	0,25	0,5	0,75	0,95
$\beta_0$	0,07	0,45	0,40	0,00*	0,87
$\beta_1$	0,07	0,78	0,71	0,00*	0,00*
$\beta_2$	0,00*	0,23	0,57	0,81	0,71
$\beta_3$	0,00*	0,16	0,41	0,00*	0,00*
$\beta_4$	0,03*	0,28	0,15	0,00*	0,00*
$\beta_5$	0,16	0,55	0,12	0,00*	0,01*

Berdasarkan hasil signifikasni parameter terlihat bahwa variabel-variabel prediktor yang berpengaruh terhadap DBD untuk tiap kuantil berbeda. Pada kuantil 0,05, variabel prediktor yang berpengaruh signifikan terhadap persentase DBD adalah variabel persentase rumah tangga miskin, variabel rasio tenaga medis (dokter umum) dan variabel rasio sarana kesehatan Puskesmas. Pada kuantil 0,25 dan 0,50 tidak ada variabel yang berpengaruh signifikan terhadap persentase DBD. Pada kuantil 0,75 dan 95 semua variabel prediktor berpengaruh signifikan terhadap persentase DBD, kecuali variabel persentase rumah tangga miskin.

Model regresi kuantil juga dapat dilihat *goodness of fit* pada tiap kuantil yang disajikan pada tabel 5 berikut ini.

**Tabel 5.** Nilai  $R^2$  Pada Tiap Kuantil

Kuantil	0,05	0,25	0,5	0,75	0,95
$R^2$	22%	11%	7%	8%	40%

Koefisien determinasi ( $R^2$ ) merupakan besaran yang digunakan untuk mengukur kelayakan model regresi dan menunjukkan besar kontribusi variabel  $x$  terhadap perubahan variabel  $y$ . Semakin tinggi nilai  $R^2$  mendekati 1 semakin baik model regresi yang terbentuk. Berdasarkan tabel 5 dapat dilihat bahwa pada model kuantil  $Q_{0,95}$ , memiliki nilai koefisien determinasi ( $R^2$ ) sebanyak 40% lebih besar dibandingkan dengan model kuantil yang lain. Artinya model ini dapat menjelaskan kasus DBD sebesar 40%.

Berdasarkan Tabel 4 dapat disimpulkan bahwa dalam pengambilan suatu kebijakan dalam mengatasi DBD tidak dapat digunakan satu model yang sama untuk seluruh kecamatan di Surabaya karena sebaran data yang heterogen dapat menghasilkan error yang besar. Selain itu diperoleh variabel prediktor yang berbeda untuk tiap kuantil yang signifikan mempengaruhi DBD sehingga treatment dalam mengatasi DBD akan berbeda untuk tiap kuantil.

**Interpertasi Model Regresi Kuantil Terbaik**

Berdasarkan model terbaik yaitu model estimasi regresi kuantil  $Q_{0,95}$ , menunjukkan bahwa variabel presentase rumah/bangunan bebas jentik nyamuk *Aedes Aegypti* ( $X_1$ ) berpengaruh negatif dan signifikan terhadap persentase kasus DBD di Kota Surabaya. Pada kuantil  $Q_{0,95}$  menunjukkan bahwa koefisien variabel presentase rumah/bangunan bebas jentik nyamuk *Aedes Aegypti* ( $X_1$ ) sebanyak minus 0,10957. Artinya jika presentase rumah/bangunan bebas jentik nyamuk *Aedes Aegypti* di suatu daerah naik sebanyak 1 persen, maka akan menurunkan tingkat persentase DBD sebanyak 0,11 per 1000 penduduk. Hal ini sesuai dengan hipotesis yang diharakan bahwa kondisi rumah/bangunan bebas jentik nyamuk *Aedes Aegypti* dapat mengurangi persentase DBD. Kondisi rumah dengan dinding yang terbuat dari kayu memungkinkan lebih

banyak lubang untuk masuknya nyamuk. Demikian pula lantai rumah dari tanah menyebabkan kondisi rumah menjadi lembab sehingga menjadi tempat yang ideal bagi nyamuk.

