PENDUGAAN AREA KECIL MENGGUNAKAN PENDEKATAN PENALIZED SPLINE PADA PENDUGAAN PENGELUARAN PERKAPITA TINGKAT KECAMATAN DI KABUPATEN SUMENEP

Nama Mahasiswa : Zulhan Widya Baskara

NRP : 1311 201 015

Pembimbing : Dr. Agnes Tuti Rumiati, M.Sc

ABSTRAK

Pendugaan pada subpopulasi berdasarkan hasil survei pada populasi kemungkinan akan menghasilkan pendugaan yang bias dan varians yang besar yang disebabkan oleh jumlah sampel yang kurang representatif untuk mewakili data (subpopulasi). Salah satu cara untuk menduga subpopulasi tersebut adalah dengan *small area estimation* (SAE). SAE umumnya menggunakan pemodelan parametik untuk menghubungkan statistik area kecil dengan variabel-variabel pendukungnya. Pemodelan ini kurang fleksibel dalam menyesuaikan dengan pola data hasil survei yang mungkin saja tidak mirip sama sekali dengan distribusi formal yang ada. Sehingga pendekatan nonparametrik menjadi alternatif pilihan. Pada laporan ini peneliti akan mencoba menggunakan pemodelan nonparametrik untuk mengatasi masalah ini yaitu dengan menggunakan metode nonparametrik *P-Spline.* Pendugaan dengan menggunakan pendekatan *P-Spline* ini kemudian digunakan untuk menduga pengeluaran perkapita pada masing-masing kecamatan di Kabupaten Sumenep. Evaluasi hasil pendugaan dilakukan melihat nilai MSE (*Mean Square Error*) hasil pendugaan dengan menggunakan metode resampling *jackknife.*

Kata Kunci : *Jackknife*, *P-spline*, Pengeluaran Perkapita, *Small area estimation* (SAE), Subpopulasi

SMALL AREA ESTIMATION USING P-SPLINE TO ESTIMATING EXPENDITURE PER CAPITA DISTRICT LEVEL IN SUMENEP

Name : Zulhan Widya Baskara

NRP : 1311 201 015

Supervisor : Dr. Agnes Tuti Rumiati, M.Sc

ABSTRACT

Estimation on subpopulations based on population surveys will produce bias estimation and large of variance values due to less number of samples to represent the data (subpopulation). One way to estimate the subpopulation is by using small area estimation (SAE). SAE generally uses parametric modeling to connect the small area statistics with variables supporters. This modeling is less flexible in adjusting to the pattern of survey data that may not resemble at all the existing formal distribution. So the non-parametric approach to be an alternative option. In this report, the researchers will try to use non-parametric modeling to overcome this problem is by using *Penalized-spline* nonparametric methods. Estimation using *P-spline* approach is then used to estimate the percapita expenditure in each district in Sumenep. Evaluation of the test results by looking at the value of MSE (Mean Square Error) prediction results by using the jackknife resampling method.

Key Words : *Jackknife*, *P-spline*, percapita expenditure, *Small area estimation* (SAE), Subpopulation

BAB II
TINJAUAN PUSTAKA

2.1 Model Pendugaan Area Kecil

Pelaksanaan survei dilakukan untuk melakukan pendugaan parameter populasi. Namun seiring berjalannya waktu, seringkali ketertarikan pada suatu kejadian/fenomena tidak hanya pada populasi tersebut, tapi juga pada subpopulasinya. Menurut Rao (2003), subpopulasi (domain) didefinisikan sebagai wilayah atau area geografi atau kelompok sosio-demografi yang lebih kecil dari suatu wilayah populasi. Parameter subpopulasi dapat berupa peubah yang diskrit, seperti proporsi pengangguran, banyak kelahiran bayi maupun dapat berupa peubah yang kontinu, seperti pendapatan perkapita.

Pendugaan secara langsung (*direct estimation*) pada subpopulasi seringkali menyebabkan tingkat keakuratan yang rendah yang dikarenakan ukuran sampel pada subpopulasi yang kecil atau bahkan tidak ada sama sekali (jika survei yang diambil berskala besar). Sehingga solusi yang digunakan agar dugaan menghasilkan tingkat keakuratan yang tinggi adalah dengan melakukan pendugaan tak langsung (*indirect estimation*), yaitu pendugaan dengan menghubungkan informasi pada dareah tersebut dengan informasi dari luar area ataupun dari luar survei dengan model yang sesuai atau biasa disebut dengan pendugaan area kecil atau *small area estimation*.

Berdasarkan data pendukungnya, pendugaan tidak langsung dapat di bedakan menjadi tiga, yaitu *domain indirect, time indirect,* dan *domain and time indirect. Domain indirect* adalah pendugaan tidak langsung dengan menggunakan informasi dari area yang lain pada satu waktu*, time indirect* adalah pendugaan tidak langsung dengan menggunakan informasi dari area yang sama pada waktu berbeda*,* dan *domain and time indirect* yaitu pendugaan tidak langsung dengan menggunakan informasi dari area dan waktu yang berbeda.

Sampel suatu area dikatakan besar jika ukuran sampelnya cukup banyak sehingga jika diestimasi dengan menggunakan *direct estimation* menghasilkan dugaan yang cukup akurat. Sedangkan dikatakan kecil jika ukuran sampelnya tidak cukup besar jika diestimasi menggunakan *direct estimation* untuk menghasilkan dugaan yang cukup akurat. Domain yang kecil ini disebut juga dengan *Small Area* (area kecil).

Terdapat dua ide utama yang digunakan dalam mengembangkan model *small area estimation* seperti yang dikembangkan oleh Fay & Heriot (1979), yaitu:

1. Model pengaruh tetap (*fixed effect model*) dimana asumsi bahwa keragaman di dalam *small area* variabel respon dapat diterangkan seluruhnya oleh hubungan keragaman yang bersesuaian pada informasi tambahan.
2. Model pengaruh acak (*random effect*) area kecil dimana asumsi keragaman spesifik *small area* tidak dapat diterangkan oleh informasi tambahan.

Gabungan antara *fixed* dan *random effect* diatas membentuk model campuran (*mixed model*). Karena variabel respon diasumsikan berdistribusi normal maka pendugaan area kecil yang dikembangkan merupakan bentuk khusus dari *General Linear Mixed Model* (GLMM).

Secara esensial, terdapat dua tipe model pada pendugaan area kecil ,yaitu:

1. Model berbasis area,yaitu model yang didasarkan pada ketersediaan data pendukung terdapat untuk level area tertentu, misalkan :

   Diberikan *M* area kecil pada populasi dengan sebanyak *m* area yang terambil sebagai sampel. Asumsikan bahwa parameter yang diperhatikan dalam area kecil ke-i, misal θ_i dapat dinyatakan dengan sebuah fugsi yang menghubungkan parameter tersebut dengan variabel bantu, $\boldsymbol{z_i} = (z_{1i}, \dots, z_{pi})^T$ dimana z_i adalah suatu vektor, i adalah banyaknya area dan p adalah banyaknya peubah penyerta. Model linear yang menjelaskan hubungan tersebut adalah :

   $$\theta_i = g(\bar{Y}_i) = z_i^T \beta + b_i v_i \qquad (1)$$

   dengan $i = 1,\dots,m$,dan $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ merupakan vektor koefisien regresi untuk data pendukung z_i , b_i merupakan konstanta dan $v_i \sim N(0, \sigma_v^2)$ sebagai pengaruh acak area yang diasumsikan menyebar normal.

   Parameter θ_i, dapat diketahui dengan mengasumsikan bahwa model penduga langsung $\hat{\theta}_i$ telah diketahui, yaitu:

$$\hat{\theta}_i = g\left(\hat{\bar{Y}}_i\right) = \theta_i + e_i \;,\; i = 1,,m \tag{2}$$

dengan $E(e_i|\theta_i) = 0$ dan $V_p(e_i|\theta_i) = \sigma_{ei}^2$

Rao (2003) menjelaskan bahwa model SAE berbasis area terdiri dari dua komponen model yaitu komponen model pendugaan langsung (persamaan (2)) dan pendugaan tak langsung (persamaan (1)).

persamaan (1) dan (2) digabung sehingga didapat model campuran

$$\hat{\theta}_i = z_i^T \beta + b_i v_i + e_i \tag{3}$$

Model area kecil pada persamaan tersebut merupakan bentuk khusus dari model linear campuran (*general linear mixed model*) yang dikenal sebagai model Fay-Heriot, dimana keragaman peubah respon didalam area kecil diasumsikan dapat diterangkan oleh hubungan antara peubah respon dengan informasi tambahan yang disebut model pengaruh tetap (*fixed effect*) yaitu *β*. Selain itu terdapat komponen keragaman spesifik area kecil yang tidak dapat diterangkan oleh informasi tambahan yang disebut dengan komponen pengaruh acak area kecil (*random effect*) yaitu v_i.

2. Model berbasis unit, dimana data-data pendukung yang tersedia bersesuaian secara individu dengan data respon.

   Misal terdapat variabel penyerta $\boldsymbol{x_{ij}} = (x_{ij1}, \dots, x_{ijp})^T$ tersedia untuk setiap elemen ke-*j* pada area ke-*i* sehingga didapatkan suatu model :

   $$y_{ij} = x_{ij}^T \beta + v_i + e_{ij} \tag{4}$$

   dengan *j=1,...,n* , *i=1,..,m* , $v_i \sim N(0, \sigma_v^2)$ dan $e_i \sim N(0, \sigma_{ei}^2)$.

   Bentuk model (4) dapat dinyatakan dalam bentuk matriks:

   $$\boldsymbol{y_i} = \begin{bmatrix} \boldsymbol{y_i} \\ \boldsymbol{y_i^*} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X_i} \\ \boldsymbol{X_i^*} \end{bmatrix} \beta + v_i \begin{bmatrix} \boldsymbol{1_i} \\ \boldsymbol{1_i^*} \end{bmatrix} + \begin{bmatrix} \boldsymbol{e_i} \\ \boldsymbol{e_i^*} \end{bmatrix}$$

Superscript * menyatakan area yang tidak terambil dalam sampel.

2.2 Model Regresi *Penalized-Spline*

Regresi nonparametrik digunakan apabila pola hubungan antar variabel respon dengan variabel prediktor diasumsikan tidak diketahui. Kurva regresi nonparametrik hanya diasumsikan smooth (mulus) dalam arti termuat di dalam suatu ruang fungsi tertentu sehingga mempunyai tingkat fleksibilitas yang tinggi (Eubank, 1988).

Spline merupakan salah satu model regresi nonparametrik yang merupakan model polinomial terputus (*truncated polinomial*) atau tersegmen dimana sifat segmen inilah yang memberikan fleksibilitas yang lebih baik dibanding model polinomial biasa. Sifat ini memungkinkan model regresi spline menyesuaikan diri secara efektif terhadap karakteristik lokal dari data.

Model nonparametrik secara umum adalah sebagai berikut:

$$y_i = m(x_i) + e_i \tag{5}$$

Dengan y_i adalah variabel respon, fungsi $m(x_i)$ adalah fungsi yang *smooth* yang tidak diketahui bentuknya. Variabel x_i sebagai variabel prediktor dengan $e_i \sim N(0, \sigma^2)$. Agar dapat menangani struktur $m(x_i)$ yang tidak linear, Didefinisifkan K buah knot $k_1,..,k_K$ dan dengan mengambil basis fungsi polinomial terputus diperoleh model berikut :

$m(x_i) = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + \sum_{j=1}^{K} \gamma_j (x_i - k_j)_+^p$, dimana:

p adalah derajat *spline*, $(x_i - k_j)_+ = maks\{0, (x_i - k_j)\}$, k_j dimana $j=1,..,K$ merupakan himpunan titik knot. Dengan menetapkan $\boldsymbol{\beta} = (\beta_0 \ldots \beta_p)^T$ adalah vektor koefisien parametrik , $\boldsymbol{\gamma} = (\gamma_1 \ldots \gamma_k)^T$ adalah vektor koefisien *spline*, $X = \left[1 \;\; x_i \ldots x_i^p\right]_{1 \le i \le n}$, $Z = \left[(x_i - k_1) \ldots . (x_i - k_K)_+^p\right]_{1 \le i \le n}$,

dengan $(x_i - k_j)_+^p \begin{cases} = (x_i - k_j)_+^p & \text{untuk } x_i \ge k_j \\ = 0 & \text{untuk } x_i < k_j \end{cases}$

sehingga model (5) dapat ditulis :

$$y_i = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + \sum_{j=1}^{K} \gamma_j (x_i - k_j)_+^p + e_i$$

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{e} \tag{6}$$

Dimana $\boldsymbol{Y} = (y_1 \ldots y_n)^T$.

Model (6) disebut sebagai regresi *spline smoothing* (Opsomer, 2004). Dari bentuk matematis fungsi *spline* pada model tersebut menunjukkan bahwa *spline* merupakan model polinomial terputus, tetapi masih bersifat kontinu pada knot-knotnya.

Knot dapat diartikan sebagai suatu titik fokus dalam fungsi *spline* sedemikian sehingga kurva yang dibentuk tersegmen pada titik tersebut. Titik knot merupakan salah satu hal yang sangat penting dalam pendekatan *spline*.

Strategi yang digunakan untuk memilih dan menentukan lokasi knot dengan tepat sangat dibutuhkan agar diperoleh model *spline* yang optimal. Jika jumlah knot terlampau banyak maka model yang dihasilkan akan *overfitting.*

Salah satu metode pemilihan titik knot optimal adalah dengan menggunakan *Generalized Cross Validation* (GCV).

Definisi GCV dapat ditulis sebagai berikut:

$$\text{GCV}(\boldsymbol{K}) = \frac{MSE(K)}{[n^{-1} trace\ (I - A(K))]^2}$$

Dimana $MSE(\boldsymbol{K}) = n^{-1}y^T(\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{K}))^T(\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{K}))\boldsymbol{y}, \boldsymbol{K} = (K_1, K_2, .., K_N)$ adalah titik knots dan martiks $\boldsymbol{A}(\boldsymbol{K})$ diperoleh dari persamaan $\hat{y} = A(\boldsymbol{K})\boldsymbol{y}$.

