E-Jurnal Matematika Vol. 7(3), Agustus 2018, pp. 259-263 ISSN: 2303-1751
DOI: https://doi.org/10.24843/MTK.2018.v07.i03.p212

PENDEKATAN REGRESI *SPLINE* UNTUK MEMODELKAN POLA PERTUMBUHAN BERAT BADAN BALITA

Ni Luh Sukerni[1], I Komang Gde Sukarsa[2], Ni Luh Putu Suciptawati[3]

[1]Program Studi Matematika, FMIPA – Universitas Udayana [Email: sukeerni@gmail.com]
[2]Program Studi Matematika, FMIPA – Universitas Udayana [Email:gedesukarsa@unud.ac.id]
[3]Program Studi Matematika, FMIPA – Universitas Udayana [Email:suciptawati@unud.ac.id]

ABSTRACT

The study is aimed to estimate the best spline regression model for toddler's weight growth patterns. Spline is one of the nonparametric regression estimation method which has a high flexibility and is able to handle data that change in particular subintervals so thus resulting in model which fitted the data. This study uses data of toddler's weight growth at Posyandu Mekar Sari, Desa Suwug, Kabupaten Buleleng. The best spline regression model is chosen based on the minimum Generalized Cross Validation (GCV) value. The study shows that the best spline regression model for the data is quadratic spline regression model with six optimal knot points. The minimum GCV value is 0,900683471925 with the determination coefficient (R^2) *equals to 0,954609.*

Keywords: *GCV, Knot points, Nonparametric Regression, Spline, Toddler's Weight*

1. PENDAHULUAN

Pendekatan nonparametrik digunakan untuk mengetahui pola hubungan antara variabel respon dengan variabel prediktor yang bentuk fungsinya tidak diketahui. Pendekatan nonparametrik tidak terikat oleh asumsi-asumsi tertentu seperti sisaan berdistribusi normal dan memiliki variansi yang konstan. Bentuk kurva hanya diasumsikan bersifat mulus (*smoothness*) di mana data akan mencari sendiri bentuk estimasinya. Terdapat beberapa pendekatan dalam regresi nonparametrik (Hardle, 1994) di antaranya histogram, penduga kernel, penduga *spline*, deret *Fourier*, dan lain-lain.

Regresi *spline* mempunyai fleksibilitas yang tinggi dan mampu menangani data yang perilakunya berubah–ubah pada sub–sub interval tertentu (Eubank, 1999). *Spline* merupakan potongan polinomial *(piecewise polynomial)* dengan sifat tersegmen yang terbentuk pada titik-titik *knot*. Titik *knot* merupakan titik perpaduan bersama yang terjadi karena terdapat perubahan pola perilaku data. Terdapat beberapa metode untuk memilih titik *knot* yang optimal dalam regresi nonparametrik *spline* antara lain metode *cross validation* (CV), *unbiassed risk* (UR), *generalized cross validation* (GCV), dan *generalized maximum likelihood* (GML).

Metode yang digunakan untuk menentukan titik *knot* optimal pada penelitian ini adalah *generalized cross validation* (GCV) dengan kriteria GCV minimum (Wahba, 1990). Penggunaan titik *knot* pada penelitian ini dibatasi sebanyak sepuluh titik knot.

Salah satu kasus yang menggunakan penduga *spline* dalam menduga model regresi adalah kasus pertumbuhan balita. Pemantauan pertumbuhan pada balita dapat dilakukan dengan cara mengukur berat badan pada balita. Pola pertumbuhan pada balita cenderung memiliki perubahan perilaku pada umur-umur tertentu, sehingga penduga *spline* merupakan metode yang sesuai untuk melihat dan memantau pertumbuhan berat badan pada balita. Penelitian ini bertujuan untuk mengestimasi model regresi *spline* terbaik pada pola pertumbuhan berat badan balita di Posyandu Mekar Sari, Desa Suwug, Kabupaten Buleleng.

Secara umum, model regresi nonparametrik dapat dituliskan sebagai berikut:

$y_i = f(x_i) + \varepsilon_i$, i = 1,2,…,$n$ (1)

dengan y_i menyatakan variabel respon dari data ke-i , x_i menyatakan variabel prediktor dari data ke-*i*, $f(x_i)$ menyatakan fungsi regresi yang tidak diketahui bentuknya, dan ε_i menyatakan error ke-*i* yang diasumsikan menyebar $N\sim(0,\sigma^2)$.

