



**APLIKASI GAMLSS DENGAN PEMULUSAN LOESS DAN
ALGORITMA RIGBY-STASINOPOULOS PADA
DATA CACAHAN**

SKRIPSI

Oleh

**Laily Fauziah
NIM 101810101029**

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2015**



**APLIKASI GAMLSS DENGAN PEMULUSAN LOESS DAN
ALGORITMA RIGBY-STASINOPOULOS PADA
DATA CACAHAN**

SKRIPSI

diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat untuk
menyelesaikan Program Studi Matematika (S1)
dan mencapai gelar Sarjana Sains

Oleh

Laily Fauziah
NIM 101810101029

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2015

PERSEMBAHAN

Skripsi ini saya persembahkan untuk:

1. Ayahanda Syamsul Arifin dan Ibunda Nurhayati yang memberikan seluruh semangat dan doanya selama perjalanan studi,
2. Adikku Rika Wulandari yang tersayang,
3. Prof. Drs. I Made Tirta M.Sc., Ph. D selaku Dosen Pembimbing Utama dan Dian Anggraeni, S.Si., M.Si selaku Dosen Pembimbing Anggota yang telah membimbing secara intensif dalam menyempurnakan skripsi ini;
4. Almamater Jurusan Matematika FMIPA Universitas Jember, SMAN 1 Genteng, SMPN 1 Srono, dan SDN 1 Rejoagung;
5. Seluruh teman-teman Mathgic '10 yang tercinta.

MOTTO

“Setiap insan yang kuat hari ini, datangnya dari kelemahan di masa lalu”

(Anonim)*)

“Dan bersabarlah dalam menunggu ketetapan Tuhanmu, maka sesungguhnya kamu berada dalam penglihatan Kami, dan bertasbihlah dengan memuji Tuhanmu ketika kamu bangun berdiri”

(Ath-Thur, Ayat 48)**)

*[http://lelimentarikecil.blogspot.com/2015/06/setiap-insan-yang-kuat-hari-
ini.html?m=1](http://lelimentarikecil.blogspot.com/2015/06/setiap-insan-yang-kuat-hari-
ini.html?m=1) [23 Juni 2015]

**Departemen Agama Republik Indonesia. 2005. Alhikmah, Al-Qur'an dan terjemahannya. Bandung : CV Penerbit Diponegoro.

PERNYATAAN

Saya yang bertandatangan di bawah ini:

Nama : Laily Fauziah

NIM : 101810101029

menyatakan dengan sebenarnya bahwa karya ilmiah yang berjudul “Aplikasi GAMLSS dengan Pemulusan Loess dan Algoritma Rigby-Stasinopoulos pada Data Cacahan” adalah benar-benar hasil karya sendiri, kecuali kutipan yang sudah saya sebutkan sumbernya, belum pernah diajukan pada institusi manapun, dan bukan karya jiplakan. Saya bertanggungjawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenarnya, tanpa ada tekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik apabila ternyata di kemudian hari pernyataan ini tidak benar.

Jember, Juni 2015

Yang menyatakan,

Laily Fauziah

NIM 101810101029

SKRIPSI

**APLIKASI GAMLSS DENGAN PEMULUSAN LOESS DAN
ALGORITMA RIGBY-STASINOPOULOS PADA
DATA CACAHAN**

Oleh

Laily Fauziah

NIM 101810101029

Pembimbing:

Dosen Pembimbing Utama : Prof. Drs. I Made Tirta, M.Sc, Ph.D

Dosen Pembimbing Anggota : Dian Anggraeni, S.Si, M.Si

PENGESAHAN

Skripsi berjudul “Aplikasi GAMLSS dengan Pemulusan Loess dan Algoritma Rigby-Stasinopoulos pada Data Cacahan” telah diuji dan disahkan pada:

Hari, tanggal :

Tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Tim Penguji:

Dosen Pembimbing Utama,

Dosen Pembimbing Anggota,

Prof. Drs. I Made Tirta, M.Sc., Ph.D.
NIP. 195912201985031002

Dian Anggraeni, S.Si, M.Si.
NIP. 198202162006042002

Penguji I,

Penguji II,

Dr. Alfian Futuhul Hadi, S.Si, M.Si.
NIP. 197407192000121001

Ika Hesti Agustin, S.Si., M.Si
NIP. 198408012008012006

Mengesahkan

Dekan,

Prof. Drs. Kusno, DEA., Ph.D.
NIP. 196101081986021001

RINGKASAN

Aplikasi GAMLSS dengan Pemulusan Loess dan Algoritma Rigby-Stasinopoulos pada Data Cacahan ; Laily Fauziah,101810101029; 2015; 36 halaman; Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

GAMLSS atau *Generalized Linear Model for Location, Scale and Shape* merupakan salah satu metode pencocokan model yang dapat memodelkan 4 parameter distribusi sekaligus, yaitu parameter lokasi (μ), parameter skala (σ), *skewness* (ν) dan kurtosis (τ). Metode ini lebih fleksibel karena mencakup perluasan dari distribusi keluarga eksponensial untuk menangani data overdispersi, termasuk data dengan kemencengan yang tinggi (*highly skew*), baik data kontinu maupun cacahan. GAMLSS juga merupakan perluasan dari GLM dan GAM dengan cakupan distribusi yang lebih luas. Karena merupakan perluasan dari GAM, metode ini juga menyediakan *smoothing* saat respon data yang akan dimodelkan tidak memenuhi asumsi kelinieran

Pada penelitian ini, jenis data yang digunakan merupakan data diskrit atau cacahan, yaitu jumlah kematian balita akibat penyakit pneumonia di seluruh provinsi di Indonesia (kecuali Provinsi Bengkulu dan Jawa Barat). Tujuannya adalah mengaplikasikan GAMLSS dengan algoritma Rigby-Stasinopoulos (RS) dan pemulusan loess untuk data cacahan dengan contoh datanya adalah data pneumonia. Untuk selanjutnya akan dianalisis bagaimana bentuk model terbaik dari aplikasi GAMLSS dengan metode pemulusan loess, untuk data jumlah kematian balita akibat pneumonia. Jumlah kematian ini diklasifikasikan sebagai variabel respon, sedangkan jumlah kasus gizi buruk, cakupan imunisasi dasar yang diberikan, jumlah rumah tangga yang sehat dan bersih, serta jumlah puskesmas tiap provinsi termasuk variabel eksplanatori. Selain itu, dari model terbaik akan dilihat hasil estimasi dari parameter

lokasi, skala, dan ukuran. Analisis ini menggunakan *software* R dengan paket *gamlss* dan *ggplot2*.

