

# Analisis Regresi *Spline* Multivariabel untuk Pemodelan Kematian Penderita Demam Berdarah Dengue (DBD) di Jawa Timur

Reza Mubarak dan I Nyoman Budiantara

Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111

*E-mail*: i\_nyoman\_b@statistika.its.ac.id

**Abstrak**— Secara global di seluruh dunia, Demam Berdarah Dengue (DBD) telah tumbuh dengan cepat dalam beberapa dekade terakhir. Di Indonesia sendiri, angka penderita penyakit DBD ini sendiri masih tergolong tinggi tiap tahunnya dan hal ini juga terjadi di Jawa Timur. Pada penelitian ini akan diselidiki faktor-faktor yang mempengaruhi persentase kematian pada penderita DBD di Jawa Timur dengan menggunakan analisis regresi *Spline*. Dari hasil analisis didapatkan bahwa model *Spline* linier dengan 1 titik knot merupakan model terbaik daripada model *spline* linier dengan 2 dan 3 titik knot berdasarkan *Generalized Cross Validation* (GCV) minimum dengan dengan R-Sq sebesar 91.62% dan nilai MSE adalah 0.1823. Variabel-variabel yang signifikan berpengaruh terhadap presentase kematian penderita DBD adalah persentase sarana kesehatan, persentase tenaga kesehatan, rata-rata usia perkawinan pertama, persentase daerah yang berstatus desa, rata-rata lama pemberian ASI eksklusif, persentase penduduk miskin, rata-rata curah hujan, persentase penduduk wanita dibanding total, rata-rata lama sekolah penduduk umur 15 tahun ke atas, angka bebas jentik.

**Kata Kunci**—DBD, *Spline*, Knot, GCV.

## I. PENDAHULUAN

TINGKAT angka kematian di Indonesia setiap tahun meningkat. Dalam kurun waktu 5 tahun terakhir dari tahun 2003 – 2008 angka kematian (*Case Fatality Rate*) di Indonesia meningkat [1]. Hal ini disebabkan karena berbagai faktor, salah satunya disebabkan menderita penyakit. Penyakit itu sendiri terdiri dari penyakit menular dan penyakit tidak menular. Contoh beberapa dari penyakit menular antara lain TBC, Hepatitis, Demam Berdarah Dengue (DBD) dan lain-lain [2].

DBD sendiri secara global di seluruh dunia telah tumbuh dengan cepat dalam beberapa dekade terakhir. DBD telah terjadi di lebih dari 100 negara dan mengancam kesehatan lebih dari 2,5 miliar orang di perkotaan, pinggiran perkotaan dan daerah pedesaan di daerah tropis dan subtropis. Penyakit ini endemik di Afrika, Amerika, Mediterania Timur, Asia Tenggara dan Pasifik Barat. Bahkan penyakit ini merupakan penyakit utama di Asia Tenggara dan Pasifik Barat [3]. Di Indonesia sendiri, angka penderita penyakit DBD ini sendiri masih tergolong tinggi tiap tahunnya. Dari data Kementerian Kesehatan RI, angka rata-rata penderita DBD pada 2009-2010 masih sebesar 65,57 kasus per 100 ribu penduduk dengan jumlah

penderita mencapai 150 ribu. Khusus di daerah Jawa Timur, jumlah kasus Demam Berdarah Dengue (DBD) tergolong tinggi. Selama Tahun 2009, sebanyak 18.631 penderita atau 50 orang per 100 ribu penduduk. Jumlah kasus selama 2010 juga tergolong tinggi. Bahkan melebihi target nasional dimana targetnya 55 penderita per 100 ribu penduduk [4].

Pada penelitian sebelumnya yang membahas tentang faktor-faktor penyebab terjangkitnya penyakit DBD adalah Erkhadifa, Wati dan Widiyanto dengan metode yang berbeda satu dengan lainnya [5]-[7]. Sedangkan penelitian yang menggunakan pemodelan analisis regresi *spline* pernah dilakukan penelitian oleh Arif dan Pratiwi yang keduanya menggunakan data pertumbuhan balita. Untuk penelitian DBD yang menggunakan metode analisis regresi *Spline* masih belum pernah dilakukan [8],[9]. Maka karena dari itu, pada penelitian ini akan diselidiki faktor-faktor yang mempengaruhi persentase kematian DBD di Jawa Timur menggunakan model *Spline*. Metode *Spline* sendiri sangat baik dalam memodelkan data yang memiliki pola yang berubah-ubah pada sub-sub interval tertentu. Selanjutnya, *Spline* mempunyai kemampuan yang sangat tepat baik untuk digeneralisasikan pada pemodelan Statistika yang kompleks dan rumit [9].