Fungsi *spline* berorde *p* adalah sebarang fungsi yang dapat ditulis dalam bentuk:

$$y_i = \beta_0 + \beta_1 x_i^1 + \cdots + \beta_p x_i^p + \sum_{k=1}^{K} \beta_{p+k}(x_i - \kappa_k)_+^p + \varepsilon_i$$
$$= \sum_{j=0}^{p} \beta_j x_i^j + \sum_{k=1}^{K} [\beta_{p+k}.\ (x_i - \kappa_k)_+^p] + \varepsilon_i \qquad (2)$$

dengan fungsi sepenggal *(truncated)* sebagai berikut:

$$(x_i - \kappa_k)_+^p = \begin{cases} (x_i - \kappa_k)^p, \\ untuk\ x_i \geq \kappa_k \\ 0,\ x_i < \kappa_k \end{cases}$$

dengan β_j adalah konstanta real dan $\kappa_1, \kappa_2, \dots, \kappa_k$ adalah titik–titik *knot*.

Fungsi *spline* dapat disajikan dalam bentuk matriks yang dituliskan sebagai berikut:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad (3)$$

dengan

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \\ \beta_{(p+1)} \\ \vdots \\ \beta_{(p+k)} \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1^1 & x_1^2 & \dots & x_1^p & (x_1 - \kappa_1)_+^p \\ 1 & x_2^1 & x_2^2 & \dots & x_2^p & (x_2 - \kappa_1)_+^p \\ & & & & & \vdots \\ 1 & x_n^1 & x_n^2 & \dots & x_n^p & (x_n - \kappa_1)_+^p \end{bmatrix}.$$

Untuk mencari nilai penduga β digunakan metode *least square*. Metode ini dilakukan dengan meminimumkan $\varepsilon^T\varepsilon$ terhadap β (Draper & Smith, 1992).

$$\varepsilon^T\varepsilon = (Y - X\beta)^T(Y - X\beta)$$
$$= Y^TY - 2\beta^TX^TY - X^T\beta^TX\beta. \qquad (4)$$

Kemudian meminimumkan persamaan (4) dengan menurunkan terhadap β sama dengan 0.

$$\frac{\partial\varepsilon^T\varepsilon}{\partial\beta} = 0$$
$$X^TX\beta = X^TXY$$

Sehingga diperoleh nilai $\widehat{\beta}$ sebagai berikut:

$$\widehat{\beta} = (X^TX)^{-1}X^TY \qquad (5)$$

Bentuk penduga dari fungsi $\hat{Y}$ dapat dituliskan sebagai berikut:

$$\hat{Y} = X(X^TX)^{-1}X^T \qquad (6)$$

Titik *knot* optimal dipilih berdasarkan kriteria *generalized cross validation* (GCV) yang minimum. Rumus untuk menghitung GCV adalah sebagai berikut:

$$GCV(k) = \frac{MSE\ (k)}{(n^{-1}tr[I - A(k)])^2} \qquad (7)$$

dengan $MSE(k) = n^{-1}\sum_{i=1}^{n}(y_i - \hat{y})^2$, n adalah jumlah data, *I* adalah matriks identitas, *k* adalah titik *knot* $(k_1, k_2, k_3, \dots, k_n)$ dan $\boldsymbol{A(k)} = \boldsymbol{X(X^TX)^{-1}X^T}$ (Eubank, 1999).

Setelah diperoleh model regresi *spline* terbaik, selanjutnya dilakukan pengujian parameter model. Pengujian parameter model terdiri dari dua tahapan, yaitu pengujian secara serentak kemudian dilanjutkan dengan pengujian secara individu. Uji serentak dilakukan untuk mengetahui signifikansi parameter model regresi secara bersama-sama (Neter et al., 1997a). Hipotesis yang digunakan pada uji serentak adalah sebagai berikut:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_{p+k} = 0$,

H_1: $minimal\ ada\ satu\ \beta_j \neq 0, j = 1,2,\dots,p+k$

dengan nilai *p*+*k* adalah jumlah parameter dalam model regresi *spline*, *p* adalah derajat pada *spline* dan *k* adalah jumlah *knot*. Statistik uji yang digunakan adalah uji *F*:

$$F_{hitung} = \frac{SSR/_{db}}{SSE/db} \qquad (8)$$

Tolak H_0 jika nilai $F_{hitung} > F_{\alpha;((p+k),n-(p+k)-1)}$.

