ISBN: 978-979-097-142-4

UJI HIPOTESIS DALAM REGRESI NONPARAMETRIK SPLINE

Stefanus Notan Tupen [1], I Nyoman Budiantara [2]

[1]Mahasiswa Magister Jurusan Statistika ITS

[2]Dosen Jurusan Statistika ITS

Abstrak

Diberikan model regresi $y_i = f(x_i) + \varepsilon_i,\quad i = 1, 2, \ldots, n$ dengan $f(x_i)$ merupakan kurva regresi. Kurva regresi f dihampiri dengan fungsi spline, sehingga diperoleh regresi spline $y_i = S(x_i) + \varepsilon_i,$ dengan $S(x_i)$ adalah fungsi spline. Estimasi kurva regresi diperoleh dari optimasi *Weighted Least Square* (*WLS*). Sedangkan pemilihan titik knot menggunakan metode *Generalized Cross Validation* (*GCV*). Inferensi statistik khususnya uji hipotesis untuk kurva f dengan pendekatan spline dapat dilakukan dengan metode *Likelihood RatioTest* (*LRT*). Estimator diperoleh dari membandingkan fungsi *likelihood* dibawah populasi dan fungsi *likelihood* dibawah H_0. Selanjutnya uji hipotesis yang diperoleh dengan spline diaplikasikan pada data berat badan dan umur balita di Jawa Timur.

Kata Kunci: Regresi spline, *Weighted Least Square*, *GCV*, Uji hipotesis

1. Pendahuluan

Dalam regresi parametrik bentuk kurva regresi diasumsikan diketahui, untuk dapat menggunakan pendekatan regresi parametrik, diperlukan pengetahuan masa lalu tentang karakteristik data yang akan diselidiki. Berbeda dengan pendekatan regresi nonparametrik, dalam regresi nonparametrik bentuk kurva regresi diasumsikan tidak diketahui. Kurva regresi nonparametrik hanya diasumsikan *smooth* (mulus) dalam arti termuat di dalam suatu ruang fungsi tertentu. Data diharapkan mencari sendiri bentuk estimasinya, tanpa dipengaruhi oleh faktor subyektifitas dari perancang penelitian. Salah satu regresi nonparametrik yang penting dan mempunyai sifat lokal,

ISBN: 978-979-097-142-4

osilasi rendah dan *smooth* adalah Spline (Agarwal dan Studen, 1980). Dengan demikian, pendekatan regresi nonparametrik memiliki fleksibilitas yang tinggi (Eubank,1988).

Penelitian yang menyelidiki tentang pengujian hipotesis dalam model spline *truncated*, belum pernah ada. Oleh karena itu dalam penelitian ini akan diturunkan pengujian hipotesis untuk model spline *truncated*. Untuk mendapatkan estimasi kurva regresi spline *truncated* digunakan metode *Weighted Least Square* (WLS). Selanjutnya hasil penurunan yang diperoleh diaplikasikan pada data pertumbuhan balita di Jawa Timur.

2. Tinjauan Pustaka

2.1. Regresi Nonparametrik

Regresi nonparametrik merupakan suatu metode Statistika yang digunakan untuk mengetahui hubungan antara variabel respon dan prediktor yang tidak diketahui bentuk fungsinya, hanya diasumsikan fungsi *smooth* (mulus) dalam arti termuat dalam suatu ruang fungsi tertentu, sehingga regresi nonparametrik memiliki fleksibilitas yang tinggi (Eubank, 1988).

Model regresi nonparametrik secara umum dapat disajikan sebagai berikut (Eubank, 1988):

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

2.2. Estimasi Titik Untuk Kurva Regresi Spline

Diberikan suatu basis untuk ruang Spline berorde *m* (Budiantara,2001) dengan bentuk:

$$\{1, x, \ldots, x^m, (x-\lambda_1)_+^m, \ldots, (x-\lambda_K)_+^m\},$$

dengan fungsi *truncated* sebagai berikut:

ISBN: 978-979-097-142-4

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

Untuk setiap fungsi f dalam ruang Spline dapat dinyatakan menjadi:

$$f(x_i) = \sum_{j=0}^{m} \gamma_j x_i^j + \sum_{k=1}^{K} \gamma_{k+m} (x_i - \lambda_k)_+^m$$

