

PENENTUAN MODEL REGRESI NONPARAMETRIK SPLINE PADA DATA PERTUMBUHAN BALITA DI DESA NANIA PROVINSI MALUKU TAHUN 2013-2014

F. Y. Rumlawang¹, S. N. Aulele², N. Kasim³

^{1,2,3}Jurusan Matematika FMIPA, Universitas Pattimura
Jl. Ir. M. Putuhena, Kampus Unpatti, Poka-Ambon, Indonesia
e-mail: ¹rumlawang@yahoo.com; ²salmon.aulele@yahoo.com

Abstrak

Regresi spline merupakan salah satu model dengan pendekatan nonparametrik, yang merupakan modifikasi dari fungsi polinomial tersegmen. Bentuk estimator spline sangat dipengaruhi oleh nilai parameter penghalus λ yang pada hakekatnya adalah penentuan lokasi titik-titik knot. Penentuan lokasi titik knot optimal merupakan persoalan yang sangat penting dalam estimasi regresi spline. Data yang digunakan dalam penelitian ini adalah data tentang berat badan balita di Desa Nania tahun 2013 sampai tahun 2014. Penelitian ini diperoleh titik knot yang optimal yaitu 6 dan 42. Nilai GCV yang diperoleh dari model regresi spline kubik optimal yaitu sebesar 0,0021643. Model spline kubik yang terbentuk yaitu: $\hat{f}(x) = 3,615333 + 0,675711x - 0,01848x^2 + 0,000204x^3 + 0,42987(x - 6)^3 - 0,19933(x - 42)^3$.

Kata Kunci: Regresi Nonparametrik, Spline, Titik Knot, GCV.

DETERMINATION OF NONPARAMETRIC SPLINE REGRESSION MODEL ON DATA GROWTH OF CHILDREN IN THE NANIA VILLAGE, PROVINCE OF MALUKU FROM 2013-2014

Abstract

Spline regression is one of the models with non parametric approach, which is a modification of the segmented polynomial functions. Shape spline estimator is strongly influenced by the value of smoothing parameter λ that is essentially a determination of the location of the points knots. Selection of the optimal λ in spline regression means the election of the location of dots knots. Therefore, the determination of the optimal location of the knots is a very important issue in the spline regression estimation. The data used in this research is data about weight infants in the village of Nania from 2013 to 2014. The research derived knots optimal point is; 6 and 42. GCV Value cubic derived from spline regression model optimized by 0,0021643. Model cubic spline is $\hat{f}(x) = 3,615333 + 0,675711 x - 0,01848x^2 + 0,000204x^3 + 0,42987(x - 6)^3 - 0,19933(x - 42)^3$

Keywords: *Nonparametric Regression, Spline, Point Knot, GCV.*

1. Pendahuluan

Analisis regresi merupakan salah satu metode statistika yang sering digunakan oleh para peneliti pada berbagai bidang keilmuan untuk mengidentifikasi sejumlah variabel prediktor dalam memprediksi suatu variabel respon. Untuk mengestimasi fungsi regresi, terdapat tiga macam pendekatan yaitu pendekatan regresi parametrik, regresi nonparametrik, dan regresi semiparametrik. Regresi parametrik digunakan jika terdapat informasi bentuk fungsi regresi berdasarkan teori maupun pengalaman masa lalu. Namun tidak semua pola data mengikuti model regresi parametrik, karena tidak adanya informasi tentang hubungan antara variabel respon dan variabel prediktor. Jika dipaksakan maka akan memperoleh kesimpulan yang menyesatkan. Solusi dari masalah tersebut adalah mengestimasi fungsi regresi menggunakan pendekatan

regresi nonparametrik. Regresi nonparametrik tidak bergantung pada model tertentu sehingga dapat dikatakan regresi nonparametrik memiliki fleksibilitas yang tinggi terhadap suatu data, Gabungan antara regresi parametrik dan nonparametrik akan diperoleh model regresi semiparametrik [1].

