

KONSTRUKSI MODEL HUBUNGAN DUA VARIABEL DENGAN ANALISIS REGRESI SPLINE

Hazmira Yozza^{1*}, Afrimayani¹

¹*Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Andalas
Jalan Limau Manis, Kec. Pauh, Kota Padang, Sumatera Barat*

**Email: hazmirayozza@unand.sci.ac.id*

ABSTRACT

This paper discussed about the construction of a model the relationship between a dependent variabel and a independent variabel using spline analysis regression. This methods is usually used when the relationship between variables is unknown. This methods is applied to construct the growth curve for children under 3 years of age in Padang.

Keywords: spline regression, growth curve, children under 3 years of age

PENDAHULUAN

Analisis regresi adalah suatu metode statistika yang digunakan untuk memodelkan hubungan antara dua atau lebih variabel yang menjadi perhatian. Dalam suatu kondisi tertentu, analisis regresi dapat digunakan untuk mengetahui adanya hubungan kausal antara dua jenis variabel yakni variabel bebas (independen) dan variabel tak bebas (dependen).

Misalkan suatu pengamatan terhadap variabel tak bebas kontinu Y diambil pada n nilai peubah bebas x_i . Selanjutnya, misalkan (x_i, y_i) adalah nilai x_i dan Y yang diperoleh dari suatu pengambilan sampel dan diasumsikan bahwa kedua variabel memiliki hubungan yang dapat dinyatakan dalam suatu model yang dinamakan model regresi sebagai berikut.

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

dengan ε_i dinamakan komponen galat (error) yang merupakan variabel acak yang saling bebas dan memiliki nilai tengah 0 dan ragam yang konstan, σ^2 . Kemudian $f(x_i)$ adalah nilai dari fungsi f pada nilai x_i ($i = 1, 2, \dots, n$)

Pendekatan parametrik adalah pendekatan yang umum digunakan untuk

mendapatkan model regresi (1) Pendekatan ini mengasumsikan bahwa bentuk fungsi f diketahui, dimana model tersebut tergantung secara linier ataupun nonlinier pada satu atau lebih parameter. Analisis regresi dilakukan untuk menduga nilai parameter dari model yang telah diketahui bentuknya tersebut.

Pendekatan lain yang dapat digunakan untuk menemukan suatu model yang sesuai data adalah pendekatan nonparametrik. Berbeda dengan pendekatan parametrik, pada pendekatan ini tidak ada asumsi mengenai bentuk parametrik dari model regresi. Oleh karena ini, dapat digunakan sebagai alternatif lain dari pendekatan parametrik ketika bentuk hubungan antara variabel tidak diketahui. Salah satu pendekatan regresi nonparametrik adalah pendekatan regresi spline.

Analisis regresi spline adalah analisis regresi polynomial yang tersegmentasi. Dengan pendekatan ini, range data terlebih dahulu dibagi menjadi $k+1$ segmen yang dipisahkan oleh k titik yang dinamakan *knot*. Selanjutnya, pada setiap segmen dibentuk fungsi polynomial berorde m dengan suatu cara yang menjamin kekontinuan model regresi yang terbentuk.

Pada tulisan ini, dibahas mengenai analisis regresi spline dan diterapkan untuk membentuk kurva pertumbuhan anak berusia

bawah tiga tahun (batita) di Kota Padang.

Analisis Regresi Kuadrat Terkecil

Analisis regresi merupakan analisis statistika yang digunakan untuk menganalisis hubungan antara variabel tak bebas Y dengan satu atau beberapa variabel bebas, sebut saja X_1, X_2, \dots, X_k . Hubungan tersebut dinyatakan dalam suatu persamaan yang dinamakan sebagai persamaan regresi yang secara umum dapat dinyatakan sebagai

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

dengan y_i adalah nilai pengamatan dari variabel tak bebas Y dan $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{ki})$ adalah nilai dari variabel bebas X_1, X_2, \dots, X_k .