Dengan $\gamma_j, \quad j = 0, 1, \ldots, m, m+1, \ldots, m+K$

Model regresi spline dapat ditulis menjadi:

$$y_i = f(x_i) + \varepsilon_i$$

$$= \sum_{j=0}^{m} \gamma_j x_i^j + \sum_{k=1}^{K} \gamma_{k+m} (x_i - \lambda_k)_+^m + \varepsilon_i$$

Apabila diasumsikan sesatan random ε_i berdistribusi normal independen dengan mean nol dan varians σ^2, maka y_i juga berdistribusi normal dengan mean $f(x_i)$ dan varians σ^2 akibatnya diperoleh fungsi *likelihood*:

$$L(y, f) = \prod_{i=1}^{n} \left((2\pi\sigma^2)^{-1/2} Exp(-\frac{1}{2\sigma^2}(y_i - f(x_i))^2) \right)$$

2.3. Pengujian Hipotesis

Diberikan model regresi:

$$y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i \,.$$

Uji hipotesis dapat dilakukan dengan menggunakan metode *Likelihood Ratio Test* (Srivastava,1994).

Prosedur uji hipotesis parameter adalah:

$$H_0 : \mathbf{C}\underset{\sim}{\beta} = \underset{\sim}{\alpha} \text{ lawan } H_1 : \mathbf{C}\underset{\sim}{\beta} \neq \underset{\sim}{\alpha}$$

Statistik penguji untuk hipotesis H_0 lawan H_1 diperoleh dari menyelesaikan rasio:

ISBN: 978-979-097-142-4

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})}.$$

2.4. Pertumbuhan Balita

Pertumbuhan adalah bertambahnya ukuran dan jumlah sel serta jaringan interseluler, yang berarti bertambahnya ukuran fisik dan struktur tubuh dalam arti sebagian atau keseluruhan (Narendra, dkk.,2002). Pertumbuhan bersifat kuantitatif dan dapat diukur dengan menggunakan satuan panjang (cm, meter), satuan berat (gram, pound, kilogram), keseimbangan metabolik (retensi kalsium dan nitrogen tubuh) dan umur tulang (Soetjiningsih, 1995).

2.5. Berat Badan Balita

Berat badan merupakan hasil peningkatan/penurunan semua jaringan yang ada pada tubuh, antara lain: tulang, otot, lemak, cairan tubuh, dan sebagainya. Berat badan dipakai sebagai indikator yang terbaik pada saat ini untuk mengetahui keadaan gizi dan tumbuh kembang anak. Selain itu, berat badan memiliki beberapa kelebihan yaitu: sensitif terhadap perubahan sedikit saja, pengukurannya objektif dan dapat diulang, dapat menggunakan timbangan apa saja yang relatif murah, mudah, dan tidak memerlukan banyak waktu. Pengukuran berat badan dapat dilakukan dengan tepat menggunakan timbangan elektronik, ketika balita dalam keadaan telanjang atau dengan memakai pakaian dalam saja.

2.6.Berat Badan Menurut Umur

Berat badan adalah salah satu parameter yang memberikan gambaran massa tubuh (Supariasa, dkk., 2002). Massa tubuh sangat sensitif terhadap perubahan-perubahan yang mendadak, misalnya terkena penyakit infeksi, menurunnya nafsu makan atau menurunnya jumlah makanan yan dikonsumsi. Berat badan adalah ukuran antopometri yang sangat labil. Dalam keadaan normal, dimana keadaan kesehatan baik dan keseimbangan antara konsumsi dan kebutuhan gizi terjamin, maka berat badan berkembang mengikuti pertambahan umur.

3. Metodologi

Data yang digunakan pada penelitian ini adalah data tentang berat badan anak balita usia 0- 60 bulan yang berasal dari Dinas Kesehatan Jawa Timur tahun 2009. Variabel yang digunakan dalam penelitian ini adalah berat badan menurut umur. Variabel prediktor (*x*) yang digunakan adalah usia anak balita 0-60 bulan sedangkan variabel respon (*y*) adalah berat badan.

