

ESTIMASI KURVA REGRESI SEMIPARAMETRIK PADA DATA LONGITUDINAL BERDASARKAN ESTIMATOR POLINOMIAL LOKAL

Tiani Wahyu Utami

¹Program Studi S1 Statistika Universitas Muhammadiyah Semarang, Jl. Kedung Mundu Raya no 18 Semarang;

Email: tiani.utami88@gmail.com

ABSTRAK

Diberikan model regresi semiparametrik untuk data longitudinal $\mathbf{y} = \mathbf{T}\boldsymbol{\alpha} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}$, dengan komponen parametrik $\mathbf{T}\boldsymbol{\alpha}$ dan komponen nonparametrik $\boldsymbol{\eta}$ yang didekati dengan Polinomial Lokal. Estimator Polinomial Lokal diperoleh dengan metode WLS (*Weighted Least Square*). Estimasi model regresi semiparametrik pada data longitudinal adalah $\hat{\mathbf{y}} = \mathbf{A}^*(\mathbf{h})\mathbf{y}$, dengan $\mathbf{A}^*(\mathbf{h}) = \mathbf{T}[\mathbf{T}^T(\mathbf{I} - \mathbf{A})^T(\mathbf{I} - \mathbf{A})\mathbf{T}]^{-1}\mathbf{T}^T(\mathbf{I} - \mathbf{A})^T(\mathbf{I} - \mathbf{A}) + \mathbf{A}(\mathbf{I} - \mathbf{T}[\mathbf{T}^T(\mathbf{I} - \mathbf{A})^T(\mathbf{I} - \mathbf{A})\mathbf{T}]^{-1}\mathbf{T}^T(\mathbf{I} - \mathbf{A})^T(\mathbf{I} - \mathbf{A}))$, $\mathbf{A} = \mathbf{C}(\mathbf{C}^T\mathbf{K}_h\mathbf{C})^{-1}\mathbf{C}^T\mathbf{K}_h$. Estimator Polinomial Lokal sangat tergantung pada *bandwidth* (h) optimal. Penentuan *bandwidth* optimal dengan menggunakan metode GCV (*Generalized Cross Validation*). Selanjutnya model regresi semiparametrik Polinomial Lokal pada data longitudinal diaplikasikan untuk memodelkan hubungan pengaruh antara kadar trombosit penderita Demam Berdarah Dengue terhadap kadar hematrokrit dan waktu pemeriksaan, dimana kadar hematrokrit sebagai komponen nonparametrik dan waktu pemeriksaan sebagai komponen parametrik. Hasil estimasi menunjukkan bahwa waktu pemeriksaan berpola kuadratik untuk setiap subjek sedangkan kadar hematrokrit pada subjek 1 mengikuti pola polinomial lokal berorde 1, sedangkan kadar hematrokrit pada subjek 2 mengikuti pola polinomial lokal berorde 4 dan pada subjek 3 kadar hematrokrit mengikuti pola polinomial lokal berorde 4. Model ini mempunyai nilai MSE sebesar 146.7636 dan koefisien determinasi $R^2 = 93,9249\%$.

Kata kunci : *Data Longitudinal, Estimator Polinomial Lokal, GCV, Regresi Semiparametrik, WLS.*

PENDAHULUAN

Analisis regresi merupakan metode statistika yang telah lama dikembangkan untuk menyelidiki pola hubungan dan pengaruh variabel prediktor terhadap variabel respon, melalui estimasi kurva regresinya. Berkaitan dengan pengestimasi kurva regresi, terdapat tiga model regresi yang dapat digunakan yaitu model regresi parametrik, model regresi nonparametrik, dan model regresi semiparametrik [2]. Dalam beberapa kasus, variabel respon diketahui pola hubungannya dengan salah satu variabel

prediktor, tetapi dengan variabel prediktor yang lain tidak diketahui bentuk pola hubungannya. Dalam keadaan seperti ini, maka digunakan pendekatan regresi semiparametrik [1].

Dalam regresi, terdapat jenis data longitudinal yang pengamatan dilakukan sebanyak n subjek saling independen dengan setiap subjek diamati secara berulang dalam kurun waktu yang berbeda [3]. Data longitudinal memiliki kelebihan diantaranya dalam jumlah subjek yang sama, hasil pengukuran error menghasilkan penaksir efek perlakuan yang lebih efisien dikarenakan

pada data longitudinal dilakukan estimasi untuk setiap pengamatan dan lebih *powerfull* walaupun hanya menggunakan subjek yang lebih sedikit [6].