Hasil analisis pada penelitian ini menunjukkan bahwa dengan mengaplikasikan GAMLSS dengan metode *smoothing loess*, didapatkan distribusi yang terbaik untuk data tersebut adalah Distribusi Negatif Binomial II dengan parameter distribusi (μ, σ) , *span* (rentang) terbaik yaitu 0,5 dan derajat polinomial 2. Variabel yang di-*smoothing* adalah variabel imunisasi. Berdasarkan model terbaik yang didapat, dapat diketahui bahwa cakupan imunisasi dasar yang diberikan pada balita berpengaruh secara signifikan pada jumlah kematian balita akibat pneumonia. Adapun model terbaik untuk variabel jumlah kematian balita pada data pneumonia adalah:

- a. $\log(\hat{\mu}_Y) = \hat{\gamma}_1$ loess (Imunisasi, span=0,5, degree=2) dengan $\hat{\gamma}_1 = -15,84$ sehingga parameter lokasi $\hat{\mu}_Y = e^{-15,84}$
- b. Untuk parameter *scale* (σ) diperoleh: $\log(\hat{\sigma}_Y)=2,495$ atau
$$\hat{\sigma}_Y = e^{2,495}$$

PRAKATA

Alhamdulillah, segala puji hanya bagi Allah atas segala rahmat dan karunia-Nya sehingga dapat terselesaikannya skripsi dengan judul “Aplikasi GAMLSS dengan Pemulusan Loess dan Algoritma Rigby-Stasinopoulos pada Data Cacahan”.

Penyusunan skripsi ini tidak lepas dari bantuan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Prof. Drs. I Made Tirta, M.Sc., Ph.D selaku Dosen Pembimbing Utama dan Dian Anggraeni, S.Si, M.Si selaku Dosen Pembimbing Anggota yang telah menyisihkan sedikit waktunya untuk membimbing penulis selama menyusun skripsi ini;
2. Dr. Alfian Futuhul Hadi, S.Si, M.Si selaku Dosen Penguji I dan Ika Hesti Agustin, S.Si., M.Si selaku Dosen Penguji II yang telah banyak memberikan kritik dan saran dalam penulisan skripsi ini;
3. Teman-teman angkatan 2010 yang telah memberikan motivasinya;
4. Aiko, Kak Yeni, serta adikku Rika Wulandari yang telah memberikan dukungan dan motivasi;
5. Teman-teman kos Ganxter 41A (Nia, Latifah, Lut, Aini) yang selalu memberikan semangat dalam kondisi apapun.

Penulis menerima segala kritik dan saran demi kesempurnaan skripsi ini. Akhirnya penulis berharap, semoga skripsi ini dapat bermanfaat.