## II. TINJAUAN PUSTAKA

### A. Analisis Regresi *Spline*

*Spline* merupakan model polinomial yang tersegmen. Polinomial tersegmen memegang peranan penting dalam teori dan aplikasi statistika. *Spline* sangat tergantung pada titik knots. Titik knots merupakan titik perpaduan bersama dimana terjadi pola perubahan perilaku dari suatu fungsi pada selang yang berbeda [10]. Secara umum fungsi  $G$  dalam ruang *spline* berorde  $m$  dengan titik knots  $k_1, k_2, \dots, k_j$  adalah sembarang fungsi yang dapat dinyatakan menjadi.

$$G(x_i) = \sum_{j=0}^m \beta_j x_i^j + \sum_{j=1}^m \beta_{j+m} (x_i - k_j)_+^{j-1} \quad (1)$$

dengan

$$(x_i - k_j)_+^m = \begin{cases} (x_i - k_j)^m & x_i - k_j \geq 0 \\ 0 & x_i - k_j \leq 0 \end{cases}$$

merupakan parameter-parameter model dengan  $m$  merupakan orde *spline* [11].

**B. Estimasi Parameter Regresi Spline**

Model regresi *spline* dapat dinyatakan sebagai berikut.

$$y_i = G(x_i) + \quad (2)$$

maka model pada persamaan (2) untuk setiap pengamatan  $i = 1, 2, \dots, n$  dapat dituliskan sebagai berikut.

$$y_i = \beta_0 + \beta_1 x_i + \dots + \beta_m x_i^m + \beta_{m+1} (x_i - k_1)_+^m + \beta_{m+2} (x_i - k_2)_+^m + \dots + \beta_{m+j} (x_i - k_j)_+^m + \quad (3)$$

Estimator *spline* dalam regresi nonparametrik mampu menangani data/fungsi yang mulus (smooth) [6]. Kemampuan ini ditunjukkan oleh fungsi truncated (potongan-potongan) yang melekat pada estimator tersebut. Apabila pada persamaan (2) diasumsikan error  $\epsilon_i \sim N(0, \sigma^2)$ , maka  $y_i$  juga berdistribusi normal dengan mean  $G(x_i)$  dan varians  $\sigma^2$ . Dengan sedikit penjabaran, maka diperoleh estimasi Likelihood untuk model regresi  $B(k)$  dinyatakan sebagai berikut.

$$B(k) = X(k) [X(k)^T X(k)]^{-1} X(k)^T \quad (4)$$

dengan

$$k = (k_1, k_2, \dots, k_j)^T$$

Titik knots itu sendiri merupakan titik perpaduan bersama dimana terdapat perubahan perilaku pola pada interval yang berlainan.

**C. Pemilihan Titik Knot Optimal**

Metode yang baik untuk memilih titik knot optimal yaitu metode Generalized Cross Validation (GCV)[12]. Secara teoritis bahwa metode GCV mempunyai sifat optimal asimotik [13]. Generalized Cross Validation (GCV) dapat didefinisikan sebagai berikut.

$$GCV(k) = \frac{MSE(k)}{(n^{-1} \text{trace}[I - B(k)])^2} \quad (5)$$

Pada persamaan (5)  $I$  adalah matriks identitas,  $n$  adalah jumlah pengamatan dan  $B(k)$  adalah matriks  $X(X^T X)^{-1} X^T$  sedangkan nilai  $MSE(k)$  dinyatakan sebagai berikut.

$$MSE(k) = n^{-1} \sum_{i=1}^n (y_i - \hat{G}_m(x_i))^2 \quad (6)$$

**III. METODOLOGI PENELITIAN**

**A. Sumber Data**

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari data SUSENAS tahun 2009. Unit penelitian yang diamati adalah masing-masing kabupaten/kota di Provinsi Jawa Timur pada tahun 2009. Provinsi Jawa Timur sendiri terdiri dari 38 kabupaten/kota.

**B. Variabel Penelitian**

Variabel yang digunakan dalam penelitian ini terbagi menjadi 2 yaitu variabel  $y$  (respon) dan variabel  $x$  (prediktor). Variabel respon pada penelitian ini adalah Persentase kematian Penderita DBD di 38 kabupaten/kota di Provinsi Jawa Timur. Sedangkan, variabel prediktor dalam penelitian ini adalah sebagai berikut.

- i. Persentase sarana kesehatan
- ii. Persentase tenaga kesehatan
- iii. Rata-rata usia perkawinan pertama
- iv. Persentase daerah yang berstatus desa
- v. Rata-rata lama pemberian ASI eksklusif
- vi. Persentase penduduk miskin
- vii. Tinggi rata-rata ibukota dari permukaan laut
- viii. Rata-rata curah hujan
- ix. Persentase keluarga berumah tak layak huni
- x. Persentase penduduk wanita dibanding total penduduk
- xi. Rata-rata lama sekolah penduduk umur 15 tahun ke atas
- xii. Persentase penduduk yang mengalami keluhan kesehatan yang melakukan pengobatan sendiri
- xiii. Angka bebas jentik di tiap kabupaten/kota di Jawa Timur

**C. Langkah Penelitian**

Berdasarkan tujuan dalam penelitian ini, maka langkah-langkah analisis adalah sebagai berikut.

