ESTIMASI INTERVAL UNTUK KURVA REGRESI CAMPURAN SPLINE TRUNCATED DAN KERNEL DALAM REGRESI NONPARAMETRIK

SYISLIAWATI
1314201214

DOSEN PEMBIMBING :
Dr. Wahyu Wibowo, S.Si, M.Si
Prof. Dr. Drs. I Nyoman Budiantara, M.Si

PROGRAM MAGISTER
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER SURABAYA
2014

Agenda Hari ini :

1 Pendahuluan → 2 Tinjauan Pustaka → 3 Metodelogi Penelitian → 4 Hasil dan Pembahasan → 5 Kesimpulan dan Saran

1. Pendahuluan

Latar Belakang

Rumusan Masalah

Tujuan Penelitian

Manfaat Penelitian

Batasan Masalah

1. Pendahuluan

Latar Belakang

Analisis Regresi

Regresi Parametrik

Regresi Nonparametrik

Regresi Semiparametrik

1. Pendahuluan

Latar Belakang

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 (Eubank, 1988).

Pendekatan regresi nonparametrik diantanya yaitu spline, kernel, k-nearest neigborhood dan lain-lain.

spline

kernel

1. Pendahuluan

Loader (2000)

Ada beberapa data penelitian yang tidak menunjukkan suatu pola hubungan yang mudah untuk digambarkan dengan fungsi tertentu. Untuk mengatasi kesulitan tersebut digunakan model regresi nonparametrik.

Eubank (1999)

Pendekatan kernel memiliki beberapa kelebihan diantaranya bentuknya lebih fleksibel dan perhitungan matematisnya mudah, sedangkan pendekatan spline memiliki fleksibel yang tinggi dan mampu menangani data yang perilakunya berubah-ubah pada subinterval-subinterval tertentu

Aydin (2007)

Membandingkan teknik smoothing spline dengan kernel pada data produk nasional bruto Turki

Syaranamual (2011)

Pernah melakukan Penelitian tentang Konfidensi Interval untuk Regresi Spline Terboboti dengan data Berat Badan Balita di Kota Blitar

1. Pendahuluan

Rumusan Masalah

1. Bagaimana mendapatkan bentuk interval konfidensi untuk parameter-parameter pada model spline dalam estimator campuran spline dan kernel.
2. Bagaimana mengaplikasikan model estimator campuran spline dan kernel pada data Angka Kematian Bayi (AKB) di Provinsi Jawa Timur.

1. Pendahuluan

Tujuan Penelitian

1. Mendapatkan bentuk interval konfidensi untuk parameter-parameter pada model spline dalam estimator campuran spline dan kernel.
2. Mengaplikasikan model estimator camuran spline dan kernel pada data Angka Kematian Bayi (AKB) di Provinsi Jawa Timur.

1. Pendahuluan

Manfaat Penelitian

1. Mengembangkan wawasan keilmuan dan pengetahuan tentang regresi nonparametrik spline *truncated*.
2. Mengembangkan wawasan keilmuan dan pengetahuan tentang regresi nonparametrik kernel.
3. Mengembangkan wawasan keilmuan dan pengetahuan tentang Estimator Campuran Kernel dan Spline dalam regresi nonparametrik.
4. Hasil penelitian diharapkan menjadi bahan masukan atau acuan dalam penelitian selanjutnya yang akan dilakukan.

1. Pendahuluan

Batasan Masalah

1. Data yang digunakan adalah data tahun 2011.
2. Model spline yang digunakan adalah spline truncated.
3. Estimator kernel yang digunakan adalah estimator kernel Naedaraya Watson.
4. Pemilihan titik knot dan bandwith optimal menggunakan metode GCV.
5. Dalam aplikasi digunakan spline truncated linear dengan satu, dua, tiga.
6. Estimasi interval yang dilakukan dalam penelitian ini hanya untuk komponen spline dalam regresi nonparametrik campuran.

2. Tinjauan Pustaka

Analisis Regresi

Regresi parametrik merupakan metode statistik yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dengan variabel respon, dengan asumsi bahwa telah diketahui bentuk fungsi regresinya.

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i$$

Regresi nonparametrik merupakan pendekatan metode regresi dimana bentuk kurva dari fungsi regresinya tidak diketahui. Kurva fungsi diasumsikan termuat dalam ruang fungsi tertentu.