Pengamatan longitudinal diberikan oleh $\{(y_{ij}, x_{ij}, t_{ij})\}; i = 1, 2, \dots, n; j = 1, 2, \dots, m$. Hubungan antara x_{ij} , t_{ij} dan y_{ij} mengikuti model regresi semiparametrik. Bentuk kurva regresi yang tidak diketahui diasumsikan *smooth* atau mulus. Estimator kurva regresi diperoleh dengan mengestimasi parameternya [7]. Terdapat beberapa pendekatan untuk mengestimasi kurva regresi salah satunya adalah dengan estimator Polinomial Lokal. Salah satu kelebihan estimator Polinomial Lokal adalah dapat mengurangi asimtotik bias dan menghasilkan estimasi yang baik [5]. Estimator Polinomial Lokal dapat diperoleh dengan optimasi WLS (*Weighted Least Square*). Sedangkan untuk mengestimasi parameter penghalus (*bandwidth*) dengan metode *Generalized Cross Validation (GCV)*.

Penelitian sebelumnya telah banyak dikembangkan mengenai data longitudinal yang menggunakan pendekatan spline untuk estimasi kurva regresi nonparametrik [11], maupun model regresi semiparametrik [8]. Estimator Polinomial Lokal dalam mengestimasi model regresi nonparametrik diteliti pertama kali menggunakan data *cross-section* [4]. Selain itu telah dilakukan juga penelitian tentang estimasi selang kepercayaan kurva regresi nonparametrik dengan error lognormal berdasarkan estimator Polinomial Lokal, akan tetapi dalam penelitiannya menggunakan data *cross-section* [9].

Penyakit Demam Berdarah Dengue atau *Dengue Haemorrhagic Fever (DHF)* adalah penyakit yang disebabkan oleh virus dengue yang ditularkan melalui gigitan nyamuk *Aedes aegypti* dan *Aedes albopictus*. Salah satu kriteria

laboratorium non spesifik untuk menegakkan diagnosis DBD yang ditetapkan oleh WHO adalah adanya Trombositopenia (trombosit < 100.000/ml). Menurut ahli penyakit infeksi tropis yang banyak meneliti tentang demam berdarah, dr. Leonard Nainggolan, Sp.PD-KPT, penderita DBD jumlah trombosit menjadi rendah atau kurang dari 100.000 per mm^3 merupakan akibat adanya kebocoran plasma di pembuluh darah kapiler sehingga tubuh berupaya menutup celah itu dengan bantuan trombosit. Kebocoran itu menyebabkan darah menjadi pekat yang dapat dilihat dari naiknya kadar hematokrit.

Pada tulisan ini akan dibahas bagaimana mengestimasi kurva regresi semiparametrik pada data longitudinal menggunakan estimator Polinomial Lokal. Selanjutnya menerapkan model regresi tersebut pada data kadar trombosit penderita DBD selama dirawat di Rumah Sakit Haji tahun 2011 berdasarkan waktu pemeriksaan, dan kadar hematokrit pasien.

METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari rekam medik Rumah Sakit Haji Surabaya tahun 2011. Data tersebut mengenai kadar trombosit penderita DBD (Demam Berdarah Dengue) yang dipengaruhi oleh kadar hematokrit dan waktu pemeriksaan selama dirawat di rumah sakit. Banyak subjek dari penelitian ini ada tiga subjek yang masing-masing disebut subjek 1 untuk penderita DBD derajat/grade I, subjek 2 untuk penderita DBD derajat/grade II dan subjek 3 untuk penderita DBD derajat/grade III. Masing-masing subjek dilakukan pengukuran berulang sebanyak 7 kali pengamatan.