- i. Membuat statistika deskriptif variabel dependen ( $y$ ) dan variabel independen ( $x_1 - x_{13}$ ).
- ii. Membuat scatter plot antarapersentase penderita DBD ( $y$ ) dengan masing-masing variabel prediktor ( $x_1 - x_{13}$ ).
- iii. Memodelkan persentase penderita DBD di Jawa Timur dengan menggunakan *spline* linier dengan 1 titik knot.
- iv. Memodelkan persentase penderita DBD di Jawa Timur dengan menggunakan *spline* linier dengan 2 titik knot.
- v. Memodelkan persentase penderita DBD di Jawa Timur dengan menggunakan *spline* linier dengan 3 titik knot.
- vi. Memilih titik knots optimal berdasarkan GCV minimum pada langkah (c), (d), dan (e).
- vii. Memodelkan persentase penderita DBD dengan variabel-variabel prediktornya menggunakan *spline* dengan knots optimal.
- viii. Melakukan pengujian signifikansi parameter dan pengujian asumsi residual *spline* terbaik.
- ix. Membuat kesimpulan

**IV. ANALISIS DAN PEMBAHASAN**

**A. Karakteristik Penduduk di Provinsi Jawa Timur**

Wilayah Jawa Timur tidak hanya terdapat di Pulau Jawa saja, tapi juga meliputi berbagai pulau lainnya seperti Pulau Madura, Pulau Bawean, Pulau Kangean serta sejumlah pulau-pulau kecil di Laut Jawa dan Samudera Hindia. Provinsi Jawa Timur terdiri dari 38 Kota/ Kabupaten dengan rincian 29 Kabupaten dan 9 Kota.

Tabel 1. Statistika Deskriptif Variabel Penelitian

Variabel	Mean	Varians	Minimum	Maksimum
Y	1,061	0,587	0	3,88
x <sub>1</sub>	0,007466	0,000006	0,0044	0,0159
x <sub>2</sub>	0,1068	0,0132	0,02	0,45
x <sub>3</sub>	20,368	0,284	19,19	21,35
x <sub>4</sub>	53,1	966,59	0	91,48
x <sub>5</sub>	2,912	0,525	0,72	4,26
x <sub>6</sub>	17,93	52,6	6,18	34,53
x <sub>7</sub>	101,8	32615,5	2	871
x <sub>8</sub>	16,243	9,589	11,44	27,78
x <sub>9</sub>	3,008	9,685	0,07	14,31
x <sub>10</sub>	48,95	0,915	47,04	50,25

Tabel 1.  
Statistika Deskriptif Variabel Penelitian (Lanjutan)

Variabel	Mean	Varians	Minimum	Maksimum
x <sub>11</sub>	7,268	1,651	4,19	9,38
x <sub>5</sub>	2,912	0,525	0,72	4,26
x <sub>6</sub>	17,93	52,6	6,18	34,53

Dari Tabel 1 dapat dilihat bahwa rata-rata persentase kematian pada penderita DBD sebesar 1.061% atau bisa dikatakan bahwa setiap 100 orang yang menderita DBD terjadi kematian sebanyak 1.06 orang. Nilai varians persentase kematian yang dialami penderita DBD tidak begitu besar yaitu 0.587. Ini bisa dilihat pada nilai minimum dan maksimumnya yang nilainya antara 0% – 3.88%. Ternyata dari 13 variabel prediktor penelitian ini yang memiliki nilai varians paling besar adalah variabel tinggi rata-rata ibukota dari permukaan laut pada tiap kabupaten/kota dengan nilai varians sebesar 32615.5. Sedangkan varians terkecil bisa dilihat pada variabel persentase sarana kesehatan di tiap kabupaten/kota dengan besar nilai varians 0.000006. Sehingga bisa dikatakan semakin besar range data, maka nilai varians cenderung ikut semakin membesar.

B. Model Spline Linier dengan 1 Titik Knot

Model terbaik dalam *spline* diawali dengan pemilihan nilai GCV minimum sehingga memberikan titik knots optimum. Berikut adalah model *spline* linier secara umum dengan 1 knot yang terbentuk.

$$f = \beta_0 + \beta_1 x_1 + \beta_2 (x_1 - k_1)_+ + \beta_3 x_2 + \beta_4 (x_2 - k_2)_+ + \dots + \beta_{22} x_{12} + \beta_{23} (x_{12} - k_{12})_+$$

Dengan percobaan 100 kali dalam membentuk model *spline* linier dengan 1 knot diperoleh titik knot yang optimum berdasarkan GCV minimum dengan nilai GCV sebesar 0.5726. Titik-titik knot dan estimasi parameter yang terbentuk disajikan dalam Tabel 2.

Tabel 2.

Titik-Titik Knot dan Estimasi Parameter Regresi *Spline* Linier dengan 1 Knot

X <sub>i</sub>	Knot ke-i	Estimasi Parameter	X <sub>i</sub>	Knot ke-i	Estimasi Parameter
		$\beta_0 = -87.4$			
1	0.0134	$\beta_1 = -150.5$	8	25.6027	$\beta_{15} = 0.1$
		$\beta_2 = 2920.7$			$\beta_{16} = -0.9$
2	0.3898	$\beta_3 = 7.9$	9	9.0602	$\beta_{17} = 1.7$
		$\beta_4 = -100.7$			$\beta_{18} = -2$
3	20.2822	$\beta_5 = 1.2$	10	48.1798	$\beta_{19} = -1.3$
		$\beta_6 = -2.4$			$\beta_{20} = -2$
4	58.1503	$\beta_7 = 0$	11	9.3644	$\beta_{21} = -1.3$
		$\beta_8 = -0.1$			$\beta_{22} = -109.9$
5	4.0862	$\beta_9 = 1.1$	12	58.0838	$\beta_{23} = -0.1$
		$\beta_{10} = -6.7$			$\beta_{24} = 0$
6	18.7664	$\beta_{11} = 0.4$	13	84.68	$\beta_{25} = 0$
		$\beta_{12} = 0.6$			$\beta_{26} = -0.1$
7	54.1564	$\beta_{13} = 0$			
		$\beta_{14} = 0$			

Dari Tabel 2 dapat dilihat bahwa knot yang terbentuk

sebanyak 13 knot. Sedangkan untuk banyaknya estimasi parameter yang terbentuk adalah 27 termasuk didalamnya terdapat intercept.