Estimasi fungsi pada regresi nonparametrik dilakukan berdasarkan data pengamatan dengan beberapa teknik *smoothing* tertentu, antara lain dengan menggunakan histogram, estimator kernel, deret orthogonal, estimator *spline*, estimator k-NN, estimator deret *fourier*, dan estimator *wavelet* [2].

Regresi spline adalah suatu pendekatan dengan tetap memperhitungkan kemulusan kurva. spline merupakan model polinomial yang tersegmen. Sifat tersegmen inilah sehingga memberikan fleksibilitas yang lebih baik daripada model polinomial biasa. Sifat ini juga memungkinkan model regresi spline menyesuaikan diri secara efektif terhadap karakteristik lokal dari data. Penggunaan spline difokuskan kepada adanya perilaku atau pola data, yang pada daerah tertentu mempunyai karakteristik yang berbeda dengan daerah lain.

Regresi spline linier biasanya diaplikasikan pada data dengan pola masih sederhana sedangkan spline kuadrat dan kubik biasanya diaplikasikan pada data dengan pola data yang lebih kompleks. Bentuk estimator spline sangat dipengaruhi oleh nilai parameter penghalus λ dan dipengaruhi oleh lokasi dan banyaknya titik - titik knot [3].

2. Metodologi Penelitian

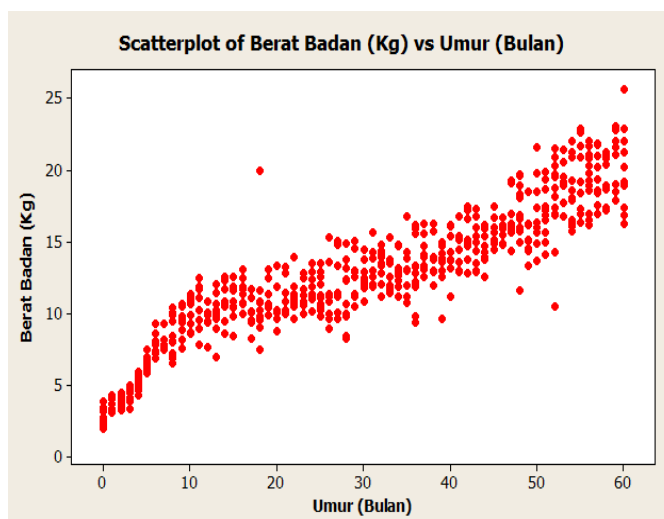
Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Puskesmas Nania. Data yang diteliti berupa data pertumbuhan balita menurut umur tahun 2013 sampai tahun 2014 dan beberapa literatur yang berhubungan pertumbuhan balita. Variabel yang digunakan yaitu variabel respon (y) berupa berat badan balita (kg) dan variabel prediktor (x) berupa umur balita (bulan). Adapun prosedur penelitian pada penulisan ini yaitu :

- Mendeskripsikan data untuk melihat pola hubungan antara variabel respon (y) dengan variabel prediktor (x)
- Membuat *scatter plot* antara variabel respon (y_i) dengan variabel prediktor (x_i) dengan $i = 1,2,3, \dots, 60$
- Menentukan titik knots dan orde knots yang menghasilkan nilai GCV terkecil untuk setiap variabel prediktor (x) dengan variabel respon (y).
- Membuat model regresi nonparametrik spline berdasarkan titik knots dan orde knots yang menghasilkan nilai GCV terkecil.
- Melakukan pengujian signifikansi parameter regresi nonparametrik spline.
- Melakukan diagnostik residual.