Smoothing *spline*s memakai knot untuk nilai yang berbeda dari variabel prediktor dan mengendalikan overfitting dengan menambahkan suatu *penalty*. Regresi *Penalized spline* merupakan suatu pendekatan smoothing yang populer karena kesederhanaan dan fleksibilitasnya. Secara komputasi umum, pemodelan *Penalized spline* juga sangat murah dan memberikan pemilihan knot yang fleksibel (hall dan Opsomer, 2005)

Sehingga salah satu alternatif untuk mengoptimalkan fit model terhadap data adalah dengan menambahkan *penalty* pada parameter *spline.* Dengan cara ini dapat dipilih jumlah knot yang banyak dan mencegah *overfitting* dengan menempatkan kendala (*constraint*). Krivobokova (2006) mengemukakan salah satu contoh kendala yang diberikan yaitu meminimumkan $||y - X\beta - Z\gamma||^2$, dengan kendala $||\gamma||^2 \leq c$, untuk konstanta *c* yang nonnegatif. Dengan menggunakan pengganda Lagrange λ, diperoleh persamaan:

$$J(\beta, \gamma) = \{||y - X\beta - Z\gamma||^2 + \lambda\gamma^T\gamma\},$$

untuk mendapatkan titik kritis sebagai penyelesaiannya, maka persamaan Lagrange perlu diturunkan terhadap parameter yang akan diestimasi.

$$J(\theta) = \{||y - C\theta||^2 + \lambda\theta^T D\theta$$

$$J(\theta) = (y - C\theta)^T(y - C\theta) + \lambda\theta^T D\theta$$

$$J(\theta) = y^T y - 2(C\theta)^T y + (C\theta)^T C\theta + \lambda\theta^T D\theta$$

Dengan $C = [X\ \ Z]$, $\theta = [\beta^T\ \ \gamma^T]^T$, $D = diag(0_{(p+1)\times(p+1)}, \boldsymbol{I_k})$, dan $\lambda \geq 0$

Bentuk ini kemudian diturunkan terhadap θ akan menghasilkan :

$$\frac{J(\theta)}{\partial\theta} = -2C^T y + 2C^T C\theta + 2\lambda D\theta = 0$$

$$C^T C\theta + \lambda D\theta = C^T y$$

$$\hat{\theta} = (C^T C + \lambda D)^{-1} C^T y$$

Sehingga

$$\hat{y} = C\hat{\theta} = C(C^T C + \lambda D)^{-1} C^T y \quad (7)$$

(Kale, 2008)

Model (6) dapat dianggap sebagai model linear campuran (*linear mixed model*), yaitu dengan menetapkan β sebagai pengaruh tetap dan γ sebagai pengaruh acak. Handerson (1950) dalam Robinson (1991) mengemukakan bahwa estimasi *best linear unbiased predictor* (BLUP) dari β dan γ dalam model linear campuran pada persamaan (6) dapat diperoleh dengan memaksimumkan fungsi kepadatan gabungan antara **y** dan γ. Diketahui $\boldsymbol{Y} \sim N\ (\boldsymbol{X\beta} + \boldsymbol{Z\gamma}, \boldsymbol{R})$, dan $\boldsymbol{\gamma} \sim N\ (\boldsymbol{0}, \boldsymbol{G})$, dengan $\mathbf{G} = \boldsymbol{\sigma_\gamma^2 I_K}$ dan $\boldsymbol{R} = \boldsymbol{\sigma_e^2 I_n}$, maka fungsi likelihood dari fungsi kepadatan gabungannya adalah:

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{G}, \boldsymbol{R}|\boldsymbol{y}) = (2\pi)^{-\frac{n}{2}}(|\boldsymbol{R}|)^{-\frac{1}{2}}(|\boldsymbol{G}|)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\begin{pmatrix}\boldsymbol{\gamma}\\ \boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{Z\gamma}\end{pmatrix}^T\begin{pmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{\gamma}\\ \boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{Z\gamma}\end{pmatrix}\right\}$$

Memaksimumkan fungsi likelihood di atas adalah sama dengan meminimumkan Q terhadap β dan γ.

Dimana $Q = \begin{pmatrix}\gamma\\ y - X\beta - Z\gamma\end{pmatrix}^T \begin{pmatrix}G & 0\\ 0 & R\end{pmatrix}^{-1} \begin{pmatrix}\gamma\\ y - X\beta - Z\gamma\end{pmatrix}$, sehingga :

$$Q = \gamma^T G^{-1}\gamma + (y - X\beta - Z\gamma)^T R^{-1}(y - X\beta - Z\gamma)$$

$$Q = (y - C\theta)^T R^{-1}(y - C\theta) + \theta^T J\theta$$

$$Q = y^T R^{-1} y - y^T R^{-1} C\theta - (C\theta)^T R^{-1} y + (C\theta)^T R^{-1} C\theta + \theta^T J\theta$$

$$Q = y^T R^{-1} y - 2(C\theta)^T R^{-1} y + (C\theta)^T R^{-1} C\theta + \theta^T J\theta$$

$$\frac{\partial Q}{\partial \theta} = -2C^T R^{-1} y + 2C^T R^{-1} C\theta + 2J\theta = 0$$

$$C^T R^{-1} C\theta + J\theta = C^T R^{-1} y$$

$$\hat{\theta} = (C^T R^{-1} C + J)^{-1}\ C^T R^{-1} y$$

Karena $R = \sigma_e^2 I_n$ dan $J = diag(0_{(p+1)(p+1),}\sigma_\gamma^2 I_K)$, maka

$$\hat{\theta} = (C^T \frac{1}{\sigma_e^2} I_\gamma C + \frac{1}{\sigma_\gamma^2} D)^{-1}\ C^T \frac{1}{\sigma_e^2} I_n y$$

$$\hat{\theta} = (\frac{1}{\sigma_e^2} C^T C + \frac{1}{\sigma_\gamma^2} D)^{-1} \frac{1}{\sigma_e^2}\ C^T y$$

$\hat{\theta} = (C^T C + \frac{\sigma_e^2}{\sigma_\gamma^2} D)^{-1}\ C^T y$, dengan $D = diag(0_{(p+1)(p+1),}\sigma_\gamma^2, I_K)$.

Maka diperoleh:

$$\hat{y} = C\hat{\theta} = C(C^T C + \frac{\sigma_e^2}{\sigma_\gamma^2} D)^{-1}\ C^T y \tag{8}$$

Dari persamaan (6) dan (7) dan dengan menganggap γ pada persamaan (6) sebagai pengaruh acak, maka estimasi parameter pada model *p-spline* dapat dilakukan dengan mengikuti kerangka kerja model linear campuran. Nilai parameter penghalus $\lambda = \frac{\sigma_e^2}{\sigma_\gamma^2}$ dapat diperoleh dengan mengestimasi parameter σ_e^2 dan σ_γ^2 dalam kerangka kerja model linear campuran (Kale, 2008)

Estimasi nilai komponen error σ_e^2 dan σ_γ^2 dalam kerangka kerja model linear campuran dapat dilakukan dengan metode *restricted maximum likelihood* (REML) yang merupakan modifikasi dari metode *maximum likelihood* standar untuk memperoleh estimasi varians yang tidak bias.

2.3 Pendugaan Area Kecil dengan Menggunakan Regresi *Penalized-Spline*

Pendugaan area kecil (SAE) adalah pendekatan yang digunakan untuk mengungkapkan hubungan antar variabel *interest* dengan variabel pendukung sebagai model linear dengan tambahan pengaruh acak area kecil.

T adalah small area, $U_1,\dots,U_T$ parameter yang akan diestimasi.

Definisikan $d_{it} = I_{\{i\epsilon U_t\}}$, $\boldsymbol{d}_i = (d_{i1},\dots,d_{iT})$,

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},\ \boldsymbol{X} = \begin{bmatrix} 1 & x_1 & \cdots & x_1^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^p \end{bmatrix}$$

$$\boldsymbol{Z} = \begin{bmatrix} (x_1 - k_1)_+^p & \cdots & (x_1 - k_K)_+^p \\ \vdots & \ddots & \vdots \\ (x_n - k_1)_+^p & \cdots & (x_n - k_K)_+^p \end{bmatrix},$$

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

Opsomer (2008) menggunakan *P-spline* untuk mengestimasi area kecil dengan menambahkan pengaruh acak area kecil pada persamaan (6), sehingga diperoleh persamaan :

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{Du} + e \tag{9}$$

Persamaan diatas terdiri dari fungsi *spline* yang merupakan fungsi nonparametrik $\boldsymbol{X\beta} + \boldsymbol{Z\gamma}$ dan pengaruh acak area kecil ($\boldsymbol{Du}$).

Dimana $\boldsymbol{D} = (\boldsymbol{d}_1, \dots, \boldsymbol{d}_n)^T$ adalah matriks covarian dan $\boldsymbol{u}$ adalah vektor pengaruh area kecil.

Dan $\gamma \sim \left(0, \sum_\gamma\right)$ dengan $\sum_\gamma \equiv \sigma_\gamma^2 I_K$

$u \sim (0, \sum_u)$ dengan $\sum_u \equiv \sigma_u^2 I_T$

$e \sim (0, \sum_e)$ dengan $\sum_e \equiv \sigma_e^2 I_n$ (10)

Estimasi pengaruh tetap $\boldsymbol{\beta}$ dapat dilakukan dengan metode *Maximum Likelihood* (ML) dengan menanggap $\boldsymbol{\gamma}$ dan $\boldsymbol{u}$ sebagai pengaruh acak. Model (9) dapat ditulis menjadi :

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e*}$$

Dimana $\boldsymbol{\varepsilon*} = \boldsymbol{Z\gamma} + \boldsymbol{Du} + \boldsymbol{e}$

Sehingga $Var\,(\mathbf{Y}) = Var\,(\boldsymbol{X\beta} + \boldsymbol{e*})$

$= Var\ (\boldsymbol{Z\gamma} + \boldsymbol{Du} + \boldsymbol{e})$

$= \boldsymbol{Z}\, Var\,(\boldsymbol{\gamma})\, \boldsymbol{Z}^T + \boldsymbol{D}\, Var(\boldsymbol{u})\, \boldsymbol{D}^T + Var(\boldsymbol{e})$

$= \boldsymbol{Z} \sum_\gamma \boldsymbol{Z}^T + \boldsymbol{D} \sum_u \boldsymbol{D}^T + \sum_e$

$\equiv \boldsymbol{V}$

Atau diperoleh

$$Var(\boldsymbol{Y}) \equiv \boldsymbol{V} = \boldsymbol{Z}\textstyle\sum_{\boldsymbol{\gamma}}\boldsymbol{Z}^{\boldsymbol{T}} + \boldsymbol{D}\sum_{\boldsymbol{u}}\boldsymbol{D}^{\boldsymbol{T}} + \sum_{\boldsymbol{e}} \tag{11}$$

Dimana *V* adalah matrix varian covarian dari *Y.*

Dengan menganggap nilai $\boldsymbol{V}$ diketahui maka fungsi likelihood (L) pada persamaan (9) adalah :

$$L(\boldsymbol{\beta}|\boldsymbol{Y}) = (2\pi)^{-\frac{\sum_{h=1}^{m} n_h}{2}} (|\boldsymbol{V}|)^{-\frac{1}{2}} \left[\exp\left(-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X\beta})^T \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{X\beta})\right)\right]$$

Log *likelihood*nya adalah:

$$Ln\, L\, (\boldsymbol{\beta}|\boldsymbol{Y}) = -\frac{\sum_{h=1}^{m} n_h}{2} \ln(2\pi) - \frac{1}{2}\ln(|\boldsymbol{V}|) - \frac{1}{2}(\mathbf{Y} - \mathbf{X\beta})^{\mathrm{T}} \mathbf{V}^{-\mathbf{1}} (\mathbf{Y} - \mathbf{X\beta})$$

Untuk memperoleh estimasi parameter β adalah dengan memaksimumkan fungsi *likelihood* atau *log likelihood*nya dengan cara menurunkannya terhadap $\boldsymbol{\beta}$.

$$\frac{\partial\, Ln\, L\,(\boldsymbol{\beta}|\boldsymbol{Y})}{\partial \boldsymbol{\beta}^{\boldsymbol{T}}} = -\frac{1}{2}\frac{\partial(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{V}^{-\mathbf{1}}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{\boldsymbol{T}}}$$

$$= -\frac{1}{2}\frac{\partial(\mathbf{Y}^{\mathbf{T}}\mathbf{V}^{-\mathbf{1}}\mathbf{Y}-2\boldsymbol{\beta}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-\mathbf{1}}\mathbf{Y}+\boldsymbol{\beta}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-\mathbf{1}}\mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{\boldsymbol{T}}}$$

$$= -\frac{1}{2}(-2\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-\mathbf{1}}\mathbf{Y}+2\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-\mathbf{1}}\mathbf{X}\boldsymbol{\beta}) = 0$$

Maka $\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-\mathbf{1}}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^{\mathbf{T}}\mathbf{V}^{-\mathbf{1}}\mathbf{Y}$

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-\mathbf{1}}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-\mathbf{1}}\mathbf{Y} \qquad (12)$$

Selanjutnya akan di uraikan mengenai prediktor terbaik dari suatu vektor acak $\boldsymbol{\gamma}$. Jika $\widehat{\boldsymbol{\gamma}}$ adalah prediktor terbaik dari $\boldsymbol{\gamma}$ maka MSE dari prediktor ini adalah:

$$E[(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})^{\boldsymbol{T}}\boldsymbol{A}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})] = \int\int[(\boldsymbol{\gamma}-\boldsymbol{\gamma})^{\boldsymbol{T}}\boldsymbol{A}(\boldsymbol{\gamma}-\boldsymbol{\gamma})f(\boldsymbol{\gamma},\boldsymbol{y})]d\boldsymbol{y}d\boldsymbol{\gamma} \qquad (13)$$

dengan $\boldsymbol{A}$ matriks simetris definit positif dengan $f(\boldsymbol{\gamma},\boldsymbol{y})$ adalah fungsi kepadatan peluang gabungan γ dan $\boldsymbol{y}$. Dengan meminimumkan persamaan diatas maka diperoleh prediktor terbaik untuk $\boldsymbol{\gamma}$ yang juga merupakan prediktor tak bias pada sample $\boldsymbol{y}$ adalah:

$$\widehat{\boldsymbol{\gamma}} = E(\boldsymbol{\gamma}|\boldsymbol{y})$$

$$= \boldsymbol{\mu}_{\boldsymbol{\gamma}} + cov(\boldsymbol{\gamma},\boldsymbol{y}^{\boldsymbol{T}})\big(cov(\boldsymbol{y})\big)^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_{\boldsymbol{y}})$$

$$= \mathbf{0} + \textstyle\sum_{\boldsymbol{\gamma}}\boldsymbol{Z}^{\boldsymbol{T}}\boldsymbol{V}^{-\mathbf{1}}\big(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\big)$$

$$\widehat{\boldsymbol{\gamma}} = \textstyle\sum_{\boldsymbol{\gamma}}\boldsymbol{Z}^{\boldsymbol{T}}\boldsymbol{V}^{-\mathbf{1}}(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}) \qquad (14)$$

Prediktor diatas merupakan prediktor linear tak bias terbaik (BLUP) dari $\boldsymbol{\gamma}$**.** Linear artinya linear dalam **y**, tak bias artinya $E(\widehat{\boldsymbol{\gamma}}) = E(\boldsymbol{\gamma})$, dan terbaik artinya prediktor yang meminimumkan MSEnya.