C. Model Spline Linier dengan 2 Titik Knot

Seperti pada model *spline* linier dengan 1 knot, penentuan model terbaik dalam *spline* linier dengan 2 knot diperoleh dengan melakukan percobaan 100 kali. Berikut adalah model *spline* linier dengan 2 knot yang terbentuk.

$$f = \beta_0 + \beta_1 x_1 + \beta_2 (x_1 - k_1)_+ + \beta_3 (x_1 - k_2)_+ + \beta_4 x_2 + \beta_5 (x_2 - k_2)_+ + \beta_6 (x_2 - k_3)_+ + \dots + \beta_{26} x_{12} + \beta_{27} (x_{12} - k_{12})_+ + \beta_{28} (x_{12} - k_{13})_+$$

Dari 100 percobaan tersebut kemudian dipilih kombinasi knot yang memiliki GCV minimum dengan nilai GCV sebesar 16.1516. Titik-titik knot dan estimasi parameter yang terbentuk disajikan dalam Tabel 3.

Tabel 3.

Titik-Titik Knot dan Estimasi Parameter Regresi *Spline* Linier dengan 2 Knot

X <sub>i</sub>	Knot ke-i	Estimasi Parameter	X <sub>i</sub>	Knot ke-i	Estimasi Parameter
		$\beta_0 = -28.996$			
1	0.0067	$\beta_1 = 961.003$	8	23.391	$\beta_{22} = -0.384$
		$\beta_2 = -312.91$			$\beta_{23} = -12.289$
	0.0096	$\beta_3 = -381.96$		25.105	$\beta_{24} = 22.5$
		$\beta_4 = -7.428$			$\beta_{25} = -1.702$
2	0.2047	$\beta_5 = -12.579$	9	5.199	$\beta_{26} = -7.838$
		$\beta_6 = -195.88$			$\beta_{27} = 11.842$
	0.4394	$\beta_7 = 1.597$		6.5401	$\beta_{28} = 0.55$
3	20.5293	$\beta_8 = 20.461$	10	48.2803	$\beta_{29} = 1.003$
		$\beta_9 = -17.399$			$\beta_{30} = -4.956$
	20.692	$\beta_{10} = 0.441$			$\beta_{31} = 0.428$
		$\beta_{11} = -0.618$		6.4231	$\beta_{32} = 0.774$
4	31.734	$\beta_{12} = 0.332$	11	8.0009	$\beta_{33} = 3.063$
		$\beta_{13} = -7.047$			$\beta_{34} = -0.362$
	65.8807	$\beta_{14} = 3.021$			$\beta_{35} = 1.711$
5	2.5502	$\beta_{15} = 7.856$	12	69.9689	$\beta_{36} = -7.709$
		$\beta_{16} = -0.143$			$\beta_{37} = -0.003$
	2.6907	$\beta_{17} = 0.34$		76.3334	$\beta_{38} = -0.198$
6	10.617	$\beta_{18} = 0.341$	13	87.0216	$\beta_{39} = 0.696$
		$\beta_{19} = -0.0002$			
	22.114	$\beta_{20} = 0.0014$		90.489	
7	372.59	$\beta_{21} = -0.0055$			
	605.78				

Dari Tabel 3 dapat dilihat bahwa knot yang terbentuk sebanyak 26. Sedangkan untuk banyaknya estimasi parameter yang terbentuk adalah 38 termasuk didalamnya terdapat intercept.

D. Model Spline Linier dengan 3 Titik Knot

Seperti pada model *spline* linier dengan 1 dan 2 knot, penentuan model terbaik dalam *spline* linier dengan 3 knot

juga diperoleh dengan melakukan percobaan 100 kali. Berikut adalah model *spline* linier dengan 3 knot yang terbentuk.

$$f = \beta_0 + \beta_1 x_1 + \beta_2 (x_1 - k_1)_+ + \beta_3 (x_1 - k_2)_+ + \beta_4 (x_1 - k_3)_+ + \beta_5 x_2 + \beta_6 (x_2 - k_4)_+ + \beta_7 (x_2 - k_5)_+ + \beta_8 (x_2 - k_6)_+ + \dots + \beta_{40} x_{20} + \beta_{20} (x_{20} - k_{27})_+ + \beta_{21} (x_{20} - k_{28})_+ + \beta_{22} (x_{20} - k_{29})_+$$

Dari 100 percobaan tersebut kemudian dipilih kombinasi knot yang memiliki GCV minimum dan akhirnya ditentukan kombinasi knot dari percobaan ke-88 dengan nilai GCV sebesar 9.3806. Titik-titik knot dan estimasi parameter yang terbentuk disajikan dalam Tabel 4.