3. Hasil Dan Pembahasan

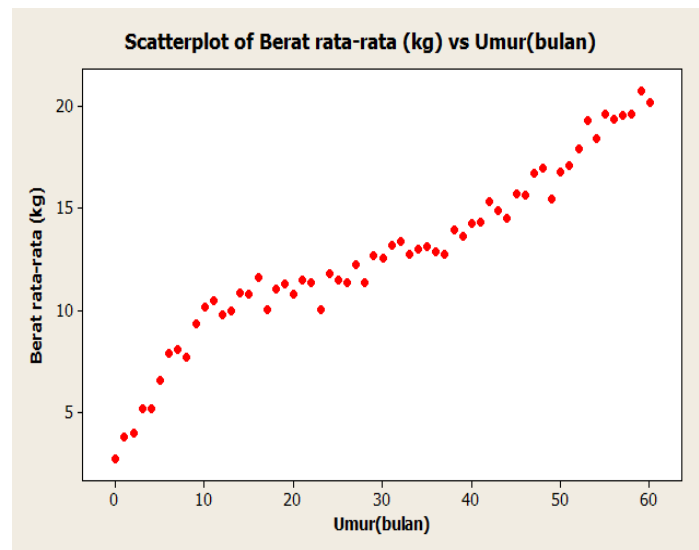
3.1 Pola Pertumbuhan Balita di Desa Nania

Data pertumbuhan balita di desa Nania pada tahun 2013 sampai 2014 disajikan dalam gambar sebagai berikut :



Gambar 1. Scatter Plot Untuk Data Pertumbuhan Balita di Desa Nania

Berdasarkan Gambar 1, terlihat bahwa semakin usia bertambah, sebaran data semakin besar, dan terdapat data pencilon. Selanjutnya rata-rata berat badan balita berdasarkan umur disajikan dalam gambar sebagai berikut :



Gambar 2. Scatter Plot Untuk Data Rata-Rata Berat Badan Balita Menurut Umur

Berdasarkan Gambar 2, terlihat bahwa terjadi pola pertumbuhan balita menurut umur di Desa Nania pada interval tertentu. Pola pertumbuhan balita saat lahir sampai umur sepuluh bulan umumnya sangat cepat, namun setelah umur sepuluh bulan dan sampai umur dua tahun pertumbuhannya mulai melambat, kemudian setelah umur dua tahun pertumbuhannya mulai cepat. Oleh sebab itu, digunakan model spline polinomial truncated untuk memodelkan pola hubungan antara usia dan berat badan balita yaitu :

$$f(x) = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \dots + \beta_mx^m + \sum_{j=1}^k \beta_{j+m-1}(x - \lambda_j)_+^{m-1}$$

Nilai m menunjukkan orde spline dan k menunjukkan banyaknya titik knot. Pemilihan titik knot optimal dalam model spline dapat menggunakan metode GCV. Knot yang optimal berkaitan dengan nilai GCV yang terkecil.

3.2 Model Regresi Nonparametrik Spline Data Pertumbuhan Balita

Untuk mendapatkan model spline yang baik dipilih dari nilai GCV yang minimum, yang dilakukan dengan cara coba-coba mulai dari bentuk spline linier ($m = 1$), kuadrat ($m = 2$) maupun kubik ($m = 3$), dengan satu titik knot sampai tiga titik knot. Dari ketiga model spline untuk pertumbuhan balita di Desa Nania tahun 2013 sampai 2014 disajikan dalam Tabel 1 dibawah ini :

Tabel 1. Model Spline Linier, Kuadrat dan Kubik Untuk Satu Titik Knot Dan Nilai GCV

Orde Spline	Titik knot	Nilai GCV
Spline linier	9	0,197974
Spline kuadrat	9	0,4297471
Spline kubik	42	0,0027593

Berdasarkan Tabel 1, terlihat bahwa model spline yang sesuai untuk satu titik knot adalah spline kubik karena memiliki nilai GCV terendah yaitu 0,0027593.

Tabel 2. Model spline linier, kuadrat dan kubik untuk dua titik knot dan nilai GCV

Orde Spline	Titik knot	Nilai GCV
Spline linier	9,26	0,0109425
Spline kuadrat	9,49	0,0173453
Spline kubik	6,42	0,0021643

Berdasarkan Tabel 2, terlihat bahwa model spline yang sesuai untuk dua titik knot adalah spline kubik karena memiliki nilai GCV terendah yaitu 0,0021643.

