ESTIMASI INTERVAL UNTUK KURVA REGRESI CAMPURAN SPLINE TRUNCATED DAN KERNEL DALAM REGRESI NONPARAMETRIK
(Studi Kasus Angka Kematian Bayi Di Provinsi Jawa Timur)

Syisliawati[1], Wahyu Wibowo[2] , dan I Nyoman Budiantara[3]
[123]Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail: syisliailasamad123@gmail.com , i_nyoman_b@statistika.its.ac.id,

Abstrak— Salah satu bagian penting dalam ilmu statistika adalah persoalan inferensi yaitu penarikan kesimpulan secara statistik. Pada umumnya analisis regresi digunakan untuk menyelidiki adanya hubungan antara variabel respon dan variabel prediktor dengan mengamati pola kecenderungan hubungannya. Apabila kurva regresi tidak diketahui maka dipergunakan metode regresi nonparametrik. Penelitian ini bertujuan mencari interval konfidensi terpendek untuk angka kematian bayi Propinsi Jawa Timur menggunakan estimator campuran *kernel* dan *spline* pada regresi nonparametrik berdasarkan hasil penelitian di peroleh model terbaik yaitu pada tiga titik knot dimana MSE dan pada tiga titik knot dimana nilai MSE dan R^2 berturut-turut adalah 4,11 dan 88,97% dan semua parameternya $\hat{\beta}$ termuat dalam interval konfidensi.

Kata kunci : *Regresi Nonparametrik Kernel dan Spline, GCV dan angka Kematian Bayi (AKB)*

I. PENDAHULUAN

Pemodelan regresi dibagi menjadi tiga yaitu metode regresi parametrik, regresi nonparametrik, dan regresi semiparametrik. Regresi semiparametrik digunakan jika salah satu kurva regresinya tidak diketahui sedangkan yang lainnya diketahui [1]. 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. Dengan demikian, pendekatan regresi nonparametrik memiliki fleksibilitas yang tinggi [2]. Regresi nonparametrik tidak mengasumsikan bentuk kurva regresi, kurva regresi diasumsikan termuat di dalam suatu ruang fungsi tertentu misalnya *Ruang Sobolev* [2]. Dalam realita, tidak semua model data dapat diduga dengan pendekatan regresi parametrik karena tidak adanya informasi yang lengkap tentang bentuk kurva regresi. Dalam keadaan seperti itu dapat digunakan pendekatan regresi nonparametrik [3]. Pendekatan regresi nonparametrik diantaranya *spline*, *kernel*, *k-nearest neigborhood* dan lain-lain. Diantara metode-metode pendekatan tersebut, regresi nonparametrik dengan pendekatan spline dan kernel merupakan metode yang sering digunakan. Inferensi statistik yang sangat penting dalam regresi *spline* adalah interval konfidensi dengan konsep interval konfidensi terpendek dengan *Pivotal Quantity*. Estimasi interval merupakan estimasi pengembangan estimasi titik. Bahwa nilai taksiran parameter tidak berfokus pada satu titik tetapi berdasarkan pada range tertentu. Dalam penelitian ini akan dilakukan estimasi campuran regresi *spline* dan *kernel* untuk memodelkan fungsi Angka Kematian Bayi (AKB) di Indonesia.

Kematian bayi adalah kematian yang terjadi pada bayi pada selang waktu antara saat setelah bayi lahir sampai dengan bayi belum berumur tepat satu tahun. Besaran yang menyatakan kemungkinan terjadinya bayi mati setelah lahir sampai mencapai usia satu tahun dalam perseribu kelahiran hidup dinamakan Angka Kematian Bayi (AKB). Berdasarkan Survei Demografi dan Kesehatan Indonesia (SDKI, 2012), AKB di Indonesia mencapai angka 32/1000 kelahiran hidup bayi, angka ini cukup tinggi dibandingkan dengan standar dari *Millennium Development Goals* (MDGs) yaitu 23/1000 untuk AKB. Pemerintah indonesia diharapkan mampu menekan nilai AKB melalui program-program yang dicanangkan atau dengan mengetahui dan mengatasi faktor-faktor yang berpengaruh secara signifikan terhadap tingginya nilai AKB. Provinsi Jawa Timur pada tahun yang sama memiliki AKB hingga 32,43 kematian per 1.000 kelahiran hidup. Pada tahun 2009 AKB mengalami penurunan menjadi 31,41, sedangkan tahun 2010 AKB menjadi 29,99 kematian per 1.000 kelahiran hidup [4]. Hal ini menunjukkan bahwa Jawa Timur belum mampu mencapai target MDG's [4]. Data AKB berdasarkan diagram pencar dengan masing-masing prediktornya mempunyai pola yang cenderung tidak mengikuti dengan pola tertentu (linier atau polinomial berderajat tertentu), sehingga pendekatan model regresi yang sesuai adalah model regresi nonparametrik. Karakteristik dari pola data AKB adalah sebagian tidak mempunyai pola sehingga metode *spline* digunakan untuk memodelkan pola data.