Hal yang sama dilakukan untuk mendapatkan prediktor untuk $\widehat{\boldsymbol{u}}$.

$$\widehat{\boldsymbol{u}} = \textstyle\sum_{\boldsymbol{u}}\boldsymbol{Z}^{\boldsymbol{T}}\boldsymbol{V}^{-\mathbf{1}}(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}) \qquad (15)$$

Anggap akan mengestimasi:

$$\overline{\boldsymbol{y}}_{\boldsymbol{t}} = \overline{\boldsymbol{x}}_{\boldsymbol{t}}\boldsymbol{\beta} + \overline{\boldsymbol{z}}_{\boldsymbol{t}}\boldsymbol{\gamma} + \boldsymbol{u}_{\boldsymbol{t}} \qquad (16)$$

Dimana $\bar{x}_t$ adalah nilai rata-rata dari x_i , $\bar{z}_t$ adalah basis fungsi *spline*, dan u_t adalah efek acak small area dengan $u_t = \overline{\boldsymbol{d}}_{\boldsymbol{t}}\boldsymbol{u} = \boldsymbol{e}_{\boldsymbol{t}}\boldsymbol{u}$, dan $\boldsymbol{e}_{\boldsymbol{t}}$ adalah vektor dengan nilai 1 saat ke-*t*, dan bernilai 0 untuk *t* lainnya. Sehingga prediktor untuk $\bar{y}_t$ yaitu:

$$\widehat{\boldsymbol{y}}_{\boldsymbol{t}} = \overline{\boldsymbol{x}}_{\boldsymbol{t}}\widehat{\boldsymbol{\beta}} + \overline{\boldsymbol{z}}_{\boldsymbol{t}}\widehat{\boldsymbol{\gamma}} + \boldsymbol{e}_{\boldsymbol{t}}\widehat{\boldsymbol{u}}_{\boldsymbol{t}} \qquad (17)$$

yang merupakan kombinasi linear estimator GLS pada persamaan (11) dan BULP pada persamaan (13) dan (14), sehingga $\hat{y}_t$ merupakan BLUP untuk $\bar{y}_t$.

Jika komponen varians tidak diketahui, maka setelah estimator β dan prediktor γ diperoleh, selanjutnya adalah mengestimasi komponen varians $\boldsymbol{V} = \boldsymbol{Z} \sum_{\gamma} \boldsymbol{Z}^T + \boldsymbol{D} \sum_{u} \boldsymbol{D}^T + \sum_{e}$. Estimasi komponen varians adalah $\sigma_{\gamma}^2, \sigma_u^2,$ dan σ_e^2. Karena estimasi komponen varians berdasarkan ML bias, maka di digunakan metode REML.

Estimasi komponen varians dengan REML (*Restricted Maksimum Likelihood*) didasarkan pada residual yang dihitung setelah β dihitung. Untuk mengestimasi komponen varians dengan REML dapat digunakan beberapa metode, yaitu metode Newton Raphson dan algoritma EM (*expectation and maximization)*. Apabila perhitungan komponen varians dengan menggunakan maksimum likelihood biasa akan menghasilkan komponen yang bias, sehingga untuk mengatasi masalah bias tersebut digunakanlah REML (Kale, 2008).

2.4 BLUP (*Best Linear Unbiased Predictors*) dan EBLUP (*Empirical Best Linear Unbiased Predictors*)

Pendugaan parameter pada model SAE umumnya mengaplikasikan metode BLUP yang meminimumkan MSE. Metode BLUP ini tidak tergantung pada kenormalan dari efek acak tetapi tergantung pada varians atau covarians dari efek acak. Namun pada prakteknya, komponen ragam ini tidak diketahui. Oleh karena itu, diperlukan pendugaan terhadap komponen ragam tersebut melalui data sampel. Metode EBLUP mensubtitusikan komponen ragam yang tidak diketahui ini dengan penduganya.

Misalkan data memenuhi model linear campuran berikut:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zv} + \boldsymbol{e} \tag{18}$$

Dimana:

$\boldsymbol{y}$ adalah vektor dara observasi berukuran $n \times 1$.

$\boldsymbol{X}$ dan $\boldsymbol{Z}$ adalah matriks berukuran $n \times p$ dan $n \times h$ yang diketahui.

$\boldsymbol{v}$ dan $\boldsymbol{e}$ adalah berdistribusi saling bebas dengan rataan 0 dan ragam **G** dan **R** yang tergantung pada parameter $\delta = (\delta_1, \dots, \delta_q)^T$, diasumsikan bahwa δ adalah himpunan bagian dari ruang euclidean sedemikian sehingga:

$$\boldsymbol{Var(y) = V = V(\delta) = R + ZGZ^T}$$

adalah non singular untuk semua $\boldsymbol{\delta}$ yang terdapat dalam himpunan bagian tersebut, dimana Var ($\boldsymbol{y}$) adakah martiks varians covarians dari **y**.

Parameter yang akan diduga merupakan kombinasi linear $\boldsymbol{\mu = 1_i^T\beta + m^T v}$ (rao, 2003). Vektor $\mathbf{1}$ dan $\boldsymbol{m}$ adalah konstan. Penduga linear dari μ adalah $\boldsymbol{\hat{\mu} = a^T\beta + b}$ untuk $\boldsymbol{a}$ dan $\boldsymbol{b}$ diketahui. Sehingga penduga tak bias $\boldsymbol{\mu}$

$$E(\hat{\mu}) = E(\mu)$$

E adalah ekspektasi berdasarkan model (18).

Selanjutnya MSE $\hat{\mu}$ didefinisikan sebagai

$$MSE(\hat{\mu}) = E(\hat{\mu} - \mu)^2$$

dan jika $\hat{\mu}$ adalah penduga tak bias dari μ, maka

$$MSE(\hat{\mu}) = E(\hat{\mu} - \mu)^2 = var(\hat{\mu} - \mu).$$

Penduga BLUP μ yang meminimumkan MSE ,$\boldsymbol{\delta}$ diketahui adalah sebagai berikut:

$$\boldsymbol{\tilde{\mu}^H = t(\delta(y)) = I_i^T\tilde{\beta} + m^T\tilde{v} = 1^T\tilde{\beta} + m^T GZ^T V^{-1}(y - X\tilde{\beta})} \qquad (19)$$

Superscript *H* adalah Henderson, yang mengusulkan persamaan (18).

Dimana

$$\boldsymbol{\tilde{\beta} = \tilde{\beta}(\delta) = (X^T V^{-1} X)^{-1} X^T V^{-1} y} \qquad (20)$$

adalah *best linear unbiased estimator* (BLUE) dari β dan

$$\boldsymbol{\tilde{v} = \tilde{v}(\delta) = GZ^T V^{-1}(y - X\tilde{\beta})} \qquad (21)$$

Penduga BLUP tergantung pada ragam δ yang biasanya tidak diketahui. Jika δ diduga dengan $\tilde{\delta} = \tilde{\delta}(y)$, maka akan diperoleh empirical best linear unbiased predictors (EBLUP) yang tetap merupakan penduga tak bias bagi μ. Penduga δ diperoleh melalui metode ML atau REML.

(Rumiati, 2012)

Untuk model berbasis area dengan model (persamaan 3):

$\hat{\theta}_i = x_i^T\beta + b_i v_i + e$, dimana x_i adalah vektor covariat level area berukuran p×1, $v_i \sim (0, \sigma_v^2)$ dan independen terhadap sampling eror $e_i \sim (0, \sigma_e^2)$ dimana variansi σ_e^2 diketahui. $\hat{\theta}_i$ adalah estimastor langsung untuk parameter area ke-*i* $\hat{\theta}_i = g(\bar{Y}_i)$.

Selanjutnya, $\boldsymbol{G_i} = \sigma_v^2$ $\quad ,dan \quad$ $\boldsymbol{R_i} = \sigma_e^2$

Sehingga $\boldsymbol{V_i} = \sigma_e^2 + \sigma_v^2 b_i^2$, Dengan $\mu_i = \theta_i = x_i^T\beta + b_i v_i,\ I_i = x_i$, dan $m_i = b_i$ dan disubstitusikan ke persamaan BLUP $\boldsymbol{\mu} = \mathbf{1}^{\boldsymbol{T}}\boldsymbol{\beta} + \boldsymbol{m}^{\boldsymbol{T}}\boldsymbol{v}$, sehingga diperoleh rataan area kecil:

$$\tilde{\theta}_i^H = z_i^T\tilde{\beta} + \gamma_i(\tilde{\theta}_i - z_i^T\tilde{\beta})$$
$$= \gamma_i\tilde{\theta}_i + (1-\gamma_i)z_i^T\tilde{\beta},$$

Dimana $\gamma_i = \sigma_v^2 b_i^2/(\sigma_e^2 + \sigma_v^2 b_i^2)$

$\tilde{\beta} = \tilde{\beta}(\sigma_v^2) = \left[\sum_{i=1}^m \boldsymbol{z_i}\boldsymbol{z_i^T}/(\sigma_e^2 + \sigma_v^2 b_i^2\right]^{-1}\left[\sum_{i=1}^m \boldsymbol{z_i}\widehat{\boldsymbol{\theta}}_{\boldsymbol{i}}/(\sigma_e^2 + \sigma_v^2 b_i^2\right].$

Estimator BLUP $\tilde{\theta}_i^H$ dapat diartikan sebagai rata-rata terboboti pada pendugaan langsung θ_i. Sedangkan $z_i^T\tilde{\beta}$ merupakan estimator sintetik dimana $\gamma_i(0 \le \gamma_i \le 1)$. Model varians $\sigma_v^2 b_i^2$ relatif terhadap total varians $\sigma_e^2 + \sigma_v^2 b_i^2$. Jika model $\sigma_e^2 + \sigma_v^2 b_i^2$ relatif kecil maka nilai γ_i akan relatif kecil juga (Rao, 2003)

2.5 REML (*Restricted Maximum Likelihood*)

Estimasi komponen varian dengan REML didasarkan pada residual yang dihitung setelah efek tetap β dihitung (Searle dkk, 1992). REML menggunakan kombinasi linear dari *y* yang dipilih sedemikian sehingga kombinasi linear ini tidak mengandung efek tetap. Estimasi komponen varians dengan REML adalah dengan memaksimumkan fungsi *marginal likelihood.* Salah satu cara untuk menurunkan fungsi *marginal likelihood* adalah dengan menggunakan estimasi bayes (searle dkk, 1992).

Jika $\boldsymbol{\beta}$ dan $\boldsymbol{v}$ diketahui maka distribusi dari $\boldsymbol{y}$ adalah :
$(y|\beta,v) \sim N(X\beta + Zv, R)$, dengan distribusi prior $(v|G) \sim N(0,G)$. Fungsi *Generalized likelihood* berdasarkan model bayes ini adalah:
$L(\beta,v,G,R|y) = f(y|\beta,v) \times f(v|G)$, dimana

$$f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{v}) = (2\pi)^{-\frac{\sum_{h=1}^m n_h}{2}}(|\boldsymbol{R}|)^{-\frac{1}{2}}\left[\exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{Zv})^T\boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{Zv})\right)\right]$$

Dan $f(\boldsymbol{v}|G) = (2\pi)^{-\frac{K+m}{2}}(|\boldsymbol{G}|)^{-\frac{1}{2}}\left[\exp\left(-\frac{1}{2}\boldsymbol{v}^{\boldsymbol{T}}\boldsymbol{G}^{-1}\boldsymbol{v}\right)\right]$

Dengan mengintegralkan fungsi diatas terhadap $\boldsymbol{v}$ dan $\boldsymbol{\beta}$ diperoleh *marginal Likelihood* sebagai berikut:

$$\boldsymbol{L(G,R|y)} = \int\int \boldsymbol{L(\beta,v,G,R|y)\partial v\partial\beta}$$

$$= \frac{(2\pi)^{-\frac{p}{2}}(\left|\boldsymbol{X^T V^{-1} X}\right|)^{-\frac{1}{2}}}{(2\pi)^{-\frac{\sum_{h=1}^{m} n_h}{2}}(|\boldsymbol{V}|)^{\frac{1}{2}}}\left[\exp\left(-\frac{1}{2}\boldsymbol{y^T P y}\right)\right]$$

Dimana $\boldsymbol{P} = \boldsymbol{V^{-1}} - \boldsymbol{V^{-1}X}(\boldsymbol{X^T V^{-1} X})^{-1}\boldsymbol{X^T V^{-1}}$.

Estimasi komponen varians dengan REML adalah dengan memaksimumkan fungsi *likelihood* atau *log likelihood* pada persamaan diatas (Kale, 2008).

2.6 Pendugaan MSE dengan menggunakan Metode Jackknife

Fay herriot (1979) mengembangkan model $y_i = x_i^T \beta + v_i + e_i$ sebagai dasar dalam pengembangan SAE. Dengan mengasumsikan bahwa β dan σ_v^2 tidak diketahui, tetapi $\sigma_{\varepsilon i}^2$ diketahui, dengan $B_i = \sigma_{ei}^2 \,/\, (\sigma_v^2 + \sigma_{ei}^2)$maka

$$\text{MSE}(\hat{\theta}_i) = \frac{\sigma_v^2 \sigma_{ei}^2}{\sigma_v^2 + \sigma_{ei}^2} = (1 - B_i)\, \sigma_{ei}^2 = g_{1i}(\sigma_v^2)$$

Dengan menduga β dengan $\hat{\beta}$ dan σ_v^2 dengan S_v^2 maka MSE dari $\hat{\theta}_i^{EBLUP}$ adalah

$$MSE\left(\hat{\theta}_i^{EBLUP}\right) = E(\hat{\theta}_i^{EBLUP} - \theta_i)^2$$
$$= Var\left(\hat{\theta}_i^{EBLUP}\right) + (bias\left(\hat{\theta}_i^{EBLUP}\right))^2$$

Persamaan tersebut dapat diuraikan menjadi:

$$MSE\left(\hat{\theta}_i^{EBLUP}\right) = MSE\left(\hat{\theta}_i^{BLUP}\right) + E\left(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP}\right)^2 \qquad (22)$$

Metode jackknife pertama kali diperkenalkan oleh Tukey pada tahun 1958 dan kemudian berkembang sebagai suatu metode untuk mengkoreksi bias pada suatu estimator. Dengan melakukan penghapusan terhadap observasi ke-*i* untuk *i=1,2,..,m* dan kemudian dilakukan pendugaan parameter misal $\hat{\theta}_{(i)}$, maka penduga bias diduga dengan

$$bias(\hat{\theta}) = (m-1)[\hat{\theta}_{(.)} - \hat{\theta}]$$

dengan $\hat{\theta}_{(.)} = m^{-1}\sum_i^m \hat{\theta}_{(i)}$.