Metodologi yang dilakukan pada penelitian ini adalah

1. Mengestimasi kurva regresi semiparametrik untuk data longitudinal berdasarkan estimator Polinomial Lokal dengan langkah-langkah sebagai berikut :
 - a. Diberikan data observasi (y_{ij}, x_{ij}, t_{ij}) , yang memenuhi model regresi semiparametrik.
 - b. Menyatakan model dalam bentuk matriks $\mathbf{y}^* = \boldsymbol{\eta} + \boldsymbol{\varepsilon}$ dengan $\mathbf{y}^* = \mathbf{y} - \mathbf{T}\boldsymbol{\alpha}$.
 - c. Fungsi $y(x_{ij})$ dapat didekati oleh polinomial berderajat p .
 - d. Mengestimasi $\boldsymbol{\beta}$ dengan meminimumkan kriteria *Weighted Least Square*.
 - e. Mendapatkan bentuk matriks \mathbf{A} berukuran $N \times N$ dengan cara menyelesaikan persamaan $\hat{\boldsymbol{\eta}} = \mathbf{A}\mathbf{y}^*$, atau $\hat{\boldsymbol{\eta}} = \mathbf{A}(\mathbf{y} - \mathbf{T}\boldsymbol{\alpha})$.
 - f. Mensubstitusikan $\hat{\boldsymbol{\eta}}$ untuk memperoleh estimasi $\boldsymbol{\alpha}$ yaitu $\hat{\boldsymbol{\alpha}}$.
2. Menerapkan hasil estimasi untuk menganalisis pengaruh kadar trombosit penderita DBD terhadap waktu pemeriksaan dan kadar hematokrit dengan langkah sebagai berikut :
 - a. Membuat plot data berpasangan (y_{ij}, x_{ij}, t_{ij}) , $i = 1, 2, 3; j = 1, 2, \dots, 7$.
 - b. Menentukan matriks $\mathbf{A}(h)$ berukuran $N \times N$.
 - c. Memilih orde polinomial p dan nilai *bandwidth* optimal yang meminimumkan $GCV(h)$.
 - d. Memodelkan orde polinomial p dan nilai *bandwidth* optimal dari langkah c secara simultan.
 - e. Menghitung nilai $MSE(h)$.

HASIL DAN PEMBAHASAN

Estimasi Kurva Regresi Semiparametrik pada Data Longitudinal Berdasarkan Estimator Polinomial Lokal

Diberikan data longitudinal sebanyak n subjek yang setiap subjek diukur

sebanyak m , $\{(y_{ij}, x_{ij}, t_{ij}); i = 1, 2, \dots, n; j = 1, 2, \dots, m\}$ dengan y_{ij} menyatakan variabel respon pengamatan untuk subjek ke- i pada waktu ke- j , x_{ij} dan t_{ij} menyatakan variabel prediktor. Hubungan y_{ij} dengan x_{ij} tidak diketahui bentuk fungsinya, sedangkan antara y_{ij} dengan t_{ij} bentuk fungsinya diketahui, maka kita dapat memodelkan hubungan antara x_{ij}, t_{ij} dan y_{ij} menggunakan model semiparametrik seperti pada persamaan (1). Model tersebut dapat ditulis dalam bentuk :

$$\mathbf{y} = \mathbf{T}\boldsymbol{\alpha} + \boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad (1)$$

dengan

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{0i} \\ \alpha_{1i} \\ \alpha_{2i} \\ \vdots \\ \alpha_{qi} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1m} \\ y_{21} \\ \vdots \\ y_{2m} \\ \vdots \\ y_{n1} \\ \vdots \\ y_{nm} \end{bmatrix}, \boldsymbol{\eta} = \begin{bmatrix} \eta(t_{11}) \\ \vdots \\ \eta(t_{1m}) \\ \eta(t_{21}) \\ \vdots \\ \eta(t_{2m}) \\ \vdots \\ \eta(t_{n1}) \\ \vdots \\ \eta(t_{nm}) \end{bmatrix}$$

$$\mathbf{T} = \begin{bmatrix} 1 & t_{11} & t_{11}^2 & \dots & t_{11}^q \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_{1m} & t_{1m}^2 & \dots & t_{1m}^q \\ 1 & t_{21} & t_{21}^2 & \dots & t_{21}^q \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_{2m} & t_{2m}^2 & \dots & t_{2m}^q \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_{n1} & t_{n1}^2 & \dots & t_{n1}^q \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_{nm} & t_{nm}^2 & \dots & t_{nm}^q \end{bmatrix} \text{ dan}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} e_{11} \\ \vdots \\ e_{1m} \\ e_{21} \\ \vdots \\ e_{2m} \\ \vdots \\ e_{n1} \\ \vdots \\ e_{nm} \end{bmatrix}$$