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

2. Tinjauan Pustaka

Spline dalam Regresi Nonparametrik

$$f(x_i) = \sum_{j=0}^{p} \beta_j x_i^j + \sum_{j=1}^{k} \beta_{j+p}(x_i - k_j)_+^p$$

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

2. Tinjauan Pustaka

Estimator Campuran Spline dan Kernel dalam Regresi Nonparametrik

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

Estimator Kernel
$\hat{g}_\alpha(t_i) = D(\alpha)y$

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

spline

kernel

Estimator Spline
$\hat{f}_{\alpha,\lambda}(x_i) = A(\lambda, \alpha)y$

Estimator campuran Spline dan kernel

$$\hat{\mu}_{\alpha,\lambda}(x_i t_i) = \hat{f}_{\alpha,\lambda}(x_i) + \hat{g}_\alpha(t_i)$$
$$= B\,(\lambda, \alpha)y$$

2. Tinjauan Pustaka

Koefisien Determinasi R^2

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

2. Tinjauan Pustaka

Interval Konfidensi

sampel random $X_1, X_2, \dots, X_n$ yang diambil dari populasi yang berdistribusi $N(\mu, \sigma^2)$ dengan σ^2 diketahui. Untuk mendapatkan interval konfidensi $1-\alpha$ untuk parameter μ dapat menggunakan *Pivotal Quantity :*

$$Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Jika σ tidak diketahui, diganti dengan s yaitu standar deviasi sampel. Dengan demikian, untuk dapat mendapatkan interval konfidensi $1-\alpha$ untuk μ digunakan *Pivotal Quantity :*

$$T = \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{(n-1)}$$

3. Metodelogi Penelitian

Sumber Data

Data yang akan digunakan dalam penelitian ini adalah 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 Penelitia n

Variabel	Nama Variabel	Tipe Variabel
y	Angka Kematian Bayi (AKB)	Kontinu
t_1	Persentase persalinan yang dilakukan dengan bantuan non medis	Kontinu
t_2	Persentase wanita perkawinan pertama kurang dari 17 tahun	Kontinu
t_3	Rata-rata jumlah pengeluaran rumah tangga perkapita sebulan	Kontinu
t_4	Persentase bayi usia 0-11 bulan di beri ASI	Kontinu
t_5	Jumlah sarana kesehatan	Disk

3. Metodelogi Penelitian

Tabel 3.2 Struktur Data Variabel

Kabupaten / Kota	t_1	t_2	t_3	t_4	t_5	y
1	t_{11}	t_{21}	t_{31}	t_{41}	t_{51}	y_1
2	t_{12}	t_{22}	t_{32}	t_{42}	t_{52}	y_2
3	t_{13}	t_{23}	t_{33}	t_{43}	t_{53}	y_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
38	t_{138}	t_{238}	t_{338}	t_{438}	t_{538}	y_{38}

3. Metodelogi Penelitian

Mencari interval konfidensi untuk β_v, $v = 1,2,\ldots,m+r$ pada estimasi campuran regresi nonparametrik spline dan kernel

Menghampiri fungsi $f(x)$ dengan spline truncated dan titik-titik knot $k_1,\ldots,k_r$

→ Menghampiri fungsi $g(t_i)$ dengan fungsi kernel

→ Mencari estimasi untuk β, f dan g dalam estimator campuran spline dan kernel.