Jember, Juni 2015

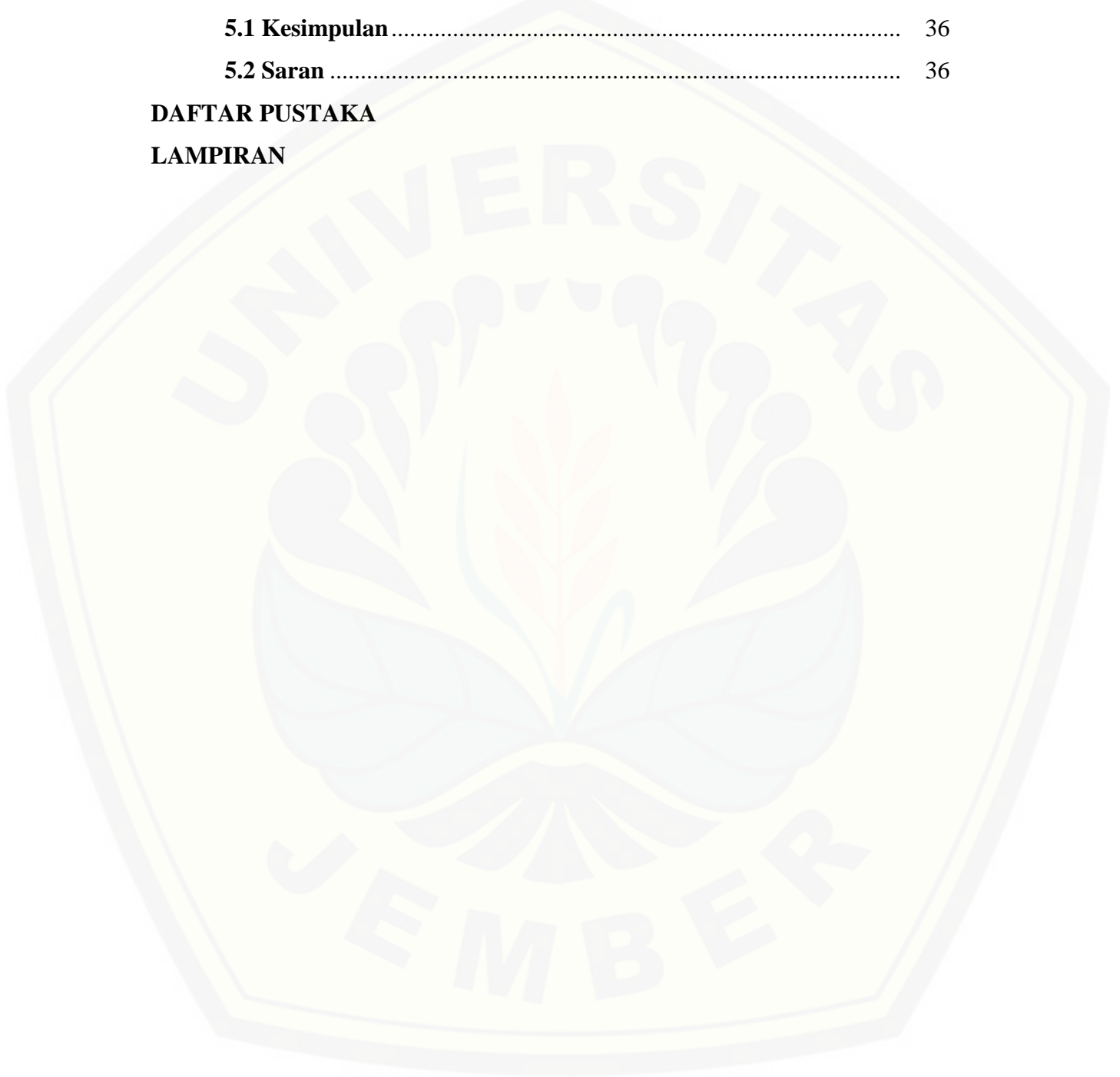
Penulis

DAFTAR ISI

| | Halaman |
|--|---------|
| HALAMAN JUDUL | i |
| HALAMAN PERSEMBAHAN | ii |
| HALAMAN MOTTO | iii |
| HALAMAN PERNYATAAN | iv |
| HALAMAN PEMBIMBINGAN | v |
| HALAMAN PENGESAHAN | vi |
| RINGKASAN | vii |
| PRAKATA | ix |
| DAFTAR ISI | x |
| DAFTAR TABEL | xiii |
| DAFTAR GAMBAR | xiv |
| DAFTAR LAMPIRAN | xv |
| BAB 1 PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 3 |
| 1.3 Tujuan Penelitian | 3 |
| 1.4 Manfaat Penelitian | 3 |
| BAB 2 TINJAUAN PUSTAKA | 4 |
| 2.1 Model Linier | 4 |
| 2.2 <i>Generalized Additive Model (GAM)</i> | 6 |
| 2.3 <i>Skewness dan kurtosis</i> | 6 |
| 2.3.1 <i>Skewness</i> | 7 |
| 2.3.2 <i>Kurtosis</i> | 7 |

| | |
|---|----|
| 2.4 Generalized Additive Model for Location and Shape (GAMLSS) | 8 |
| 2.4.1 Bentuk dan Asumsi GAMLSS | 9 |
| 2.4.2 Algoritma Rigby & Stasinopoulos (RS) | 10 |
| 2.4.3 Metode Pemulusan Loess..... | 12 |
| 2.4.4 Distribusi untuk Data Cacahan yang Dapat Dipakai pada GAMLSS | 14 |
| 2.5 Akaike's Information Criterion (AIC) | 16 |
| 2.6 Data Cacahan | 16 |
| BAB 3 METODE PENELITIAN | 18 |
| 3.1 Data | 18 |
| 3.2 Pengolahan Data | 19 |
| 3.3 Analisis Data | 20 |
| BAB 4 HASIL DAN PEMBAHASAN | 22 |
| 4.1 Eksplorasi Data | 22 |
| 4.2 Pengepasan Model pada Data Menggunakan GAMLSS dengan Algoritma RS dan Metode Pemulusan Loess | 24 |
| 4.2.1 Menentukan Distribusi Terbaik | 24 |
| 4.2.2 Menganalisis Data Menggunakan Pemulusan Loess | 26 |
| 4.2.1.a Menentukan memilih derajat polinomial..... | 26 |
| 4.2.2.b Memilih span yang cocok | 27 |
| 4.3 Memvisualisasikan Perbandingan Rentang (<i>Span</i>) dalam Bentuk Plot | 28 |
| 4.4 Meng-<i>update</i> Distribusi Model Terbaik Setelah Dilakukan Smoothing | 29 |
| 4.5 Menentukan Variabel yang Berpengaruh pada Data Pneumonia | 30 |
| 4.6 Menentukan Model Terbaik dengan Menggunakan stepGAIC () | 31 |

| | |
|--|----|
| 4.7 Menentukan Nilai Estimasi Parameter dari Model Terbaik ... | 34 |
| BAB 5 PENUTUP | 36 |
| 5.1 Kesimpulan | 36 |
| 5.2 Saran | 36 |
| DAFTAR PUSTAKA | |
| LAMPIRAN | |



DAFTAR TABEL

| | Halaman |
|--|---------|
| Tabel 3.1 Variabel Prediktor (x) yang digunakan | 18 |
| Tabel 4.1 Nilai AIC Distribusi Model | 26 |
| Tabel 4.2 Tabel <i>Degree</i> Terbaik | 27 |
| Tabel 4.3 Tabel Nilai AIC <i>Span</i> terbaik | 28 |
| Tabel 4.4 Tabel <i>Update</i> Distribusi | 30 |
| Tabel 4.5 Nilai AIC Model GAMLSS dengan Variasi Variabel | 31 |
| Tabel 4.6 Nilai Estimasi, Nilai t dan p -value | 34 |

DAFTAR GAMBAR

| | Halaman |
|---|---------|
| Gambar 3.1 Skema Metode Penelitian..... | 21 |
| Gambar 4.1 <i>Scatterplotmatrix</i> Hubungan Antara Variabel Respon dan Variabel Eksplanatori..... | 23 |
| Gambar 4.2 Histogram Masing-Masing Pilihan Distribusi | 25 |
| Gambar 4.3 Plot Data Sebelum di- <i>smoothing</i> | 28 |
| Gambar 4.4 Visualisasi <i>Smoothing</i> dengan Beberapa Pilihan <i>Span</i> | 29 |
| Gambar 4.5 Plot Residual Model ms5b | 33 |