Tabel 3. Model spline linier, kuadrat dan kubik untuk tiga titik knot dan nilai GCV

Orde Spline	Titik knot	Nilai GCV
Spline linier	9;18;53	0,148758
Spline kuadrat	4;9;49	1,804485
Spline kubik	6;9;49	0,121249

Berdasarkan Tabel 3, terlihat bahwa model spline yang sesuai untuk tiga titik knot adalah spline kubik karena memiliki nilai GCV terendah yaitu 0,121249. Sehingga secara keseluruhan model spline yang sesuai untuk data pertumbuhan balita di Desa Nania adalah model spline kubik orde tiga dengan kombinasi dua titik knot yaitu knot 6 dan 42 serta nilai GCV minimum sebesar 0,0021643.

Dengan menggunakan estimasi model regresi spline kubik dengan dua titik knot yang disajikan dalam fungsi terpotong (*truncated*), maka diperoleh :

Tabel 4. Fungsi terpotong model regresi spline kubik dengan 2 titik knot

Interval	x	$\hat{f}(x)$
$x \leq 6$	3	5,481654
$6 < x < 42$	7	8,03929
$x \geq 42$	45	23,261

3.3 Estimasi Parameter Untuk Model Spline Kubik

Selanjutnya dilakukan estimasi parameter untuk model spline kubik. Dengan menggunakan software R diperoleh hasil sebagai berikut :

Tabel 5. Estimasi Parameter Untuk Spline Kubik

Parameter	Estimasi
β_0	3,615333
β_1	0,675711
β_2	-0,01848
β_3	0,000203

Model spline yang diperoleh untuk data pertumbuhan balita di Desa Nania Tahun 2013 sampai 2014 adalah model spline kubik dengan kombinasi dua titik knot yaitu 6 dan 42 dengan nilai GCV sebesar 0,0021643. Berdasarkan Tabel 5, maka model yang diperoleh adalah sebagai berikut :

$$\hat{f}(x) = 3,615333 + 0,675711x - 0,01848x^2 + 0,000204x^3 + 0,42987(x - 6)^3 - 0,19933(x - 42)^3$$

Selanjutnya estimasi model regresi spline kubik dengan dua titik knot disajikan dalam bentuk fungsi terpotong (*truncated*) sebagai berikut :

$$\hat{f}(x) = \begin{cases} 3,615333 + 0,675711x - 0,01848x^2 + 0,000204x^3 & ; x \leq 6 \\ -110,76293 + 57,864843x - 9,550002x^2 + 0,529733x^3 & ; 6 < x < 42 \\ -22149,438 + 1583,0366x - 37,69374x^2 + 0,299214x^3 & ; x \geq 42 \end{cases}$$

3.4 Pengujian Asumsi Residual

Selanjutnya akan dilakukan pengujian asumsi residual terhadap model regresi spline yang diperoleh.

a. Pengujian Residual Identik

Hipotesis yang digunakan dalam pengujian residual identik yaitu :

$$H_0 : \sigma_0^2 = \sigma_1^2 = \dots = \sigma_n^2 = \sigma_n$$

$$H_1 : \text{minimal terdapat satu } \sigma_i^2 = \sigma^2, \quad i = 0, 1, 2, \dots, n$$

Dengan menggunakan software SPSS, diperoleh hasil bahwa nilai signifikan sebesar 0,101 lebih besar dari 0,05 sehingga dapat disimpulkan bahwa tidak terdapat heteroskedastisitas.

b. Pengujian Residual Independen

Hipotesis yang digunakan dalam pengujian residual independen yaitu :

$H_0 : \rho = 0$ (tidak ada korelasi antar residual)

$H_1 : \rho \neq 0$ (ada korelasi antar residual)

Dengan menggunakan software SPSS, diperoleh hasil bahwa nilai $DW = 2,085$. Berdasarkan Tabel d Durbin Watson untuk $n = 61$ dan $K = 1$, nilai $dL = 1,5524$; $dU = 1,6189$ dan nilai $4 - dU = 2,3811$. Karena nilai $DW = 2,085$ kurang dari $4 - dU = 2,3811$ maka tidak ada korelasi antar residual.