Langkah-langkah Analisis

1. Mengkaji estimasi spline dalam regresi nonparametrik dengan langkah-langkah:
   a. Membuat model $y_i = f(x_i) + \varepsilon_i$
   b. Membuat pendekatan fungsi f dengan model spline:
   $$f(x) = \sum_{j=0}^{m} \gamma_j x^j + \sum_{k=1}^{K} \gamma_{k+m} (x-\lambda_k)_+^m$$
   c. Membuat model regresi spline: $y_i = \sum_{j=0}^{m} \gamma_j x_i^j + \sum_{k=1}^{K} \gamma_{k+m} (x_i-\lambda_k)_+^m + \varepsilon_i$
   d. Menyajikan model regresi spline dalam bentuk: $\underset{\sim}{y} = B\underset{\sim}{\gamma} + \underset{\sim}{\varepsilon}$
   e. Menyelesaikan optimasi *WLS* yang meminimumkan:
   $$\underset{\sim}{\varepsilon}^T V^{-1} \underset{\sim}{\varepsilon} = \left\{ (\underset{\sim}{y} - B\underset{\sim}{\gamma})^T V^{-1} (\underset{\sim}{y} - B\underset{\sim}{\gamma}) \right\}$$
2. Menguji hipotesis untuk regresi spline dengan langkah-langkah:
   a. Merumuskan uji hipotesis untuk parameter:
   $$H_0 : \mathbf{C}\underset{\sim}{\gamma} = \underset{\sim}{\alpha}$$
   $$H_1 : \mathbf{C}\underset{\sim}{\gamma} \neq \underset{\sim}{\alpha}$$
   b. Membuat fungsi *likelihood* dibawah Ω ruang parameter populasi: $L(\Omega)$.
   c. Membuat fungsi *likelihood* dibawah H_0: $L(\omega)$.
   d. Membuat rasio *likelihood* Hipotesis: $\lambda = \dfrac{L(\hat{\omega})}{L(\hat{\Omega})}$.

ISBN: 978-979-097-142-4

e. Menentukan daerah penolakan hipotesis H_0:

3. Menerapkan model spline untuk estimasi pola hubungan berat badan dan usia anak balita
   a. Membuat *scatter plot* antara usia anak balita (x) dan berat badan (y) untuk mengetahui hubungan antara kedua variabel.
   b. Memodelkan berat badan dan usia anak balita dengan menggunakan spline linear, spline kuadratik, dan spline kubik dengan menggunakan satu titik *knots,* dua titik *knots,* dan tiga titik *knots*.
   c. Memilih model spline terbaik dengan memilih titik *knots* optimum dilihat dari nilai *GCV* yang paling minimum.
   d. Berdasarkan model spline yang terbaik langkah berikutnya adalah menguji signifikansi parameter model untuk parameter fungsi polinomial dan fungsi potongan (*truncated)*.
   e. Melakukan pengujian normalitas.
   f. Menghitung nilai koefisien determinasi (R^2).

4. Hasil dan Pembahasan

4.1. Estimasi Titik Untuk Kurva Regresi *f*

Diberikan suatu basis untuk ruang Spline berorde m (Budiantara,2001(b)) dengan bentuk:

$$\{1, x, ..., x^m, (x-\lambda_1)_+^m, ..., (x-\lambda_K)_+^m\},$$

dengan fungsi *truncated* sebagai berikut:

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

dan $\lambda_1, ..., \lambda_K$ merupakan titik-titik knots

Untuk setiap fungsi f dalam ruang Spline dapat dinyatakan menjadi:

$$f(x) = \sum_{j=0}^{m} \gamma_j x^j + \sum_{k=1}^{K} \gamma_{k+m} (x-\lambda_k)_+^m$$

ISBN: 978-979-097-142-4

Dengan γ_j, $j = 0,1,...,m,m+1,...,m+K$, merupakan konstanta yang bernilai real. Model regresi spline dapat ditulis menjadi:

$$y_i = f(x_i) + \varepsilon_i$$

$$= \sum_{j=0}^{m} \gamma_j x_i^j + \sum_{k=1}^{K} \gamma_{k+m} (x_i - \lambda_k)_+^m + \varepsilon_i$$

Dari regresi spline ini dapat ditulis:

$$\varepsilon_i = y_i - \sum_{j=0}^{m} \gamma_j x_i^j - \sum_{k=1}^{K} \gamma_{k+m} (x_i - \lambda_k)_+^m \text{ untuk setiap } i = 1,2,\cdots,n$$