II. TINJAUAN PUSTAKA

A. Analisis Regresi

Analisis regresi merupakan analisis statistika yang digunakan untuk memodelkan pola hubungan antara variabel prediktor dan variabel respon. Pendekatan regresi parametrik digunakan jika bentuk kurva regresi diketahui. Jika pola hubungan data membentuk pola linear maka digunakan pendekatan regresi parametrik linear. Jika pola hubungan data membentuk pola kuadrat maka digunakan pendekatan regresi kuadratik, dan lain-lain [2]. Bentuk pola hubungan dapat diidentifikasi berdasarkan pada informasi masa lalu atau *scatterplot* data [3]

B. Regresi Nonparametrik

Regresi nonparametrik merupakan pendekatan metode regresi dimana bentuk kurva dari fungsi regresinya tidak diketahui. Kurva fungsi diasumsikan termuat dalam ruang fungsi tertentu [2]. Model regresi nonparametrik diberikan oleh persamaan berikut.

$$y_i = f\left(x_i\right) + \varepsilon,\ i = 1, 2, ..., n$$

dengan y_i adalah variabel respon, x_i merupakan variabel prediktor, $f(x_i)$ adalah fungsi regresi, dimana bentuk kurva tidak diketahui dan ε_i adalah *error* random berdistribusi normal, dengan mean nol dan variansi σ^2.

C.Regresi Nonparametrik Spline

Spline dalam regresi nonparametrik mempunyai sifat fleksibilitas yang tinggi dan mempunyai kemampuan mengestimasi perilaku data yang cenderung berbeda pada interval yang berlainan [2,5]. Kemampuan mengestimasi perilaku data ini ditunjukkan oleh fungsi *truncated* (potongan-potongan) yang melekat pada estimator dan potongan-potongan tersebut yang disebut titik knot. Titik knot merupakan titik perpaduan bersama yang menunjukkan perubahan pola perilaku fungsi pada selang yang berbeda. Spline merupakan salah satu jenis *piecewise* polinomial, yaitu polinomial yang memiliki sifat tersegmen. Sifat tersegmen ini memberikan fleksibilitas lebih dari polinomial biasa, sehingga memungkinkan untuk menyesuaikan diri secara lebih efektif terhadap karakteristik lokal suatu fungsi atau data. Fungsi spline berderajat *p* adalah sebarang fungsi yang secara umum dapat disajikan dalam bentuk sebagai berikut :

$$f\left(x_i\right) = \sum_{j=0}^{p} \beta_j x_i^j \sum_{j=}^{k} \beta_{j+p} \left(x_i - k_j\right)_+^p$$

dengan β_j adalah konstanta riil, dan

$$\left(x_i - k_j\right)_+^p = \begin{matrix} \left(x_i - k_j\right)^p & ; x \geq k_j \\ 0 & ; x < k_j \end{matrix}$$

Jika *p* = 1, 2, dan 3 diperoleh berturut-turut spline linear, spline kuadratik dan spline kubik serta k_j adalah titik knot.

Apabila diasumsikan *error* ε_i berdistribusi normal independen dengan rata-rata nol dan variansi σ^2, maka y_i pada model regresi juga berdistribusi normal dengan rata-rata $f\left(x_i\right)$ dan variansi σ^2. Estimasi untuk parameter β dengan menggunakan metode *least square*, yaitu dengan meminimumkan jumlah kuadrat *error* adalah sebagai berikut,

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_j - f\left(x_i\right)\right)^2$$

Estimator β dapat di peroleh dari meminimumkan

$$Q\left(\beta\right) = \left(y - X\beta\right)^T \left(y - X\beta\right)$$

D.Regresi Nonparametrik Kernel

Estimator kernel mempunyai kelebihan yaitu fleksibel, bentuk matematisnya mudah dan dapat mencapai tingkat kekonvergenan yang relative cepat. Kurva regresi $g(t_i)$ yang dihampiri fungsi kernel, estimasi kurva regresi dapat disajikan dalam bentuk :

$$\hat{g}_\alpha\left(t\right) = n^{-1} \sum_{i=1}^{n} \left[\frac{K_\alpha\left(t - t_i\right)}{n^{-1} \sum_{j=1}^{n} K_\alpha\left(t - t_j\right)} \right] y_i = n^{-1} \sum_{i=1}^{n} W_{\alpha i}\left(t\right) y_i \quad (2)$$

dimana :

$$W_{\alpha i}\left(t\right) = \frac{K_\alpha\left(t - t_i\right)}{n^{-1} \sum_{j=1}^{n} K_\alpha\left(t - t_j\right)}; \quad K_\alpha\left(t - t_i\right) = \frac{1}{\alpha} K\left(\frac{t - t_i}{\alpha}\right)$$

dengan K merupakan fungsi kernel. Fungsi kernel K berupa Kernel Gaussian : $K(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}) I_{[-\infty,\infty]}(z)$