Tabel 4. Titik-Titik Knot dan Estimasi Parameter Regresi *Spline* Linier dengan 3 Knot

X <sub>i</sub>	Knot ke-i	Estimasi Parameter	X <sub>i</sub>	Knot ke-i	Estimasi Parameter
		$\beta_0=17.9$			
	0.0089	$\beta_1=666.5$	14.3502		$\beta_{22}=2.2$
1	0.0094	$\beta_2=-8930.7$	8	24.9829	$\beta_{23}=0.7$
	0.014	$\beta_3=9852.9$		26.9506	$\beta_{24}=-0.8$
	0.055	$\beta_4=-5554.4$		0.1936	$\beta_{25}=0$
2	0.272	$\beta_5=-12.7$	9	2.9149	$\beta_{26}=-0.4$
	0.398	$\beta_6=-11.1$		3.172	$\beta_{27}=1.1$
	19.394	$\beta_7=60.7$		47.9211	$\beta_{28}=-3$
3	19.6	$\beta_8=-54.3$	10	48.005	$\beta_{29}=1$
	19.637	$\beta_9=0.1$		48.9225	$\beta_{30}=-0.9$
	57.725	$\beta_{10}=5.9$		5.1574	$\beta_{31}=-6$
4	61.094	$\beta_{11}=-19.6$	11	6.2287	$\beta_{32}=24.6$
	67.0019	$\beta_{12}=12.9$		6.2481	$\beta_{33}=-1.7$
	1.8328	$\beta_{13}=-0.1$		56.2002	$\beta_{34}=2.3$
5	2.3774	$\beta_{14}=0.4$	12	56.3228	$\beta_{35}=-4$
	2.9908	$\beta_{15}=-0.5$		73.4756	$\beta_{36}=3$
	9.7203	$\beta_{16}=0.1$		56.5764	$\beta_{37}=0$
6	22.1135	$\beta_{17}=2$	13	85.8744	$\beta_{38}=6.3$
	24.9967	$\beta_{18}=-3.8$		88.984	$\beta_{39}=-5.7$
	587.7296	$\beta_{19}=3.5$			
7	593.1905	$\beta_{20}=-2.6$			
	810.27	$\beta_{21}=-2.1$			

Dari Tabel 4 dapat dilihat bahwa knot yang terbentuk sebanyak 39. Sedangkan untuk banyaknya estimasi parameter yang terbentuk adalah 52 termasuk didalamnya terdapat intercept.

E. Pemilihan Model Terbaik Berdasarkan GCV Minimum

Dari ketiga model terbaik dari masing-masing 1 knot, 2 knot dan 3 knot dipilih model terbaik berdasarkan GCV minimum. Perbandingan GCV dari ketiga model tersebut dapat dilihat pada Tabel 5.

Tabel 5. Perbandingan Nilai GCV dari Ketiga Model *Spline*

Model	GCV
1 Knot	0.573
2 Knot	16.152
3 Knot	9.381

Berdasarkan perbandingan GCV yang ada pada Tabel 5 maka dapat diperoleh model terbaik yaitu dengan menggunakan *spline* linier dengan 1 knot yaitu dengan nilai GCV sebesar 0.573. Nilai GCV ini paling kecil dibandingkan dengan nilai GCV dari model *spline* linier dengan 2 knot dan 3 knot yang masing-masing nilai GCV-nya 16.152 dan 9.381. Selain itu, model *spline* linier dengan titik 1 knot juga memiliki kriteria model yang baik, yaitu dengan memiliki R-Sq besar dan MSE yang kecil, yaitu masing-masing 91.62% dan 0.1823. Dengan demikian, model *spline* yang digunakan untuk memodelkan data presentase kematian penderita DBD di Jawa Timur adalah model *spline* linier dengan 1 titik knot.

F. Pengujian Signifikansi Parameter Model Terbaik

Terdapat dua tahap pengujian parameter regresi, yaitu pengujian secara simultan (uji serentak) dan secara parsial. Uji simultan merupakan pengujian parameter model regresi secara bersamaan. Sedangkan uji parsial adalah pengujian parameter model regresi secara satu persatu. Dalam uji simultan dilakukan pengujian hipotesis sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{26} = 0$$

$$H_1 : \text{minimal terdapat satu } \beta_j \neq 0, j = 1, 2, \dots, 26.$$

Nilai *k* adalah jumlah parameter yang terdapat dalam model regresi dan *n* merupakan jumlah observasi.