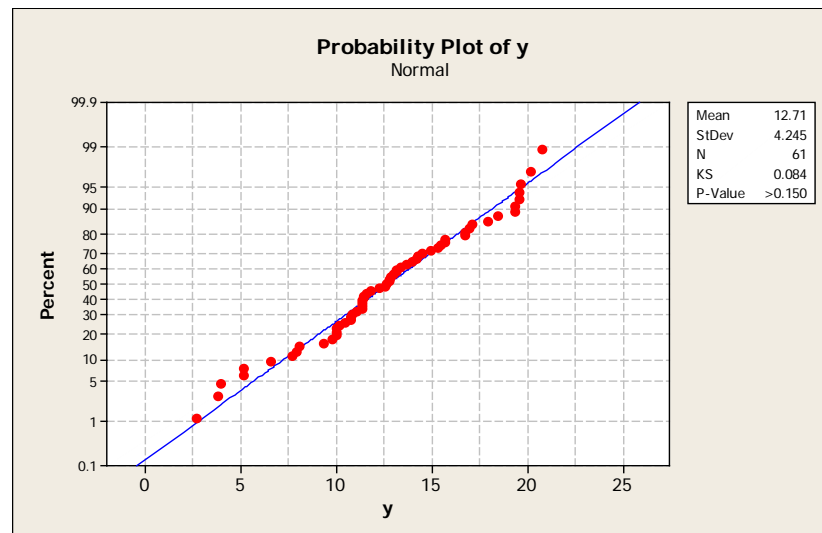
c. Pengujian residual berdistribusi normal

Hipotesis yang digunakan dalam pengujian residual berdistribusi normal yaitu :

$H_0 =$ Residual berdistribusi normal

$H_1 =$ Residual tidak berdistribusi normal

Dengan menggunakan software SPSS diperoleh *Probability Plot* sebagai berikut :



Gambar 3. *Probability Plot*

Berdasarkan Gambar 3, terlihat bahwa nilai p-value lebih dari 0,05. Hal ini menunjukkan bahwa residual berdistribusi normal.

4. Kesimpulan

Berdasarkan hasil dan pembahasan, diperoleh kesimpulan sebagai berikut :

1. Data pertumbuhan balita di Desa Nania tahun 2013 sampai 2014 dapat dimodelkan dengan menggunakan regresi nonparametrik spline dengan model matematis sebagai berikut :

$$\hat{f}(x) = 3,615333 + 0,675711x - 0,01848x^2 + 0,000204x^3 + 0,42987(x - 6)^3 - 0,19933(x - 42)^3$$

2. Model regresi spline kubik adalah model terbaik yang dapat memodelkan pertumbuhan balita di Desa Nania tahun 2013 sampai 2014, dengan dua titik knot yaitu 6 dan 42 dan nilai GCV yang diperoleh sebesar 0,0021643.

Daftar Pustaka

- [1] Wand, M. P, "Smoothing and mixed models," *a journal of smoothing methods*, 2002.
- [2] Eubank, R. L, "Spline Smoothing And Nonparametric Regression," 1998.
- [3] Budiantara, I. N, "Penentuan Titik-Titik Knots dalam Regresi Spline," *Jurnal Jurusan Statistika FMIPA-ITS*, 2005.
- [4] Hardle, W, "Applied Nonparametric Regression," *Cambridge University Press*, 1990.
- [5] Tripena, Agustini, "Penentuan Model Regresi Spline Terbaik," *Jurnal Program Studi Matematika FMIPA dan Teknik Jendral Soedirman*, 2011.
- [6] Gujarati, D, "Basic Econometric," 2004.