E. Estimator Campuran Spline dan Kernel

Penelitian tentang model regresi nonparametrik additif yang memiliki dua komponen variabel prediktor. Komponen prediktor pertama, kurva regresinya dihampiri menggunakan regresi spline, sedang komponen prediktor kedua kurva regresi dihampiri dengan regresi kernel [10]. Data berpasangan (x_i, t_i, y_i) dengan diasumsikan hubungan antar variabel prediktor x_i, t_i dan variabel respon y_i mengikuti model regresi nonparametrik :

$$y_i = \mu(x_i, t_i) + \varepsilon_i\ ,\ \text{i=1,2,…,n}$$

Bentuk kurva regresi diasumsikan tidak diketahui dan hanya diasumsikan smooth dalam arti kontinu dan *differensiabel*. Error random ε_i berdistribusi normal dengan mean nol dan $E(\varepsilon_i^2) = \sigma^2$. Selanjutnya kurva regesi $\mu(x_i, t_i)$ diasumsikan additif, dalam arti $\mu(x_i, t_i)$ dapat ditulis dalam bentuk :

$$\mu(x_i, t_i) = f(x_i) + g(t_i)$$

Dengan $f(x_i)$ dan $g(t_i)$ merupakan fungsi-fungsi yang *smooth*. Persoalan utama dalam estimator campuran kurva regresi nonparametrik adalah mendapatkan bentuk estimasi kurva regresi $\mu(x_i, t_i)$ yaitu :

$$\hat{\mu}(\mathbf{x}_i, t_i) = \sum_{p=1}^{q} \hat{f}_p\left(x_{pi}\right) + \hat{g}\left(t_i\right)$$

Estimator campuran regresi Spline dan Kernel $\hat{\mu}_{\alpha,\lambda}\left(x, t\right)$ sangat bergantung kepada banyak dan letak titik knot $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_r)^T$ dan parameter *bandwitdh* α . Untuk memperoleh estimator campuran regresi Spline dan Kernel yang terbaik perlu dilakukan pemilihan titik knot dan parameter *bandwidth* yang optimal. Metode yang biasa digunakan adalah *Generalized Cross Validation* (GCV). Fungsi GCV yang diberikan [6] adalah :

$$GCV(\lambda,\alpha)=\frac{n^{-1}\|\mathbf{y}-\hat{\boldsymbol{\mu}}_{\alpha,\lambda}(x,t)\|^2}{\left(n^{-1}trace\left[\mathbf{I}-\mathbf{A}(\lambda,\alpha)-\mathbf{D}(\alpha)\right]\right)^2}$$

Titik knot optimal $\lambda_{opt}=(\lambda_{1(opt)},\lambda_{2(opt)},...,\lambda_{r(opt)})^T$ dan parameter *bandwidth* optimal α_{opt} diperoleh dari optimasi :

$$G(\lambda_{opt},\alpha_{opt})=\underset{\lambda,\alpha}{Min}\{G(\lambda,\alpha)\}$$

F. Koefisien Determinasi ($\mathbf{R^2}$)

Koefisien determinasi ($\mathbf{R^2}$) merupakan ukuran ketelitian atau ketepatan model regresi, atau besarnya kontribusi x terhadap perubahan y. Semakin tinggi nilai R2 akan semakin baik

$$R^2=\frac{SSR}{SST}=\frac{(\hat{Y}-\bar{Y})'(\hat{Y}-\bar{Y})}{(Y-\bar{Y})'(Y-\bar{Y})}$$

G. Interval Konfidensi

Estimasi interval adalah estimasi model ini merupakan pengembangan estimasi titik. Bahwa nilai taksiran parameter tidak berfokus pada satu titik tetapi berdasarkan pada range tertentu, sehingga estimasinya memiliki nilai tertinggi (max) dan nilai terendah (min). interval ini lebih dikenal dengan inteval konfidensi. Nilai yang muncul adalah nilai yang didasarkan probabilitas tertentu, dalam prakteknya biasanya dipilih 90%, 95% atau 99%. Misalkan $X_1,X_2,...,X_n$ adalah sampel random yang diambil dari populasi dengan parameter θ, maka interval konfidensi $1-\alpha$ untuk θ adalah $P(a\le\theta\le b)=1-\alpha$.

III. METODOLOGI PENELITIAN

Penelitian ini menggunakan data sekunder yang berasal dari Badan Pusat Statistika (BPS) tahun 2011. Unit observasi pada penelitian ini adalah 38 Kabupaten/Kota yang ada di Propinsi Jawa Timur pada tahun 2011. Variabel yang digunakan adalah Angka Kematian Bayi (AKB), persentase pertolongan terakhir persalinan yang dilakukan dengan bantuan non medis, persentase perkawinan pada wanita yang usia kawin kurang dari 17 tahun, rata-rata jumlah pengeluaran rumah tangga perkapita sebulan, persentase bayi usia 0-11 bulan yang di beri ASI, dan jumlah sarana kesehatan (rumah sakit dan puskemas). Tahapan analisis yang dilakukan adalah sebagai berikut.

Tahapan Penelitian

Diberikan data berpasangan $(x_i,t_i,y_i),i=1,2,3,...,n$ yang mengikuti regresi nonparametrik

$$y_i=f(x_i)+g(t_i)+\varepsilon$$

Kurva regresi $f(x_i)$ dihampiri dengan spline truncated:

$$f(x_i)=\sum_{v=0}^{m}\beta_v x_i^v+\sum_{v=1}^{r}\beta_{(v+m)}(x_i-k_v)_+^m$$

1. Mencari interval konfidensi untuk $\beta_v, v=1,2,...,m+r$ pada estimasi campuran regresi nonparametrik spline dan kernel.
   Langkah-langkahnya adalah sebagai berikut:
   1) Menghampiri fungsi $f(x)$ dengan spline truncated dan titik-titik knot $k_1,...,k_r$.
   2) Menghampiri fungsi $g(t_i)$ dengan fungsi kernel
   3) Mencari estimasi untuk β, f dan g dalam estimator campuran spline dan kernel.
   4) Mencari distribusi dari $\hat{\beta}$
   5) Dengan menggunakan distribusi dari $\hat{\beta}$ dicari suatu statistik T yang merupakan Pivotal Quantity untuk β
   6) Mencari interval koefisien terpendek untuk parameter β.
2. Memodelkan data angka kematian bayi Propinsi Jawa timur, dengan langkah-langkah sebagai berikut:
   1) Membuat scatter plot antara variabel respon dengan masing-masing variabel prediktor.
   2) menentukan variabel-variabel prediktor yang menggunakan komponen spline dan komponen kernel.
   3) Memodelkan data menggunakan estimasi campuran spline dan kernel dengan berbagai knot (satu knot, dua knot dan tiga knot).
   4) Memilih titik knot dan bandwitdh optimal dengan metode GCV
   5) Menetapkan model terbaik dari nilai GCV terkecil.
   6) Membuat interval konfidensi terpendek 95% untuk parameter β.
   7) Menghitung MSE dan R^2 dari data.

IV. HASIL DAN PEMBAHASAN

A. Interval konfidensi terpendek untuk parameter model komponen spline truncated variansi diketahui

Berdasarkan estimator yang diperoleh pada persamaan di peroleh estimasi untuk β dalam estimator campuran spline dan kernel yaitu:

$$\hat{\tilde{\beta}}(\hat{\lambda},\alpha)=\left[\left(X(\hat{\lambda})\right)^T X(\hat{\lambda})\right]^{-1}\left(X(\hat{\lambda})\right)^T\left(I-D(\alpha)\right)y$$

dimana

$$\left(\hat{\tilde{\beta}}(\hat{\lambda},\alpha)\right)\sim N\left[E\left(\hat{\tilde{\beta}}(\hat{\lambda},\alpha)\right),Var\hat{\tilde{\beta}}(\hat{\lambda},\alpha)\right]$$

$$\left(\hat{\tilde{\beta}}(\hat{\lambda},\alpha)\right)\sim N\left[\beta,\sigma^2W(\hat{\lambda},\alpha)\right]$$

Langkah-langkah interval konfidensi

$$Z=\frac{\hat{\tilde{\beta}}_j-E\left(\hat{\tilde{\beta}}\right)}{\sqrt{Var\left(\hat{\tilde{\beta}}_j\right)}}\sim N(0,1)$$

$$Z=\frac{[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1}X(\tilde{\lambda})^T(I-D(\alpha))\tilde{y}]_j-\beta_j}{[\sqrt{\sigma^2\omega_j(\tilde{\lambda},\alpha)}]_{jj}}\sim N(0,1)$$

Dimana ω_j merupakan elemen diagonal ke-j dari suatu matriks $W\left(\tilde{\lambda},\alpha\right)$. variabel random Z_j berdistribusi $N(0,1)$. Dengan demikian, Z_j merupakan pivotal Quantity untuk kurva regresi β_j . Interval konfidensi $1-\alpha$ diperoleh dari menyelesaikan persamaan probabilitas:

$$P\left(a \le Z_j \le b\right) = 1-\alpha$$

$$a \in R, b \in R,\ a < b,\ i = 1,2,...,n$$

persamaan diatas dapat dinyatakan menjadi:

$$P\left(a \le \frac{[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I - D(\alpha))\tilde{y}]_j}{[\sqrt{\sigma^2 \omega_j(\tilde{\lambda},\alpha)}]_{jj}} \le b \right) = 1-\alpha$$

Dengan sedikit penjabaran, sehingga diperoleh interval konfidensi $1-\alpha$ untuk $\beta_j,\ j = 1,2,...,n$

$$P[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I - D(\alpha))\tilde{y}]_j + b[\sqrt{\sigma^2 \omega_j(\tilde{\lambda},\alpha)}]_{jj} \le \beta_j \le$$
$$[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I - D(\alpha))\tilde{y}]_j + a[\sqrt{\sigma^2 \omega_j(\tilde{\lambda},\alpha)}]_{jj}] = 1-\alpha$$

Dengan konsep interval terpendek harus ditentukan nilai $a \in R$ dan $b \in R$ sehingga panjang interval $\lambda\left(a,b\right)$ pada persamaan diatas terpendek. Untuk tujuan ini dicari penyelesaian optimasi bersyarat berikut.

$$\underset{a\in R, b\in R}{Min}\{\lambda(a,b)\} = \underset{a\in R, b\in R}{Min}\left\{(b-a)\left[\sqrt{MSE\omega_i\left(\tilde{\lambda},\alpha\right)}\right]_{jj}\right\}$$

Dengan syarat $\int_a^b \varphi\left(z\right)dz = 1-\alpha$, atau $\varphi\left(b\right)-\varphi\left(a\right)-\left(1-\alpha\right)=0$

Fungsi φ merupakan distribusi probabilitas N(0,1) dan φ merupakan distribusi probabilitas kumulatif N(0,1). Optimasi 1 dan 2 dapat diselesaikan dengan menggunakan metode lagrange multiple dibentuk fungsi lagrange multiple.

$$\Omega\left(a,b,c\right) = \left(b-a\right)\left[\sqrt{\sigma^2\omega_j\left(\tilde{\lambda},\alpha\right)}\right]_{jj} + c\left[\int_a^b \varphi\left(z\right)dz - \left(1-\alpha\right)\right]$$