DAFTAR LAMPIRAN

| | Halaman |
|---|---------|
| A. Data Jumlah Kematian Akibat Penyakit Pneumonia pada Balita Selama Tahun 2013 di Seluruh Indonesia..... | 39 |
| B. Tabel Daftar Distribusi Diskrit yang Dapat Diimplementasikan dalam Paket GAMLSS..... | 41 |
| C. Pencocokan <i>Generalized Additive Model for Location Scale and Shape</i> (GAMLSS) Secara Umum | 42 |
| D. Program..... | 44 |
| D1. Eksplorasi Data | 44 |
| D2. Analisis Distribusi Terbaik | 45 |
| D3. Analisis Derajat Polinomial (<i>degree</i>) Terbaik | 47 |
| D4. Analisis <i>Span</i> Terbaik | 47 |
| D5. Visualisasi Model Sesudah dilakukan <i>Smoothing</i> dengan Beberapa Pilihan <i>Span</i> | 49 |
| D6. Melihat Update Distribusi Terbaik Setelah di- <i>Smoothing</i> | 50 |
| D7. Analisis Variabel yang Berpengaruh pada Data Pneumonia | 51 |
| D8. Ringkasan Variabel Paling Berpengaruh | 53 |
| D9. Analisis Model Terbaik dengan Menggunakan <i>stepGAIC()</i> | 53 |
| D10. Plotting dan Ringkasan Residual Model ms5b | 55 |

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Teori regresi guna menganalisis data baik berupa cacahan maupun kontinu semakin berkembang selama beberapa dekade terakhir. Selain digunakan untuk melakukan prediksi, analisis regresi juga digunakan untuk mencari hubungan antar dua variabel atau lebih. Penerapan model regresi pun saat ini diharapkan semakin fleksibel dengan berbagai jenis data dan algoritma yang memungkinkan untuk menganalisis data secara cepat dan akurat.

Awal tahun 1990, Galton (1822-1911) memperkenalkan model regresi dengan mencari hubungan tinggi badan ayah dan anaknya. Model regresi paling sederhana adalah Model Linier Normal atau biasa disebut *Normal Linear Model* (NLM), tapi NLM hanya mampu menganalisis data yang berdistribusi normal. Faktanya, banyak data yang tidak memenuhi asumsi kenormalan sehingga berkembanglah model regresi linier umum, atau *Generalized Linear Model* (GLM). Model ini dapat menganalisis data berdistribusi eksponensial, termasuk distribusi normal. Namun, saat respon data yang dimodelkan tidak memenuhi asumsi kelinieran, maka akan dilakukan *smoothing* yang tercakup dalam model tambahan dengan mengganti prediktor linier pada GLM dengan prediktor aditif, sehingga model ini disebut dengan *Generalized Additive Model* (GAM). GAM pun hanya memodelkan parameter lokasi (μ) dan parameter skala (σ), namun tidak memodelkan parameter distribusi yang lain seperti parameter ukuran yaitu *skewness* (ν) dan kurtosis (τ). Parameter lokasi (mean) diperlukan untuk menggambarkan letak pemusatan data, parameter skala (varian) sebagai parameter dispersi atau penyebaran data, sedangkan parameter ukuran yang terdiri dari *skewness* (ν) dan kurtosis (τ) digunakan untuk mendefinisikan bentuk kurva dari sebaran data.

Namun, pada kedua model regresi diatas *skewness* dan kurtosis tidak dimodelkan secara eksplisit dalam variabel eksplanatori, tapi secara implisit melalui ketergantungan *skewness* (kemencengan) dan kurtosis (keruncingan) terhadap μ . *Skewness* dan kurtosis sendiri diperlukan untuk melihat bagaimana bentuk kurva tersebut, sangat menceng, curam, atau landai, sehingga diharapkan pemodelan statistik yang dilakukan bisa lebih baik.

Berdasarkan kekurangan yang terdapat pada model-model diatas, Rigby dan Stasinopoulos, *et al.*, (2007) memperkenalkan sebuah metode pencocokan model yang disebut *Generalized Additive Model for Location, Scale & Shape* (GAMLSS) yang dapat memodelkan 4 parameter distribusi diatas. Selain itu, metode ini lebih fleksibel karena mencakup perluasan dari distribusi keluarga eksponensial untuk menangani data overdispersi, termasuk data dengan kemencengan yang tinggi (*highly skew*), baik data kontinu maupun cacahan. Model regresi ini juga merupakan perluasan dari GLM dan GAM.

Terdapat beberapa penelitian terkait GAMLSS, diantaranya adalah Wandri (2012) yang telah meneliti data agroklimatologi berupa data kontinu pada sub proyek Pekalen-Sampean Kabupaten Jember dengan menggunakan paket GAMLSS, algoritma RS (Rigby-Stasinopoulos), dan pemulusan *cubic spline*. Hasil penelitiannya menunjukkan distribusi yang sesuai untuk data kontinu tersebut adalah *Generalized Gamma* dengan memodelkan parameter μ, σ dan ν beserta hasil estimasinya. Disamping itu, Gilchrist, *et al.*, (2011) juga telah menerapkan GAMLSS dengan algoritma RS dan pemulusan loess untuk memodelkan parameter distribusi pada data pendapatan film *box-office* di tahun 1930 yang *highly skewness* berupa data kontinu. Hasilnya model *Box Cox Power Eksponensial* (BCPE) adalah model pengepasan terbaik untuk data tersebut.

Berdasarkan pada penelitian diatas, maka penelitian kali ini akan mengaplikasikan GAMLSS dengan algoritma RS dan pemulusan loess untuk data cacahan. Data yang digunakan merupakan data sekunder dari Dinas Kesehatan

Republik Indonesia mengenai tingkat kematian balita akibat penyakit pneumonia dari seluruh provinsi di Indonesia selama tahun 2013. Untuk selanjutnya akan dianalisis hubungan antara jumlah kematian akibat pneumonia dengan jumlah kasus gizi buruk, cakupan imunisasi dasar yang diberikan, jumlah rumah tangga yang sehat dan bersih, serta jumlah puskesmas tiap provinsi sebagai variabel eksplanatorinya.