Persamaan (1) mengandung dua bagian, yaitu komponen parametrik $\mathbf{T}\boldsymbol{\alpha}$ dan komponen nonparametrik $\boldsymbol{\eta}$. Asumsikan bahwa nilai $\boldsymbol{\alpha}$ diketahui, sehingga dapat dibentuk persamaan :

$$\mathbf{y}^* = \mathbf{y} - \mathbf{T}\boldsymbol{\alpha}$$

dengan

$$y^* = [y_{11}^* \dots y_{1m}^* \ y_{21}^* \dots y_{2m}^* \dots y_{n1}^* \dots y_{nm}^*]^T.$$

Fungsi $y(x_{ij})$ tidak diketahui bentuknya maka diestimasi menggunakan estimator Polinomial Lokal. Untuk menduga $y(x_{ij})$ pada pengukuran ke- j objek ke- i maka didekati dengan polinomial derajat p

$$y(x_{ij}) = S_{0i} + (x_{ij} - x)S_{1i} + (x_{ij} - x)^2 S_{2i} + \dots + (x_{ij} - x)^p S_{pi} \quad (2)$$

Persamaan (2) dapat ditulis menjadi :

$$\eta = C\beta$$

dengan $C = [S_{0i}, S_{1i}, S_{2i}, \dots, S_{pi}]^T$ dan

$$C = \begin{pmatrix} \mathbf{X}(h_1) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}(h_2) & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}(h_n) \end{pmatrix},$$

$$\mathbf{X}(h_i) = \begin{pmatrix} 1 & (x_{i1} - x) & \dots & (x_{i1} - x)^p \\ 1 & (x_{i2} - x) & \dots & (x_{i2} - x)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (x_{im} - x) & \dots & (x_{im} - x)^p \end{pmatrix}$$

$\hat{\beta}$ didapatkan dengan cara meminimumkan kriteria *Weighted Least Square* (WLS)

$$\phi(\beta) = (y^* - C\beta)^T K_h (y^* - C\beta) \quad (3)$$

dengan $K_h = \text{diag}(K_{1h}, K_{2h}, \dots, K_{nh})$ dan $K_{ih} = \text{diag}(K_h(x_{i1} - x), K_h(x_{i2} - x), \dots, K_h(x_{im} - x))$.

Matriks K_h adalah matriks yang berisi fungsi pembobot, dengan $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$. $K(\cdot)$ adalah fungsi Kernel dan h adalah *bandwidth*.

Nilai dugaan untuk $\hat{\alpha}$ adalah $\hat{\alpha}$ yang bila disubstitusikan ke dalam persamaan (3) akan meminimumkan $\phi(\beta)$ diperoleh :

$$\hat{\beta} = (C^T K_h C)^{-1} C^T K_h y \quad (4)$$

Dengan demikian η dapat dinyatakan sebagai berikut :

$$\eta = Ay = A(y - T\alpha)$$

dengan $A = C(C^T K_h C)^{-1} C^T K_h$.

Selanjutnya dicari $\hat{\alpha}$ dengan meminimumkan *Weighted Least Square* (WLS) sebagai berikut : $\Omega(\alpha) = ((I - A)y - (I - A)T\alpha)^T ((I - A)y - (I - A)T\alpha)$.

Nilai dugaan α adalah $\hat{\alpha}$. Nilai dugaan tersebut dapat diperoleh dengan cara meminimumkan persamaan $\Omega(\alpha)$ sehingga diperoleh

$$\hat{\alpha} = [T^T (I - A)^T (I - A) T]^{-1} T^T (I - A)^T (I - A) y$$

Berdasarkan $\hat{\alpha}$ diperoleh persamaan :

$$T\hat{\alpha} = A_{par} y \quad (5)$$

dengan,

$$A_{par} = T[T^T (I - A)^T (I - A) T]^{-1} T^T (I - A)^T (I - A).$$

Selanjutnya mendapatkan estimator Polinomial Lokal :

$$\hat{\eta} = A(y - T\hat{\alpha}) = A_{nonpar} y \quad (6)$$

dengan, $A_{nonpar} = A(I - A_{par})$.

Berdasarkan persamaan (5) dan (6), maka estimasi model regresi semiparametrik pada persamaan (1) diperoleh :

$$\hat{y} = A^*(h)y \quad (7)$$

dengan, $A^*(h) = A_{par} + A(I - A_{par})$.