Selanjutnya dengan menderivatifkan fungsi $\Omega\left(a,b,c\right)$ terhadap a, b dan c diperoleh:

$$\frac{\Omega\left(a,b,c\right)}{\partial a} = 0 \Rightarrow \left[\sqrt{\sigma^2\omega_j\left(\tilde{\lambda},\alpha\right)}\right]_{jj} - c\varphi\left(a\right) = 0$$

$$\frac{\Omega\left(a,b,c\right)}{\partial b} = 0 \Rightarrow \left[\sqrt{\sigma^2\omega_j\left(\tilde{\lambda},\alpha\right)}\right]_{jj} + c\varphi\left(b\right) = 0$$

$$\frac{\Omega\left(a,b,c\right)}{\partial c} = 0 \Rightarrow \Omega\left(b\right) - \Omega\left(a\right) - \left(1-\alpha\right) = 0$$

Persamaan 4 dan 5 menghasilkan penyelesaian $\Omega\left(a\right) = \Omega\left(b\right)$

Menginggat persamaan 6 dan $Z \sim N\left(0,1\right)$ maka penyelesaian persamaan 7 adalah $a = b$ atau $a = -b$. Akan tetapi, persamaan $a = b$ tidak memenuhi. Jadi,agar diperoleh interval konfidensi terpendek harus harus diambil nilai a dan b yang memenuhi persamaan

$$\int_{-\infty}^{a} \varphi\left(z\right)dz = \frac{\alpha}{2} = \int_{b}^{\infty} \varphi\left(z\right)dz$$

Jika tingkat konfidensi $1-\alpha$ diberikan maka nilai a dan b dapat dilihat pada tabel distribusi $N\left(0,1\right)$. Interval konfidensi $1-\alpha$ untuk kurva regresi $\beta_j,\ j = 1,2,...,n$, diberikan oleh persamaan (1) dengan a dan b memenuhi persamaan

B. Interval konfidensi terpendek untuk parameter model komponen spline truncated variansi tidak diketahui

Umumnya, dalam analisis regresi spline, varians σ^2 tidak diketahui. Jika demikian σ^2 dapat diganti dengan estimasinya,yaitu:

$$\hat{\sigma}^2 = MSE = \frac{1}{n-m-r-1}\sum_{i=1}^{n}\left(\hat{y} - y_i\right)^2$$

$$T = \frac{\hat{\tilde{\beta}}_j - E\left(\hat{\tilde{\beta}}\right)}{\sqrt{Var\left(\hat{\tilde{\beta}}_j\right)}} \sim t_{(n-m-r-1)}$$

$$T = \frac{[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I - D(\alpha))\tilde{y}]_j - \beta_j}{[\sqrt{MSE\omega_j(\tilde{\lambda},\alpha)}]_{jj}} \sim t_{(n-m-r-1)}$$

Dengan ω_j elemen diagonal ke-j dari matriks $W\left(\tilde{\lambda},\alpha\right)$. Variabel random T berdistribusi $t_{(n-m-r-1)}$. Dengan demikian,T merupakan pivotal Quantity untuk kurva regresi β_j. Interval konfidensi $1-\alpha$ diperoleh dari menyelesaikan persamaan probabilitas:

$$P\left(a \le T \le b\right) = 1-\alpha$$

$$a \in R, b \in R,\ a < b,\ i = 1,2,...,n$$

persamaan diatas dapat dinyatakan menjadi:

$$P\left(a^* \le \frac{[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I - D(\alpha))\tilde{y}]_j - \beta_j}{[\sqrt{MSE\omega_j(\tilde{\lambda},\alpha)}]_{jj}} \le b^* \right) = 1-\alpha$$

Setelah dilakukan penjabaran, maka diperoleh interval konfidensi $1-\alpha$ untuk $\beta_j,\ j = 1,2,...,n$

$$P[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I - D(\alpha))\tilde{y}]_j + b^*[\sqrt{MSE\omega_j(\tilde{\lambda},\alpha)}]_{jj} \le \beta_j \le$$
$$[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I - D(\alpha))\tilde{y}]_j + a^*[\sqrt{MSE\omega_j(\tilde{\lambda},\alpha)}]_{jj}] = 1-\alpha$$

Dengan konsep interval terpendek harus ditentukan nilai $a \in R \, dan \, b \in R$ sehingga panjang interval $\lambda(a^*, b^*)$ pada persamaan diatas terpendek. Untuk tujuan ini dicari penyelesaian optimasi bersyarat berikut.

$$\underset{a\in R, b\in R}{Min}\left\{\lambda\left(a^*,b^*\right)\right\} = \underset{a\in R, b\in R}{Min}\left\{\left(b^*-a^*\right)\left[\sqrt{MSE\omega_i\left(\tilde{\lambda},\alpha\right)}\right]_{jj}\right\}$$

Dengan syarat

$$\int_{a^*}^{b^*} \varphi(z)\,dz = 1-\alpha$$

Fungsi φ merupakan distribusi probabilitas $t_{(n-m-r-1)}$ dan φ merupakan distribusi probabilitas kumulatif $t_{(n-m-r-1)}$ .Optimasi 1 dan 2 dapat diselesaikan dengan menggunakan metode lagrange multiple dibentuk fungsi lagrange multiple

$$\Omega\left(a^*,b^*,c\right) = \left(b^*-a^*\right)\left[\sqrt{MSE\omega_j\left(\tilde{\lambda},\alpha\right)}\right]_{jj} + c\left[\int_{a^*}^{b^*}\varphi(z)\,dz-(1-\alpha)\right]$$