Pengujian parameter secara individu bertujuan untuk mengetahui mana variabel yang berpengaruh secara signifikan terhadap model

E-Jurnal Matematika Vol. 7(3), Agustus 2018, pp. 259-263 ISSN: 2303-1751
DOI: https://doi.org/10.24843/MTK.2018.v07.i03.p212

(Neter et al., 1997a). Hipotesis yang digunakan dalam uji individu adalah sebagai berikut:

H_0: $\beta_j = 0$

H_1: $\beta_j \neq 0\,;\ j = 1,2,\dots,p+k$

Statistik uji yang digunakan adalah uji *t*:

$$t_{hitung} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \qquad (9)$$

dengan daerah kritis tolak H_0 jika $|t_{hitung}| > t_{(\frac{\alpha}{2},n-(p+k)-1)}$ dengan n merupakan jumlah pengamatan dan (p+k) merupakan jumlah parameter dalam model regresi *spline*.

2. METODE PENELITIAN

Jenis dan Sumber Data

Data yang digunakan adalah data rekam berat badan balita yang diambil di Posyandu Mekar Sari, Desa Suwug, Kabupaten Buleleng. Variabel respon (*Y)* pada penelitian ini adalah berat badan balita saat ditimbang dalam satuan kg. Variabel prediktor dalam penelitian ini adalah berat badan balita saat lahir dalam satuan kilogram (X_1) dan umur balita saat ditimbang dalam satuan bulan (X_2).

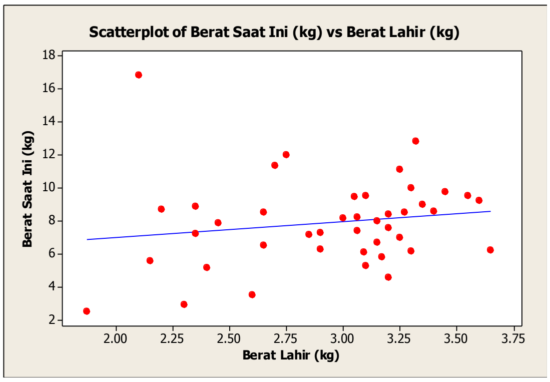
Metode Analisis Data

Langkah analisis dalam penelitian ini adalah sebagai berikut:

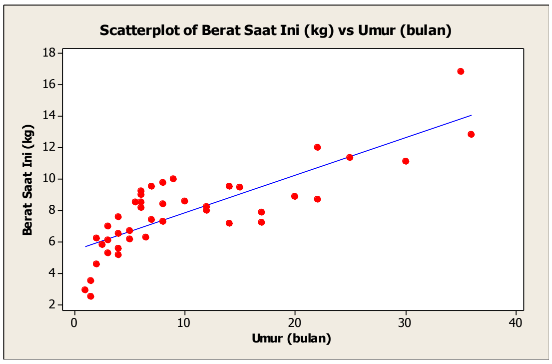
1. Mengidentifikasi pola hubungan antara variabel respon (*Y*) dengan variabel prediktor (X_1) serta (X_2) menggunakan *scatter plot.*
2. Memodelkan hubungan variabel prediktor (X_1) dan (X_2) terhadap variabel respon (*Y*) dengan regresi *spline* satu *knot* sampai sepuluh *knot* untuk masing-masing orde linearr, kuadratik, kubik.
3. Menentukan titik *knot* optimal dengan kriteria *Generalized Cross Validation* (GCV) minimum.
4. Menetapkan model regresi *spline* terbaik.
5. Melakukan uji signifikansi parameter secara serentak dan individu.
6. Menginterpretasikan model regresi *spline* terbaik dan menarik kesimpulan dari model.

3. HASIL DAN PEMBAHASAN

Scatterplot untuk Variabel Respon dengan Variabel Prediktor



Gambar 1. Scatterplot antara Berat Badan Balita saat Lahir (X_1) dengan Berat Badan Balita saat Ditimbang (Y)



Gambar 2. Scatterplot antara Umur Balita saat Ditimbang (X_2) dengan Berat Badan Balita saat Ditimbang (Y)

Pola hubungan antara setiap variabel prediktor dengan variabel respon memiliki pola yang tidak mengikuti pola tertentu sehingga sulit didekati dengan pendekatan regresi parametrik.