Berikut ini diberikan sifat-sifat yang dimiliki oleh estimator Polinomial Lokal. Persamaan (1) dapat diubah kedalam bentuk sebagai berikut :

$$y = \psi(x, t) + \varepsilon$$

$$\text{dengan } \psi(x, t) = \begin{pmatrix} \psi(x_{11}, t_{11}) \\ \psi(x_{12}, t_{12}) \\ \vdots \\ \psi(x_{nm}, t_{nm}) \end{pmatrix},$$

$$\psi(x_{ij}, t_{ij}) = \Gamma_{0i} + t_{ij} \Gamma_{1i} + \dots + t_{ij}^q \alpha_{qi} + \gamma_j(x_{ij}).$$

Estimasi kurva regresi $\psi(x, t)$ adalah

$$\hat{\psi}(x, t) = [T^T [T^T (I - A)^T (I - A) T]^{-1} (T^T (I - A)^T (I - A) y - A) + A[I - T[T^T (I - A)^T (I - A) T]^{-1} T^T (I - A)^T (I - A)] y = H[h]y \quad (8)$$

dengan

$$H[h] = [T[T^T (I - A)^T (I - A) T]^{-1} (T^T (I - A)^T (I - A) + A[I - T[T^T (I - A)^T (I - A) T]^{-1} T^T (I - A)^T (I - A)]]$$

Berdasarkan persamaan (8) terlihat estimator Polinomial Lokal $\hat{\psi}(x, t)$ merupakan estimator linier dalam observasi y . Selanjutnya akan diperlihatkan estimator Polinomial Lokal mempunyai sifat bias untuk kurva regresi $\psi(x, t)$.

Berikut akan dicari ekspektasi dari $\hat{\psi}(x, t)$ adalah

$$E(\hat{\psi}(x, t)) = E(\mathbf{H}[\mathbf{h}]\mathbf{y})$$

$$= \mathbf{H}[\mathbf{h}] \begin{pmatrix} E(\psi(x_{11}, t_{11})) \\ E(\psi(x_{12}, t_{12})) \\ \vdots \\ E(\psi(x_{nm}, t_{nm})) \end{pmatrix}$$

$$E(\hat{\psi}(x, t)) \neq \begin{pmatrix} E(\psi(x_{11}, t_{11})) \\ E(\psi(x_{12}, t_{12})) \\ \vdots \\ E(\psi(x_{nm}, t_{nm})) \end{pmatrix}$$

Karena $E(\hat{\psi}(x, t)) \neq \psi(x, t)$, maka estimator $\hat{\psi}$ bias untuk kurva regresi.

Kriteria GCV merupakan salah satu metode untuk memilih *bandwidth* yang optimal. Fungsi GCV didefinisikan sebagai berikut :

$$GCV(h) = \frac{N^{-1}(\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}})}{(N^{-1} \text{trace}[\mathbf{I} - \mathbf{A}^*(\mathbf{h})])^2}$$

$$= \frac{N^{-1}\mathbf{y}^T(\mathbf{I} - \mathbf{A}^*(\mathbf{h}))^T(\mathbf{I} - \mathbf{A}^*(\mathbf{h}))\mathbf{y}}{(N^{-1} \text{trace}[\mathbf{I} - \mathbf{A}^*(\mathbf{h})])^2} \quad (9)$$

dengan, $N=nxm$.

Selanjutnya untuk pemilihan *bandwidth* optimal dengan menggunakan GCV, dilakukan dengan cara mensubstitusikan nilai *bandwidth* (h (0,)) kedalam matrik $\mathbf{A}(\mathbf{h})$ sehingga diperoleh nilai minimum $GCV(h)$.

Aplikasi Model Regresi Semiparametrik Polinomial Lokal untuk Data Longitudinal pada Kadar Trombosit Penderita DBD

Estimasi kurva regresi semiparametrik Polinomial Lokal untuk data longitudinal, diterapkan pada data kadar trombosit penderita Demam Berdarah Dengue selama dirawat di Rumah Sakit Haji. Pada penelitian ini, difokuskan untuk mengetahui hubungan antara waktu pemeriksaan selama dirawat di rumah sakit dan kadar hematokrit terhadap kadar trombosit penderita Demam Berdarah Dengue (DBD). Terdapat tiga derajat/grade DBD yang diamati, dimana di dalam setiap subjek (derajat/grade DBD) yang teramati diukur secara berulang dari hari ke hari. Kadar trombosit sebagai variabel respon (y), waktu pemeriksaan (t)

merupakan variabel prediktor komponen parametrik, sedangkan kadar hematokrit (x) merupakan variabel prediktor komponen nonparametrik pada model regresi semiparametrik.