1.2 Rumusan Masalah

Adapun rumusan masalah yang didapat dari latar belakang diatas adalah sebagai berikut.

- a. Bagaimana model terbaik dari aplikasi GAMLSS dengan pemulusan loess dan algoritma RS untuk data cacahan?
- b. Bagaimana hasil estimasi dari parameter lokasi, skala, dan ukuran dari model terbaik dengan pemulusan loess dan algoritma RS terhadap data cacahan?

1.3 Tujuan Penelitian

Tujuan yang dicapai dalam penulisan tugas akhir ini adalah untuk mendapatkan model terbaik yang sesuai untuk data cacahan dengan menggunakan GAMLSS dengan pemulusan loess dan algoritma RS. Selain itu, tujuan lainnya adalah mendapatkan hasil estimasi parameter lokasi, skala, dan ukuran dari model yang terbaik.

1.4 Manfaat Penelitian

Manfaat yang diharapkan dari penyusunan tugas akhir ini adalah sebagai bahan perbandingan metode statistika bagi peneliti selanjutnya terutama yang berhubungan dengan GAMLSS dan pemulusan loess serta sebagai informasi atas variabel yang berpengaruh pada jumlah kematian akibat pneumonia di Indonesia pada tahun 2013.

BAB 2 TINJAUAN PUSTAKA

2.1 Model Linier

Perkembangan model linier dimulai dengan adanya kajian oleh Galton (1822-1911) yang membahas tentang hubungan tinggi badan ayah dan anaknya, dilanjutkan dengan perkembangan analisis regresi pada abad ke-19 oleh Pearson yang kemudian bersambung dengan pengembangan korelasi setelah itu. Berlanjutnya perkembangan model linier tidak bisa dilepaskan dengan adanya teori matriks atau aljabar linier yaitu determinan, invers, dan perkalian matriks. Melalui teori matriks tersebut pembahasan model linier dapat didekati secara umum.

Tirta (2009) menyatakan bahwa bentuk umum dari pemodelan linier ini adalah:

$$y_i = \sum_{j=0}^p x_{ij} \beta_j + \epsilon_i \quad (2.1)$$

untuk $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, p$ dengan x_i merupakan peubah tetap, β_j adalah parameter yang menentukan koefisien dari x_i dan ϵ_i adalah galat yang diasumsikan merupakan peubah acak yang mengikuti distribusi tertentu. Pada pemodelan stokastik, kesalahan atau galat yang diasumsikan merupakan peubah acak yang berasal dari distribusi tertentu, misalnya distribusi normal. Jika galatnya diasumsikan berdistribusi normal dan respon saling independen, maka model persamaan (2.1) disebut model linier normal (NLM) atau dapat disebut dengan model linier klasik. Bentuk model linier normal dapat dituliskan dalam bentuk matriks sebagai berikut:

$$Y = X\beta + \epsilon \quad (2.2)$$

dengan asumsi \mathbf{X} adalah peubah tetap dan diukur tanpa galat, sedangkan ϵ adalah komponen kesalahan dimana galat antar respon saling bebas dan berdistribusi $N(0, \sigma^2)$.

Dalam beberapa kasus di lapangan menyatakan bahwa distribusi respon tidak selalu berdistribusi normal, meskipun masih saling bebas. Kondisi seperti ini dapat diatasi dengan melakukan transformasi dari respon. Namun, hal ini dapat berdampak pada syarat ketidakbergantungan dimana akan terdapat kerancuan dalam menafsirkan hasil penelitian karena efek yang diuji adalah dalam skala logaritma, bukan dalam skala aslinya. Nelder dan Wedderburn mengembangkan model linier yang dikenal dengan *Generalized Linear Model* (GLM) atau model linier tergeneralisir. Model ini mempunyai cakupan distribusi yang lebih luas, yaitu menggunakan asumsi bahwa respon berdistribusi keluarga eksponensial dimana distribusi normal termasuk di dalamnya.

Menurut Nelder dan Wedderburn (1972) GLM terdiri dari 3 komponen, yaitu:

1. Komponen acak, menentukan distribusi bersyarat dari variabel respon, dimana Y_i saling bebas atau independen. Distribusi dari Y_i adalah anggota dari keluarga eksponensial, seperti Distribusi Normal, Binomial, Poisson, Gamma, atau distribusi dari keluarga Invers-Gaussian.
2. Prediktor linier, yaitu sebuah fungsi dari $\eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}$ dimana x_i adalah variabel eksplanatori untuk unit sebanyak j dengan efek acak β .
3. Fungsi *link*, yaitu fungsi yang mentransformasikan ekspektasi dari variabel respon, $\mu_i = E(Y_i)$ dengan prediktor linier:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}$$

Berdasarkan uraian diatas, GLM memiliki keterbatasan yaitu hanya mampu mencocokkan model dengan variabel respon berdistribusi eksponensial dengan prediktor linier. Untuk data yang tidak linier, misalnya mengandung pencilan,

GLM tidak tepat lagi diaplikasikan pada data tersebut. Oleh karena itu, dibutuhkan model baru yang fungsi dari variabel-variabelnya tidak harus linier, melainkan mencakup pemulusan (*smoothing*) yang kemudian disebut model aditif.

2.2 *Generalized Additive Model* (GAM)

Hastie dan Tibshirani (1990) menjelaskan bahwa GAM atau model aditif umum yang mencakup fungsi pemulusan dari variabel eksplanatori. Secara umum model ini memiliki struktur sebagai berikut:

$$Y = \sum_{j=0}^p f_j(X_j) + \epsilon \quad (2.3)$$

dengan $E(\epsilon) = 0$, $\text{var}(\epsilon) = \sigma^2$ dan ϵ berdistribusi keluarga eksponensial. GAM disebut juga sebagai perluasan dari GLM karena mewadahi prediktor linier $\sum_{j=0}^p \beta_j x_j$, prediktor nonlinier dan *smoothing*, dimana ketiga komponen tersebut tercakup dalam fungsi aditif $\sum_{j=0}^p f_j(x_j)$, sehingga prediktor GAM disebut dengan prediktor aditif.