Pendekatan parametrik adalah suatu pendekatan yang dapat digunakan untuk mendapatkan hubungan antar variabel tersebut. Dengan pendekatan tersebut, diasumsikan bahwa bentuk hubungan antar variabel tersebut diketahui dan melalui beberapa parameter yang tidak diketahui. Secara khusus, terdapat suatu vector parameter $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k) \in \mathbf{B}$, suatu subset dari \mathbb{R}^p , dan suatu fungsi yang diketahui bentuknya, yakni $f(\cdot|\boldsymbol{\beta})$ sedemikian sehingga $f(\cdot) = f(\cdot|\boldsymbol{\beta})$ (Eubank, 1988).

Model parametrik bergantung terhadap $\boldsymbol{\beta}$ baik secara linier maupun secara non linier. Secara linier, hubungan tersebut dinyatakan sebagai

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ji} + \varepsilon_i \quad (2)$$

dengan $i = 1, 2, \dots, n$, ε_i adalah galat (error) model yang diasumsikan saling bebas dengan nilai tengah 0 dan varians yang konstan (Myers, 1990). Model pada Persamaan (2) dinamakan model regresi linier. Dalam catatan matriks, model tersebut dapat dinyatakan sebagai

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3)$$

dengan vector nilai $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$, vector parameter $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)^T$, X adalah matriks rancangan dan dan vektor galat $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$.

Metode Kuadrat Terkecil (Least Square Methods) adalah salah satu metode yang digunakan untuk menduga parameter regresi yang selanjutnya dapat digunakan untuk menduga $f(x)$. Dengan metode ini, penduga

bagi vektor parameter diperoleh dengan meminimumkan jumlah kuadrat sisaan, sehingga diperoleh penduga

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y}) \quad (4)$$

(Draper & Smith, 1992).

Analisis Regresi Spline

Analisis regresi non parametrik adalah pendekatan regresi yang digunakan ketika tidak terdapat informasi yang lengkap tentang bentuk pola hubungan antara variabel. Metode ini digunakan dengan mengabaikan asumsi-asumsi yang biasanya terdapat dalam regresi parametrik. Pendekatan ini memiliki fleksibilitas yang tinggi karena data diharapkan mencari sendiri bentuk dugaan kurva regresinya. Bentuk umum model regresi ini adalah

$$y = f(x) + \varepsilon$$

dengan y adalah nilai variabel tak bebas, x adalah nilai variabel bebas, ε adalah *error* dan $f(x)$ adalah fungsi regresi yang tidak diketahui bentuk atau polanya (Wahba, 1990). Salah satu metode regresi non parametrik adalah regresi *spline*.

Spline merupakan model polinom yang memiliki sifat tersegmen atau terpotong-potong (*piecewise polynomial*) (Eubank, 1988) dimana kurva dibentuk secara individu pada setiap segmen dan selanjutnya digabungkan (Ryan, 1997). Fungsi *spline* merupakan suatu gabungan fungsi polinomial dimana penggabungan beberapa polinomial tersebut pada knot-knot dengan suatu cara yang menjamin sifat kontinuitas (Eubank, 1988). Metode ini adalah pendekatan yang sangat baik dalam memodelkan data yang memiliki pola berubah-ubah pada sub-interval (segmen) tertentu. Sifat segmen memberikan fleksibilitas yang lebih baik dibanding model polinomial biasa.

Knot adalah nilai variabel bebas yang menjadi batas antar segmen. *Knot* diartikan sebagai suatu titik fokus dalam fungsi *spline*, sehingga kurva yang dibentuk tersegmen pada titik tersebut. Setiap garis regresi mendefinisikan satu segmen sehingga *knot* dapat didefinisikan sebagai akhir dari satu segmen sekaligus merupakan awal dari segmen

yang lain. Fungsi *spline* berorde m dengan titik-titik knot $\xi_1, \xi_2, \dots, \xi_L$ didefinisikan sebagai sembarang fungsi f yang disajikan dalam bentuk (Eubank, 1988) :

$$S(x) = \sum_{i=1}^m \theta_i x^i + \sum_{j=1}^H \eta_j (x - \xi_j)_+^m \quad (5)$$

dengan fungsi potongan $(x - \xi_j)_+^m$ didefinisikan sebagai

$$(x - \xi_j)_+^m = \begin{cases} (x - \xi_j)^m & x \geq \xi_j \\ 0 & x < \xi_j \end{cases} \quad (6)$$