Selanjutnya estimasi regresi kuantil  $Q_{0,95}$ , menunjukkan bahwa variabel presentase rumah tangga miskin ( $X_2$ ) berpengaruh negatif dan tidak signifikan terhadap persentase kasus DBD. Kuantil  $Q_{0,95}$  menunjukkan bahwa koefisien variabel presentase rumah tangga miskin ( $X_2$ ) sebanyak minus 0,01. Hal ini berarti bahwa daerah dengan rumah tangga miskin tinggi cenderung memiliki tingkat kerentanan terhadap penyakit DBD lebih sedikit. Hal ini tidak sesuai dengan hipotesis yang diharapkan yaitu rumah tangga miskin dapat meningkatkan persentase kasus DBD. Ketidaksesuaian hasil ini, diduga disebabkan karena kurang bervariasinya data banyaknya kasus penderita DBD.

Selanjutnya pada kuantil  $Q_{0,95}$  menunjukkan bahwa variabel rasio tenaga medis (dokter umum) ( $X_3$ ), berpengaruh positif dan signifikan terhadap kasus DBD. Koefisien rasio tenaga medis (dokter umum) ( $X_3$ ) sebanyak 0,1302. Artinya jika rasio tenaga medis (dokter umum) di suatu daerah naik sebanyak 1, maka akan meningkatkan persentase kasus DBD sebanyak 0,13 per 1000 penduduk. Hal ini tidak sesuai dengan hipotesis yang diharapkan bahwa banyaknya tenaga medis dapat menekan tingkat persentase kasus DBD. Ketidaksesuaian hubungan ini, diduga disebabkan karena jumlah tenaga medis yang banyak tidak menjamin kualitas dalam menangani jumlah penyakit DBD yang terus meningkat. Maka untuk menghasilkan tenaga medis yang berkualitas dibutuhkan juga sumber daya manusia yang berkualitas pula. Permasalahan besar tentang SDM adalah inefisiensi dan inefektivitas SDM dalam menanggulangi masalah kesehatan. Walaupun rasio SDM kesehatan telah meningkat, tetapi masih jauh dari target Indonesia Sehat 2010 dan variasinya antar daerah masih tajam (Bappeda Jatim, 2012)

Selanjutnya pada kuantil  $Q_{0,95}$  menunjukkan bahwa variabel rasio banyaknya puskesmas ( $X_4$ ) berpengaruh positif dan signifikan terhadap persentase DBD. Koefisien rasio banyaknya

puskesmas ( $X_4$ ) sebanyak 0,0435. Artinya bahwa ada kecenderungan daerah dengan rasio banyaknya puskesmas relatif tinggi, akan diikuti dengan tingkat persentase DBD yang relatif tinggi pula. Hal ini tidak sesuai dengan hipotesis yang diharapkan bahwa banyaknya puskesmas dapat menekan tingkat persentase DBD. Ketidaksesuaian hubungan ini, diduga disebabkan karena jumlah puskesmas yang banyak tidak menjamin akses ke sarana kesehatan yang lebih baik bagi masyarakat. Faktor utama yang menjadi kendala adalah upaya peningkatan derajat kesehatan masyarakat yang masih belum merata disetiap kecamatan. Selain itu, jumlah puskesmas yang banyak namun tidak diikuti dengan pelayanan yang maksimal akan menyebabkan ketimpangan dalam akses sarana kesehatan bagi masyarakat.

Selanjutnya pada kuantil  $Q_{0,95}$  menunjukkan bahwa variabel persentase rumah tangga yang memiliki tempat sampah sehat ( $X_5$ ) berpengaruh negatif dan signifikan terhadap persentase kasus DBD. Koefisien variabel persentase rumah tangga yang memiliki tempat sampah sehat ( $X_5$ ) sebanyak minus 5,25682. Artinya jika persentase rumah tangga yang memiliki tempat sampah sehat naik sebanyak 1 persen, maka akan menurunkan tingkat persentase DBD sebanyak 5,25 per 1000 penduduk. Artinya ada kecenderungan daerah dengan persentase rumah tangga yang memiliki tempat sampah sehat relatif tinggi akan diikuti tingkat persentase DBD yang relatif lebih rendah. Hal ini sesuai dengan hipotesis yang diharapkan bahwa rumah tangga yang memiliki tempat sampah sehat dapat menurunkan tingkat persentase kasus DBD. Kondisi rumah dengan tempat sampah yang tidak sehat menjadi tempat yang ideal bagi nyamuk.