Jika persamaan di atas dinyatakan dalam bentuk matriks, maka diperoleh:

$$\underset{\sim}{\varepsilon} = \underset{\sim}{y} - \mathbf{X}(x,\lambda)\underset{\sim}{\gamma}$$

Selanjutnya dibentuk suatu fungsi:

$$Q(\underset{\sim}{\gamma}) = \underset{\sim}{\varepsilon}' V^{-1} \underset{\sim}{\varepsilon}$$

Dengan $\underset{\sim}{\gamma} = \left(\gamma_0, \gamma_1, ..., \gamma_p, \gamma_{p+1}, ..., \gamma_{p+K}\right)'$, $\underset{\sim}{y} = \left(y_1, y_2 ..., y_n\right)'$, dan $\mathbf{X}(x,\lambda)$ matriks berukuran nx(m+K+1), diberikan oleh:

$$\mathbf{X}(x,\lambda) = \begin{pmatrix} 1 & x_1 & \cdots & x_1^m & (x_1-\lambda_1)_+^m & \cdots & (x_1-\lambda_K)_+^m \\ 1 & x_2 & \cdots & x_2^m & (x_2-\lambda_1)_+^m & \cdots & (x_2-\lambda_K)_+^m \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_n & \cdots & x_n^m & (x_n-\lambda_1)_+^m & \cdots & (x_n-\lambda_K)_+^m \end{pmatrix}$$

$$\frac{\partial Q(\underset{\sim}{\gamma})}{\partial \underset{\sim}{\boldsymbol{\gamma}}} = \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\gamma}}}\left[\left(\underset{\sim}{y} - \mathbf{X}(x,\lambda)\underset{\sim}{\gamma}\right)' \mathbf{V}^{-1} \left(\underset{\sim}{y} - \mathbf{X}(x,\lambda)\underset{\sim}{\gamma}\right)\right]$$

$$\hat{\underset{\sim}{\gamma}} = (\mathbf{X}'(x,\lambda)\mathbf{V}^{-1}\mathbf{X}(x,\lambda))^{-1}\mathbf{X}'(x,\lambda)\mathbf{V}^{-1}\underset{\sim}{y}$$

mengingat $\mathbf{X}(x,\lambda)$ merupakan matriks dengan *rank* penuh, maka diperoleh estimasi $\underset{\sim}{\gamma}$ adalah :

$$\hat{\underset{\sim}{\gamma}} = (\mathbf{X}'(x,\lambda)\mathbf{V}^{-1}\mathbf{X}(x,\lambda))^{-1}\mathbf{X}'(x,\lambda)\mathbf{V}^{-1}\underset{\sim}{y}$$

Estimator kurva regresi $f(x)$ diberikan oleh:

$$\begin{aligned}\hat{f}(x,\lambda) &= \mathbf{X}(x,\lambda)\left[\mathbf{X}'(x,\lambda)\mathbf{V}^{-1}\mathbf{X}(x,\lambda)\right]^{-1}\mathbf{X}'(x,\lambda)\mathbf{V}^{-1}\underset{\sim}{y} \\ &= B(x,\lambda)\,y\end{aligned}$$

4.2. Uji Hipotesis

Diberikan model regresi spline:

$$y_i = \sum_{j=0}^{m}\gamma_j x_i^{\,j} + \sum_{k=1}^{K}\gamma_{k+m}(x_i-\lambda_k)_+^m + \varepsilon_i, \quad \text{dengan } \varepsilon_i \sim N(0,\sigma^2)$$

Untuk menurunkan uji hipotesis H_0 lawan H_1 dapat menggunakan metode LRT. Perhatikan model regresi spline, dengan ε_i berdistribusi independen identik $N(0,\sigma^2)$.