Penduga jackknife diperoleh dari:

$\hat{\theta}_{jack} = \hat{\theta} - bias\,(\hat{\theta})$ dan $v_{jack}\ (\hat{\theta}) = \frac{(n-1)}{n}\sum_i^m [\hat{\theta}_i - \hat{\theta}]^2$

Penerapan jackknife pada SAE dilakukan untuk mengkoreksi pendugaan MSE akibat adanya pendugaan β dan σ_v^2. Persamaan (21) setara dengan $g_{1i}(\sigma_v^2)$ + (*bias)*^{2} jika σ_v^2 diduga.

Dengan j adalah banyak replikasi jackknife dan i adalah banyak data, maka prosedur *jackknife* $MSE\left(\hat{\theta}_i^{EBLUP}\right)$ berdasarkan persamaan (22) adalah sebagai berikut:

1. $MSE\left(\hat{\theta}_i^{EBLUP}\right)$ didekati oleh:

   $MSE_j\left(\hat{\theta}_i\right) = h_{1i} + h_{2i}$

2. Menduga variansi $MSE\left(\hat{\theta}_i^{BLUP}\right)$ dengan menghitung:

   $h_{1i} = g_{1i}(S_v^2) - \left(\frac{m-1}{m}\right)\sum_{u=1}^{m}\left[g_{1i}\left(S_{v(-u)}^2\right) - g_{1i}(S_v^2)\right]$

   dimana $g_{1i}\left(S_{v(-u)}^2\right)$ diperoleh dengan menghapus pengamatan ke-u pada himpunan data $g_{1i}(S_v^2)$ dan u = 1,2,...m. Dengan :

   $S_v^2 = (m-1)^{-1}\sum_i (y_i - \bar{y})^2 - \sigma_e^2$

   $S_{v(-u)}^2 = (m-2)^{-1}\sum_{i(-u)}\left(y_i - \bar{y}_{(-u)}\right)^2 - \sigma_e^2$

3. Menduga $E\left(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP}\right)^2$dengan menghitung:

   $h_{2i} = \left(\frac{m-1}{m}\right)\sum_{u=1}^{m}\left[\left(\hat{\theta}_{i(-u)}\right) - \left(\hat{\theta}_i\right)\right]^2$

   dimana $(\hat{\theta}_{i(-u)})$ diperoleh dengan menghapus pengamatan ke-u pada himpunan data $(\hat{\theta}_i)$.

2.7 Pengeluaran Perkapita

Besarnya pendapatan perkapita dapat menggambarkan kesejahteraan suatu masyarakat. Namun data pendapatan yang akurat sulit diperoleh, sehingga dalam kegiatan susenas data ini didekati melalui data pengeluaran perkapita (Jawa Tengah dalam angka 2009, publikasi BPS 2010).

Pengeluaran perkapita menunjukkan besarnya pengeluaran setiap anggota rumah tangga dalam kurun wakti satu bulan (BPS, 2008). Maksud dari rumah tangga sendiri adalah sekelompok orang yang mendiami sebagian atau seluruh bangunan fisik dan biasanya tingga bersama serta makan dari satu dapur. Dimana pengeluaran perkapita dipengaruhi oleh pendapatan perkapitanya. Asumsi ini menjelaskan pada saat pendapatan seseorang semakin meningkat maka semakin tinggi pula pengeluarannya.

Berdasarkan pedoman pencacah modul konsumsi SUSENAS 2002**,** Dalam sensus, pengeluaran perkapita merupakan pengeluaran untuk rumah tangga / anggota rumah tangga saja, tidak termasuk pengeluaran untuk keperlauan usaha rumah tangga, atau yang diberikan kepada orang lain. Untuk konsumsi makanan, baik banyaknya (kuantitas) maupun nilainya yang dicatat adalah yang betul-betul telah dikonsumsi selama referensi waktu survei (*consumption approach*), sedangkan untuk bukan makanan konsep yang dipakai pada umumnya adalah kosep pembelian (*delivery approach*), yaitu dicatat sebagai pengeluaran pada waktu barang tersebut dibeli/diperoleh, asalkan tujuannya untuk kebutuhan rumah tangga. Data pengeluaran perkapita diperoleh dari jumlah pengeluaran rumah tangga sebulan dibagi dengan jumlah anggota keluarga tersebut.

(halaman ini sengaja dikosongkan)

BAB III
METODELOGI PENELITIAN

3.1 Variabel Penelitian

Variabel dalam penelitian ini adalah sebagai berikut:

1. Variabel respon

   Variabel respon yang digunakan dalam penelitian ini adalah pengeluaran perkapita yang diperoleh dari data SUSENAS 2009. Pengeluaran perkapita diperoleh dengan cara pengeluaran rumah tangga dibagi dengan jumlah anggota rumah tangganya.

2. Variabel penyerta

   Variabel penyerta yang digunakan pada penelitian ini adalah kepadatan penduduk, karena berdasarkan penelitian terdahulu (Fausi, 2011) diketahui bahwa variabel yang paling berpengaruh pada pengeluaran perkapita di Kabupaten Sumenep adalah kepadatan penduduk.

Tabel 3.1 Variabel - Variabel Penelitian

No.	Variabel	Keterangan	Definisi Operasional
1	X	Kepadatan penduduk	Jumlah penduduk tiap satuan luasan 1 km2
2	Y	Rata-rata pengeluaran perkapita	Jumlah rata-rata pengeluaran rumah tangga sebulan dibagi dengan jumlah anggota keluarga (× Rp. 100.000)

3.2 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari BPS yang berupa data pengeluaran perkapita yang berasal dari modul SUSENAS 2009 yang merupakan hasil dari survey pada tahun yang sama yaitu 2009 dan data variabel pendukungnya diperoleh dari data Kabupaten Sumenep dalam angka 2010 dimana data ini merupakan data yang diperoleh pada survey satu tahun sebelumnya , yaitu 2009 dan dipublikasikan pada tahun 2010.

Dalam SUSENAS tahun 2009 semua kecamatan di wilayah Kabupaten Sumenep yang terambil sebagai sampel, namun pada Kecamatan Batuan, Giligenteng, Nonggunong, Ra'as, Kangayan dan Kecamatan Masalembu jumlah sampel yang terambil hanya sedikit yaitu hanya 16 RT. Sehingga keenam kecamatan ini dapat dikategorikan sebagai *small area* (area kecil). Estimasi area ini dilakukan menganganggap daerah tersebut memiliki karakteristik yang sama dengan domainnya.

Tabel 3.2 Struktur Data

No.	**Kecamatan**		**Y**	**X**
1	1	y_1	$y_{1.1}, y_{1.2}, y_{1.3} \cdots, y_{1.n1}$	x_1
2	2	y_2	$y_{2.1}, y_{2.2}, y_{2.3} \cdots, y_{2.n2}$	x_2
.	.	.		.
.	.	.		.
.	.	.		.
m	m	y_m	$y_{m.1}, y_{m.2}, y_{m.3} \cdots, y_{m.nm}$	x_m

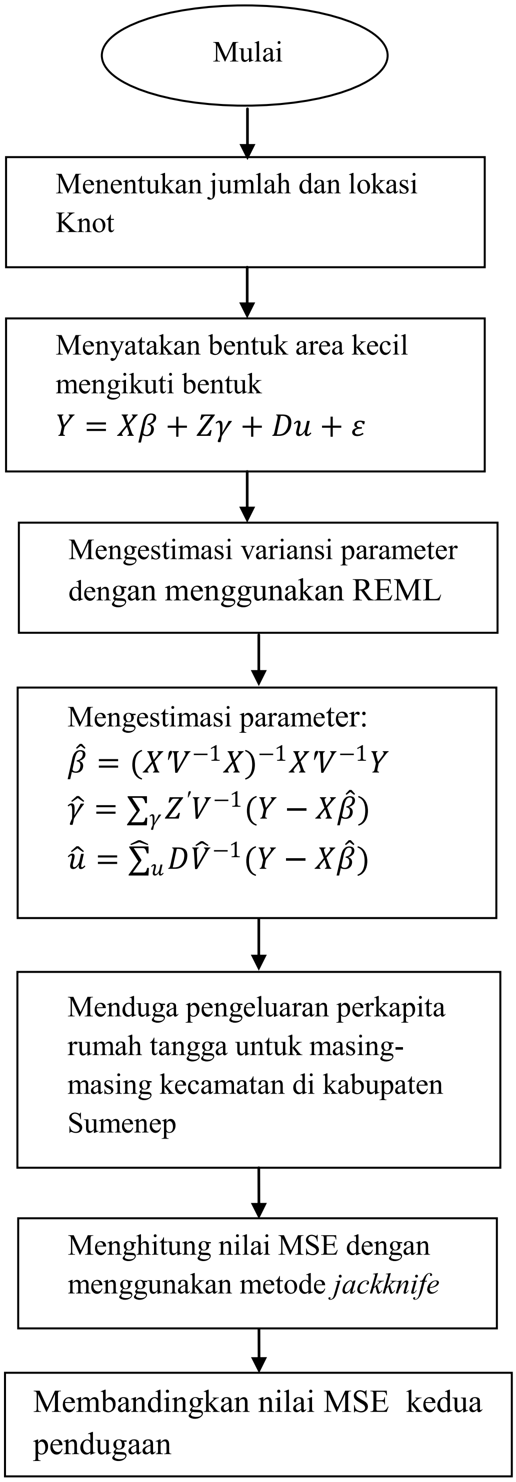
3.3 Metode Analisis

Tahapan-tahapan dalam penelitian ini adalah:

1. Menduga pengeluaran perkapita untuk masing-masing kecamatan di Kabupaten Sumenep menggunakan estimasi *P-Spline* dengan variabel bantu yang digunakan adalah kepadatan dengan langkah-langkah sebagai berikut:
   i. Menentukan orde (linear, kuadratik, atau kubik) dan titik knot yang menghasilkan GCV minimum.
   ii. Menyatakan bentuk area kecil mengikuti suatu fungsi *P-Spline* berbasis polinomial terputus.
   $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{u}$
   iii. Menganggap parameter $\boldsymbol{X\beta}$ sebagai sebagai pengaruh tetap (*Fixed effect*) dan $\boldsymbol{Z\gamma}$ sebagai pengaruh acak dan $\boldsymbol{u}$ sebagai pengaruh acakarea kecil.
   iv. Mengestimasi parameter regresi dalam kerangka kerja model linear campuran.
   v. Mengestimasi $\beta, \gamma,$ dan u dengan menggunakan metode REML.

vi. Menduga pengeluaran perkapita rumah tangga untuk masing-masing kecamatan di kabupaten Sumenep.

2. Menghitung nilai *mean square error* (MSE) hasil pendugaan pengeluaran perkapita untuk masing-masing kecamatan dari metode *P-Spline*.



Gambar 3.1 Flow Chart Metode Analisis

BAB 4
ANALISIS DAN PEMBAHASAN

Berdasarkan hasil dari SUSENAS 2009 dimana data pengeluaran perkapita yang tersedia untuk masing-masing Kecamatan di Kabupaten Sumenep adalah hanya untuk 21 kecamatan sedangkan untuk 6 kecamatan lainnya yaitu Kecamatan Giligenteng, Batuan, Nonggunong, Ra'as, Kangayan, dan Masalembu sampel yang tersedia hanya sedikit yaitu sebanyak 16 RT. Secara geografis letak masing-masing kecamatan dapat dilihat pada gambar 4.1.

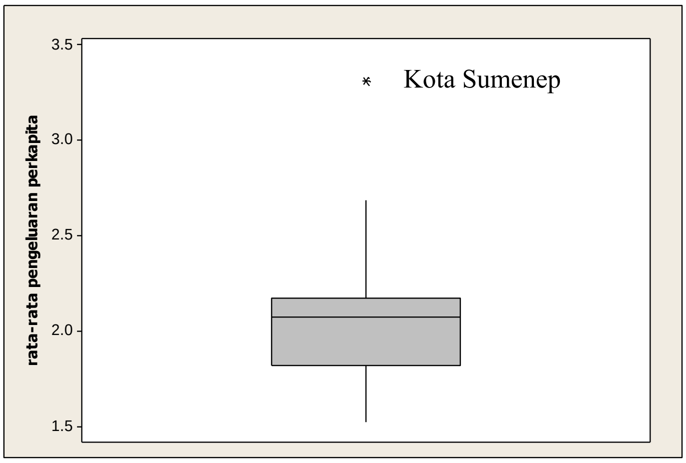
Gambar 4.1 Peta Kabupaten Sumenep

Eksplorasi data dilakukan terhadap data pengeluaran perkapita dan variabel penyerta di tiap kecamatan di Kabupaten Sumenep. Berdasarkan Tabel 4.1, diketahui bahwa rata-rata pengeluaran perkapita penduduk di 21 kecamatan tersurvei di Kabupaten Sumenep adalah sebesar Rp. 206.825,00. Pengeluaran terbesar terjadi di Kota Sumenep yang merupakan ibukota kabupaten dimana pusat kegiatan pemerintahan dan perekonomian berlangsung dengan jumlah pengeluaran perkapita Rp. 331.510,00 dan kecamatan dengan jumlah pengeluaran terendah yaitu sebesar Rp.152.799,00 adalah kecamatan Bluto. Tingkat keragaman pengeluaran perkapita di Kabupaten Sumenep tidak terlalu beragam yang ditunjukkan dengan nilai koefisien variansi sebesar 18.8 persen.

Tabel 4.1 Statisika Deskriptif Pengeluaran Perkapita di Kabupaten Sumenep

Statistik	Pengeluaran Perkapita
Rata-rata	2.0683
Variansi	0.1515
Koefisien Variansi	18.82
Jangkauan	1.7871
Minimum	1.528
Maximum	3.3151

Pada gambar 4.2 terlihat bahwa pengeluaran perkapita di Kota Sumenep menjadi pencilan, dimana pengeluarannya jauh melebihi kecamatan-kacamatan lainnya di wilayah Kabupaten Sumenep. Pada boxplot, pola rata-rata pengeluaran perkapita masing-masing kecamatan di Kabupaten Sumenep lebar pada bagian bawah yang berarti persebaran pengeluaran perkapita lebih banyak yang berada di bawah rata-rata nilai pengeluaran perkapita di Kabupaten Sumenep.