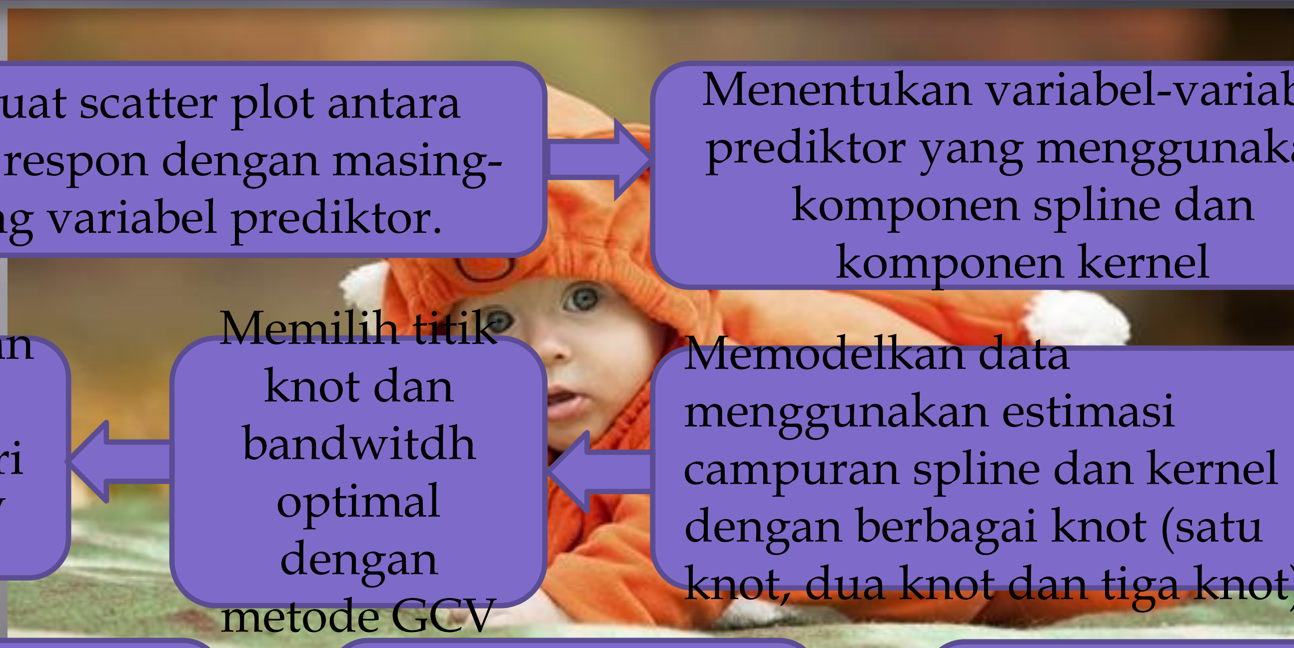
→ Mencari distribusi dari $\hat{\beta}$

→ dicari suatu statistik T yang merupakan Pivotal Quantity untuk β

→ Mencari interval koefisien terpendek untuk parameter β

3. Metodelogi Penelitian

Memodelkan data angka kematian bayi Propinsi Jawa timur, dengan langkah-langkah sebagai berikut:



Membuat scatter plot antara variabel respon dengan masing-masing variabel prediktor.

Menentukan variabel-variabel prediktor yang menggunakan komponen spline dan komponen kernel

Memodelkan data menggunakan estimasi campuran spline dan kernel dengan berbagai knot (satu knot, dua knot dan tiga knot).

Memilih titik knot dan bandwitdh optimal dengan metode GCV

Menetapkan model terbaik dari nilai GCV terkecil

Membuat interval konfidensi terpendek 95% untuk parameter β

Membuat interval konfidensi terpendek 95% untuk interval f.

Menghitung MSE dan R^2

4. Hasil dan Pembahasan

4.1 Interval konfidensi terpendek untuk parameter model komponen spline truncated variansi diketahui

Untuk mendapatkan estimator campuran regresi Spline dan Kernel, pertama kurva regresi $f(x_i)$ dihampiri dengan fungsi Spline truncated derajat m dan titik knot λ.

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

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

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

Dari model diatas di peroleh

$$Y = (I - D(\alpha))^{-1} X(\tilde{\lambda})\hat{\tilde{\beta}}(\tilde{\lambda},\alpha) + (I - D(\alpha))^{-1} + \varepsilon$$

Karena $Y \sim N\left(E(\tilde{Y}),\ Var\,(\tilde{Y})\right)$

$$E(\tilde{Y}) = E\left[\left(I - D(\alpha)\right)^{-1} X(\tilde{\lambda})\hat{\tilde{\beta}}(\tilde{\lambda},\alpha) + \left(I - D(\alpha)\right)^{-1}\varepsilon\right]$$

$$= \left(I - D(\alpha)\right)^{-1} X(\tilde{\lambda})\hat{\tilde{\beta}}(\tilde{\lambda},\alpha) + 0$$

$$Var\,(\tilde{Y}) = Var\left(\left(I - D(\alpha)\right)^{-1} X(\tilde{\lambda})\hat{\tilde{\beta}}(\tilde{\lambda},\alpha) + \left(I - D(\alpha)\right)^{-1}\varepsilon\right)$$

$$= 0 + Var[(I - D)^{-1}\varepsilon]$$

$$= \left(I - D(\alpha)\right)^{-1} Var(\varepsilon)\left(I - D(\alpha)\right)^{-1T}$$

$$= \left(I - D(\alpha)\right)^{-1} (\sigma^2 I)\left(I - D(\alpha)\right)^{-1T}$$