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

dengan

$$f(x_i) = \sum_{j=0}^{m}\gamma_j x^j + \sum_{k=1}^{K}\gamma_{k+m}(x-\lambda_k)_+^m$$

$$= \gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + \cdots + \gamma_m x_i^m + \gamma_{m+1}(x_i-\lambda_1)_+^m + \cdots + \gamma_{m+k}(x_i-\lambda_K)_+^m + \varepsilon_i$$

Karena $\varepsilon_i \sim N(0,\sigma^2)$ maka $y_i \sim N(f(x_i),\sigma^2)$ fungsi *likelihood* diberikan oleh:

$$L(\underset{\sim}{\gamma},\sigma^2) = \prod_{i=1}^{n}\left((2\pi\sigma^2)^{-\frac{1}{2}} Exp(-\frac{1}{2\sigma^2}(y_i - f(x_i))^2)\right)$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} Exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - f(x_i))^2$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} Exp(-\frac{1}{2\sigma^2}(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma})'(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}))$$

Dengan $\underset{\sim}{y} = \left(y_1, y_2 ..., y_n\right)'$, $\underset{\sim}{\gamma} = \left(\gamma_0, \gamma_1, ..., \gamma_m, \gamma_{m+1}, ..., \gamma_{m+K+1}\right)'$

Pertama diperhatikan Ruang ω:

Fungsi *likelihood* diberikan oleh:

$$L(\underset{\sim}{\gamma}_\omega, \sigma^2_\omega) = (2\pi\sigma^2_\omega)^{-\frac{n}{2}} Exp(-\frac{1}{2\sigma^2_\omega}(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}_\omega)'(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}_\omega))$$

$$\frac{\partial \log L(\underset{\sim}{\gamma}_\omega, \sigma^2_\omega)}{\partial \underset{\sim}{\gamma}_\omega} = \frac{\partial}{\partial \underset{\sim}{\gamma}_\omega}[-\frac{n}{2}\log(2\pi\sigma^2_\omega) - \frac{1}{2\sigma^2_\omega}(\underset{\sim}{y}'\underset{\sim}{y} - 2\underset{\sim}{\gamma}_\omega'\mathbf{X}'\underset{\sim}{y} + \underset{\sim}{\gamma}_\omega'\mathbf{X}'\mathbf{X}\underset{\sim}{\gamma}_\omega))]$$

$$= 0 - \frac{1}{2\sigma^2_\omega}(0 - 2\mathbf{X}'\underset{\sim}{y} + 2\mathbf{X}'\mathbf{X}\underset{\sim}{\gamma}_\omega) = 0$$

$$\hat{\underset{\sim}{\gamma}}_\omega = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\underset{\sim}{y}$$

$$\frac{\partial \log L(\underset{\sim}{\gamma}_\omega, \sigma^2_\omega)}{\partial \sigma^2_\omega} = \frac{\partial}{\partial \sigma^2_\omega}[-\frac{n}{2}\log(2\pi\sigma^2_\omega) - \frac{1}{2\sigma^2_\omega}((\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}_\omega)'(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}_\omega))]$$

$$\hat{\sigma}^2_\omega = \frac{(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}_\omega)'(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}_\omega)}{n}$$

$$\underset{\omega}{Max}\, L(\gamma_\omega, \sigma^2_\omega) = (2\pi\hat{\sigma}^2_\omega)^{-\frac{n}{2}} Exp\left[-\frac{1}{2\hat{\sigma}^2_\omega}(\underset{\sim}{y} - \mathbf{X}\hat{\underset{\sim}{\gamma}}_\omega)'(\underset{\sim}{y} - \mathbf{X}\hat{\underset{\sim}{\gamma}}_\omega)\right]$$

$$= (2\pi\hat{\sigma}^2_\omega)^{-\frac{n}{2}} e^{-\frac{n}{2}}$$

Selanjutnya diperhatikan Ruang Ω:

Fungsi *likelihood* diberikan oleh:

$$\frac{\partial \log L(\underset{\sim}{\gamma}_\Omega, \sigma^2_\Omega)}{\partial \sigma^2_\Omega} = \frac{\partial}{\partial \sigma^2_\Omega}[-\frac{n}{2}\log(2\pi\sigma^2_\Omega) - \frac{1}{2\sigma^2_\Omega}((\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}_\Omega)'(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}_\Omega))]$$

$$\hat{\sigma}^2_{\Omega} = \frac{(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}_{\Omega})'(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}_{\Omega})}{n}$$