Estimasi Model Regresi Spline Terbaik

Pemilihan model regresi *spline* terbaik dipengaruhi oleh lokasi dan banyaknya titik *knot*. Lokasi titik *knot* yang berbeda akan menghasilkan model yang berbeda. Pemilihan titik *knot* optimal berdasarkan kriteria GCV minimum. Berikut ini merupakan tabel yang menunjukkan titik-titik *knot* yang optimal, nilai GCV minimum serta orde yang optimal untuk setiap variabel.

Tabel 1. Nilai GCV Minimum untuk Masing-masing Titik *Knot*

Titik *knot*	Nilai GCV minimum	Orde	
		X_1	X_2
1 titik *knot*	1,455221	kubik	kubik
2 titik *knot*	1,328506	kuadratik	kuadratik
3 titik *knot*	0,9947627	kubik	kubik
4 titik *knot*	0,9910701	kuadratik	kuadratik
5 titik *knot*	0,9173034	kuadratik	kuadratik
6 titik *knot*	0,9006834	kuadratik	kuadratik
7 titik *knot*	1,0140267	kuadratik	kuadratik
8 titik *knot*	1,1362618	linear	linear
9 titik *knot*	1,2691636	linear	linear
10 titik *knot*	1,5345317	linear	linear

Tabel 1 menunjukkan bahwa nilai GCV minimum diperoleh untuk model dengan enam titik *knot* pada orde kuadratik. Titik *knot* pada X_1 adalah $K_1 = 1{,}8878$; $K_2 = 2{,}081622$; $K_3 = 2{,}275444$; $K_4 = 2{,}6630889$; $K_5 = 2{,}85691111$;$K_6 = 3{,}05073333$. Titik *knot* pada X_2 adalah $K_1 = 1{,}35$; $K_2 = 5{,}16111$; $K_3 = 8{,}9722222$; $K_4 = 16{,}594444$; $K_5 = 20{,}4055556$; $K_6 = 24{,}2166667$.

Nilai estimasi parameter $\hat{\beta}$ yang diperoleh untuk regresi *spline* orde kuadratik dengan 6 titik *knot* adalah sebagai berikut:

$$\hat{\beta} = \begin{bmatrix} -0{,}156245062553979 \\ 0{,}100550364394037 \\ -1{,}108564002538701 \\ 1{,}52722226563398 \\ -0{,}577515588303437 \\ 0{,}100833839676168 \\ -0{,}045746458863910 \\ 0{,}0016664516812001 \\ -0{,}086243375517550 \\ 0{,}032517164203101 \\ -0{,}032660375867682 \\ 0{,}000082183735415 \\ 0{,}000079441422929 \\ 0{,}000090794244481 \\ -0{,}000346005761766 \\ 0{,}000291060190130 \end{bmatrix}$$

Selanjutnya dilakukan uji signifikansi parameter regresi nonparametrik *spline* secara serentak dengan hipotesis sebagai berikut:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_{16} = 0$,

H_1: $minimal\ ada\ satu\ \beta_j \neq 0, j = 1,2,\ldots,16$

Tabel 2. Analisis Varians Regresi Nonparametrik *Spline*

Sumber	db	SS	MS	F_{hitung}
Regresi	15	280,90	18,72	37,45
Error	26	13,12	0,5	
Total	41	294,02		

Dengan menggunakan α = 0,05, maka diperoleh nilai F tabel sebesar 2,07 sehingga Tabel 2 menunjukkan bahwa nilai $F_{hitung} > F_{tabel}$. Hal ini mengindikasikan bahwa H_0 ditolak, yang artinya terdapat pengaruh yang signifikan secara bersama-sama antara variabel bebas terhadap variabel respon pada model regresi nonparametrik *spline*. Kemudian dilakukan uji parameter secara individu dengan menggunakan uji *t*. Hasil pengujian signifikansi parameter model secara individu yang disajikan pada Tabel 3.