Selanjutnya dengan menderivatifkan fungsi $\Omega\left(a^*,b^*,c\right)$ terhadap a^*,b^* dan c diperoleh:

$$\frac{\Omega\left(a^*,b^*,c\right)}{\partial a^*} = 0 \Rightarrow \left[\sqrt{MSE\omega_j\left(\tilde{\lambda},\alpha\right)}\right]_{jj} - c\varphi\left(a^*\right) = 0$$

$$\frac{\Omega\left(a^*,b^*,c\right)}{\partial b^*} = 0 \Rightarrow \left[\sqrt{MSE\omega_j\left(\tilde{\lambda},\alpha\right)}\right]_{jj} + c\varphi\left(b^*\right) = 0$$

$$\frac{\Omega\left(a^*,b^*,c\right)}{\partial c} = 0 \Rightarrow \Omega\left(b^*\right) - \Omega\left(a^*\right) - (1-\alpha) = 0$$

Persamaan 4 dan 5 menghasilkan penyelesaian

$$\Omega\left(a^*\right) = \Omega\left(b^*\right)$$

Menginggat persamaan (4.14) dan $T \sim t_{(n-m-r-1)}$ maka penyelesaian persamaan (4.15) adalah $a^* = b^*$ atau $a^* = -b^*$. Akan tetapi, persamaan $a^* = b^*$ tidak memenuhi. Jadi,agar diperoleh interval konfidensi terpendek harus harus diambil nilai a^* dan b^* yang memenuhi persamaan

$$\int_{-\infty}^{a^*} \varphi(t)dt = \frac{\alpha}{2} = \int_{b^*}^{\infty} \varphi(t)\,dt$$

Jika tingkat konfidensi $1-\alpha$ diberikan maka nilai a dan b dapat dilihat pada tabel distribusi $t_{(n-m-r-1)}$. Interval konfidensi $1-\alpha$ untuk kurva regresi β_j, $j = 1,2,\dots,n$ diberikan oleh persamaan (9) dengan a^* dan b^* memenuhi persamaan 16.

C. Analisis Deskriptif Angka Kematian Bayi Propinsi Jawa Timur

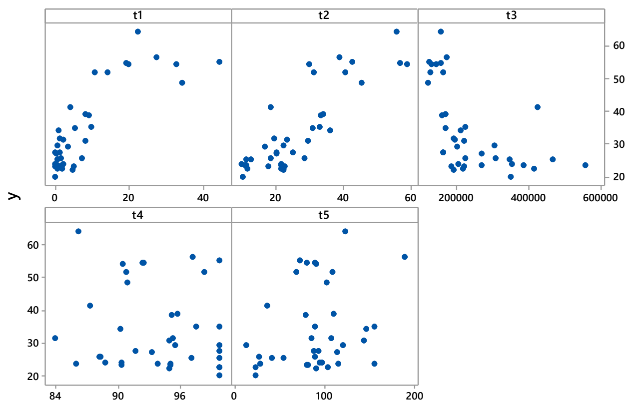
a. Karakteristik Data Angka Kematian Bayi dan Faktor-Faktor yang Mempengaruhinya

Data angka kematian bayi dan faktor-faktor yang mempengaruhinya dideskripsikan menggunakan *mean*, standar deviasi, nilai minimum dan nilai maksimum. Berikut adalah karakteristik data angka kematian bayi dan faktor-faktor yang mempengaruhi disajikan dalam Tabel 1.

Tabel 1Karakteristik Angka Kematian Bayi dan Faktor-Faktor yang Mempengaruhi

Variabel	*Mean*	Standar Deviasi	Minimal	Maksimal
Y	34,18	12,68	20,02	64,19
X_1	8,31	11,19	0,00	44,99
X_2	27,00	13,00	10,07	59,09
X_3	2392	105660	122240	558590
X_4	93,92	4,640	83,920	100,000
X_5	90,11	39,94	13,00	189,00

Gambar 1 Scatterplot Data Angka Kematian Bayi dan Faktor yang Mempengaruhinya



Berdasarkan Gambar 1 terlihat bahwa hubungan antara variabel angka kematian bayi dengan masing-masing faktornya tidak mengikuti pola tertentu karena plot data tersebar secara acak. Maka, variabel independen merupakan komponen nonparametrik. Pada gambar x_1, x_2 dan x_3 terlihat bahwa pola hubunganya berubah-ubah pada sub-sub interval tertentu sehingga dapat didekati dengan regresi spline *truncated* linier, sedangkan pada gambar x_4 dan x_5 terlihat tidak memiliki pola hubungan tertentu, sehingga pola hubungannya tidak jelas dapat didekati dengan regresi kernel.

Pemilihan Titik Knot Optimal *Spline Truncated* dan kernel yang terbaik dilihat pada nilai GCV terkecil. Adapun nilai GCV pada satu knot, dua knot dan tiga knot dapat dilihat pada tabel 2.