$$= \sigma^2\left(I - D(\alpha)\right)^{-1}\left(I - D(\alpha)\right)^{-1T}$$

$$\tilde{Y} \sim N\left(\left(I - D(\alpha)\right)^{-1} X(\tilde{\lambda})\hat{\tilde{\beta}}(\tilde{\lambda},\alpha),\quad \sigma^2\left(I - D(\alpha)\right)^{-1}\left(I - D(\alpha)\right)^{-1}\right)$$

$$E(\hat{\tilde{\beta}}(\tilde{\lambda},\alpha)) = E(A\tilde{y})$$

$$= A\,E(y)$$

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

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

$$= \beta$$

$$Var\left(\hat{\tilde{\beta}}\right) = Var\,(Ay) = A\,Var\,(y)A^T$$

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

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

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

$$= \left[\left(X\left(\hat{\lambda}\right)\right)^T X\left(\hat{\lambda}\right)\right]^{-1}\left(X\left(\hat{\lambda}\right)\right)^T (I - D(\alpha))\sigma^2\left(I - D(\alpha)\right)^{-1}$$

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

$$= \sigma^2\left[\left(X\left(\hat{\lambda}\right)\right)^T X\left(\hat{\lambda}\right)\right]^{-1}\left(X\left(\hat{\lambda}\right)\right)^T (I - D(\alpha))\sigma^2\left(I - D(\alpha)\right)^{-1}$$

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

$$= \sigma^2\left[\left(X\left(\hat{\lambda}\right)\right)^T X\left(\hat{\lambda}\right)\right]^{-1}$$

$$= \sigma^2 W\left(\hat{\lambda}, \alpha\right)$$

$$W = \left[\left(X\left(\hat{\lambda}\right)\right)^T X\left(\hat{\lambda}\right)\right]^{-1}$$

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

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

Langkah-langkah interval konfidensi

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

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

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

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

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

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

$$\min_{a\epsilon R\ , b\epsilon R}\{\lambda(a,b)\} = \min_{a\epsilon R\ , b\epsilon R}\left\{(b-a) \quad \left[\sqrt{\sigma^2 \omega_i(\tilde{\lambda}, \alpha)}\right]_{jj}\right\}$$

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

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

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

$$\frac{\Omega(a,b,c)}{\partial a} = 0 \Rightarrow -\left[\sqrt{\sigma^2 \omega_i(\tilde{\lambda}, \alpha)}\right]_{jj} - c\varphi(a) = 0 \qquad (4.4)$$

$$\frac{\Omega(a,b,c)}{\partial b} = 0 \Rightarrow \left[\sqrt{\sigma^2 \omega_i(\tilde{\lambda}, \alpha)}\right]_{jj} + c\ \varphi(b) = 0 \qquad (4.5)$$

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

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

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

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

$$P = (a^* \leq T \leq b^*) = 1-\alpha$$

$$a^* \in R, \qquad b \in R \quad a < b^*, \qquad j = 1,2,\ldots,n$$

persamaan diatas dapat dinyatakan menjadi:

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

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

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

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

Dengan syarat

$$\int_{a^*}^{b^*}\varphi(z)dz = 1-\alpha, \text{ atau } \phi(b^*)-\phi(a^*)-(1-\alpha)=0 \qquad (4.11)$$

fungsi lagrange multiple

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

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

$$\frac{\Omega(a^*,b^*,c)}{\partial a^*} = 0 \Rightarrow -\left[\sqrt{MSE\omega_i(\tilde{\lambda},\alpha)}\right]_{jj} - c\varphi(a^*) = 0 \qquad (4.12)$$

$$\frac{\Omega(a^*,b^*,c)}{\partial b^*} = 0 \Rightarrow \left[\sqrt{MSE\omega_i(\tilde{\lambda},\alpha)}\right]_{jj} + c\,\varphi(b^*) = 0 \qquad (4.13)$$

$$\frac{\Omega(a^*,b^*,c)}{\partial c} = 0 \Rightarrow \Omega(b^*) - \Omega(a^*) - (1-\alpha) = 0 \qquad (4.14)$$