Namun baik GLM maupun GAM tidak dapat memodelkan *skewness* dan kurtosis secara eksplisit ke dalam variabel eksplanatori. Keduanya hanya disebutkan secara implisit melalui ketergantungan *skewness* dan kurtosis terhadap μ . Selain itu, GAM hanya mampu mengakomodasi distribusi keluarga eksponensial, oleh karena itu dibutuhkan sebuah model baru yang mampu memodelkan keempat parameter distribusi (termasuk *skewness* dan kurtosis) dan mencakup distribusi yang lebih luas dari keluarga eksponensial.

2.3 *Skewness* dan *Kurtosis*

Meskipun μ dan σ adalah ukuran deskriptif yang signifikan untuk menggambarkan pusat lokasi dan dispersi dari fungsi kepadatan peluang $f(x)$, keduanya tidak memberikan karakterisasi yang unik dari distribusi. Adanya nilai *skewness* (kemiringan) dan kurtosis (keruncingan) diharapkan dapat memberikan

gambaran lebih lengkap dalam memahami data yang terkumpul, sehingga diharapkan model statistik yang dilakukan bisa lebih valid.

2.3.1 *Skewness*

Skewness merupakan pengukuran tingkat ketidaksimetrisan (kemencengan) sebaran data di sekitar rata-ratanya. *Skewness* digunakan untuk menunjukkan kesimetrisan bentuk kurva yang dihasilkan dari distribusi suatu gugusan data. Untuk mengukur derajat kemencengan suatu distribusi dinyatakan dengan koefisien kemencengan (koefisien *skewness*). Menurut Ramachandran dan Tsokos (2009) *skewness* didefinisikan sebagai momen ke-3 standar terhadap mean, yaitu:

$$v_i = \frac{E[(X-\mu)]^3}{\sigma^3} \quad (2.4)$$

Ukuran kemiringan (*skewness*) atau ukuran ketidaksimetrisan suatu distribusi data dibagi dalam 3 jenis, yaitu:

- a. Simetris : menunjukkan letak nilai rata-rata, median, dan modus berimpit (berkisar di satu titik). Salah satu contoh distribusi yang simetris adalah distribusi normal, sehingga nilai *skewness*nya sama dengan nol, dengan mean = median = modus.
- b. Menceng ke kanan : *skewness* bernilai positif dimana ujung dari kecondongan menjulur ke arah positif (ekor kurva sebelah kanan lebih panjang), dimana modus < median < mean.
- c. Menceng ke kiri : *skewness* bernilai negatif : ujung dari kecondongan menjulur ke arah nilai negatif (ekor kurva sebelah kiri lebih panjang), dengan mean < median < modus.

2.3.2 Kurtosis

Menurut Ramachandran dan Tsokos (2009), kurtosis adalah ukuran untuk menggambarkan keruncingan (*peakness*) atau kerataan (*flatness*) suatu distribusi data dibandingkan distribusi normal. Terdapat 3 jenis kurtosis, yaitu: leptokurtis,

mesokurtis, dan platikurtis. Kurva yang lebih runcing dari distribusi normal dinamakan leptokurtis, sedangkan yang lebih datar disebut platikurtik. Sebagai pembandingan, distribusi normal sendiri memiliki bentuk mesokurtis dengan nilai koefisien kurtosis sama dengan 3.

Bohm dan Zech (2010) menjelaskan kurtosis dihitung dengan menggunakan koefisien Pearson (β_2) keempat terhadap mean. Selanjutnya dijabarkan sebagai berikut.

$$\tau = \beta_2 = \frac{E[(X-\mu)]^4}{\sigma^4} \quad (2.5)$$

Adapun 3 jenis kurtosis dapat diklasifikasikan sebagai berikut:

- a. Leptokurtik jika $\tau > 3$
- b. Platikurtik jika $\tau < 3$
- c. Mesokurtik (Normal) jika $\tau = 3$

2.4 Generalized Additive Model for Location & Shape (GAMLSS)

GAMLSS adalah sebuah kelas dalam model statistik yang dikembangkan oleh Rigby & Stasinopoulos yang menyediakan perluasan kemampuan dari GLM dan GAM. Pada model yang lebih sederhana menyediakan parameter lokasi yang hanya mendeskripsikan aspek yang terbatas dari distribusi dari variabel respon. Pendekatan melalui GAMLSS menyediakan parameter lain dari distribusi yang berrelasi dengan variabel eksplanatori, dimana parameter lain tersebut diinterpretasikan sebagai parameter skala (*scale*) dan bentuk (*shape*) dari distribusi, dari variabel respon (y) yang berupa fungsi linier maupun nonlinier, bersifat parametrik maupun fungsi non-parametrik aditif dari variabel eksplanatori dan efek acak.

Pada GAMLSS variabel respon berasal dari distribusi keluarga eksponensial dan tambahan distribusi-distribusi lain termasuk untuk distribusi diskrit dan kontinu dengan *highly skewed* dan kurtosis. Untuk jenis respon cacahan, metode

ini cocok untuk data yang mengalami overdispersi dengan menggunakan distribusi overdispersi untuk data diskrit.

2.4.1 Bentuk dan Asumsi GAMLSS

GAMLSS mengasumsikan variabel tak bebas y_i untuk $i = 1, 2, \dots, n$ dengan fungsi kepadatan peluang $f(y_i | \theta^i)$ dengan $\theta^i = \theta_{i1}, \theta_{i2}, \theta_{i3}, \dots, \theta_{ip}$. θ^i merupakan vektor dari 4 parameter distribusi yaitu μ, σ, ν, τ yang dapat disebut sebagai fungsi dari variabel eksplanatori. Parameter μ dan σ dikarakteristikan sebagai parameter lokasi (*location*) dan skala (*scale*), sedangkan dua parameter lainnya yaitu ν dan τ disebut sebagai parameter *skewness* (ν) dan kurtosis (τ) yang tergabung dalam parameter ukuran (*shape*).