Parameter $\theta_1, \theta_2, \dots, \theta_m, \eta_1, \eta_2, \dots, \eta_k$ selanjutnya diduga dengan meminimumkan jumlah kuadrat sisaan model.

Dalam catatan matriks, model (5) dapat dituliskan sebagai

$$\hat{y}_\lambda = \mathbf{X}_\lambda (\mathbf{X}_\lambda' \mathbf{X}_\lambda)^{-1} (\mathbf{X}_\lambda' \mathbf{y}) = \mathbf{A}_\lambda \mathbf{y} \quad (7)$$

dengan $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ dan

$$\mathbf{X}_\lambda = \begin{pmatrix} 1 & t_1 & \dots & t_1^{m-1} & (t_1 - \xi_1)_+^{m-1} & \dots & (t_1 - \xi_L)_+^{m-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & \dots & t_n^{m-1} & (t_n - \xi_1)_+^{m-1} & \dots & (t_n - \xi_L)_+^{m-1} \end{pmatrix} \quad (8)$$

Parameter λ merupakan pengontrol keseimbangan antara kesesuaian kurva terhadap data dan kemulusan kurva. Memasangkan nilai λ yang sangat kecil atau besar akan memberikan bentuk fungsi penyelesaian yang sangat kasar atau sangat mulus (Eubank, 1988) (Butchon & Behav, 2017), sehingga perlu didapatkan λ yang optimal. Penentuan model *spline* terbaik pada dasarnya adalah menentukan titik *knot* optimum mencakup penentuan banyaknya titik *knot* dan pemilihan lokasi titik-titik *knot* tersebut. Penentuan model *spline* terbaik juga mencakup pemilihan orde dari fungsi polynomial yang digunakan pada setiap segmen.

Salah satu kriteria yang banyak dikembangkan untuk menentukan model regresi *spline* terbaik adalah kriteria *Generalized Cross Validation* (GCV). Kriteria ini dikembangkan di antaranya oleh (Wahba, 1990) dan (Budiantara, 1999). Pada model regresi *spline* terbobot kriteria GCV didefinisikan sebagai

$$GCV_\lambda = \frac{MSE}{[n^{-1} \text{trace}(I - A_\lambda)]^2}$$

dengan MSE (*mean square error*) adalah jumlah kuadrat galat, \mathbf{I} adalah matriks identitas dan n adalah banyaknya pengamatan. Model terbaik adalah model dengan GCV minimum.

Konstruksi Kurva Pertumbuhan Balita Dengan Analisis Regresi Spline

Tahun-tahun pertama kehidupan adalah fase yang sangat penting dalam kehidupan seorang manusia, karena masa ini memiliki pengaruh yang sangat penting terhadap kualitas seorang manusia di masa yang akan datang. Pada masa tersebut, terjadi perubahan yang sangat cepat dalam pertumbuhan dan perkembangan fisik anak. Pada masa tersebut juga terjadi penambahan yang sangat cepat dari jumlah-jumlah sel tubuh yang terlihat dari bertambahnya berat badan, tinggi badan dan lingkaran kepala anak. Dalam fase tersebut proses penambahan dan pematangan sel-sel otak serta sambungannya secara progresif. Demikian juga halnya dengan perkembangan motorik kasar dan motorik halus. Tidak salah jika periode ini sering dinamakan periode emas dalam hidup manusia.