Persamaan 4 dan 5 menghasilkan penyelesaian

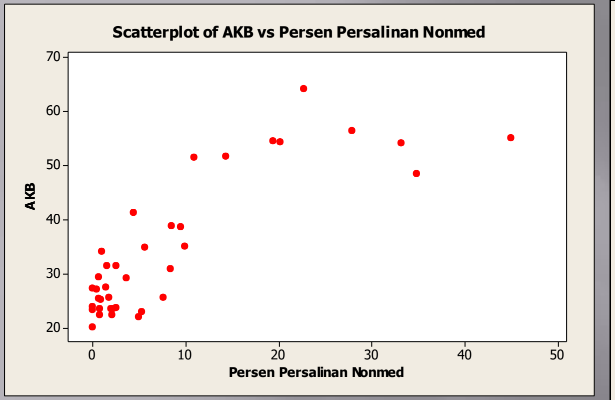
$$\varphi(a^*) = \varphi(b^*) \qquad (4.15)$$

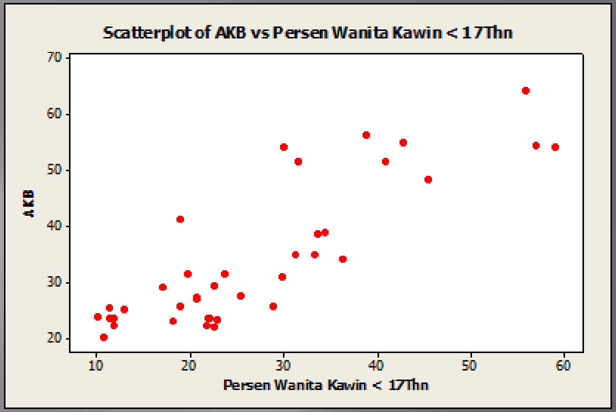
4.3. Analisis Deskriptif Angka Kematian Bayi Propinsi Jawa Timur

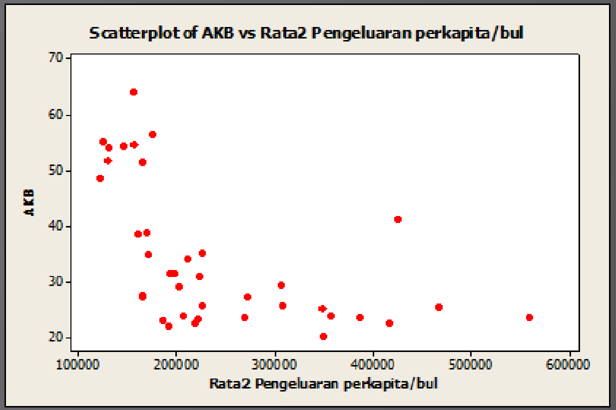
Tabel 4.1 Statistika Deskriptif AKB dengan 5 Variabel Prediktor

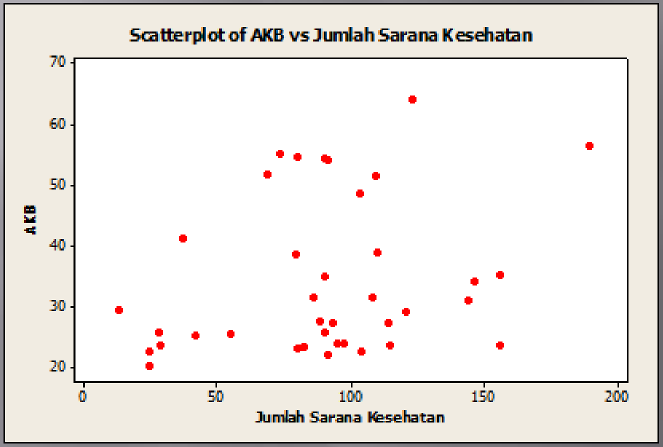
Variabel	Mean	Maksimum	Minimum	Varians
Y	34,18	64,19	20,02	160,7
t_1	8,31	44,99	0	125,29
t_2	27,01	59,09	10,07	169,11
t_3	239.322,00	558.590	122.240	11.164.079.726
t_4	93,95	100	83,92	21,532
t_5	90,10	189	13	1595,39

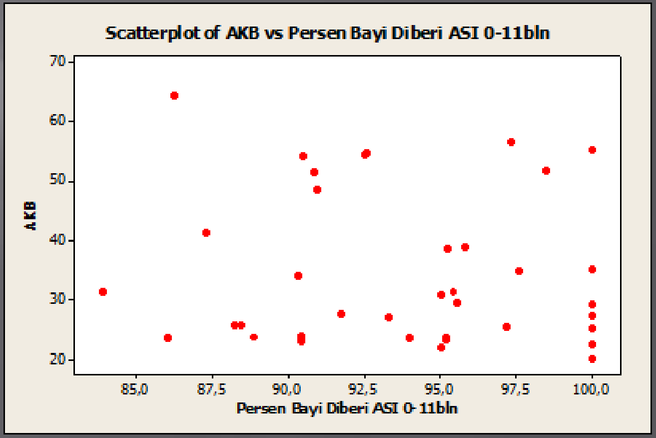
Scatter Plot AKB Prop.Jawa Timur











Pemilihan Titik Knot Optimal *Spline Truncated* dan kernel 1 Titik Knot

Tabel 4.2 Skor GCV untuk 1 *Knot* pada Model campuran spline truncated dan kernel pada provinsi Jawa Timur