Estimasi model semiparametrik Polinomial Lokal sangat bergantung pada pemilihan *bandwidth* optimal dan derajat polinomial, dimana penentuan *bandwidth* optimal dan derajat polinomial dapat dilihat pada nilai GCV minimum. Langkah awal dalam pemilihan *bandwidth* optimal dan orde polinomial adalah menentukan nilai awal *bandwidth* dan orde polinomial. Adapun cara untuk menentukan nilai awal *bandwidth* dan orde polinomial adalah dengan melakukan pemilihan *bandwidth* optimal dan orde polinomial pada masing-masing subjek secara parsial. Dalam penelitian ini fungsi kernel yang digunakan adalah fungsi Kernel Gaussian, Kernel Epachenikov, dan Kernel Kuadrat. Setelah dicobakan pada berbagai fungsi kernel tersebut diperoleh model yang paling baik adalah model regresi menggunakan fungsi Kernel Gaussian dengan $GCV=1011,295$ dan $MSE=146,7639$ serta $R^2 = 93,92485\%$.

Oleh karena itu, dalam pemodelan simultan akan digunakan fungsi Kernel Gaussian. Sebelum melakukan estimasi, dilakukan terlebih dahulu pemilihan *bandwidth* optimum dan orde polinomial secara simultan. Nilai GCV minimum sebesar 1011,2931, dimana subjek 1 mempunyai *bandwidth* optimal sebesar 0,005 dan orde polinomial $p_1=1$. Subjek 2 mempunyai *bandwidth* optimal sebesar 63,3 dan orde polinomial $p_2=4$. Subjek 3 mempunyai *bandwidth* optimal sebesar 30,6 dengan orde polinomial $p_3=4$.

Setelah diperoleh *bandwidth* optimal dan orde polinomial masing-masing subjek secara simultan maka estimasi model semiparametrik Polinomial Lokal untuk masing-masing subjek diberikan oleh :

Subjek 1 :

$$\hat{y}_1 = -61.1241t_1 + 9.59263t_1^2 + 148.805$$

Subjek 2 :

$$\hat{y}_2 = -61.124t_2 + 9.59263t_2^2 + 138.7746 - 2.4065(x_2 - 41.6) - 17.125(x_2 - 41.6)^2 + 4.688(x_2 - 41.6)^3 - 0.318(x_2 - 41.6)^4$$

Subjek 3 :

$$\hat{y}_3 = -61.124t_3 + 9.5926t_3^2 + 64.5623 - 25.1(x_3 - 36.4) - 21.983(x_3 - 36.4)^2 - 3.392(x_3 - 36.4)^3 + 0.145(x_3 - 36.4)^4$$

Model tersebut mempunyai nilai MSE = 146.7636 dan koefisien determinasi $R^2 = 93,925\%$.

KESIMPULAN

Berdasarkan hasil penelitian diperoleh kesimpulan bahwa melalui pendekatan estimator Polinomial Lokal,

1. Estimasi model regresi semiparametrik untuk data longitudinal adalah sebagai berikut:

$$\hat{y} = A^*(h)y,$$

dengan

$$A^*(h) = A_{par} + A(I - A_{par}),$$

$$A_{nonpar} = A(I - A_{par}),$$

$$A_{par} = T[T^T(I - A)^T(I - A)T]^{-1}T^T(I - A)^T(I - A),$$

$$\text{dan } A = C(C^TK_hC)^{-1}C^TK_h$$

2. Pola hubungan antara kadar trombosit penderita DBD terhadap kadar trombosit secara bersama-sama dapat dibentuk dalam model regresi semiparametrik Polinomial Lokal, dengan kadar hematokrit sebagai komponen nonparametrik dan waktu pemeriksaan sebagai komponen parametrik. Model semiparametrik kadar trombosit yang diestimasi dengan pendekatan Polinomial Lokal diberikan oleh :