Tabel 3. Hasil Uji Individu Parameter Model Regresi Nonparametrik *Spline*

Variabel	Parameter	*t*-hitung	*t*-tabel	Keputusan
X_1	β_1	-5,18	2,05954	Tolak
	β_2	6,48		Tolak
	β_3	-7,02		Tolak
	β_4	6,84		Tolak
	β_5	-5,96		Tolak
	β_6	2,14		Tolak
	β_7	-0,94		Gagal Tolak
	β_8	0,07		Gagal Tolak
X_2	β_9	-3,17		Tolak
	β_{10}	3,19		Tolak
	β_{11}	-3,18		Tolak
	β_{12}	0,65		Gagal Tolak
	β_{13}	1,22		Gagal Tolak
	β_{14}	0,82		Gagal Tolak
	β_{15}	-2,15		Tolak
	β_{16}	3,02		Tolak

Dengan menggunakan tingkat signifikansi α = 5% didapatkan parameter-parameter yang signifikan yaitu β_1 , β_2 , β_3 , β_4 , β_5 , β_6 , β_9 , β_{10} , β_{11} , β_{15} ,β_{16} . Meskipun terdapat beberapa parameter yang tidak signifikan pada model, namun semua variabel bebas berpengaruh terhadap pola pertumbuhan berat badan balita.

E-Jurnal Matematika Vol. 7(3), Agustus 2018, pp. 259-263 ISSN: 2303-1751
DOI: https://doi.org/10.24843/MTK.2018.v07.i03.p212

Berikut merupakan estimasi model regresi nonparametrik *spline* pada orde kuadratik dengan 6 titik *knot* optimal yaitu:

$$
\begin{aligned}
\hat{Y} = & -0{,}1562450625539\, x_1 + 0{,}100550364394037\, x_1^2 \\
& -1{,}108564002538701\, (x_1 - 1{,}8878)_+^2 \\
& +1{,}52722226563398\, (x_1 - 2{,}0816222)_+^2 \\
& -0{,}577515588303437\, (x_1 - 2{,}2754444)_+^2 \\
& +0{,}100833839676168\, (x_1 - 2{,}66308889)_+^2 \\
& -0{,}04574645886391\, (x_1 - 2{,}85691111)_+^2 \\
& +0{,}0016664516812001\, (x_1 - 3{,}05073333)_+^2 \\
& -0{,}08624337551755\, x_2 + 0{,}0325171642031\, x_2^2 \\
& -0{,}032660375867682\, (x_2 - 1{,}35)_+^2 \\
& +0{,}000082183735415\, (x_2 - 5{,}161111)_+^2 \\
& +0{,}000079441422929\, (x_2 - 8{,}9722222)_+^2 \\
& +0{,}000090794244481\, (x_2 - 16{,}5944444)_+^2 \\
& -0{,}000346005761766\, (x_2 - 20{,}4055556)_+^2 \\
& +0{,}00029106019013\, (x_2 - 24{,}2166667)_+^2
\end{aligned}
$$

4. KESIMPULAN DAN SARAN

Estimasi model regresi nonparametrik *spline* yang terbaik pada kasus pertumbuhan berat badan balita adalah model regresi spline dengan orde optimal kuadratik dengan 6 titik *knot* pada masing-masing variabel prediktor. Nilai GCV minimum yang dihasilkan adalah 0,900683471925 serta koefisien determinasi *(*R^2*)* sebesar 0,954609. Hal ini menunjukkan bahwa regresi nonparametrik *spline* dapat memodelkan hubungan antara berat badan balita saat lahir (X_1) dan umur balita saat ditimbang (X_2) terhadap berat badan balita saat ditimbang (Y) dengan baik.

Untuk penelitian selanjutnya saran yang dapat penulis sampaikan untuk penelitian lebih lanjut dengan pendekatan regresi nonparametrik yang lain regresi kernel. Penentuan orde pada setiap variabel dalam regresi *spline* tidak harus sama melainkan bisa berbeda-beda untuk setiap variabel.

DAFTAR PUSTAKA

Draper, N. R., & Smith, H. (1992). *Analisis Regresi Terapan, diterjemahkan oleh Bambang Sumantri*. Jakarta: PT. Gramedia Pustaka Utama.

Eubank, R. (1999). *Spline Smoothing and Nonparametric Regression.* New York: Marcell Dekker.

Hardle,W. (1994). *Applied Nonparametric Regression.*New York: Cambridge University Press.

Neter, J., Wasserman, W., & Kutner, M. H. (1997a). Model Linier Terapan I : Analisis Regresi Linier Sederhana. Terjemahan Bambang Sumantri. Bogor: Jurusan Statistika FMIPA IPB.

Wahba, G. (1990). Spline Models for Observational Data. *CBMS-NSF Regional Conference Series in Applied Mathematics, 59.*