Banyak knot	Lokasi knot			Bandwith		GCV	R^2
	x_1	x_2	x_3	α_1	α_2		
	k_1	k_1	k_1				
Knot 1	10.100	21.074	220196.100	3.620	39.520	33.678	84,98%
	11.018	22.075	229101.200	3.948	43.112	33.873	
	9.182	20.074	211291.000	3.292	35.928	33.915	
	11.936	23.075	238006.300	4.276	46.704	34.011	
	12.854	24.076	246911.400	4.604	50.296	34.348	

Estimasi model campuran regresi nonparametrik spline truncated dan kernel dengan 1 titi knot

$$\hat{y} = \hat{\beta}_{10} + \hat{\beta}_{11}x_{1i} + \hat{\beta}_{12}\left(x_{1i} - k_{1i}\right)_{+} + \hat{\beta}_{20} + \hat{\beta}_{21}x_{2i} + \hat{\beta}_{22}\left(x_{2i} - k_{2i}\right)_{+} - \hat{\beta}_{30} + \hat{\beta}_{31}x_{3i} + \hat{\beta}_{32}\left(x_{3i} - k_{3i}\right)_{+} + \sum_{i=1}^{R}\left(\frac{\frac{1}{\alpha_1}K\left(\frac{t_1 - t_{1i}}{\alpha_1}\right)}{\sum_{j=1}^{R}\frac{1}{\alpha_1}K\left(\frac{t_1 - t_{1i}}{\alpha_1}\right)}\right)y_i + \sum_{i=1}^{R}\left(\frac{\frac{1}{\alpha_2}K\left(\frac{t_2 - t_{2i}}{\alpha_2}\right)}{\sum_{j=1}^{R}\frac{1}{\alpha_2}K\left(\frac{t_2 - t_{2i}}{\alpha_2}\right)}\right)y_i$$

$$\hat{y} = 0,0156 + 0,3569x_{1i} + 0,0156\left(x_{1i} - 10,100\right)_{+} + 0,6209 + 0,0156x_{2i} - 0,0001\left(x_{2i} - 21,074\right)_{+} - 0,0522 - 0,1227x_{3i} + 0,0001\left(x_{3i} - 220196,100\right)_{+} + \sum_{i=1}^{38}\left(\frac{\frac{1}{3,6197}K\left(\frac{t_1 - t_{1i}}{3,6197}\right)}{\sum_{j=1}^{38}\frac{1}{3,6197}K\left(\frac{t_1 - t_{1i}}{3,6197}\right)}\right)y_i + \sum_{i=1}^{R}\left(\frac{\frac{1}{39,5202}k\left(\frac{t_2 - t_{2i}}{39,5202}\right)}{\sum_{j=1}^{R}\frac{1}{39,5202}k\left(\frac{t_2 - t_{2i}}{39,5202}\right)}\right)y_i$$

Tabel 4.3 Skor GCV untuk 2 *Knot* pada Model campuran spline truncated dan kernel pada provinsi Jawa Timur

Banyak knot	Lokasi knot						Bandwith		GCV	R^2
	x_1		x_2		x_3		α_1	α_2		
	k_1	k_2	k_1	k_2	k_1	k_2				
Knot 2	12.85	31.21	24.07	44.08	246911.4	425013.5	7.88	86.21	28.83	87,51%
	12.85	28.46	24.07	41.08	246911.4	398298.2	7.84	85.74	28.87	
	12.85	29.38	24.07	42.08	246911.4	407203.3	7.85	85.90	28.91	
	12.85	27.54	24.07	40.08	246911.4	389393.1	7.82	85.58	28.97	
	12.85	31.21	24.07	44.08	246911.4	425013.5	7.88	86.05	29.0	