Subjek 1 :

$$\hat{y}_1 = -61.1241t_1 + 9.59263t_1^2 + 148.805$$

Subjek 2 :

$$\hat{y}_2 = -61.124t_2 + 9.59263t_2^2 + 138.7746 - 2.4065(x_2 - 41.6) - 17.125(x_2 - 41.6)^2 + 4.688(x_2 - 41.6)^3 - 0.318(x_2 - 41.6)^4$$

Subjek 3 :

$$\hat{y}_3 = -61.124t_3 + 9.5926t_3^2 + 64.5623 - 25.1(x_3 - 36.4) - 21.983(x_3 - 36.4)^2 - 3.392(x_3 - 36.4)^3 + 0.145(x_3 - 36.4)^4$$

Model tersebut mempunyai nilai MSE = 146.7636 dan koefisien determinasi $R^2 = 93,925\%$.

Dalam penelitian ini, permasalahan yang dikaji masih sangat terbatas hanya pada pemodelan pola hubungan antara variabel respon terhadap variabel prediktor dengan pendekatan Polinomial Lokal, belum ada menguji signifikansi parameter model yang dihasilkan. Pemilihan variabel prediktor yang digunakan dalam penelitian ini hanya satu variabel prediktor komponen parametrik dan satu variabel prediktor komponen nonparametrik untuk penelitian selanjutnya masih perlu untuk dikembangkan untuk beberapa variabel prediktor dan dilakukan penelitian tentang uji parameter model dalam regresi semiparametrik Polinomial Lokal.

DAFTAR PUSTAKA

- A.H. Welsh, and T.Y. Yee, 2005. Local Regression for Vector Responses. *Journal of Statistical Planning and Inference*. Vol. 136: 3007-3031.
- G. Wahba. 1990. *Spline Models for Observation Data*, SIAM: Philadelphia. CBMS-NSF Regional Conference Series in Applied Mathematics. Vol. 59.
- H. Basri. 2009. Estimasi Kurva Regresi Nonparametrik pada Data Longitudinal dengan Pendekatan Spline. Tesis, Jurusan Statistika, FMIPA Institut Teknologi Sepuluh November, Surabaya. R.L. Eubank. 1988. *Spline Smoothing and Nonparametric Regression*, Marcel Dekker: New York.
- H. Kuswantoro. 2005. Model Gamma-Fraily untuk Data Longitudinal dan Pendugaan Korelasi Serial dengan Metode Composite Likelihood. Tesis Jurusan Statistika FMIPA ITS: Surabaya.

-
-
- H. Wu, and J.T. Zhang. 2006. *Nonparametric Regression Methods for Longitudinal Data Analysis*, A John-Wiley and Sons Inc. Publication, New Jersey.
- I. N. Budiantara, B. Lestari, dan A. Islamiyati. 2010. Estimator Spline Terbobot dalam Regresi Nonparametrik dan Semiparametrik Heteroskedastisitas untuk Data Longitudinal. *Hibah Penelitian Kompetensi*. DP2M-Dikti: Jakarta.
- N. Chamidah. 2008. Inferensi Kurva Regresi Nonparametrik Berdasarkan Estimator Polinomial Lokal Dengan Error Lognormal. *J. Peneliti. Med. Eksakta*, Vol. 7, No. 1, hal. 61-6.
- P. J. Diggle, Y. K. Liang dan S. L. Zeger. 1994. *Analysis of Longitudinal Data*, Oxford Statistical Science Series 13, New York.
- I.N. Budiantara. dan Mulianah. 2007. Pemilihan Bandwidth Optimal Dalam Regresi Semiparametrik Kernel dan Aplikasinya, *Journal Sains dan Teknologi SIGMA*, 10 : 159-166.
- L. Laome. 2009. Model Regresi Semiparametrik Spline untuk Data Longitudinal pada Kasus Kadar CD4 Penderita HIV. *Paradigma*. Vol. 13, No.2: 189-194.
- M. P. Wand dan M. C. Jones, 1995. *Kernel Smoothing*, Chapman and Hall, London.
- S. Y. Hong. 1999. "Automatic Bandwidth Choice in a Semiparametric Regression Model", *Statistica Sinica*, 9 : 775-794.
- W. Hardle. 1990. *Applied Nonparametric Regression*. Cambridge University Press, New York.