Mengingat pentingnya periode ini, maka pertumbuhan dan perkembangan anak harus mendapatkan perhatian besar. Selain dengan memperhatikan asupan gizi yang mereka terima, harus dilakukan pemantauan berkala terhadap pertumbuhan anak. Pemantauan terhadap pertumbuhan anak mencakup pengukuran-pengukuran rutin yang dilakukan untuk mendeteksi terjadinya pertumbuhan yang tidak normal pada anak, yang dikombinasikan dengan tindakan-tindakan bila hal tersebut terjadi. Tujuannya adalah untuk memperbaiki keadaan gizi anak, mengurangi resiko kematian ataupun ketidakcukupan gizi (Garner, Panpanich, & Logan, 2000). Salah satu pengukuran yang biasa digunakan dalam memonitor pertumbuhan pada anak adalah berat badan. Hasil pengukuran tersebut selanjutnya diplotkan dalam suatu kurva pertumbuhan, untuk kemudian dinilai apakah terdapat masalah dalam pertumbuhan balita.

Di Indonesia, pengukuran berat badan anak dilakukan dalam suatu program pemerintah Indonesia yang biasa dinamakan

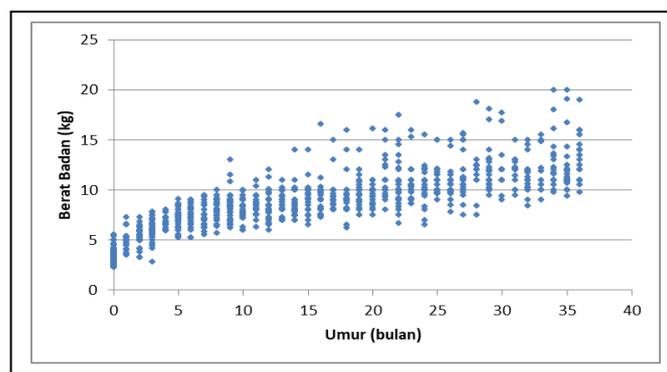
Posyandu (Pos Pelayanan Terpadu) yang dilaksanakan setiap bulan di berbagai lokasi di Indonesia. Salah satu kegiatan yang dilaksanakan adalah penimbangan badan balita. Untuk memantau pertumbuhan balita, berat badan balita kemudian dibandingkan dengan kartu KMS (Kartu Menuju Sehat) yang memuat rujukan baku standar berat badan sebagaimana yang dikeluarkan oleh WHO. Namun beberapa penelitian menunjukkan bahwa standar yang telah ditetapkan oleh WHO tidak sesuai dengan pertumbuhan balita di Indonesia sehingga perlu dibentuk kurva pertumbuhan lain yang lebih sesuai dengan pertumbuhan balita di suatu daerah.

Pada tulisan ini, akan dikonstruksi kurva pertumbuhan anak batita yang berdomisili di Kota Padang berdasarkan berat badan. Data yang digunakan untuk membentuk kurva pertumbuhan batita ini diperoleh dari suatu survey yang dilakukan di beberapa posyandu dan perumahan di enam kecamatan di Kota Padang. Variabel bebas yang digunakan untuk membentuk kurva pertumbuhan adalah umur balita (dalam bulan) sedangkan variabel bebasnya adalah berat badan batita (dalam kilogram).

Dari Gambar 1 terlihat bahwa terjadi perubahan pola pertumbuhan berat badan anak pada selang usia tertentu, dimana pada usia dibawah 5 bulan, terjadi penambahan yang cepat pada berat badan anak dan menjadi relatif stabil mendekati usia 36 bulan. Karena terdapatnya perubahan pola pertumbuhan berat badan pada selang umur tertentu, maka analisis regresi spline merupakan metode yang lebih tepat digunakan untuk memodelkan pola pertumbuhan berat badan anak batita di Kota Padang.

Selanjutnya, dibentuk model regresi spline dengan menggunakan model polinomial berorde 1 (model linier), orde 2 (model kuadratik) dan orde 3 (kubik) dengan banyaknya titik knot $k=1,2,3$ dan untuk semua kemungkinan lokasi knot. Selanjutnya, dari semua kemungkinan model, ditentukan model terbaik berdasarkan GCV.