Estimasi model campuran regresi nonparametrik spline truncated dan kernel dengan 2 titik knot

$$\begin{aligned}
\hat{y} = &\ 1,0292 + 8,3742 x_{1i} + 1,0292 \left(x_{1i} - 12,85 \right)_+ + 3,9799 \left(x_{1i} - 31,21 \right)_+ \\
&+ 1,0292 - 8,9918 x_{2i} - 4,7502 \left(x_{2i} - 24,07 \right)_+ - 6,6493 \left(x_{2i} - 44,08 \right)_+ \\
&+ 2,8409 - 9,5796 x_{3i} + 1,5121 \left(x_{3i} - 246911,4 \right)_+ - 1,0907 \left(x_{3i} - 425013,5 \right)_+ \\
&+ \sum_{i=1}^{38} \left(\frac{\frac{1}{7,8859} K \left(\frac{t_1 - t_{1i}}{7,8859} \right)}{\sum_{j=1}^{38} \frac{1}{7,8859} K \left(\frac{t_1 - t_{1i}}{7,8859} \right)} \right) y_i + \sum_{i=1}^{38} \left(\frac{\frac{1}{86,2140} k \left(\frac{t_2 - t_{2i}}{86,2140} \right)}{\sum_{j=1}^{38} \frac{1}{86,2140} k \left(\frac{t_2 - t_{2i}}{86,2140} \right)} \right) y_i
\end{aligned}$$

Tabel 4.4 Skor GCV untuk 3 *Knot* pada Model campuran spline truncated dan kernel pada provinsi Jawa Timur

Banyak knot	Lokasi knot								
	x_1			x_2			x_3		
	k_1	k_2	k_3	k_1	k_2	k_3	k_1	k_2	k_3
Knot 3	11.01	30.29	31.21	22.07	43.08	44.08	229101.2	416108.4	425013.5
	10.09	30.29	31.21	21.07	43.08	44.08	220196.1	416108.4	425013.5
	11.93	30.29	31.21	23.07	43.08	44.08	238006.3	416108.4	425013.5
	12.85	30.29	31.21	24.07	43.08	44.08	246911.4	416108.4	425013.5
	11.01	29.38	31.21	22.07	42.08	44.08	229101.2	407203.3	425013.5

Bandwith		GCV	R^2
α_1	α_2		
9.340956	102.1399	27.47449	88,97%
8.755214	95.72876	27.65107	
9.894156	108.1948	27.65179	
10.41575	113.9037	28.01197	
9.32701	101.9872	28.21371	

Estimasi model campuran regresi nonparametrik spline truncated dan kernel dengan 3 titik knot :

$$
\begin{aligned}
\hat{y} = {} & 7,7106 + 6,8103x_{1i} + 7,7106(x_{1i} - 11,01)_+ + 3,3473(x_{1i} - 30,29)_+ \\
& + 7,7106(x_{1i} - 31,21)_+ - 8,0600 - 1,1722x_{2i} - 5,5378(x_{2i} - 22,07)_+ \\
& - 5,5484(x_{2i} - 43,08)_+ + 1,4338(x_{1i} - 44,08)_+ - 8,0007 - 8,1938x_{3i} \\
& + 9,8303(x_{3i} - 229101)_+ + 1,1259(x_{3i} - 416108,4)_+ - 1,2371(x_{3i} - 425013,5)_+ \\
& + \sum_{i=1}^{38}\left(\frac{\frac{1}{9,3409}K\left(\frac{t_1 - t_{1i}}{9,3409}\right)}{\sum_{j=1}^{38}\frac{1}{9,3409}K\left(\frac{t_1 - t_{1i}}{9,3409}\right)}\right)y_i + \sum_{i=1}^{38}\left(\frac{\frac{1}{102,1398}K\left(\frac{t_2 - t_{2i}}{102,1398}\right)}{\sum_{j=1}^{38}\frac{1}{102,1398}K\left(\frac{t_2 - t_{2i}}{102,1398}\right)}\right)y_i
\end{aligned}
$$

Tabel 4.5 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

Interval Konfidensi Provinsi Jawa Timur

Dengan model *spline Truncated* dengan tiga titik *knot* , akan dibangun interval konfidensi 95% dengan persamaan sebagai berikut:

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

Table 4.5 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_{12}	0.9954	0.68103	0.36659
	β_{11}	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	0.000098303	0.000037802
	β_{33}	0.00209	0.0011259	0.00016169
	β_{34}	-0.00023	-0.001237	-0.0022356

Sehingga dapat disimpulkan bahwa model terbaik yaitu pada tiga titik knot dimana nilai MSE dan *R2* yang terbaik berturut-turut adalah 4,11 dan 88,97% dan semua parameternya memenuhi batas atas, batas bawah dan