Akibatnya

$$\underset{\Omega}{Max}\ L(\gamma_{\Omega}, \sigma^2_{\Omega}) = (2\pi\hat{\sigma}^2_{\Omega})^{-\frac{n}{2}} Exp\left[-\frac{1}{2\hat{\sigma}^2_{\Omega}} (\underset{\sim}{y} - \mathbf{X}\hat{\underset{\sim}{\gamma}}_{\Omega})'(\underset{\sim}{y} - \mathbf{X}\hat{\underset{\sim}{\gamma}}_{\Omega}) \right]$$

$$= (2\pi\hat{\sigma}^2_{\Omega})^{-\frac{n}{2}} e^{-\frac{n}{2}}$$

Selanjutnya diperoleh *Ratio Likelihood*:

$$\lambda = \frac{\underset{\omega}{Max}\ L(\gamma_{\omega}, \sigma^2_{\omega})}{\underset{\Omega}{Max}\ L(\gamma_{\Omega}, \sigma^2_{\Omega})} = \frac{\hat{\sigma}^2_{\Omega}}{\hat{\sigma}^2_{\omega}}$$

Dengan memperhatikan hipotesis:

$$H_0 : \mathbf{C}\underset{\sim}{\gamma} = \underset{\sim}{\alpha} \text{ atau}$$
$$H_0 : \mathbf{C}\underset{\sim}{\gamma} - \underset{\sim}{\alpha} = 0$$

$$\lambda^{\frac{2}{n}} = \frac{\hat{\sigma}^2_{\Omega}}{\hat{\sigma}^2_{\omega}}$$

$$\hat{\sigma}^2_{\Omega} = \frac{(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}_{\Omega})'(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}_{\Omega})}{n}$$

$$n\hat{\sigma}^2_{\Omega} = (\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}_{\Omega})'(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}_{\Omega}) = S$$

$$\frac{\partial}{\partial \underset{\sim}{\beta}} (S + \underset{\sim}{\beta}'(C\underset{\sim}{\gamma} - \alpha)) = C\underset{\sim}{\gamma} - \underset{\sim}{\alpha}$$

ISBN: 978-979-097-142-4

$$\hat{\underset{\sim}{\gamma}} = \hat{\gamma}_{\Omega} - \frac{1}{2}(\mathbf{X'X})^{-1}\mathbf{C}'\underset{\sim}{\beta}$$

$$\underset{\sim}{\beta} = 2(\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\gamma_{\Omega} - \alpha)$$

$$\hat{\underset{\sim}{\gamma}} = \hat{\gamma}_{\Omega} - (\mathbf{X'X})^{-1}\mathbf{C}'\left((\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\gamma_{\Omega} - \alpha)\right)$$

Sehingga diperoleh:

$$\hat{\sigma}_{\Omega}^2 = \frac{(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}_{\Omega})'(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\gamma}_{\Omega})}{n}$$

$$\hat{\sigma}_{\omega}^2 = \frac{(\underset{\sim}{y} - \mathbf{X}\hat{\underset{\sim}{\gamma}})'(\underset{\sim}{y} - \mathbf{X}\hat{\underset{\sim}{\gamma}})}{n}$$

$$\lambda^{\frac{2}{n}} = \frac{(\underset{\sim}{y} - \mathbf{X}\hat{\underset{\sim}{\gamma}})'(\underset{\sim}{y} - \mathbf{X}\hat{\underset{\sim}{\gamma}}) + (\mathbf{C}\underset{\sim}{\gamma} - \underset{\sim}{\alpha})'[\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{C}']^{-1}(\mathbf{C}\underset{\sim}{\gamma} - \underset{\sim}{\alpha})}{(\underset{\sim}{y} - \mathbf{X}\hat{\underset{\sim}{\gamma}})'(\underset{\sim}{y} - \mathbf{X}\hat{\underset{\sim}{\gamma}})}$$

Statistik: *sum of squares* dari hipotesis (SSH):

$$\frac{SSH}{\sigma^2} = \frac{(\mathbf{C}\hat{\gamma} - \alpha)'[C(X'X)^{-1}C']^{-1}(\mathbf{C}\hat{\gamma} - \alpha)}{\sigma^2} \sim \chi^2(m+K)$$

Dan *sum of squares residual* (SSE):

$$\frac{SSE}{\sigma^2} = \frac{y'(I - \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X}']y}{\sigma^2} \sim \chi^2(n-m-K-1)$$