### **Memodelkan Kasus DBD Kota Surabaya pada Tiap Kuantil Menggunakan Satu Titik Knot**

Pemilihan titik knot optimal diawali dengan satu titik knot. Dengan menggunakan satu titik knot pada variabel-variabel yang mempengaruhi DBD diharapkan dapat menemukan GCV yang paling minimum. GCV minimum tersebut diharapkan nantinya dapat menghasilkan model Spline *Truncated* dengan satu titik knot pada kasus DBD.

$$\hat{y} = \beta_0 + \beta_1\theta + \beta_2(\theta - K_1)_+^1$$

Nilai GCV yang dihasilkan dengan menggunakan fungsi spline *truncated* dengan satu titik knot. Lokasi satu titik knot dengan GCV minimum yang dihasilkan adalah sebagai berikut:

$$Q_{0,05} = 0,7194 \text{ dengan lokasi titik knot } K_1 = 0,5007$$

$$Q_{0,25} = 0,7396 \text{ dengan lokasi titik knot } K_1 = 0,6543$$

$$Q_{0,5} = 0,7986 \text{ dengan lokasi titik knot } K_1 = 1,1657$$

$$Q_{0,75} = 0,7363 \text{ dengan lokasi titik knot } K_1 = 1,7603$$

$$Q_{0,95} = 3,6440 \text{ dengan lokasi titik knot } K_1 = 0,6589$$

**Tabel 6.** Koefisien determinasi ( $R^2$ )

Model	$R^2$
$Q_{0,05}$	26%
$Q_{0,25}$	20%
$Q_{0,5}$	14%
$Q_{0,75}$	21%
$Q_{0,95}$	29%

**Memodelkan Kasus DBD Kota Surabaya pada Tiap Kuantil Menggunakan Dua Titik Knot**

Setelah mendapatkan knot optimal dari satu titik knot maka selanjutnya adalah mencari titik knot optimal dengan dua titik knot. Percobaan dilakukan dengan cara serupa dan dipilih GCV yang paling minimum. Berikut adalah model regresi Spline *Truncated* untuk dua titik knot.

$$\hat{y} = \beta_0 + \beta_1\theta + \beta_2(\theta - K_1)_+^1 + \beta_3(\theta - K_2)_+^1$$

Berdasarkan persamaan di atas, maka dibutuhkan dua titik knot yang optimal untuk masing-masing variabel prediktor. Untuk memperoleh knot yang optimal maka dipilih nilai GCV yang paling minimum. Lokasi dua titik knot dengan GCV minimum yang dihasilkan adalah sebagai berikut:

$$Q_{0,05} = 0,5588 \text{ dengan lokasi titik knot } K_1 = 0,7060 \text{ dan } K_2 = 0,7288$$

$$Q_{0,25} = 0,5196 \text{ dengan lokasi titik knot } K_1 = 0,9095 \text{ dan } K_2 = 0,9379$$

$$Q_{0,5} = 0,7282 \text{ dengan lokasi titik knot } K_1 = 0,9757 \text{ dan } K_2 = 1,0232$$

$$Q_{0,75} = 0,6620 \text{ dengan lokasi titik knot } K_1 = 1,6907 \text{ dan } K_2 = 1,7255$$

$$Q_{0,95} = 0,3602 \text{ dengan lokasi titik knot } K_1 = 4,0997 \text{ dan } K_2 = 4,3732$$

**Tabel 7.** Koefisien determinasi ( $R^2$ )

Model	$R^2$
$Q_{0,05}$	44%
$Q_{0,25}$	48%
$Q_{0,5}$	27%
$Q_{0,75}$	34%
$Q_{0,95}$	64%

**Memodelkan Kasus DBD Kota Surabaya pada Tiap Kuantil Menggunakan Tiga Titik Knot**

Setelah mendapatkan knot optimal dari satu titik knot dan dua titik knot, maka selanjutnya adalah mencari titik knot optimal dengan tiga titik knot. Percobaan dilakukan dengan cara serupa dan dipilih GCV yang paling minimum.

$$\hat{y} = \beta_0 + \beta_1\theta + \beta_2(\theta - K_1)_+^1 + \beta_3(\theta - K_2)_+^1 + \beta_4(\theta - K_3)_+^1$$