5. Kesimpulan

5.1 Kesimpulan

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

Interval Konfidensi untuk parameter β diberikan oleh:

$$P\left[\left[\left[X\left(\tilde{\lambda}\right)^{T} X\left(\tilde{\lambda}\right)\right]^{-1}\left(X\left(\tilde{\lambda}\right)^{T}\right)\left(I-D(\alpha)\right)\tilde{y}\right]_{j} - b\left[\sqrt{\sigma^{2} w_{j}\left(\tilde{\lambda},\alpha\right)}\right]_{jj} \leq \beta_{j}\right.$$

$$\left.\leq \left[\left[X\left(\tilde{\lambda}\right)^{T} X\left(\tilde{\lambda}\right)\right]^{-1}\left(X\left(\tilde{\lambda}\right)^{T}\right)\left(I-D(\alpha)\right)\tilde{y}\right]_{j} - a\left[\sqrt{\sigma^{2} w_{j}\left(\tilde{\lambda},\alpha\right)}\right]_{jj}\right] = 1-\alpha$$

5.1 Kesimpulan

2. Interval konfidensi terpendek untuk parameter model komponen spline truncated variansi tidak ketahui.

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

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

Interval Konfidensi untuk parameter β diberikan oleh:

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

5.1 Kesimpulan

3. Interval Konfidensi Provinsi Jawa Timur

Berdasarkan nilai GCV terkecil diperoleh model *spline Truncated* dengan tiga titik *knot* yang memiliki nilai GCV terkecil yaitu 27.47449 dan dimana nilai MSE dan R^2 yang terbaik berturut-turut adalah 4,11 dan 88,97%. kemudian akan dibangun interval konfidensi 95% dengan persamaan sebagai berikut:

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

$\hat{\beta}$

5.1 Kesimpulan

		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	0.000098303	0.000037802
	β_{33}	0.00209	0.0011259	0.00016169
	β_{34}	-0.00023	-0.001237	-0.0022356

Sehingga dapat disimpulkan bahwa model terbaik yaitu pada tiga titik knot dimana nilai MSE dan *R2* yang terbaik berturut-turut adalah 4,11 dan 88,97% dan semua parameternya memenuhi batas atas, batas bawah dan beta

5.2 Saran

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

1. Dalam pengambilan keputusan untuk menekan jumlah Angka Kematian Bayi hendaknya disesuaikan dengan wilayah kabupaten kota karena kondisi wilayah dan karakteristik budaya yang berbeda.
2. Dapat menggunakan inferensi statistik lainnya selain interval konfidensi pada penelitian berikutnya.

DAFTAR PUSTAKA

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.

Aydin, Dursun. 2007. *A Comparison of the Nonparametric Regression Models using Smoothing Spline and Kernel Regression*. World Academy of Science, Engineering and Technology, 36, 253-257, Turkey.

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

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

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

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

Eubank, R.L., 1999, *Spline Smoothing and Nonparametric Regression 2nd Edition*, Marcel Deker, New York.
Hardle, W., (1990), *Applied Nonparametric Regression*, Cambrige University Press, Boston.
KemenKes, 2012, *Data Informasi Kesehatan Provinsi Jawa Timur*, Kementrian Kesehatan Republik Indonesia, Jakarta.
Laome, L., (2008), *Estimasi Parameter Pada Regresi Semiparametrik Untuk Data Longitudinal*. http://jurnal.unhalu.ac.id. Tanggal akses : 5 Oktober 2011.
Puspitasari, Elsha., (2012), *Model Regresi Spline Knot Optimal Untuk Mengetahui Faktor-Faktor Yang Mempengaruhi Jumlah Kematian Bayi di Jawa Timur*, Skripsi, Jurusan Matematika, FMIPA UNESA, Surabaya.
Sukarsa, K.G., (2012), *Estimator Kernel Dalam Model Regresi Nonparametrik*. Jurnal Matematika Vol. 2 No. 1, Juni 2012. ISSN : 1693-1394
Tripena, (2011), *Penentuan Model Regresi Spline Terbaik.* Prosiding Seminar Nasional Statistika Universitas Diponegoro 2011. ISBN: 978-979-097-142-4
Wasono, (2014), *Analisis Model Regresi Nonparametrik Multivariabel Heteroskedastik Spline.* Thesis, Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS), Surabaya.

TERIMA KASIH