Tabel 6. Analisis Varians Model Terbaik

Sumber Variasi	Derajat Bebas	Jumlah Kuadrat	Rataan Kuadrat	F <sub>hitung</sub>
Regresi	26	19.94	0.18	
Residual	11	1.82	0.76	2.72
Total	37	21.76		

Pada Tabel 6 dapat dilihat bahwa nilai F<sub>hitung</sub> sebesar 2.72. Pada pengujian ini, taraf signifikansi adalah 5% dan F<sub>0,05(26,11)</sub> = 2,18. Daerah penolakan H<sub>0</sub> dilakukan apabila nilai F<sub>hitung</sub> > F<sub>0,05(26,11)</sub>. Berdasarkan nilai F<sub>hitung</sub> dan F<sub>0,05(26,11)</sub>, maka keputusan yang diambil H<sub>0</sub> ditolak karena F<sub>hitung</sub> > F<sub>0,05(26,11)</sub>. Sehingga, dapat ditarik kesimpulan bahwa di dalam parameter pada model ini terdapat minimal satu parameter yang signifikan.

Untuk melihat parameter yang mana yang signifikan atau tidak dalam model, maka dilakukan uji individu. Rumusan hipotesisnya sebagai berikut.

$$H_0 : \beta_1 = 0, \quad H_1 : \beta_1 \neq 0$$

$$H_0 : \beta_2 = 0, \quad H_1 : \beta_2 \neq 0$$

$$\vdots \quad \quad \quad \vdots$$

$$H_0 : \beta_{26} = 0, \quad H_1 : \beta_{26} \neq 0$$

H<sub>0</sub> ditolak apabila |t<sub>hit</sub>| > t<sub>((α/2),(n-m))</sub>. Dengan taraf signifikansi α = 5%, maka diperoleh nilai t<sub>((0.025),(11))</sub> = 2.201.

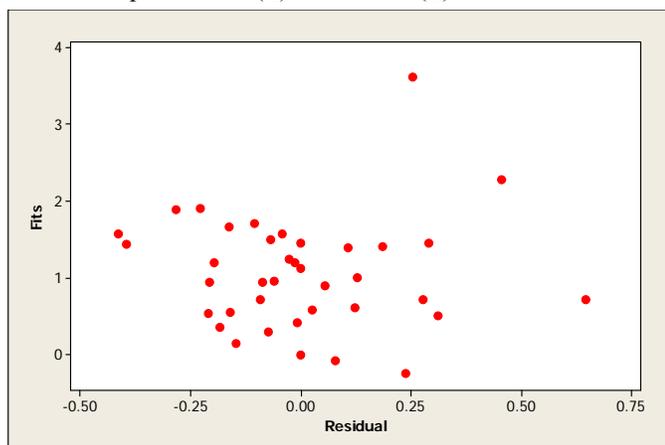
Tabel 7.  
Uji Parsial Estimasi Parameter *Spline* Linier dengan 1 Titik Knot

Parameter	Nilai Estimasi	t <sub>hitung</sub>	Keputusan
β <sub>0</sub>	-87.4	3.257	Signifikan
β <sub>1</sub>	-150.5	1.135	Tidak Signifikan
β <sub>2</sub>	2920.7	2.685	Signifikan
β <sub>3</sub>	7.9	2.047	Tidak Signifikan
β <sub>4</sub>	-100.7	3.373	Signifikan
β <sub>5</sub>	1.2	2.502	Signifikan
β <sub>6</sub>	-2.4	2.217	Signifikan
β <sub>7</sub>	0	2.303	Signifikan
β <sub>8</sub>	-0.1	4.728	Signifikan
β <sub>9</sub>	1.1	4.825	Signifikan
β <sub>10</sub>	-6.7	1.689	Tidak Signifikan
β <sub>11</sub>	-0.4	4.959	Signifikan
β <sub>12</sub>	0.6	5.743	Signifikan
β <sub>13</sub>	0	0.205	Tidak Signifikan
β <sub>14</sub>	0	0.377	Tidak Signifikan
β <sub>15</sub>	0.1	2.451	Signifikan
β <sub>16</sub>	-0.9	2.391	Signifikan
β <sub>17</sub>	-0.1	0.524	Tidak Signifikan
β <sub>18</sub>	0.4	1.196	Tidak Signifikan
β <sub>19</sub>	1.7	3.366	Signifikan
β <sub>20</sub>	-2	2.856	Signifikan
β <sub>21</sub>	-1.3	4.29	Signifikan
β <sub>22</sub>	-109.9	1.375	Tidak Signifikan
β <sub>23</sub>	-0.1	0.833	Tidak Signifikan
β <sub>24</sub>	0	0.216	Tidak Signifikan
β <sub>25</sub>	0	1.26	Tidak Signifikan
β <sub>26</sub>	-0.1	2.362	Signifikan

G. Pengujian Asumsi Residual

Pengujian asumsi residual erat kaitannya dengan kelayakan model regresi. Suatu model regresi dengan parameter signifikan dan memenuhi kriteria terbaik tetapi melanggar asumsi residual tidak disarankan untuk dipakai untuk menggambarkan pola hubungan antara variabel prediktor dan variabel respon. Asumsi residual dalam analisis regresi meliputi residual identik, independen dan distribusi Normal (0,σ<sup>2</sup>).

Tujuan dilakukan uji identik adalah mendeteksi adanya kasus heteroskedastisitas adalah upaya untuk mengurangi kerugian bagi efisiensi estimator [11]. Pelanggaran terhadap asumsi ini disebut heteroskedastisitas yaitu keadaan dimana variansi residual tidak homogen. Homogenitas variansi residual didasarkan pada sifat E(ε<sub>i</sub>)=0 dimana V(ε<sub>i</sub>) = σ<sup>2</sup>.