Tabel 2 Perbandingan Nilai GCV minimum dari titik knot

No	Titik Knot	Nilai GCV Minimum
1.	1 Titik Knot	33.678
2.	2 Titik Knot	28.83
3.	3 Titik Knot	27.47449

Pada tabel 2 terlihat bahwa nilai GCV terkecil berada pada tiga titik knot yaitu 27.47449. Adapun skor GCV untuk tiga titik knot dapat dilihat pada tabel 3.

Tabel 2 Skor GCV untuk 3 *Knot* pada Model campuran spline truncated dan kernel pada provinsi Jawa Timur

Banyak knot			**Knot 3**		
lokasi Knot	x_1	k_1	11.11	10.09	11.93
		k_2	30.29	30.29	30.29
		k_3	31.21	31.21	31.21
	x_2	k_1	22.07	21.07	23.07
		k_2	43.08	43.08	43.08
		k_3	44.08	44.08	44.08
	x_3	k_1	229101.2	220196.1	238006.3
		k_2	416108.4	416108.4	416108.4
		k_3	425013.5	425013.5	425013.5
Bandwiht	α_1		9.340956	8.755214	9.894156
	α_2		102.1399	95.72876	108.1948
GCV			27.47449	27.65107	27.65179
R^2			**88,97%**		

Berdasarkan pada Tabel 4.5 dari model *spline trubcated* yang diperoleh , diketahui bahwa pada tingkat signifikansi 95%, AKB di Provinsi Jawa Timur masih berada di dalam kisaran *mean*-nya. Dari tabel di bawah ini dapat dilihat batas atas dan bawahnya, yang dapat membantu pembaca khususnya pemerintah daerah dalam memprediksi angka kematian bayi yang baik adalah yang tidak lebih dan tidak kurang dari batas atas dan batas bawah konfidensi interval yang didapat.

Tabel 2 batas atas dan bawah konfidensi interval Provinsi Jawa Timur

		Batas atas	$\hat{\beta}$	Batas bawah
x_1	β_{10}	0.0122	0.0077106	0.0031326
	β_{11}	0.9954	0.68103	0.36659
	β_{12}	0.0122	0.0077106	0.0031326
	β_{13}	0.5459	0.33473	0.12349
	β_{14}	0.0122	0.0077106	0.0031326
x_2	β_{20}	-0.00005	-0.00008060	-0.00010802
	β_{21}	0.08141	-0.1172	-0.31585
	β_{22}	-0.1625	-0.5537	-0.94502
	β_{23}	-0.16267	-0.5484	-0.93419
	β_{24}	0.23285	0.14338	0.053923
x_3	β_{30}	0.19338	-0.08000	-0.35340
	β_{31}	0.17393	-0.08193	-0.33781
	β_{32}	0.00015	0.00009830 3	0.000037802
	β_{33}	0.00209	0.0011259	0.00016169
	β_{34}	-0.00023	-0.001237	-0.0022356

Pada tabel 2 dapat dilihat bahwa semua parameter dari $\hat{\beta}$ termuat dalam interval konfidensi, artinya variabel angka kematian bayi berpengaruh terhadap variabel persentase persalinan yang dilakukan dengan bantuan non medis, persentase wanita perkawinan pertama kurang dari 17 tahun dan rata-rata jumlah pengeluaran rumah tangga perkapita sebulan.

V.KESIMPULAN DAN SARAN

A.Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilakukan pada bab 4 dpat diambil kesimpulan sebagai berikut:

1. Interval konfidensi terpendek untuk parameter model komponen spline truncated variansi diketahui.
   Interval Konfidensi untuk parameter β_j diberikan oleh

$$P\left[\left[X(\tilde{\lambda})^T X(\tilde{\lambda})\right]^{-1}\left(X(\tilde{\lambda})^T\right)(I-D(\alpha))\tilde{y}\right]_j - Z_{\alpha/2}\left[\sqrt{\sigma^2 w_j(\tilde{\lambda},\alpha)}\right]_{jj} \le \beta_j$$
$$\le \left[\left[X(\tilde{\lambda})^T X(\tilde{\lambda})\right]^{-1}\left(X(\tilde{\lambda})^T\right)(I-D(\alpha))\tilde{y}\right]_j + Z_{\alpha/2}\left[\sqrt{\sigma^2 w_j(\tilde{\lambda},\alpha)}\right]_{jj} = 1-\alpha$$

2. Interval konfidensi terpendek untuk parameter model komponen spline truncated variansi tidak ketahui.
   Interval Konfidensi untuk parameter β_j diberikan oleh:

$$P[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I-D(\alpha))\tilde{y}]_j - t_{(n-m-r-1,\alpha/2)}[\sqrt{MSE\omega_j(\tilde{\lambda},\alpha)}]_{jj} \le \beta_j \le$$
$$[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I-D(\alpha))\tilde{y}]_j + t_{(n-m-r-1,\alpha/2)}[\sqrt{MSE\omega_j(\tilde{\lambda},\alpha)}]_{jj}] = 1-\alpha$$

3. Interval Konfidensi Provinsi Jawa Timur
   Berdasarkan nilai GCV terkecil diperoleh model *spline Truncated* dengan tiga titik *knot* yang memiliki nilai GCV yaitu 27.47449 dan dimana nilai MSE dan R^2 berturut-turut adalah 4,11 dan 88,97%. Modelnya dapat dituliskan sebagai berikut.