Untuk membentuk model seperti diatas maka dibutuhkan titik-titik knot optimal titik knot optimal tersebut didapat dari nilai GCV yang paling minimum. Lokasi Tiga titik knot dengan GCV minimum yang dihasilkan adalah sebagai berikut:

$$Q_{0,05} = 0,498130307138 \text{ dengan lokasi titik knot } K_1 = 0,7288; K_2 = 0,7516 \text{ dan } K_3 = 0,7973$$

$$Q_{0,25} = 0,499263543557870 \text{ dengan lokasi titik knot } K_1 = 0,9095; K_2 = 0,9379 \text{ dan } K_3 = 0,9662$$

$$Q_{0,5} = 0,48936664560 \text{ dengan lokasi titik knot } K_1 = 1,0944; K_2 = 1,1182 \text{ dan } K_3 = 1,1657$$

$$Q_{0,75} = 0,3039 \text{ dengan lokasi titik knot } K_1 = 1,7255; K_2 = 1,7603 \text{ dan } K_3 = 1,7951$$



$Q_{0,95} = 0,297647215134642$  dengan lokasi titik knot  
 $K_1 = 2,0036; K_2 = 4,009$  dan  $K_3 = 4,555$

**Tabel 8.** Koefisien determinasi ( $R^2$ )

Model	$R^2$
$Q_{0,05}$	53%
$Q_{0,25}$	54%
$Q_{0,5}$	55%
$Q_{0,75}$	71%
$Q_{0,95}$	72%

**Pemilihan Model Spline Truncated Terbaik**

Berdasarkan hasil dan koefisien determinasi ( $R^2$ ) pada tiap kuantil, diketahui model kuantil  $Q_{0,95}$  memiliki koefisien determinasi ( $R^2$ ) sebesar 72% dan berdasarkan ketiga hasil *trial error*, diambil knot optimum yang memiliki nilai GCV minimum. Nilai GCV minimum untuk satu, dua dan tiga titik knot disajikan pada tabel 9.

**Tabel 9.** Titik Knot Optimum Satu, Dua dan Tiga Titik Knot

Jumlah Knot	Titik Knot	GCV
1	3,6440	0,6589
2	4,0997	0,3602
	4,3732	
3	<b>2,0036</b>	<b>0,2976</b>
	<b>4,009</b>	
	<b>4,555</b>	

Berdasarkan tabel 9 didapat titik knot ( $K$ ) yang paling optimal dengan nilai GCV minimum sebesar 0,2976 terletak pada  $K_1 = 2,0036; K_2 = 4,009$  dan  $K_3 = 4,55$  sehingga model terbaik adalah model spline dengan tiga titik knot.

**G. Interpretasi Model Spline Truncated**

Berdasarkan hasil analisis, model Spline *Truncated* terbaik adalah dengan tiga titik knot. Berikut adalah model terbaik yang telah diperoleh.

$$\hat{y} = 0,035 + 0,597\theta - 1,002(\theta - 2,0036)_+ + 10,572(\theta - 4,009)_+ - 26,803(\theta - 4,55)_+$$

dengan fungsi *truncated* sebagai berikut:

$$\hat{y} = \begin{cases} 0,035 + 0,597\theta; & \theta < 2,0036 \\ 2,0426 - 0,405\theta; & 2,0036 \leq \theta < 4,009 \\ -40,3405 + 10,167\theta; & 4,009 \leq \theta < 4,55 \\ 81,61311 - 16,636\theta; & \theta \geq 4,55 \end{cases}$$

Berdasarkan model, dapat diinterpretasikan yaitu apabila model kuantil 0,95 kurang dari 2,036 maka persentase DBD akan naik 0,597. Wilayah yang termasuk dalam interval ini adalah kecamatan Pabean Cantikan, Lakarsantri, Benowo, Dukuh Pakis, Tegalsari, Genteng, Karang Pilang, Pakal, Semampir dan Tenggilis. Selanjutnya apabila model kuantil antara 2,0036 hingga 4,009 maka persentase DBD turun 0,405. Wilayah yang termasuk dalam interval ini adalah kecamatan Wiyung, Sukolilo, Simokerto, Asemrowo, Kenjeran, Sambikerep, Sukomanunggal, Wonokromo, Bubutan, Wonokoncolo, Jambangan, Gubeng, Krembangan, Tambaksari, Gayungan, Sawahan dan Bulak. Apabila model kuantil 0,95 antara 4,009 hingga 4,55 maka persentase DBD akan naik 10,167. Wilayah yang termasuk interval ini adalah kecamatan Rungkut, Tendes dan Mulyorejo. Apabila model kuantil 0,95 lebih dari 4,55 maka persentase DBD akan turun 16,636. Wilayah yang termasuk dalam interval ini adalah kecamatan Gunung Anyar.