Gambar. 1. Plot antara Residual dan F<sub>its</sub>.

Deteksi adanya heteroskedastisitas pada data jika plot antara residual dengan *fits* membentuk suatu pola dan plot tidak menyebar secara random. Dari Gambar 1 dapat dilihat bahwa plot yang terbentuk tidak membentuk pola tertentu. Untuk memastikan apakah terjadi heteroskedastisitas maka dilakukan pengujian secara statistik dengan menggunakan uji *glejser*. Pengujian ini dilakukan dengan cara meregresikan harga mutlak residual dengan variabel prediktor (x) yang signifikan dalam model. Rumusan hipotesisnya adalah sebagai berikut.

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$$

$$H_1 : \text{minimal terdapat satu } \sigma_i^2 \neq \sigma^2, j = 1, 2, \dots, n.$$

Daerah penolakan H<sub>0</sub> jika nilai F<sub>hitung</sub> > F<sub>α((m-1),(n-m))</sub>. Apabila pada kesimpulan dihasilkan penolakan H<sub>0</sub>, maka dapat dinyatakan bahwa artinya terdapat minimal satu σ<sub>i</sub><sup>2</sup> ≠ σ<sub>j</sub><sup>2</sup> dan itu berarti terdapat heteroskedastisitas.

Tabel 8.  
Analisis Varians untuk Uji *Glejser*

Sumber Variasi	DF	SS	MS	F <sub>hitung</sub>
Regression	10	0.30333	0.03033	
Residual Error	27	0.46417	0.01719	1.76
Total	37	0.7675		

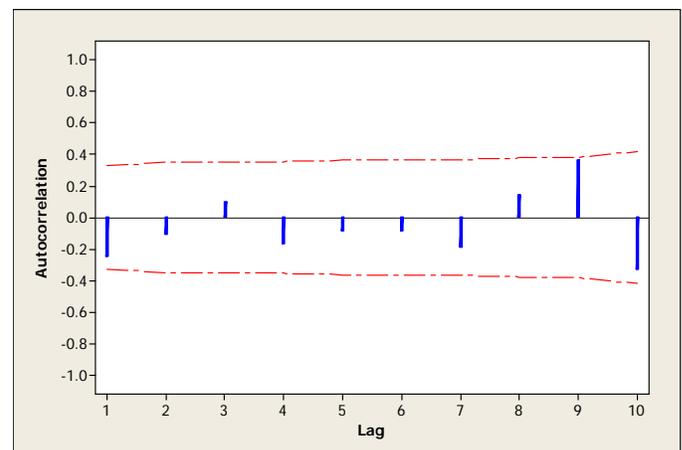
Pada Tabel 8 dapat dilihat bahwa nilai F<sub>hitung</sub> sebesar 1.48. Dengan taraf signifikansi α = 0.05 maka didapatkan juga nilai F<sub>0.05(10,27)</sub> sebesar 2.72. Sehingga keputusannya adalah Gagal Tolak H<sub>0</sub>, yang artinya variansi residual homogen.

Sedangkan dalam pengujian asumsi independen residual menggunakan ACF dengan hipotesis sebagai berikut.

$$H_0 : \rho_s = 0$$

$$H_1 : \rho_s \neq 0$$

H<sub>0</sub> ditolak jika terdapat Lag yang keluar dari batas atas ataupun batas bawah.



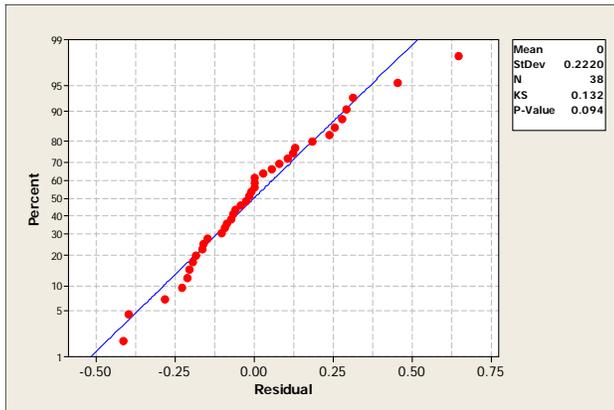
Gambar. 2. ACF Residual.

Pada Gambar 2 dapat dilihat bahwa nilai autokorelasi residual tidak ada yang keluar batas pada Lag manapun sehingga Keputusan Terima H<sub>0</sub> sehingga dapat ditarik kesimpulan bahwa residual saling independen.

Pengujian distribusi normal pada residual menggunakan *normal probability plot* dengan rumusan hipotesis sebagai berikut.

$H_0$  : Residual berdistribusi normal  
 $H_1$  : Residual tidak berdistribusi normal

Gagal tolak  $H_0$  jika plot yang dihasilkan cenderung membentuk garis lurus dengan sudut sekitar  $45^\circ$  atau  $P\text{-Value} > 0.05$ .



Gambar. 3. Normal Probability Plot untuk Residual.