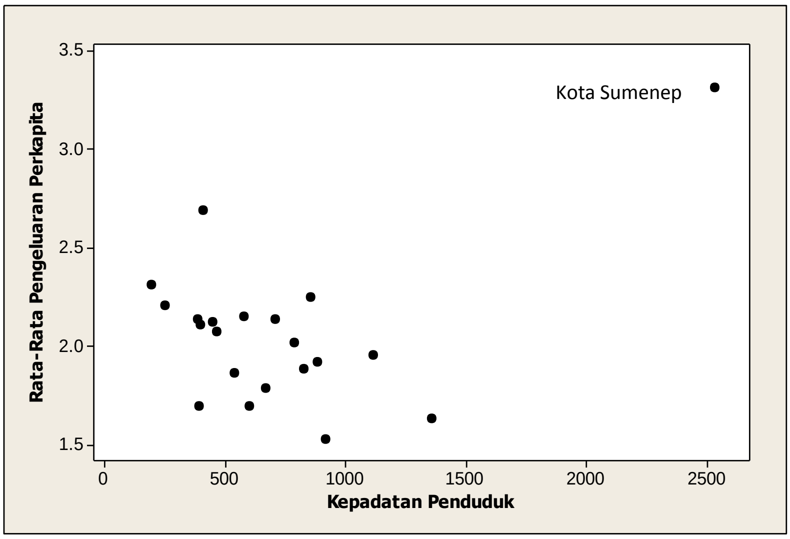
Gambar 4.2 Boxplot Pengeluaran Perkapita di Kabupaten Sumenep

Fausi (2011) menduga pengeluaran perkapita dipengaruhi oleh beberapa faktor, yaitu: peresentase penduduk yang bekerja disektor pertanian, rata-rata anggota keluarga, presentase penduduk yang berpendidikan minimal SD, presentase penduduk miskin, dan kepadatan penduduk. Namun berdasarkan nilai korelasinya diketahui bahwa hanya kepadatan penduduk yang berpengaruh signifikan terhadap pengeluaran perkapita di Kabupaten Sumenep sehingga pada penelitian ini akan menduga pengeluaran perkapita masing-masing kecamatan di Kabupaten Sumenep dengan menggunakan variabel bantu kepadatan penduduk. Statistika deskriptif mengenai kepadatan peduduk di Kabupaten Sumenep dapat dilihat pada tabel 4.2.

Tabel 4.2 Statisika Deskriptif Kepadatan Penduduk

Statistik	**Kepadatan Penduduk**
Rata-rata	686.5
Koefisien Variansi	68.19
Minimum	107.0
Maximum	2535.0

Dari tabel diatas diketahui bahwa rata-rata jumlah penduduk di Kabupaten Sumenep tiap luasan 1 km^2 dihuni oleh 687 penduduk dengan nilai keragaman antar kecamatan sebesar 68.19 persen. Kecamatan dengan penduduk paling padat adalah Kota Sumenep dengan jumlah penduduk 2535 orang/km^2 sedangkan Kecamatan Kangayan merupakan daerah dengan kepadatan paling sedikit, yaitu 107 orang/km^2. Pola hubungan antara rata-rata pengeluaran perkapita masing-masing kecamatan dengan kepadatan penduduk dapat dilihat pada gambar 4.3.



Gambar 4.3 Plot Hubungan Rata-Rata Pengeluaran Perkapita dengan Kepadatan Penduduk

Berdasarkan Gambar 4.3 diketahui bahwa pola hubungan antara pengeluaran perkapita dengan kepadatan penduduk tidak menunjukkan pola hubungan tertentu dimana persebaran datanya bergerombol. Selain itu juga terdapat data pencilan yaitu pada wilayah Kota Sumenep dimana kepadatan penduduk dan rata-rata pengeluaran perkapitannya jauh lebih banyak daripada kecamatan-kecamatan lainnya di wilayah Kabupaten Sumenep. Hal ini dapat dimengerti karena di Kota Sumenep merupakan pusat pemerintahan dan perekonomian di Kabupaten ini**.**

4.1 Pendugaan Area Kecil Menggunakan Metode *P-Spline*

Pola sebaran data antara rata-rata pengeluaran perkapita dengan kepadatan penduduk di Kabupaten Sumenep seperti pada Gambar 4.3 tidak membentuk suatu pola tertentu, dimana persebaran datanya bergerombol dan juga terdapat data pencilan, yaitu pada pengamatan di Kota Sumenep. Berdasarkan hal tersebut sehingga pada penelitian ini akan mencari dua model nonparametrik *P-Spline* yaitu pertama dengan menggunakan keseluruhan data yang diperoleh (21 data) dan yang kedua adalah dengan menggunakan data setelah menghapus pengamatan di Kota Sumenep (data pencilan).

4.1.1 Pendugaan Area Kecil untuk menduga pengeluaran Perkapita dengan menggunakan keseluruhan data

Berdasarkan penjelasan diatas maka diketahui bahwa hanya terdapat satu variabel yang berkorelasi cukup kuat dengan pengeluaran perkapita masyarakat di Kabupaten Sumenep yaitu kepadatan penduduk. Pola hubungan antara pengeluaran perkapita dengan kepadatan penduduk yang dapat dilihat pada Gambar 4.3 tidak menunjukkan pola tertentu, dimana pola sebarannya bergerombol dan terdapatnya data pencilan.

Model pendekatan nonparametrik yang digunakan untuk memodelkan varibel respon dengan variabel penyerta dapat digunakan pendekatan *spline* linear, kuadratik dan kubik dengan 1,2, hingga 5 titik knot. Langkah awal dalam pembentukan model *spline* adalah menentukan titik-titik perubahan pola perilaku data yang disebut dengan titik knot. Titik Knot yang dipilih adalah titik knot yang menghasilkan nilai GCV minimum dan dibatasi sampai 5 titik knot.

- **Pendekatan dengan *P-Spline* Orde Satu**

Model *spline* orde satu atau linear diberikan sebagai berikut:

a. *Spline* linear dengan satu titik knot

$$Y = \beta_0 + \beta_1 x + \gamma_1 (X - k_1)_+^1 + u$$

b. *Spline* linear dengan dua titik knots

$$Y = \beta_0 + \beta_1 x + \gamma_1 (X - k_1)_+^1 + \gamma_2 (X - k_2)_+^1 + u$$

c. *Spline* linear dengan tiga titik knots

$$Y = \beta_0 + \beta_1 x + \gamma_1 (X - k_1)_+^1 + \gamma_2 (X - k_2)_+^1 + \gamma_3 (X - k_3)_+^1 + u$$

d. *Spline* linear dengan empat titik knot

$$Y = \beta_0 + \beta_1 x + \gamma_1 (X - k_1)_+^1 + \gamma_2 (X - k_2)_+^1 + \gamma_3 (X - k_3)_+^1 + \gamma_4 (X - k_4)_+^1 + u$$

e. *Spline* linear dengan lima titik knot

$$Y = \beta_0 + \beta_1 x + \gamma_1 (X - k_1)_+^1 + \gamma_2 (X - k_2)_+^1 + \gamma_3 (X - k_3)_+^1 + \gamma_4 (X - k_4)_+^1 + \gamma_5 (X - k_5)_+^1 + u$$

Untuk mendapatkan model *P-Spline* terbaik dipilih titik-titik knot yang menghasilkan nilai GCV nimimum dengan maksimal jumlah knot dibatasi sebanyak 5 titik. Tabel 4.3 menyajikan titik-titik knot dan nilai GCV untuk model *P-Spline* Orde satu.

Tabel 4.3 Kombinasi Titik Knot Orde Satu

No.	**Orde**	**Jumlah Knot**	**Titik Knot**		**Lambda**	**GCV**
1	1	1	1117.41		24598.83	0.07579
2	1	2	886.10	926.27	281.7496	0.07526
3	1	3	644.77 926.85	868.87	225.512	0.07526
4	1	4	669.05 873.43	708.10 926.29	294.2274	0.07872
5	1	5	398.62 482.15 926.62	408.50 872.16	2.782753	0.05454 *

* GCV minimum

Berdasarkan Tabel 4.3 diatas, nilai GCV minimum adalah sebesar 0.0545, sehingga model *P-Spline* orde satu terbaik diperoleh dengan menggunakan lima titik knot, yaitu di titik 398.62, 408.50, 482.15, 872.16, dan 926.62. model *P-Spline* orde satu terbaik adalah sebagai berikut:

$$Y = \beta_0 + \beta_1 x + \gamma_1 (X - 398.62)_+^1 + \gamma_2 (X - 408.50)_+^1 + \gamma_3 (X - 482.15)_+^1 + \gamma_4 (X - 872.16)_+^1 + \gamma_5 (X - 926.62)_+^1 + u$$

- **Pendekatan dengan *P-Spline* orde dua**

Model *spline* orde dua diberikan sebagai berikut:

a. *Spline* kuadratik dengan satu titik knot

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \gamma_1 (X - k_1)_+^2 + u$$

b. *Spline* kuadratik dengan dua titik knot

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \gamma_1 (X - k_1)_+^2 + \gamma_2 (X - k_2)_+^2 + u$$

c. *Spline* kuadratik dengan tiga titik knot

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \gamma_1 (X - k_1)_+^2 + \gamma_2 (X - k_2)_+^2 + \gamma_3 (X - k_3)_+^2 + u$$

d. *Spline* kuadratik dengan empat titik knot

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \gamma_1 (X - k_1)_+^2 + \gamma_2 (X - k_2)_+^2 + \gamma_3 (X - k_3)_+^2 + \gamma_4 (X - k_4)_+^2 + u$$

e. *Spline* kuadratik dengan lima titik knot

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \gamma_1 (X - k_1)_+^2 + \gamma_2 (X - k_2)_+^2 + \gamma_3 (X - k_3)_+^2 + \gamma_4 (X - k_4)_+^2 + \gamma_5 (X - k_5)_+^2 + u$$

Untuk mendapatkan model *P-Spline* orde dua terbaik dipilih titik-titik knot yang menghasilkan nilai GCV minimum dengan membatasi maksimal jumlah knot sebanyak 5. Tabel 4.4 menyajikan kombinasi titik-titik knotdn GCV untuk model *P-Spline* orde dua sebagai berikut:

Tabel 4.4 Kombinasi Titik Knot Orde Dua

No.	Orde	Jumlah knot	Titik knot		Lambda	GCV
1	2	1	256.29		4.595555e+16	0.06997164
2	2	2	390.43	392.74	6.759412e+15	0.06997160*
3	2	3	389.68 416.44	392.99	4.614974e+16	0.06997174
4	2	4	389.30 410.36	392.43 516.35	3.198827e+16	0.06997363
5	2	5	389.72	392.43	3.178561e+16	0.06997673

			412.08	512.56		
			1068.78			

* GCV minimum

Dapat dilihat pada tabel 4.5 bahwa titik knot maksimum untuk *spline* orde dua terletak pada titik 390.43 dan titik 392.74 dengan nilai GCV 0.06997 sehingga model *P-Spline* orde dua terbaik adalah sebagai berikut:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \gamma_1 (X - 390.43)_+^2 + \gamma_2 (X - 392{,}72)_+^2 + u$$

- **Pendekatan dengan *P-Spline* orde tiga**

Model *spline* orde tiga diberikan sebagai berikut:

a. *Spline* orde tiga dengan satu titik knot

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \gamma_1 (X - k_1)_+^3 + u$$

b. *Spline* kuadratik dengan satu titik knot

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \gamma_1 (X - k_1)_+^3 + \gamma_2 (X - k_2)_+^3 + u$$

c. *Spline* kuadratik dengan satu titik knot

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \gamma_1 (X - k_1)_+^3 + \gamma_2 (X - k_2)_+^3 + \gamma_3 (X - k_3)_+^3 + u$$

d. *Spline* kuadratik dengan satu titik knot

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \gamma_1 (X - k_1)_+^3 + \gamma_2 (X - k_2)_+^3 + \gamma_3 (X - k_3)_+^3 + \gamma_4 (X - k_4)_+^3 + u$$

e. *Spline* 5 Knots

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \gamma_1 (X - k_1)_+^3 + \gamma_2 (X - k_2)_+^3 + \gamma_3 (X - k_3)_+^3 + \gamma_4 (X - k_4)_+^3 + \gamma_5 (X - k_5)_+^3 + u$$

Untuk mendapatkan model *P-Spline* orde tiga terbaik dipilih titik-titik knot yang menghasilkan nilai GCV minimum. Tabel 4.5 menyajikan kombinasi knot *P-Spline* orde tiga dan nilai GCVnya.

Tabel 4.5 Kombinasi Titik Knot Orde Tiga

No.	Orde	Jumlah Knot	Titik Knot		Lambda	GCV
1	3	1	252.96		2.497766e+23	0.07544704
2	3	2	255.89	925.27	2.723166e+22	0.07545217
3	3	3	254.40 545.33	418.56	3.405105e+22	0.07544979
4	3	4	252.82	395.36	2.154976e+22	0.0754537

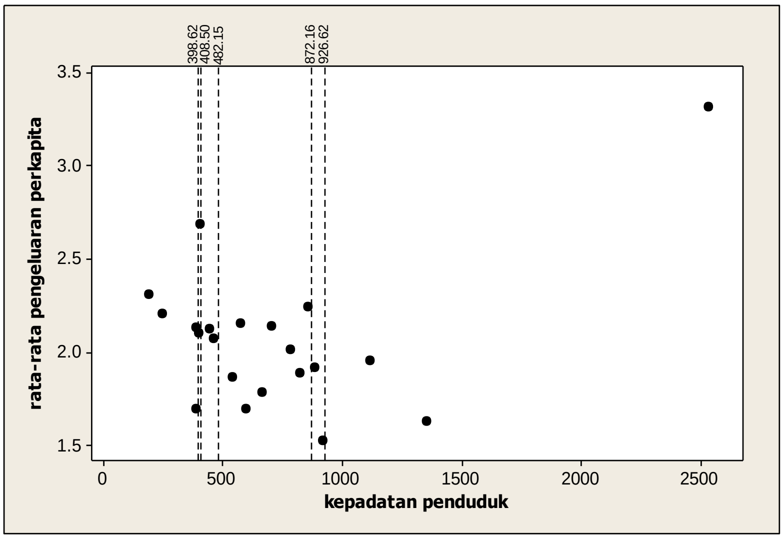
			456.08	912.06		
5	3	5	386.07	396.99	1.731112e+22	0.07541534*
			409.11	823.93		
			873.78			

* GCV minimum

dari tabel 4.6 maka diketahui model *spline* orde tiga dengan nilai GCV terkecil adalah dengan menggunakan lima titik knot, yaitu pada titik 386.07, 396.99, 409.11, 823.93 dan titik 873.78 dengan model sebagai berikut:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \gamma_1 (X - 386.07)_+^3 + \gamma_2 (X - 396.99)_+^3 + \gamma_3 (X - 409.11)_+^3 + \gamma_4 (X - 823.93)_+^3 + \gamma_5 (X - 873.78)_+^3 + u$$

Dari ketiga model *P-Spline* tersebut, yaitu *spline* orde satu, dua dan tiga, dapat disimpulkan bahwa nilai GCV minimum terbaik diperoleh dari model *P-Spline* linear dengan lima titik knot yaitu pada titik 398.62, 408.50, 482.15, 872.16, dan 926.62 dengan nilai GCV sebesar 0.0545. Berikut adalah gambaran lokasi titik-titik knotnya.