$$\hat{y} = 0,0077106 + 0,68103x_{1i} + 0,0077106(x_{1i} - 11,01)_{+} + 0,33473$$
$$(x_{1i} - 30,29)_{+} + 0,0077106(x_{1i} - 31,21)_{+} \, 0,0000806 - 0,11722x_{2i}$$
$$- 0,55378(x_{2i} - 22,07)_{+} - 0,55484(x_{2i} - 43,08)_{+} + 0,14338(x_{1i} - 44,08)_{+}$$
$$- 0,080007 - 0,081938x_{3i} + 0,000098303(x_{3i} - 229101)_{+} + 0,0011259$$
$$(x_{3i} - 416108,4)_{+} - 0,0012371(x_{3i} - 425013,5)_{+}$$
$$+ \sum_{i=1}^{38} \left(\frac{\frac{1}{9,3409} K\left(\frac{t_1 - t_{1i}}{9,3409}\right)}{\sum_{j=1}^{38} \frac{1}{9,3409} K\left(\frac{t_1 - t_{1i}}{9,3409}\right)} \right) y_i + \sum_{i=1}^{38} \left(\frac{\frac{1}{102,1398} K\left(\frac{t_2 - t_{2i}}{102,1398}\right)}{\sum_{j=1}^{38} \frac{1}{102,1398} K\left(\frac{t_2 - t_{2i}}{102,1398}\right)} \right) y_i$$

Interval konfidensi 95% variabel model dapat ditulis sebagai berikut:

Untuk persamaan koefisien x_1 (persentase persalinan dengan bantuan non medis) dapat dilihat sebagai berikut:

$P(0,00313 < \beta_{10} < 0,0122) = 0,95$

$P(0,36659 < \beta_{11} < 0,9954) = 0,95$

$P(0,00313 < \beta_{12} < 0,0122) = 0,95$

$P(0,12349 < \beta_{13} < 0,5459) = 0,95$

$P(0,00313 < \beta_{14} < 0,0122) = 0,95$

Berdasarkan interval konfidensi ini didapatkan persamaan $\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}$ sesuai. Jadi x_1 (persentase persalinan dengan bantuan non medis) sesuai dalam model.

Untuk persamaan koefisien x_2 (persentase perkawinan pertama kurang dari 17 tahun) interval konfidensinya dapat dilihat sebagai berikut:

$P(0,000108 < \beta_{10} < 0,00005) = 0,95$

$P(-0,94502 < \beta_{22} < -0,1625) = 0,95$

$P(-0,93419 < \beta_{23} < -0,16267) = 0,95$

$P(0,053923 < \beta_{24} < 0,23285) = 0,95$

Berdasarkan interval konfidensi ini didapat persamaan $\beta_{20}, \beta_{21}, \beta_{23}, \beta_{24}$ sesuai. Jadi x_2 (persentase perkawinan pertama kurang dari 17 tahun) sesuai dalam model.

Untuk persamaan koefisien x_3 (rata-rata jumlah pengeluaran rumah tangga perkapital sebulan) interval konfidensinya dapat dilihat sebagai berikut:

$P(0,000037 < \beta_{32} < 0,00015) = 0,95$

$P(0,00209 < \beta_{33} < 0,000161) = 0,95$

$P(-0,00223 < \beta_{23} < -0,00023) = 0,95$

Berdasarkan interval konfidensi ini didapat persamaan $\beta_{32}, \beta_{33}, \beta_{34}$ sesuai. Jadi x_3 (rata-rata jumlah pengeluaran rumah tangga perkapital sebulan) sesuai dalam model.

B. Saran

Berdasarkan hasil penelitian yang diperoleh, saran yang dapat diberikan penulis adalah sebagai berikut:

1. Bagi pemerintah karena variabel-variabel yang mempengaruhi adalah x_1 (persentase persalinan dengan bantuan non medis), x_2 (persentase wanita perkawinan pertama kurang dari 17 tahun) dan x_3 (rata-tata jumlah pengeluaran rumah tangga perkapital sebulan) maka untuk menurunkan angka kematian bayi perlu mempengaruhi variabel-variabel tersebut.
2. Untuk penelitian lebih lanjut dapat pula menggunakan pengujian hipotesis untuk variabel-variabel berpengaruh terhadap angka kematian bayi.

DAFTAR PUSTAKA

[1] Budiantara, I.N., (2000), *Metode U, GML, CV dan GCV Dalam Regresi Nonparametrik Spline*, Majalah Ilmiah Himpunan Matematika Indonesia (MIHMI), 6, 285-290.

[2] Eubank, R., (1988), *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York.

[3] Hardle, W., (1990), *Applied Nonparametric Regression*, Cambrige University Press, Boston.

[4] KemenKes, 2012, *Data Informasi Kesehatan Provinsi Jawa Timur*, Kementrian Kesehatan Republik Indonesia, Jakarta.

[5] Budiantara, I.N., (2006), *Model Spline dengan Knots Optimal*. Jurnal Natural FMIPA Universitas Jember, Jember.

[6] Aulele, S.N., (2010), *Model Geographically Weighted Regression (Studi Kasus : Jumlah Kematian Bayi di Propinsi Jawa Tengah Tahun 2007)*, Thesis, Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS), Surabaya

[7] Badan Pusat Statistika, (2011), *Hasil Survie Sosial Ekonomi Nasional Tahun 2011 Propinsi Jawa Timur*, BPS, Surabaya