Berdasarkan Gambar 3 bisa dilihat bahwa plot-plot residual hampir membentuk garis lurus dengan sudut sekitar  $45^\circ$  dan juga  $P\text{-Value}$  dari Normal Probability Plot untuk Residual adalah sebesar 0.094. Sehingga, keputusan yang diambil adalah Gagal Tolak  $H_0$  yang berarti residual berdistribusi normal.

V. KESIMPULAN/RINGKASAN

Berdasarkan hasil analisis, maka dapat diambil kesimpulan bahwa rata-rata persentase kematian pada penderita DBD di Jawa Timur Tahun 2009 sebesar 1,061%. Berdasarkan nilai varians sebesar 0,587, penderita DBD tersebar merata di berbagai kabupaten/kota di Jawa Timur. Model yang dihasilkan dari analisis regresi linier berganda memiliki R-Sq sebesar 48,4% dan MSE sebesar 0,683. Sedangkan dengan menggunakan model *spline* linier terbaik, yaitu model *spline* linier dengan 1 titik knot, dengan nilai GCV adalah 0,573 dengan R-Sq sebesar 91,62% dan nilai MSE sebesar 0,1823. Model yang terbentuk adalah sebagai berikut.

$$\hat{y} = -87,4 + 2920,7(x_1 - 0,0135)_+ +$$

$$-100,7(x_2 - 0,3898)_+ + 1,2x_3 +$$

$$-2,4(x_4 - 20,2822)_+ - 0,1(x_5 - 38,1308)_+ +$$

$$1,1x_6 - 0,4x_7 + 0,6(x_8 - 18,7664)_+ + 0,1x_9 +$$

$$-0,9(x_{10} - 23,6027)_+ + 1,7x_{11} +$$

$$-2(x_{12} - 48,1798)_+ - 1,3x_{13} - 0,1(x_{14} - 84,68)_+$$

Untuk variabel-variabel prediktor yang signifikan untuk model *spline* linier dengan 1 titik knot adalah sebagai berikut.

- a. persentase sarana kesehatan,
- b. persentase tenaga kesehatan,
- c. rata-rata usia perkawinan pertama,
- d. persentase daerah yang berstatus desa,
- e. rata-rata lama pemberian ASI eksklusif,
- f. persentase penduduk miskin,
- g. rata-rata curah hujan,
- h. persentase penduduk wanita dibanding total,

- i. rata-rata lama sekolah penduduk umur 15 tahun ke atas,
- j. angka bebas jentik di tiap kabupaten/kota.

DAFTAR PUSTAKA

- [1] Direktorat Jenderal P2PL, (2009), "Informasi Pengendalian Penyakit dan Kesehatan Lingkungan". Departemen Kesehatan RI, Jakarta.
- [2] Penyakit dan Pengobatannya, 2011, "Penyakit Menular dan Tidak Menular". <http://www.infopenyakit.com>. Diakses Jum'at, 10 Desember 2011 pukul 14.15.
- [3] World Health Organization, (1999), Strengthening implementation of the global strategy for Dengue Fever and Dengue Haemorrhagic Fever, prevention and control. WHO HQ, Geneva.
- [4] Detik Surabaya, (2011), Jumlah Penderita DB di Jatim Masih Tinggi, Angka Kematian Turun. <http://surabaya.detik.com>. Diakses Jum'at, 10 Desember 2011 pukul 14.12.
- [5] Erkhadifa, R., (2012), Perbandingan Geographically Weighted Poisson Regression, Geographically Weighted Poisson Regression Semiparametric (Studi Kasus: Kematian Demam Berdarah Dengue di Jawa Timur), Tugas Akhir, Jurusan Statistika ITS.
- [6] Wati, W.E. (2009). Beberapa Faktor yang Berhubungan Dengan Kejadian Demam Berdarah Dengue (DBD) di Kelurahan Ploso Kecamatan Pacitan Tahun 2009. Surakarta: Program Studi Kesehatan Masyarakat Fakultas Ilmu Kesehatan Universitas Muhammadiyah Surakarta.
- [7] Widiyanto, T. (2007). Kajian Manajemen Lingkungan Terhadap Kejadian Demam Berdarah Dengue (DBD) di Kota Purwokerto Jawa-Tengah. Semarang: Program Pasca Sarjana, Universitas Diponegoro.
- [8] Pratiwi, E.D, (2008), Pemodelan Pertumbuhan Balita Menggunakan Regresi Spline Sebagai Pendekatan Pola Kurva Kartu Menuju Sehat (KMS), Tugas Akhir, Jurusan Statistika ITS, Surabaya.
- [9] Budiantara, I. N., (2001), Estimasi Parametrik dan Nonparametrik untuk Pendekatan Kurva Regresi, Seminar Nasional Statistika V, Jurusan Statistika, FMIPA, ITS, Surabaya.
- [10] Hardle, W.,(1990), Applied Nonparametric Regression, Cambridge University Press, New York.
- [11] Budiantara, I. N., (2009), Spline Dalam Regresi Nonparametrik dan Semiparametrik : Sebuah Pemodelan Statistika Masa Kini dan Masa Datang, ITS Press, Surabaya.
- [12] Wahba, G.,(1990), Spline Models For Observasion Data, SIAM Pennsylvania.
- [13] Ryan, T.P., (1997), Modern Regression Method, John Wiley & Sons, Inc, Canada.