Gambar 4.4 Lokasi Titik Knot *P-Spline* Orde Satu dengan Menggunakan Keseluruhan Data

Setelah Jumlah dan lokasi knot diperoleh selanjutnya adalah mengestimasi parameter β dengan memaksimumkan fungsi *likelihood* atau *log likelihood*nya dan mencari prediktor $\hat{\gamma}$ dan $\hat{u}$ yang merupakan EBLUP (*Empirical Best Linear Unbiased Predictors*) dari γ dan u sebagai pengaruh acak. Tabel 4.6 berikut menyajikan nilai estimasi β untuk model *P-Spline* terbaik (orde satu). Nilai estimasi pengaruh acak dapat dilihat pada lampiran 3.

Tabel 4.6 Estimasi Efek Tetap

Parameter	**Estimasi**
β_0	2.2239
β_1	-0.00031

Sehingga diperoleh estimasi model sebagai berikut:

$$\boldsymbol{Y} = 2.2239 + (-0.00031\,\boldsymbol{X}) + \gamma_1(\boldsymbol{X} - 398.62)_+^1 + \gamma_2(\boldsymbol{X} - 408.50)_+^1 + \gamma_3(\boldsymbol{X} - 482.15)_+^1 + \gamma_4(\boldsymbol{X} - 872.16)_+^1 + \gamma_5(\boldsymbol{X} - 926.62)_+^1 + u$$

Berdasarkan model *P-Spline* terbaik diatas, maka:

a. Setiap kenaikan satu satuan x untuk $x < 398.62$ akan merubah nilai y sebesar (-0.00031) satuan.
b. Untuk nilai $x \leq 398.62$, setiap kenaikan satu satuan akan berpengaruh sebesar $(-0.00031 + \gamma_1)$ satuan terhadap *y.*
c. Untuk nilai $398.62 < x \leq 408.50$, setiap kenaikan satu satuan x akan berpengaruh sebesar $(-0.00031 + \gamma_1 + \gamma_2)$ satuan terhadap *y.*
d. Untuk nilai $408.50 < x \leq 872.16$, setiap kenaikan satu satuan x akan berpengaruh sebesar $(-0.00031 + \gamma_1 + \gamma_2 + \gamma_3)$ satuan terhadap *y.*
e. Untuk nilai $872.16 < x \leq 926.61$, setiap kenaikan satu satuan x akan berpengaruh sebesar $(-0.00031 + \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4)$ satuan terhadap *y.*
f. Untuk nilai $x > 926.61$, setiap kenaikan satu satuan akan x berpengaruh sebesar $(-0.00031 + \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 + \gamma_5)$ satuan terhadap *y.*

Berdasarkan Tabel 4.7, Hasil pendugaan dengan menggunakan *P-Spline,* rata-rata pengeluaran perkapita di Kabupaten Sumenep pada tahun 2009 adalah sebesar Rp. 206.683,00. Dengan nilai koefisien variansi sebesar 18.69 persen menunjukkan bahwa nilai perndugaan pengeluaran perkapita di Kabupaten Sumenep tidak terlalu beragam. Nilai pendugaan pengeluaran perkapita terkecil sebesar Rp. 153.220,00 dan nilai pengeluaran perkapita terbesar sebesar Rp. 331.350,00. Kecamatan dengan rata-rata pengeluaran perkapitanya paling rendah adalah Kecamatan Bluto, sedangkan Kota Sumenep merupakan daerah dengan rata-rata pengeluaran perkapitanya paling besar. Hasil pendugaan pengeluaran perkapita pada masing-masing kecamatan di Kabupaten Sumenep dengan menggunakan keseluruhan data dapat dilihat pada lampiran 4. Berikut adalah statistika deskriptif pendugaan rata-rata pengeluaran perkapita di Kabupaten Sumenep dengan menggunakan keseluruhan data.

Tabel 4.7 Statisika Deskriptif Hasil Estimasi dengan Menggunakan Keseluruhan data

Statistik	**Pendugaan Pengeluaran Perkapita**
Rata-rata	2.0683
Variansi	0.1495
Koefisien Variansi	18.69
Jangkauan	1.7812
Minimum	1.5322
Maximum	3.3135

Menurut Rao (2003), Konsep estimasi sintetik dapat digunakan untuk mengestimasi daerah yang tidak tersampel dengan menganggap daerah tersebut memiliki karakteristik yang sama dengan domainnya. Untuk mengestimasi pengeluaran perkapita untuk kecamatan-kecamatan dengan jumlah sampel yang sedikit atau bahkan tidak ada sama sekali dengan asumsi perilaku antar kecamatan di Kabupaten Sumenep sama. Nilai harapan dari model area kecil adalah

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{i}} = \boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{T}}\widehat{\boldsymbol{\beta}}.$$

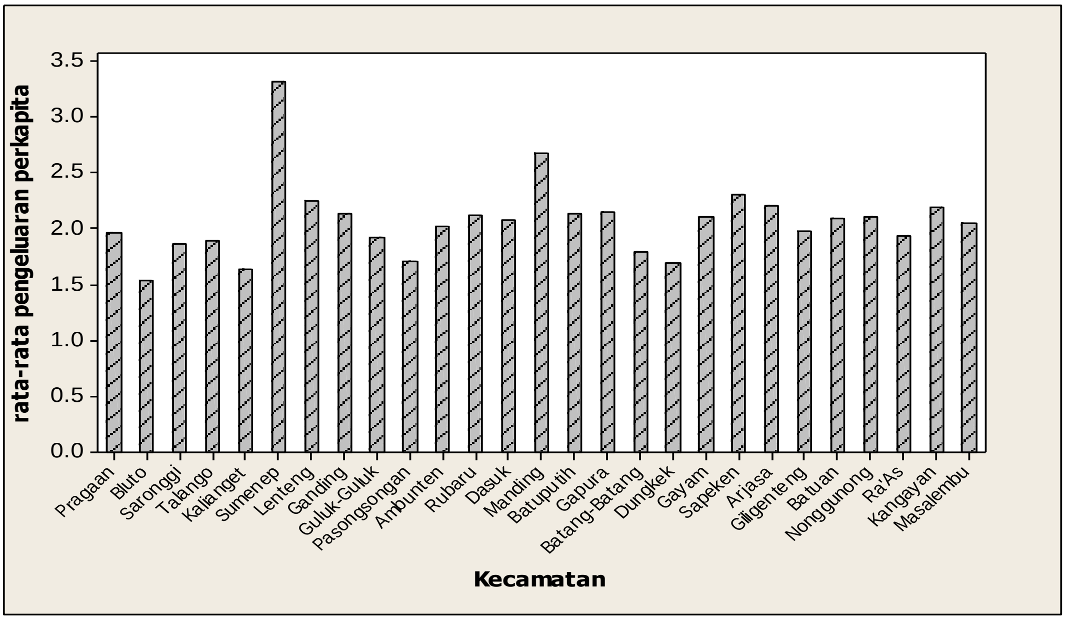
Dengan estimator $\hat{\beta} = [2.2239 \quad -0.00031]$

Indeks i adalah indeks untuk melambangkan kecamatan yang diestimasi nilai pengeluaran perkapitanya. Berikut adalah hasil estimasi pengeluaran perkapita di Kabupaten Sumenep untuk daerah dengan jumlah sampel sedikit.

Tabel 4.8 Hasil Estimasi *P-Spline* dengan Menggunakan Keseluruhan Data

Kecamatan	Pengeluaran perkapita
Giligenteng	1.972001
Batuan	2.087353
Nonggunong	2.108362
Ra'As	1.933151
Kangayan	2.19
Masalembu	2.043868

Gambar 4.4 menyajikan hasil pendugaan semua kecamatan dengan metode *P-Spline* dengan menggunakan semua data pengamatan. Dari gambar dapat terlihat bahwa Kota Sumenep merupakan daerah dengan rata-rata pengeluaran perkapitanya paling tinggi dengan selisih yang sangat jauh dengan daerah-daerah lainnya. Hasil estimasi untuk daerah dengan jumlah sampel yang sedikit yaitu Kecamatan Giligenteng, Batuan, Nonggunong, Ra'as, Kangayan, dan Masalembu berkisar diantara nilai rata-rata pengeluaran perkapita di Kabupaten Sumenep.



Gambar 4.5 Grafik Hasil Estimasi Rata-rata Pengeluaran Perkapita Semua Kecamatan

4.1.2 Pendugaan Area Kecil untuk menduga pengeluaran perkapita dengan penghapusan data pencilan

Berdasarkan penjelasan gambar 4.2 dimana pengamatan Kota Sumenep merupakan data pencilan karena rata-rata pengeluaran perkapita dan kepadatan penduduknya jauh melebihi nilai untuk daerah lainnya, sehingga pada subbab ini akan dilakukan pemodelan rata-rata pengeluaran perkapita di Kabupaten Sumenep dengan menghapus pengamatan Kota Sumenep.

Model pendekatan nonparametrik yang digunakan adalah pendekatan *P-Spline* orde 1,2 dan 3 dengan jumlah knot maksimum lima titik. Langkah awal dalam pembentukan model adalah menentukan kombinasi titik-titik knot yang optimum yaitu yang meminimumkan nilai GCV.

- **Pendekatan dengan *P-Spline* Orde Satu**

Berikut adalah tabel kombinasi titik knot *P-Spline* orde satu. Kombinasi titik knot terbaik adalah yang menghasilkan nilai GCV minimum.

Tabel 4.9 Kombinasi Titik Knot Orde Satu

No.	Orde	Jumlah Knot	Titik Knot		Lambda	GCV
1	1	1	252.78		2.34E+09	0.063475
2	1	2	252.77	392.33	2.78E+09	0.063475
3	1	3	398.80 466.76	408.47	2.953136	0.051672
4	1	4	398.74 481.45	408.50 739.96	1.394918	0.05519
5	1	5	398.75 481.85 920.98	408.47 857.88	2.525142	0.05055*

Berdasarkan tabel 4.9 diatas untuk model *P-Spline* orde satu diperoleh titik knot yang menghasilkan nilai GCV minimum adalah dengan menggunakan kombinasi 5 titik knot yang terletak di titik 398.75, 408.47, 481.85, 857.88 dan 920.98 dengan GCV sebesar 0.05055 dengan model sebagai berikut:

$$Y = \beta_0 + \beta_1 x + \gamma_1 (X - 398.75)_+^1 + \gamma_2 (X - 408.47)_+^1 + \gamma_3 (X - 481.85)_+^1 + \gamma_4 (X - 857.88)_+^1 + \gamma_5 (X - 920.98)_+^1 + u$$

- **Pendekatan dengan *P-Spline* Orde Dua**

Pada Tabel 4.10 menyajikan kombinasi titik-titik knot untuk mengestimasi pengeluaran perkapita di Kabupaten Sumenep dengan menggunakan *P-Spline* orde dua. Kombinasi titik-titik knot yang terbaik merupakan yang menghasilkan nilai GCV minimum.

Tabel 4.10 Kombinasi Titik Knot Orde Dua

No.	Orde	Jumlah Knot	Titik Knot		Lambda	GCV
1	2	1	257.33		2.68E+15	0.070784
2	2	2	253.37	393.61	2.10E+15	0.070784

3	2	3	390.60 408.70	393.02	296E+15	0.070783*
4	2	4	390.12 408.78	392.77 871.43	6.58E+14	0.070784
5	2	5	389.59 408.68 920.57	392.70 837.80	7.97E+14	0.070783

Berdasarkan tabel 4.10 untuk model *P-Spline* orde dua diperoleh titik knot yang menghasilkan nilai GCV minimum adalah dengan menggunakan 3 titik knot yang terletak di titik 390.60, 393.02 dan 408.70 dengan GCV sebesar 0.0708. model terbaik untuk untuk pendugaan dengan menggunakan *P-Spline* orde dua adalah sebagai berikut:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \gamma_1 (X - 390.60)_+^2 + \gamma_2 (X - 393.02)_+^2 + \gamma_3 (X - 408.70)_+^2 + u$$

- **Pendekatan dengan *P-Spline* Orde Tiga**

Untuk mendapatkan model *P-Spline* orde tiga terbaik dipilih titik-titik knot yang menghasilkan nilai GCV minimum. Tabel 4.11 menyajikan kombinasi knot *P-Spline* orde tiga dan nilai GCVnya.

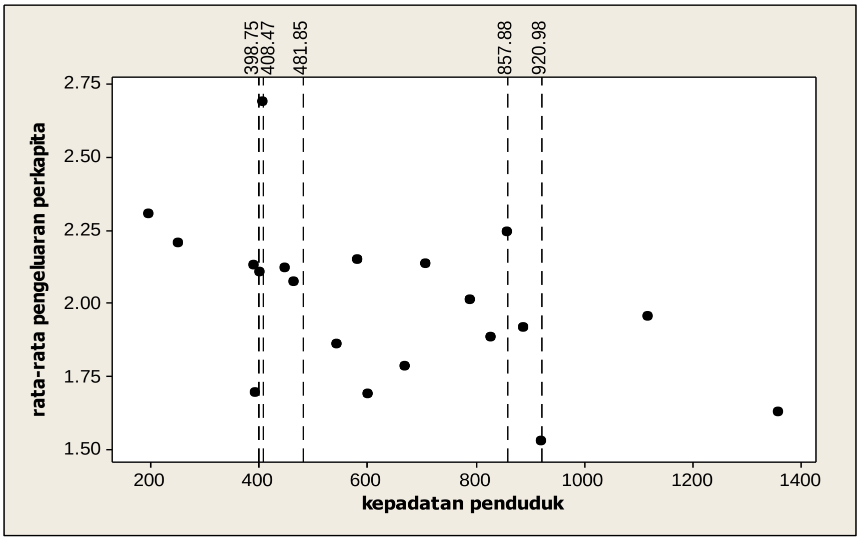
Tabel 4.11 Kombinasi Titik Knot Orde Tiga

No.	**Orde**	**Jumlah Knot**	**Titik Knot**		**Lambda**	**GCV**
1	3	1	253.82		3.60E+21	0.07816
2	3	2	257.39	416.85	1.11E+21	0.07816
3	3	3	389.50 409.65	393.05	2.67E+21	0.07814*
4	3	4	390.19 408.97	392.41 633.62	2.84E+20	0.07815
5	3	5	389.76 408.59 619.29	392.40 474.85	4.42E+20	0.07819

Berdasarkan tabel 4.11 diatas untuk model *spline* orde tiga diperoleh titik knot yang menghasilkan nilai GCV minimum adalah dengan menggunakan 3 titik knot yang terletak di titik 389.50, 393.05, dan 409.65 dengan GCV sebesar 0.07814 dengan model sebagai berikut:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \gamma_1 (X - 389.50)_+^3 + \gamma_2 (X - 393.05)_+^3 + \gamma_3 (X - 409.65)_+^3 + u$$

Dari ketiga model *P-Spline* tersebut dapat disimpulkan bahwa kombinasi titik knot terbaik untuk memodelkan rata-rata pengeluaran perkapita masing-masing kecamatan di Kabupaten Sumenep dengan menghilangkan data pencilan (Kota Sumenep) diperoleh dari model *P-Spline* orde satu dengan lima titik knot. Gambar 4.4 ilustrasi lokasi titik knot.