Tabel 1 berikut memperlihatkan lokasi dari titik knot dan nilai GCV bagi model terbaik untuk setiap kombinasi orde polinomial yang digunakan dan banyaknya titik knot pada model spline.



Gambar 1. Diagram Pencar Berat Badan Batita terhadap Umur

Tabel 1. Knot Optimal dan Nilai GCV

Model	k=1	k=2	k=3
Linier (m=1)	$\xi = 4$ GCV = 2.373	$\xi_1= 1; \xi_2 = 5$ GCV=2.368	$\xi_1=1; \xi_2 =5; \xi_3=35$ GCV = 2.373
Kuadratik (m=2)	$\xi = 6$ GCV =2,374	$\xi_1= 1; \xi_2 =7$ GCV=2.377	$\xi_1=2; \xi_2=3; \xi_3=5$ GCV = 2.380
Kubik (m=3)	$\xi = 8$ GCV =2.379	$\xi_1= 9; \xi_2 =35$ GCV=2,382	$\xi_1=1; \xi_2=11; \xi_3=30$ GCV = 2.386

Dapat dilihat bahwa GCV minimum (GCV=2.368) diperoleh pada model spline linier dengan 2 titik knot, yaitu pada umur 1 dan 5 bulan. Hal ini berarti bahwa berat badan anak berusia di bawah 3 tahun di Kota Padang mengalami perubahan pola pertumbuhan pada usia 1 dan 5 bulan, dimana pada masing-masing segmen umur tersebut (umur 0-1 bulan, 1-5 bulan dan 5-30 bulan) pertumbuhan berat badan tersebut dapat dimodelkan secara linier

Model spline linier yang terbentuk adalah:

$$y = 4,428 + 0,5540t - 0,01779(t-1)^+ + 0,000338(t-5)^+$$

dengan y = berat badan batita, t adalah umur. Dalam cara lain, model ini dapat dinyatakan dalam bentuk:

$$y = \begin{cases} 4,428 + 0,5540t & t < 1 \\ 4,4458 + 0,5362t & 1 \leq t < 1 \\ 4,4475 + 0,5366t & t \geq 5 \end{cases}$$

Koefisien determinasi dari model ini adalah 94,87% yang berarti bahwa 94,87% dari total keragaman berat badan anak dapat dijelaskan oleh model tersebut. Dengan nilai koefisien determinasi tersebut, dapat dikatakan bahwa model ini sangat baik dalam menggambarkan pola pertumbuhan berat badan batita di Kota Padang.

Selanjutnya akan dilakukan pengujian terhadap model regresi yang terbentuk. Pada tahap awal akan dilakukan pengujian terhadap keberartian model regresi. Pengujian dilakukan dengan hipotesis :

$$H_0 : \theta_1 = \eta_1 = \eta_2 = 0$$

$$H_1 : \theta_1 \neq 0 \text{ atau } \eta_j \neq 0$$

Pada pengujian tersebut, diperoleh nilai F-hitung sebesar 203,50 dengan nilai signifikansi sebesar 0,000. Dengan mengambil taraf nyata $\alpha=0,05$ diputuskan untuk menolak H_0 yang terdapat paling tidak satu parameter model yang signifikan dalam memodelkan berat badan anak bawah tiga tahun di Kota Padang.

Selanjutnya, dilakukan pengujian individual pada taraf nyata 5% untuk memeriksa parameter model mana yang signifikan. Pengujian dilakukan dengan hipotesis:

$$H_0 : \theta_1 = 0$$

$$H_1 : \theta_1 \neq 0$$

untuk $j = 1,2$ hipotesis yang diuji adalah:

$$H_0 : \eta_j = 0$$

$$H_1 : \eta_j \neq 0$$

Hasil pengujian individual adalah sebagaimana tersaji pada Tabel 2 berikut.