Distribusi Statistik uji untuk F adalah:

$$F = \frac{\frac{SSH}{m+K}}{\frac{SSE}{n-m-K-1}}$$

$$= \frac{\frac{(\mathbf{C}\hat{\gamma} - \alpha)'[C(X'X)^{-1}C']^{-1}(\mathbf{C}\hat{\gamma} - \alpha)}{m+K}}{\frac{SSE}{n-m-K-1}} \sim F_{(m+K, n-m-K-1)}$$

ISBN: 978-979-097-142-4

Hipotesis H_0 akan ditolak jika dan hanya jika $F \geq F_{(\alpha, m+K, n-m-K-1)}$

ISBN: 978-979-097-142-4

4.3. Aplikasi Uji Hipotesis Model Spline pada Pertumbuhan Balita di Jawa Timur

$$\hat{f}(x)=\begin{cases} 3,819+0,799x-0,031x^2 & ;x<9 \\ 3,819+0,799x-0,031x^2+0,023(x-9)^2 & ;9<x<14 \\ 3,819+0,799x-0,031x^2+0,023(x-9)^2+0,008(x-14)^2 & ;x>14 \end{cases}$$

$$=\begin{cases} 3,819+0,799x-0,031x^2 & ;x<9 \\ 5,682+0,385x-0,008x^2 & ;9<x<14 \\ 7,250+0,161x & ;x>14 \end{cases}$$

Perhatikan uji hipotesis:

$H_0:\mathbf{C}\underset{\sim}{\gamma}=\underset{\sim}{\alpha}$ lawan $H_1:\mathbf{C}\underset{\sim}{\gamma}\neq\underset{\sim}{\alpha}$

Dimana: $C=\begin{pmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{pmatrix}$; $\underset{\sim}{\gamma}=(\gamma_0,\gamma_1,\gamma_2,\gamma_3,\gamma_4)'$,

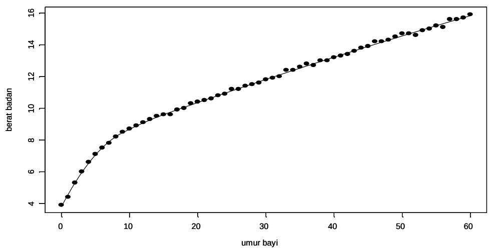
Tabel 1. Analisis Variansi Model Spline Terbobot Kuadrat Dua Knot.

Sumber Variasi	Derajat Bebas	Jumlah Kuadrat	Rata-Rata Jumlah Kuadrat	F-hitung
Regresi	4	644,63	161,16	20144,69
Residual	56	0,448	0,008	
Total	60	645,08		

ISBN: 978-979-097-142-4

Dengan menggunakan tingkat signifikansi 5%, diperoleh nilai distribusi *F* dengan derajat bebas pembilang 4 dan derajat bebas penyebut 56, sebesar 2,536. Berdasarkan Tabel diperoleh kesimpulan menolak H_0 karena nilai $F_{hitung} = 20145 > F_{tabel} = 2{,}536$. Hal ini berarti parameter-parameter $\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4$ signifikan pada model.

Model spline terbobot kuadratik dengan dua titik knot pada umur *x* = 9 bulan dan umur *x* = 14 bulan diberikan dalam Gambar 4.2. Model Spline kuadrat terbobot ini mempunyai koefisien determinasi (R^2) sebesar 99,93%. Nilai R^2 ini menunjukan bahwa model spline terbobot kuadratik dengan 2 titik knot pada umur *x* = 9 bulan dan umur *x* = 14 bulan, sangat layak digunakan sebagai model pola hubungan antara umur dan berat badan balita di Jawa Timur.