**5. KESIMPULAN**

Berdasarkan dari analisis dan pembahasan di atas maka dapat diambil kesimpulan:

1. Pemodelan kasus DBD di Kota Surabaya dengan menggunakan Regresi Kuantil dan dilakukan pada kuantil 0,05; 0,25; 0,50; 0,75; dan 0,95. Dari hasil pemodelan menggunakan Regresi Kuantil diperoleh estimasi  $\hat{\beta}$  yang berbeda untuk setiap kuantil, artinya didapatkan 5 model untuk mengetahui hubungan antara kasus DBD dengan persentase rumah/bangunan bebas jentik nyamuk *Aedes Aegypti*, persentase rumah tangga miskin, rasio tenaga medis (dokter umum), rasio sarana kesehatan puskesmas dan persentase rumah tangga yang memiliki tempat sampah sehat.
2. Pemodelan kasus DBD di kota Surabaya pada masing-masing kuantil dengan menggunakan

fungsi Spline *Truncated* diperoleh model terbaik dengan tiga titik knot pada model kuantil 0,95. Model ini memiliki nilai GCV paling minimum yaitu 0,2976 dengan  $R^2$  sebesar 72%. Berikut adalah model spline terbaik

$$\hat{y} = 0,035 + 0,597\theta - 1,002(\theta - 2,0036)_+ + 10,572(\theta - 4,009)_+ - 26,803(\theta - 4,55)_+$$

#### DAFTAR PUSTAKA

- [1] Abrevaya, J. (2001), "The Effects of Demographics and Maternal Behavior on The Distribution of Birth Outcomes", *Empirical Economics*, Springer, Vol. 26, No.1, hal: 247-257
- [2] Budiantara, I. N. dan Subanar, (1998). "Estimator Spline Terbobot", *Majalah Ilmiah Himpunan Matematika Indonesia*, hal. 35-45
- [3] Budiantara, I. N., Mariati, N.P.A.M, Ratnasari, V., Ismaini, Z., Ratna, M., Sudiarsa, I.W. , Mardianto, M.F.F., dan Hendayanti, N.P.N, (2015), "Comparison Truncated Spline and Fourier Series in Multivariable Nonparametric Regression Models (Application: Data of Poverty in Papua, Indonesia)", *International Journal of Basic & Applied Sciences IJBAS-IJENS*, Vol. 15, No. 04, hal. 9-12
- [4] Davino, Furno dan Vistocco, (2014). *Quantile regression (Theory and application)*. Stanford Weisberg.
- [5] Dinkes Jatim (2013), *Profil Kesehatan Provinsi Jawa Timur Tahun 2012*, Dinkes Provinsi Jatim, Surabaya.
- [6] Eubank, R. (1988), *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York
- [7] Goh, S. C. dan Knight, K. (2009). "Nonstandard Quantile-Regression Inference". *Econometric Theory*. Vol. 25, hal. 1415-1432
- [8] Hardle, W, (1990), *Applied Nonparametric Regression*. Cambridge University Press. New York.
- [9] Kementerian Kesehatan RI (2010), *Buletin Jendela Epidemiologi*, Vol.2, Kemen-kes RI, Jakarta
- [10] Koenker, R. dan Billias, Y., (2001). "Quantile Regression for Duration Data: A Reappraisal of The Pennsylvania Reemployment Bonus Experiments", *Empirical Economics*, Springer, Vol. 26, No. 1, hal: 199-220.
- [11] Machado, J. A. F. dan Mata, J., (2001), "Earning Functions In Portugal 1982-1994: Evidence From Quantile Regressions", *Empirical Economics*, Vol. 26, No.1, hal: 115-134.
- [12] Mubarak, R, (2012), "Analisis Regresi Spline Multivariabel untuk Pemodelan Kematian Penderita Demam Berdarah Dengue (DBD) di Jawa Timur", *Jurnal Sains dan Seni ITS*, Vol.1, No.1, hal: 224-229.
- [13] Rahmawati R., Kartono, Sulistyio R.H., Noranita B., Sarwoko E.A., dan Wardaya A.Y., (2012), "Analisis Pengaruh Karakteristik Wilayah (Kelurahan) Terhadap Banyaknya Kasus Demam Berdarah Dengue (DBD) Di Kota Semarang", *Media Statistika*, Vol. 5, No. 2, hal. 87-93.