Gambar 4.6 Lokasi Titik Knot *P-Spline* Orde Satu dengan Menghapus Data Pencilan

Setelah orde dan lokasi titik knot diketahui, langkah berikutnya adalah mengestimasi efek tetap dan efek acaknya. Berikut adalah nilai estimasi efek tetap untuk model *P-Spline* orde satu dengan lima titik knot. Nilai estimasi efek acak masing-masing kecamatan dapat dilihat pada lampiran 3.

Tabel 4.12 Estimasi Efek Tetap

Parameter	**Estimasi**
β_0	2.3088
β_1	-0.00047

Sehingga diperoleh estimasi model sebagai berikut:

$$\boldsymbol{Y} = 2.3088 + (-0.00047\,\boldsymbol{X}) + \gamma_1(\boldsymbol{X} - 398.75)_+^1 + \gamma_2(\boldsymbol{X} - 408.47)_+^1 + \gamma_3(\boldsymbol{X} - 481.85)_+^1 + \gamma_4(\boldsymbol{X} - 857.88)_+^1 + \gamma_5(\boldsymbol{X} - 920.98)_+^1 + u$$

Model *P-Spline* untuk menduga pengeluaran perkapita dengan menghapus data pencilan diatas dapat di interpretasikan sebagai berikut:

a. Setiap kenaikan satu satuan x untuk $x \leq 398.75$ akan merubah nilai y sebesar (-0.00047) satuan.
b. Untuk $398.75 < x \leq 408.47$, setiap kenaikan satu satuan x akan berpengaruh sebesar $(-0.00047 + \gamma_1)$ satuan terhadap *y.*
c. Untuk $408.47 < x \leq 481.85$, setiap kenaikan satu satuan x akan berpengaruh sebesar $(-0.00047 + \gamma_1 + \gamma_2)$ satuan terhadap *y.*
d. Untuk nilai $481.85 < x \leq 857.88$, setiap kenaikan satu satuan x akan berpengaruh sebesar $(-0.00047 + \gamma_1 + \gamma_2 + \gamma_3)$ satuan terhadap *y.*
e. Untuk nilai $857.88 < x \leq 920.98$, setiap kenaikan satu satuan x akan berpengaruh sebesar $(-0.00047 + \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4)$ satuan terhadap *y.*
f. Untuk nilai $x > 920.98$, setiap kenaikan satu satuan x akan berpengaruh sebesar $(-0.00047 + \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 + \gamma_5)$ satuan terhadap *y.*

Hasil pendugaan menggunakan *P-Spline* dengan penghapusan data pencilan diperoleh rata-rata pengeluaran perkapita sebesar Rp. 206.590,00. Dengan nilai koefisien variansi sebesar 13.52 persen menunjukkan bahwa nilai perndugaan pengeluaran perkapita di Kabupaten Sumenep tidak terlalu beragam. Nilai pendugaan rata-rata pengeluaran perkapita terkecil terjadi di Kecamatan Bluto dengan pengeluaran sebesar Rp. 152.800,00 sedangkan nilai pengeluaran perkapita terbesar yaitu Rp. 268.900,00 terjadi di Kecamatan Manding. Hasil pendugaan pengeluaran perkapita pada masing-masing kecamatan di Kabupaten Sumenep dengan penghapusan data penculan dapat dilihat pada lampiran 7. Statistika deskriptif hasil pendugaan *P-Spline* dengan penghapusan data pencilan dapat dilihat pada Tabel 4.13.

Tabel 4.13 Statistika Deskriptif Hasil Estimasi *P-Spline* Dengan Menghilangkan Data Pencilan

Statistik	***P-Spline***
Rata-rata	2.0059
Standar Deviasi	0.2712
Variansi	0.0736

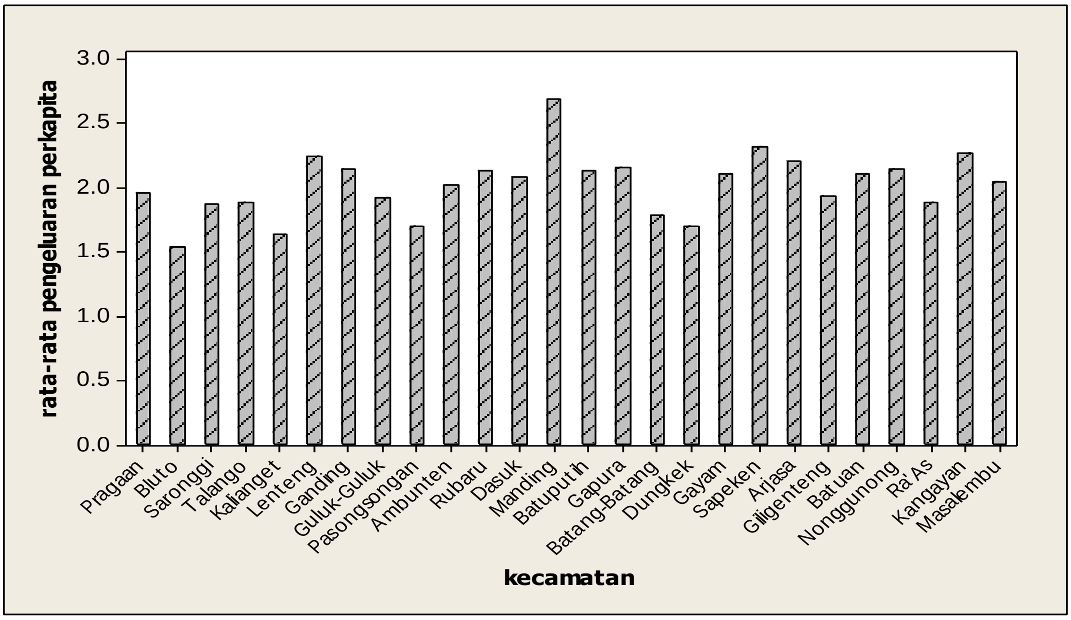
Minimum	1.528
Maximum	2.6890

Estimasi sintetik digunakan untuk mengestimasi pengeluaran perkapita untuk kecamatan-kecamatan dengan jumlah sampel yang sedikit dengan menganggap daerah tersebut memiliki karakteristik yang sama dengan domainnya dengan estimator $\hat{\beta} = [2.3088 \quad -0.00047]$. Hasil estimasi rata-rata pengeluaran perkapita untuk daerah dengan sampel sedikit dapat dilihat pada tabel 4.14.

Tabel 4.14 Hasil Estimasi dengan Menghilangkan Data Pencilan

Kecamatan	Pengeluaran perkapita
Giligenteng	1.933055
Batuan	2.105135
Nonggunong	2.136476
Ra'As	1.875099
Kangayan	2.258261
Masalembu	2.040266

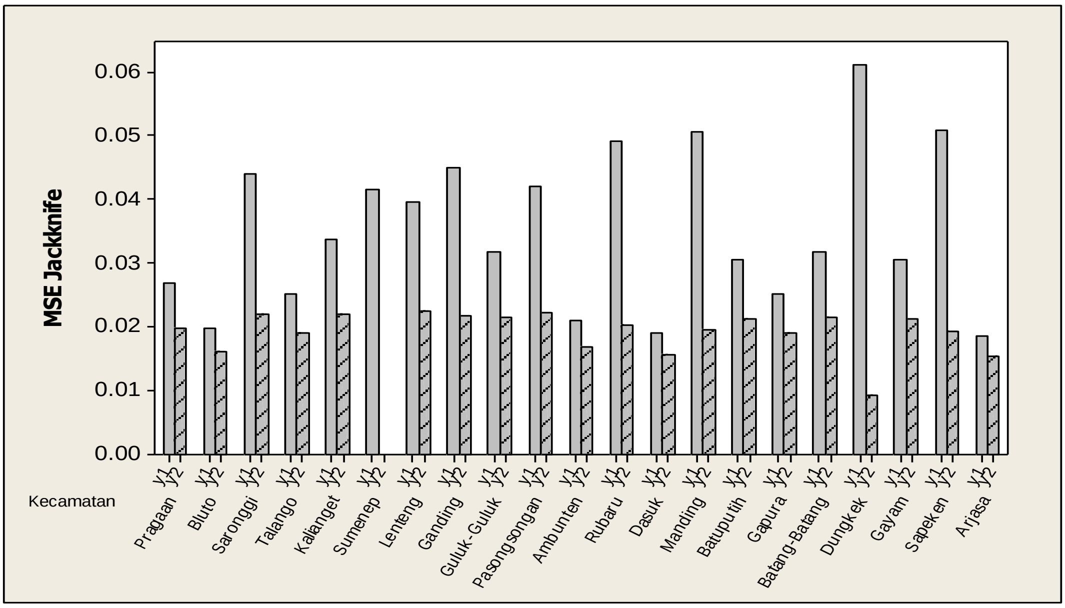
Gambar 4.5 menyajikan hasil pendugaan semua kecamatan dengan metode *P-Spline* dengan penghapusan data pencilan, yaitu data Kota Sumenep. Hasil estimasi untuk daerah dengan jumlah sampel yang sedikit yaitu Kecamatan Giligenteng, Batuan, Nonggunong, Ra’as, Kangayan, dan Masalembu memiliki rentang nilai yang tidak besar dengan kecamatan-kecamatan lainnya di Kabupaten Sumenep.



Gambar 4.7 Grafik Hasil Estimasi Rata-rata Pengeluaran Perkapita Semua Kecamatan Dengan Penghapusan Data Pencilan

4.2 Kebaikan Model Pendugaan

Setelah dilakukan estimasi terhadap rata-rata pengeluaran perkapita baik dengan menggunakan motode *P-Spline* dengan menggunakan keseluruhan data maupun dengan penghapusan data pencilan, langkah selanjutnya adalah menduga nilai MSE dari kedua pendugaan tersebut. pada pendugaan tidak langsung, dilakukan koreksi terhadap nilai MSE dengan menggunakan metode resampling *jackknife*. perbandingan MSE jackknife keduanya disajikan pada gambar berikut:



Gambar 4.8 Grafik MSE *jackknife* Hasil Pendugaan

Dari grafik MSE *jackknife* hasil pendugaan diatas, dapat terlihat bahwa hasil pendugaan dengan menggunakan keseluruhan data (y_1) menghasilkan nilai MSE jackknife yang lebih besar daripada dengan menggunakan data dengan penghapusan data pencilan (y_2) pada setiap kecamatan. Nilai MSE *jackknife* masing-masing kecamatan dapat dilihat pada lampiran 5 dan lampiran 8. Statistika deskriptif MSE jackknife dapat dilihat pada tabel 4.15.

Tabel 4.15 Deskriptif MSE *Jackknife*

Statistik	MSE *Jackknife* Tanpa Penghapusan Data Pencilan	MSE *Jackknife* Dengan Penghapusan Data Pencilan
Rata-rata	0.03518	0.019299
Variansi	0.00015	0.000011
Minimum	0.01867	0.009295
Maximum	0.06111	0.022490

Dapat dilihat pada tabel 4.15, dimana rata-rata MSE *jackknife* dengan menggunakan keseluruhan data adalah 0.035 dengan variansi 0.00015, sedangkan dengan menggunakan data dengan menghapus data pencilan diperoleh rata-rata nilai MSE *jackknife* yang lebih kecil, yaitu 0.019 dengan variansi 0.00001. Berdasarkan penjelasan tersebut, maka pendugaan pengeluaran perkapita tingkat kecamatan di Kabupaten Sumenep akan menghasilkan nilai pendugaan yang lebih baik jika menggunakan data dengan penghapusan data pencilan.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dilakukan pada bab sebelumnya, maka disimpulakan bahwa:

1. Pola sebaran data antara rata-rata pengeluaran perkapita dengan kepadatan penduduk di Kabupaten Sumenep tidak membentuk suatu pola tertentu, dimana persebaran datanya bergerombol dan juga terdapat data pencilan, yaitu pada pengamatan di Kota Sumenep, sehingga pada penelitian ini mencari dua model nonparametrik *P-Spline* yaitu pertama dengan menggunakan keseluruhan data yang diperoleh dan yang kedua adalah dengan menggunakan data setelah menghapus data pencilan.
   a. Model *P-Spline* terbaik untuk mengestimasi pengeluaran perkapita masing-masing kecamatan di Kabupaten Sumenep dengan menggunakan keseluruhan data adalah model *P-Spline* orde satu dengan lima titik knot, yaitu pada titik 398.6199, 408.4978, 482.1517, 872.1651, dan pada titik 926.6157.

      Dengan model sebagai berikut:

      $$\boldsymbol{Y} = 2.224 + (-0.00031\,\boldsymbol{X}) + \gamma_1(\boldsymbol{X} - 398.62)_+^1 + \gamma_2(\boldsymbol{X} - 408.50)_+^1 + \gamma_3(\boldsymbol{X} - 482.15)_+^1 + \gamma_4(\boldsymbol{X} - 872.16)_+^1 + \gamma_5(\boldsymbol{X} - 926.61)_+^1 + u$$

   b. Model *P-spline* terbaik dengan menghilangkan data pencilan untuk mengestimasi pengeluaran perkapita di Kabupaten Sumenep adalah *P-Spline* orde satu satu dengan kombinasi lima titik knot, yaitu pada titik 398.7486, 408.4710, 481.8522, 857.8794 dan 920.9820

      dengan model sebagai berikut:

      $$\boldsymbol{Y} = 2.3088 + (-0.00047\,\boldsymbol{X}) + \gamma_1(\boldsymbol{X} - 398{,}75)_+^1 + \gamma_2(\boldsymbol{X} - 408.47)_+^1 + \gamma_3(\boldsymbol{X} - 481.85)_+^1 + \gamma_4(\boldsymbol{X} - 857.88)_+^1 + \gamma_5(\boldsymbol{X} - 920.98)_+^1 + u$$

2. Berdasarkan nilai MSE *jackknife*, pendugaan pengeluaran perkapita tingkat kecamatan di Kabupaten Sumenep dengan menggunakan *P-spline* akan menghasilkan nilai pendugaan yang lebih baik jika menggunakan data dengan menghapus data pencilan terlebih dahulu (pengamatan Kota Sumenep)

5.2 Saran

Penelitian selanjutnya disarankan agar menggunakan metode *Penalized Spline* dengan variabel bantu lebih dari satu (*multivariabel*).

Lampiran 1.