5. Kesimpulan

Berdasarkan analisis dan pembahasan yang telah diuraikan pada bab sebelumnya, maka dapat diperoleh kesimpulan:

1. Estimasi parameter model regresi nonparametrik spline dengan menggunakan metode *Weighted Least Square* diperoleh:

$$\hat{\underset{\sim}{\gamma}} = (\mathbf{X}'(x,\lambda)\mathbf{V}^{-1}\mathbf{X}(x,\lambda))^{-1}\mathbf{X}'(x,\lambda)\mathbf{V}^{-1}\underset{\sim}{y}$$

2. Uji hipotesis dalam model regresi nonparametrik spline dapat dilakukan dengan menggunakan *Likelihood Ratio Test* dengan formulasi hipotesis sebagai berikut:

$$H_0 : \mathbf{C}\underset{\sim}{\boldsymbol{\gamma}} = \underset{\sim}{\alpha}$$
$$H_1 : \mathbf{C}\underset{\sim}{\boldsymbol{\gamma}} \neq \underset{\sim}{\alpha}$$

Statistik test untuk uji hipotesis di atas diberikan oleh:

ISBN: 978-979-097-142-4

$$F = \frac{\dfrac{SSH}{m+K}}{\dfrac{SSE}{n-m-K-1}}$$

$$= \frac{\dfrac{(\mathbf{C}\hat{\gamma}-\alpha)'[C(X'X)^{-1}C']^{-1}(\mathbf{C}\hat{\gamma}-\alpha)}{m+K}}{\dfrac{SSE}{n-m-K-1}} \sim F_{(m+K,n-m-K-1)}$$

Hipotesis H_0 ditolak jika $F \geq F_{(\alpha,m+K,n-m-K-1)}$

3. Model spline terbaik untuk pertumbuhan balita di Jawa Timur adalah spline terbobot kuadratik dengan dua titik knot ($x = 9$ dan $x = 14$). Model spline terbobot diberikan oleh:

$$\hat{f}(x) = 3,819 + 0,799x - 0,031x^2 + 0,023(x-9)_+^2 + 0,008(x-14)_+^2$$

$$= \begin{cases} 3,819 + 0,799x - 0,031x^2 & ; x < 9 \\ 5,682 + 0,385x - 0,008x^2 & ; 9 < x < 14 \\ 7,250 + 0,161x & ; x > 14 \end{cases}$$

Untuk penelitian selanjutnya perlu dikaji lagi uji hipotesis dalam model regresi nonparametrik spline, untuk model yang lebih rumit seperti multirespon dan semiparametrik.

Daftar Pustaka

Aritonang, I., 2000, *Pemantauan Pertumbuhan Balita (Petunjuk Praktis Menilai Status Gizi & Kesehatan)*, Kanisius, Yogyakarta.

Budiantara, I.N., 2001, Estimasi Parametrik dan Nonparametrik untuk Pendekatan Kurva Regresi, *Makalah Pembicara Utama pada Seminar Nasional Statistika V, Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS)*, Surabaya.

ISBN: 978-979-097-142-4

__________, 2006, Model Spline Dengan Knots Optimal, *Jurnal Ilmu Dasar, FMIPA Universitas Jember*,7,77-85.

__________, 2008, Inferensi Statistik Untuk Model Spline, *Jurnal Matematika dan Statistika Universitas Bina Nusantara*, Jakarta

Drapper, N.R , Smith, H., 1996, *Applied Regression Analysis, 2nd edition*, John Wiley & Sons, Chapman and Hall, New York.

Eubank, R.L., 1988, *Spline Smoothing and Nonparametric Regression*, Mercel Dekker, New York.

Hardle, W., 1990, *Applied Nonparametric Regression*, Cambridge University Press, New York.

Khair, A., Budiantara, I.N., dan Fitriasari, K., 2006, Spline Polinomial Truncated untuk Interval Konfidensi Kurva Regresi Nonparametrik, *Prosiding Seminar Nasional Statistika VII*, ITS, Surabaya.

Muni, S. dan Sen, A., 1994, *Regression Ananysis, Theory, Method, and Applications*, Springer-Verlag, New York.

Rencher, A.C., 2000, *Linear Models in Statistics*, John Wiley & Sons, Chapman and Hall, New York.

Syaranamual, R.D., 2011, *Interval Konfidensi Spline Kuadrat dengan Pendekatan Pivotal Quantity*, Draft Tesis, Jurusan Statistika ITS.

Wahba, G., 1983, Bayesian Confidence Interval for the Cross Validated Smoothing Parameter in the Generalized Spline Smoothing Problems, *The Annals of Statistics*, 13, 1378-1402.

Wahba, G., 1990, *Spline Models for Observasional Data*, SIAM Pensylvania.