Tabel 2. Hasil Pengujian Individual

Variabel	Koef	Standar Error	t_{hit}	sig
Umur(t)	0,55400	0,079800	6,94	0,00
$(t-1)^+$	-0,01978	0,004480	-3,97	0,00
$(t-5)^+$	0,00034	0,000098	3,47	0,00

Dari Tabel 2, diperoleh nilai $\text{sig} < \alpha=0.05$, untuk semua variabel. Dengan demikian, diputuskan untuk menolak H_0 dan disimpulkan bahwa semua parameter model signifikan dalam memodelkan pertumbuhan berat badan batita di Kota Padang. Dengan demikian, model tersebut dapat digunakan untuk memodelkan pertumbuhan anak di bawah tiga tahun di Kota Padang.

Jika pemodelan dilakukan dengan menggunakan model polynomial yang sama-sama berderajat 1 namun diperoleh dengan pendekatan analisis regresi linier biasa, akan didapat model

$$y = 6,269 + 0,17418t$$

dengan koefisien determinasi $R^2 = 89\%$. Jadi sekitar 89% dari keragaman berat badan batita dapat dijelaskan oleh model tersebut

Jika kedua pendekatan ini dibandingkan pada kasus ini, diketahui bahwa model yang diperoleh dengan pendekatan regresi spline ini lebih baik, karena menghasilkan koefisien determinasi yang lebih tinggi. Berarti bahwa model dengan pendekatan regresi spline ini dapat menjelaskan pertumbuhan berat badan batita di Kota Padang dengan lebih baik.

KESIMPULAN

Analisis regresi spline adalah pendekatan nonparametrik untuk memodelkan hubungan antara variabel bebas dan variabel tak bebas. Dengan pendekatan ini, selang variabel bebas terlebih dahulu dibagi menjadi beberapa segmen, selanjutnya sebuah fungsi polinomial berorde m dibentuk pada masing-masing segmen dengan suatu cara yang menjamin kekontinuan fungsi di sepanjang selang variabel bebasnya.

Pada tulisan ini, analisis regresi spline digunakan untuk memodelkan kurva pertumbuhan anak berusia di bawah tiga tahun (batita) di Kota Padang. Dengan pendekatan tersebut, diketahui bahwa terjadi perubahan pola pertumbuhan berat badan batita pada usia

1 dan 5 bulan. Model yang dihasilkan adalah

$$y = \begin{cases} 4,428 + 0,5540t & t < 1 \\ 4,4458 + 0,5362t & 1 \leq t < 5 \\ 4,4475 + 0,5366t & t \geq 5 \end{cases}$$

dengan koefisien determinasi sebesar $R^2=94,87\%$. Jika dibandingkan dengan model berorde sama yang dibentuk dengan analisis regresi kuadrat terkecil biasa pada kasus ini, diketahui bahwa pendekatan ini menghasilkan model yang relatif lebih baik, karena menghasilkan koefisien determinasi yang lebih tinggi.

DAFTAR KEPUSTAKAAN

- Budiantara, I. N. (1999). Estimator Spline Terbobot dalam Regresi Semiparametrik. *Majalah Ilmu Pengetahuan Dan Teknologi*, 10, 103–109.
- Butchon, R., & Behav, T. L. (2017). The Development and Growth of Children Aged under 5 years in Northeastern Thailand: a Cross-Sectional Study. *J Child Adolesc*, 5(1).
- Draper, N. R., & Smith, H. (1992). *Applied Regression Analysis (terjemahan)*. Jakarta: PT. Gramedia Utama.
- Eubank, R. L. (1988). *Nonparametric regression and spline smoothing 2nd edition*. New York: Marcel Dekker.
- Garner, P., Panpanich, R., & Logan, S. (2000). Is routine growth monitoring eVective? A systematic review of trials. *Arch Dis Child*, 82(3), 197–201.
- Myers, R. H. (1990). *Classical and modern regression with applications*. Boston: PWS and Kent Publishing Company.
- Ryan, T. P. (1997). *Modern Regression Method*. New Jersey: Wiley.
- Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia: Pennsylvania.