Data Variabel Bantu Tingkat Kecamatan di Kabupaten Sumenep Berdasakan Kabupaten Sumenep Dalam Angka 2009

No	Kecamatan	Kepadatan Penduduk
1	Pragaan	1117.53
2	Bluto	921.40
3	Saronggi	542.73
4	Giligenteng	794.95
5	Talango	828.80
6	Kalianget	1356.67
7	Kota Sumenep	2535.02
8	Batuan	430.92
9	Lenteng	857.98
10	Ganding	708.10
11	Guluk	887.58
12	Pasongsongan	392.32
13	Ambunten	788.80
14	Rubaru	447.56
15	Dasuk	465.81
16	Manding	408.45
17	Batuputih	390.61
18	Gapura	580.82
19	Batang	669.08
20	Dungkek	600.10
21	Nonggunong	364.62
22	Gayam	401.39
23	Ra'As	917.56
24	Sapeken	198.14
25	Arjasa	252.74
26	Kangayan	106.99
27	Masalembu	568.15

(*halaman ini sengaja dikosongkan*)

Lampiran 2.

Data Pengeluaran Perkapita Tingkat Kecamatan di Kabupaten Sumenep Berdasakan Data SUSENAS 2010

Kecamatan	**Jumlah Sampel**	**Rata-rata Pengeluaran perkapita**	**Di**
Pragaan	92	1.95459	0.029369
Bluto	65	1.52799	0.021063
Saronggi	65	1.86102	0.051935
Talango	66	1.88380	0.027481
Kalianget	64	1.62802	0.037923
Kota Sumenep	113	3.31510	0.048293
Lenteng	79	2.24386	0.045777
Ganding	65	2.13588	0.053314
Guluk-Guluk	73	1.91476	0.035415
Pasongsongan	65	1.69374	0.049071
Ambunten	84	2.01365	0.022516
Rubaru	62	2.12178	0.059438
Dasuk	45	2.07330	0.0203078
Manding	40	2.68897	0.061589
Batuputih	90	2.13224	0.034017
Gapura	64	2.15088	0.027349
Batang-Batang	77	1.78621	0.035437
Dungkek	71	1.69120	0.07946
Gayam	70	2.10407	0.034011
Sapeken	67	2.30692	0.062062
Arjasa	118	2.20532	0.019808

(*halaman ini sengaja dikosongkan*)

Lampiran 3.

Nilai Estimasi Efek Acak *P-Spline* Orde Satu dengan 5 Titik Knot Dengan Menggunakan Keseluruhan Data

No	γ_1	γ_2	γ_3	γ_4	γ_5	$\hat{u}$
1	2.547E-05	2.512E-05	2.251E-05	8.692E-06	6.763E-06	0.03036576
2	-1.315E-04	-1.290E-04	-1.105E-04	-1.239E-05	0	-0.2156343
3	-3.002E-05	-2.796E-05	-1.262E-05	0	0	-0.1785541
4	-2.458E-05	-2.402E-05	-1.981E-05	0	0	-0.04897977
5	-4.103E-05	-4.061E-05	-3.746E-05	-2.075E-05	-1.842E-05	-0.03671443
6	2.073E-04	2.063E-04	1.992E-04	1.613E-04	1.560E-04	0.08316445
7	9.386E-05	9.184E-05	7.679E-05	0	0	0.1751389
8	3.800E-05	3.678E-05	2.774E-05	0	0	0.1052427
9	-9.048E-06	-8.865E-06	-7.502E-06	-2.852E-07	0	-0.01586156
10	0	0	0	0	0	-0.3980288
11	1.225E-05	1.194E-05	9.624E-06	0	0	0.02690245
12	2.213E-06	1.766E-06	0	0	0	0.03876065
13	-2.284E-07	-1.948E-07	0	0	0	-0.002914208
14	6.686E-06	0	0	0	0	0.5829822
15	0	0	0	0	0	0.03149468
16	2.137E-05	2.021E-05	1.157E-05	0	0	0.1005335
17	-5.813E-05	-5.601E-05	-4.018E-05	0	0	-0.1842332
18	-7.150E-05	-6.800E-05	-4.186E-05	0	0	-0.3041999
19	2.332E-08	0	0	0	0	0.007217253
20	0	0	0	0	0	0.1429964
21	0	0	0	0	0	0.06032122

(*halaman ini sengaja dikosongkan*)

Lampiran 4.

Hasil Pendugaan Rata-Rata Pengeluaran Perkapita Tingkat Kecamatan Tersampel di Kabupaten Sumenep dengan Menggunakan Keseluruhan Data

No.	Kecamatan	Pendugaan dengan Keseluruhan Data
1	Pragaan	1.953994
2	Bluto	1.532224
3	Saronggi	1.864526
4	Talango	1.884762
5	Kalianget	1.628741
6	Kota Sumenep	3.313467
7	Lenteng	2.240421
8	Ganding	2.133813
9	Guluk-Guluk	1.915071
10	Pasongsongan	1.701556
11	Ambunten	2.013122
12	Rubaru	2.121019
13	Dasuk	2.073357
14	Manding	2.677522
15	Batuputih	2.131622
16	Gapura	2.148906
17	Batang-Batang	1.789828
18	Dungkek	1.697174
19	Gayam	2.103928
20	Sapeken	2.304112
21	Arjasa	2.204135

(*halaman ini sengaja dikosongkan*)

Lampiran 5.

MSE *Jackknife* Hasil Estimasi Rata-rata Pengeluaran Perkapita Tingkat Kecamatan Tersampel di Kabupaten Sumenep Dengan Menggunakan Keseluruhan Data

No.	Kecamatan	h1	h2	MSE jackknife
1	Pragaan	0.026856	1.19E-06	0.026857
2	Bluto	0.01977	2.35E-05	0.019794
3	Saronggi	0.044077	1.57E-05	0.044093
4	Talango	0.025281	1.64E-06	0.025282
5	Kalianget	0.033733	1.22E-06	0.033734
6	Kota Sumenep	0.041498	6.58E-07	0.041499
7	Lenteng	0.039672	2.43E-05	0.039696
8	Ganding	0.045033	1.34E-05	0.045047
9	Guluk-Guluk	0.031761	1.09E-06	0.031762
10	Pasongsongan	0.042056	7.1E-05	0.042127
11	Ambunten	0.021039	2.98E-06	0.021042
12	Rubaru	0.049146	4.52E-06	0.04915
13	Dasuk	0.019106	3.06E-06	0.019109
14	Manding	0.050538	0.000142	0.05068
15	Batuputih	0.030646	4.51E-06	0.03065
16	Gapura	0.02517	1.11E-05	0.025181
17	Batang-Batang	0.031778	1.73E-05	0.031796
18	Dungkek	0.061066	4.75E-05	0.061113
19	Gayam	0.030641	3.74E-06	0.030645
20	Sapeken	0.050841	1.65E-05	0.050857
21	Arjasa	0.018665	7.25E-06	0.018672

(*halaman ini sengaja dikosongkan*)

Lampiran 6.

Nilai Estimasi Efek Acak *P-Spline* Orde Satu dengan 5 Titik Knot Dengan Penghapusan Data Pencilan

No	γ_1	γ_2	γ_3	γ_4	γ_5	$\hat{u}$
1	3.107E-14	3.07E-14	2.75E-14	1.12E-14	8.5E-15	0.174019
2	-4.48E-14	-4.4E-14	-3.8E-14	-5.4E-15	-3.6E-17	-0.34529
3	-6.84E-15	-6.4E-15	-2.9E-15	0	0	-0.19126
4	-3.55E-15	-3.5E-15	-2.9E-15	0	0	-0.03326
5	-9.4E-15	-9.3E-15	-8.6E-15	-4.9E-15	-4.3E-15	-0.03951
6	3.885E-14	3.8E-14	3.18E-14	8.51E-18	0	0.340598
7	1.243E-14	1.2E-14	9.09E-15	0	0	0.161769
8	3.095E-15	3.03E-15	2.57E-15	1.88E-16	0	0.02549
9	0	0	0	0	0	-0.42964
10	7.527E-15	7.34E-15	5.92E-15	0	0	0.077686
11	2.972E-16	2.38E-16	0	0	0	0.02451
12	-2.56E-16	-2.2E-16	0	0	0	-0.01534
13	1.381E-15	0	0	0	0	0.573212
14	0	0	0	0	0	0.008049
15	5.274E-15	4.99E-15	2.87E-15	0	0	0.116603
16	-1.39E-14	-1.3E-14	-9.6E-15	0	0	-0.20635
17	-1.67E-14	-1.6E-14	-9.8E-15	0	0	-0.33396
18	-9.86E-18	0	0	0	0	-0.01502
19	0	0	0	0	0	0.091748
20	0	0	0	0	0	0.015957

(*halaman ini sengaja dikosongkan*)

Lampiran 7.

Hasil Estimasi Rata-rata Pengeluaran Perkapita Tingkat Kecamatan Tersampel di Kabupaten Sumenep Dengan Penghapusan Data Pencilan

No.	Kecamatan	Pendugaan Dengan Penghapusan Data Pencilan
1	Pragaan	1.95459
2	Bluto	1.52799
3	Saronggi	1.86102
4	Talango	1.8838
5	Kalianget	1.62802
6	Lenteng	2.24386
7	Ganding	2.13588
8	Guluk-Guluk	1.91476
9	Pasongsongan	1.69374
10	Ambunten	2.01365
11	Rubaru	2.12178
12	Dasuk	2.0733
13	Manding	2.68897
14	Batuputih	2.13224
15	Gapura	2.15088
16	Batang-Batang	1.78621
17	Dungkek	1.6912
18	Gayam	2.10407
19	Sapeken	2.30692
20	Arjasa	2.20532

(*halaman ini sengaja dikosongkan*)

Lampiran 8.

MSE *jackknife* Hasil Estimasi Rata-rata Pengeluaran Perkapita Tingkat Kecamatan Tersampel di Kabupaten Sumenep Dengan Penghapusan Data Pencilan

No.	Kecamatan	h1	h2	MSE jackknife
1	Pragaan	0.019784	5.45E-14	0.019784
2	Bluto	0.016133	3.64E-13	0.016133
3	Saronggi	0.021961	5.58E-14	0.021961
4	Talango	0.019089	1.38E-14	0.019089
5	Kalianget	0.021941	4.92E-14	0.021941
6	Lenteng	0.02249	2.83E-13	0.02249
7	Ganding	0.021727	2E-13	0.021727
8	Guluk-Guluk	0.021477	1.29E-14	0.021477
9	Pasongsongan	0.022312	4.25E-13	0.022312
10	Ambunten	0.016882	2.27E-14	0.016882
11	Rubaru	0.020178	1.22E-14	0.020178
12	Dasuk	0.015725	5.64E-15	0.015725
13	Manding	0.019436	5.17E-13	0.019436
14	Batuputih	0.021158	9.82E-15	0.021158
15	Gapura	0.019037	1.19E-13	0.019037
16	Batang-Batang	0.021482	1.87E-13	0.021482
17	Dungkek	0.009295	3E-13	0.009295
18	Gayam	0.021156	7.45E-15	0.021156
19	Sapeken	0.019259	6.22E-14	0.019259
20	Arjasa	0.015448	1.34E-14	0.015448

(*halaman ini sengaja dikosongkan*)

Lampiran 9.

Program dengan Menggunakan *Software* R

```
library(pspline)
library(freeknotsplines)
library(nlme)

dt=read.table("data fix.txt",header=T)
y=dt$y
x=dt$x
idnum=dt$id

xy.freekt1_1 <- fit.search.numknots(x, y, degree = 1, minknot = 1, maxknot = 1,alg="PS", seed = 555)
xy.freekt1_2 <- fit.search.numknots(x, y, degree = 1, minknot = 2, maxknot = 2,alg="PS", seed = 555)
xy.freekt1_3 <- fit.search.numknots(x, y, degree = 1, minknot = 3, maxknot = 3,alg="PS", seed = 555)
xy.freekt1_4 <- fit.search.numknots(x, y, degree = 1, minknot = 4, maxknot = 4,alg="PS", seed = 555)
xy.freekt1_5 <- fit.search.numknots(x, y, degree = 1, minknot = 5, maxknot = 5,alg="PS", seed = 555)

xy.freekt2_1 <- fit.search.numknots(x, y, degree = 2, minknot = 1, maxknot = 1,alg="PS", seed = 555)
xy.freekt2_2 <- fit.search.numknots(x, y, degree = 2, minknot = 2, maxknot = 2,alg="PS", seed = 555)
xy.freekt2_3 <- fit.search.numknots(x, y, degree = 2, minknot = 3, maxknot = 3,alg="PS", seed = 555)
xy.freekt2_4 <- fit.search.numknots(x, y, degree = 2, minknot = 4, maxknot = 4,alg="PS", seed = 555)
xy.freekt2_5 <- fit.search.numknots(x, y, degree = 2, minknot = 5, maxknot = 5,alg="PS", seed = 555)

xy.freekt3_1 <- fit.search.numknots(x, y, degree = 3, minknot = 1, maxknot = 1,alg="PS", seed = 555)
xy.freekt3_2 <- fit.search.numknots(x, y, degree = 3, minknot = 2, maxknot = 2,alg="PS", seed = 555)
xy.freekt3_3 <- fit.search.numknots(x, y, degree = 3, minknot = 3, maxknot = 3,alg="PS", seed = 555)
xy.freekt3_4 <- fit.search.numknots(x, y, degree = 3, minknot = 4, maxknot = 4,alg="PS", seed = 555)
xy.freekt3_5 <- fit.search.numknots(x, y, degree = 3, minknot = 5, maxknot = 5,alg="PS", seed = 555)

X=cbind(rep(1,length(y)),x)
knot=xy.freekt1_5@optknot
z.spline=outer(x,knot,"-")
z.spline=z.spline*(z.spline>0)

id=factor(idnum)
group <- 1:21

ZBlock <- pdBlocked(list(pdIdent(~z.spline-1),pdIdent(~id-1)))
dataFr <- groupedData(y~x|group,data=data.frame(y,X,z.spline,idnum))
fit=lme(fixed=y ~ -1+X, data=dataFr, random=ZBlock)

fit$coef

beta.hat=fit$coef$fixed
random=fit$coef$random
```

(*halaman ini sengaja dikosongkan*)

BIOGRAFI PENULIS

Zulhan Widya Baskara atau bisa dipanggil Zulhan lahir di Mataram pada tanggal 1 Agustus 1987. Anak kedua dari dua bersaudara pasangan H Wildan dan Hj. Retno Widowati A. Penulis menempuh jenjang pendidikan di SD Negeri 13 Mataram, SMPN 1 Mataram, SMA 1 Mataram, Jurusan matematika Murni Universitas Negeri Malang, dan setelah lulus di S1 penulis melanjutkan studi S2 di Jurusan Statistika FMIPA Institut Teknologi Sepuluh Nopember.

Bagi pembaca yang ingin berdiskusi, memberikan saran dan kritik tentang Tesis ini dapat disampaikan melalui email zulhan_widya_baskara@ymail.com. Semoga Tesis ini dapat bermanfaat bagi pembaca.

Terima Kasih,
Zulhan