Rigby dan Stasinopoulos (2007) mendefinisikan model dari GAMLSS sebagai berikut. Misalkan $\mathbf{y}^T = y_1, y_2, \dots, y_n$ dengan n adalah panjang vektor dari variabel respon, $k = 1, 2, 3, 4$, dan $g_k(\cdot)$ diketahui sebagai fungsi link monotonik yang menghubungkan antara parameter distribusi dengan variabel eksplanatori, maka

$$g_k(\theta_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \mathbf{Z}_{jk} \gamma_{jk} \quad (2.6)$$

Jika $\mathbf{Z}_{jk} = \mathbf{I}_n$, dengan \mathbf{I}_n adalah matriks identitas berukuran $n \times n$ dan $\gamma_{jk} = h_{jk} = h_{jk}(x_{jk})$ untuk semua kombinasi dari j dan k pada (2.6), maka didapat bentuk lain dari GAMLSS yang dapat dituliskan sebagai berikut:

$$g_k(\theta_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} h_{jk}(x_{jk}) \quad (2.7)$$

$$g_1(\boldsymbol{\mu}) = \boldsymbol{\eta}_1 = \mathbf{X}_1 \boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} h_{j1}(x_{j1})$$

$$g_2(\boldsymbol{\sigma}) = \boldsymbol{\eta}_2 = \mathbf{X}_2 \boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} h_{j2}(x_{j2})$$

$$g_3(\mathbf{v}) = \boldsymbol{\eta}_3 = \mathbf{X}_3 \boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} h_{j3}(x_{j3})$$

$$g_4(\boldsymbol{\tau}) = \boldsymbol{\eta}_4 = \mathbf{X}_4 \boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} h_{j4}(x_{j4})$$

Keterangan:

$\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{v}, \boldsymbol{\tau}, \boldsymbol{\eta}_k$ = vektor dengan panjang n ;

$\boldsymbol{\beta}_k^T$ = $\beta_{1k}, \beta_{2k}, \dots, \beta_{j'_k k}$ adalah sebuah vektor parameter

\mathbf{X}_k = matriks berukuran $n \times J'_k$

h_{jk} = fungsi *smooth* non-parametrik dari variabel eksplanatori \mathbf{x}_{jk} , $j = 1, 2, \dots, j_k$ dan $k = 1, 2, 3, 4$.

dimana \mathbf{x}_{jk} untuk $j = 1, 2, \dots, j_k$ juga vektor dengan panjang n . Fungsi h_{jk} adalah fungsi tak diketahui dari variabel eksplanatori X_k dan $h_{jk} = h_{jk}(x_{jk})$ adalah sebuah vektor yang mengevaluasi fungsi h_{jk} pada (x_{jk}) .

2.4.2 Algoritma Rigby & Stasinopoulos (RS)

Rigby (2005) mendefinisikan algoritma dasar dalam GAMLSS yaitu Algoritma Rigby & Stasinopoulos (RS) adalah sebagai berikut. Misalkan $\mathbf{u}_k = \frac{\partial l}{\partial \eta_k}$ merupakan fungsi nilai $Z_k = \eta_k + \mathbf{W}_{kk}^{-1} \eta_k$ dengan variabel bebas yang dapat disesuaikan dengan \mathbf{W}_{ks} matriks diagonal hasil dari iterasi bobot untuk $k = 1, 2, \dots, p$ dan $s = 1, 2, \dots, p$. Algoritma ini memiliki *outer cycle* yang dapat memaksimalkan *penalized likelihood* dengan keterkaitan β_k dan γ_{jk} untuk $j = 1, 2, \dots, J_k$ dalam model berturut-turut untuk $\boldsymbol{\theta}_k$ dengan $k = 1, 2, \dots, p$. Setiap kalkulasi nilai yang didapatkan nilai kuantitas yang akan selalu digunakan pada setiap iterasi. Algoritma RS bukan bentuk khusus dari algoritma Cole & Green (CG) karena dalam algoritma RS diagonal matriks berbobot \mathbf{W}_{kk} dievaluasi (di-

update) dalam pencocokan setiap parameter θ_k , sedangkan pada algoritma CG semua diagonal matriks berbobot W_{ks} untuk $s = 1, 2, \dots, p$.

Misalkan r adalah indeks iterasi dari *outer cycle*, k parameter indeks, i adalah indeks iterasi dari *inner cycle*, m indeks algoritma backfitting, dan j *random effect* (atau *nonparametric*). Misalkan $\gamma_{jk}^{r,i,m}$ merupakan nilai terbaru yang didapat dari γ_{jk} pada saat ke- r (indeks *outer cycle*), ke- i (indeks *inner cycle*) dan algoritma *backfitting* ke- m dan misalkan $\gamma_{jk}^{r,i}$ menyatakan nilai dari γ_{jk} pada saat nilai *backfitting* konvergen untuk saat ke- i dan ke- r dengan $j = 1, 2, \dots, j_k$ dan $k = 1, 2, \dots, p$.

Langkah-langkah pada Algoritma RS adalah sebagai berikut:

1. Memberikan nilai awal *fitted value* $\theta_k^{(1,1)}$ dan *random effect* $\gamma_{jk}^{1,1,1}$ untuk $j = 1, 2, \dots, j_k$ dan $k = 1, 2, \dots, p$. Lalu dievaluasi nilai prediktor aditif $\eta_k^{1,1} = g_k(\eta_k^{1,1})$ untuk $k = 1, 2, \dots, p$.
2. Memasukkan nilai r (indeks *outer cycle*) dengan $r = 1, 2, \dots$ hingga konvergen untuk $k = 1, 2, \dots, p$.
 - a. Memberikan nilai awal *inner cycle* $i = 1, 2, \dots$ hingga konvergen.
 - i. Evaluasi nilai terbaru $u_k^{(r,i)}$, $W_{kk}^{r,i}$ dan $Z_k^{r,i}$,
 - ii. Mulai pemberian nilai awal algoritma *backfitting* dengan $m = 1, 2, \dots$ hingga konvergen
 - iii. Meregresi nilai residual terbaru secara partial dari $\epsilon_{0k}^{r,i,m} = Z_k^{r,i} - \sum_{j=1}^{j_k} Z_{jk} \gamma_{jk}^{r,i,m}$ yang berupa matriks X_k , menggunakan iterasi berbobot $W_{kk}^{r,i}$ untuk mendapatkan parameter estimasi terbaru $\beta_k^{r,i,m+1}$
 - iv. Untuk $j = 1, 2, \dots, j_k$ pemulusan parsial residual $\epsilon_{0k}^{r,i,m} = Z_k^{r,i} - X_k \beta_k^{r,i,m+1} - \sum_{t=1, t \neq j}^{j_k} Z_{tk} \gamma_{tk}^{r,i,c}$ menggunakan *shrinking*

(pemulusan) matriks \mathbf{S}_{jk} diberikan oleh persamaan $\mathbf{S}_{jk} = \mathbf{Z}_{jk}^T (\mathbf{W}_{kk} \mathbf{Z}_{jk} + \mathbf{G}_{jk})^{-1} \mathbf{Z}_{jk}^T \mathbf{W}_{kk}$ untuk mendapatkan prediktor aditif terbaru $\mathbf{Z}_{jk} \boldsymbol{\gamma}_{jk}^{r,i,m}$

- v. *Backfitting* berakhir ketika didapat nilai yang konvergen dari $\boldsymbol{\beta}_k^{r,i}$ dan $\mathbf{Z}_{jk} \boldsymbol{\gamma}_{jk}^{r,i}$ dengan $\boldsymbol{\beta}_k^{r,i+1} = \boldsymbol{\beta}_k^{r,i}$ dan $\boldsymbol{\gamma}_{jk}^{r,i+1} = \boldsymbol{\gamma}_{jk}^{r,i}$ untuk $j = 1, 2, \dots, j_k$. Jika tidak, *update m* dan kembali mengulang *backfitting*.
- vi. Kalkulasi $\boldsymbol{\gamma}_{jk}^{r,i+1}$ dan $\boldsymbol{\theta}_k^{r,i+1}$ terkini.

- b. *Inner cycle* berakhir dengan didapat $\boldsymbol{\beta}_k^r$ yang konvergen dan prediktor aditif $\mathbf{Z}_{jk} \boldsymbol{\gamma}_{jk}^r$ dengan $\boldsymbol{\beta}_k^{r+1,i} = \boldsymbol{\beta}_k^{r,i}$, $\boldsymbol{\gamma}_{jk}^{r+1,i} = \boldsymbol{\gamma}_{jk}^{r,i}$ dan $\boldsymbol{\theta}_{jk}^{r+1,i} = \boldsymbol{\theta}_k^r$ untuk $j = 1, 2, \dots, j_k$.

3. *Update* nilai k

- 4. *Outer cycle* berakhir jika didapatkan (*penalized*) likelihood yang cukup kecil. Jika tidak, *update r* dan ulangi kembali ke *outer cycle*.

2.4.3 Metode Pemulusan Loess

Loess merupakan akronim dari *local regression*, menyediakan metode pengepasan regresi ke data. Proses pemulusan dikatakan lokal karena setiap nilai yang dimuluskan ditentukan oleh titik data yang berdekatan dalam suatu rentang. Misalkan dalam suatu rentang terdapat beberapa x dimana x merupakan variabel prediktor, yang permukaan regresinya akan diestimasi oleh sebuah fungsi dari spesifik kelas parametrik, yaitu polinomial berderajat 1 atau 2. (Cleveland, *et al*, 1997)

Fungsi $\text{lo}()$ yang berarti loess pada GAMLSS sama dengan implementasi loess pada GAM. $\text{lo}()$ adalah satu-satunya fungsi yang tersedia dalam GAMLSS yang menyediakan pemulusan lebih dari satu variabel prediktor.

Cleveland (1979) menyatakan ada beberapa tahapan dalam pemulusan loess, yaitu sebagai berikut

1. Menentukan rentang. Untuk setiap i , misalkan h_i adalah jarak antara x_i ke titik terdekat ke- r dari x_i , sedemikian hingga $h_i = |x_i - x_j|$ untuk setiap $j = 1, \dots, n$ dan didefinisikan pula fungsi bobot *tricube* yaitu:

$$W(x) = (1 - |x|^3)^3, \text{ untuk } |x| < 1 \\ = 0, \text{ untuk } |x| \geq 1.$$

Cleveland (1988) menyatakan bahwa pemulusan pada pengepasan loess bergantung pada nilai f dimana $f = r/n$, dengan r merupakan titik terdekat dari x_i ke x_j dan n adalah jumlah data. Semakin besar nilai f maka \hat{y}_i semakin mulus.

2. Menghitung estimasi $\hat{\beta}_m(x_i)$. Untuk setiap i , dengan $m = 0, \dots, d$. $\hat{\beta}_m$ adalah parameter pada sebuah regresi polinomial berderajat d dari y_k pada x_k , yang dipaskan oleh kuadrat terkecil terbobot dengan bobot $w_k(x_i)$ untuk (x_k, y_k) . Oleh karena itu, $\hat{\beta}_m(x_i)$ adalah nilai dari $\hat{\beta}_m$ dengan meminimalkan $\sum_{k=1}^n w_k(x_i)(y_k - \beta_0 - \beta_1 x_k - \dots - \beta_d x_k^d)^2$. Saat $d = 1$, maka polinomial akan berbentuk linier, sedangkan jika $d = 2$ maka polinomial berbentuk kuadratik. Titik yang dimuluskan pada x_i menggunakan *locally weighted regression* berderajat d adalah (x_i, \hat{y}_i) dengan \hat{y}_i adalah nilai pengepasan (*fitted value*) dari regresi diatas pada x_i , dengan demikian dapat ditulis bahwa $\hat{y}_i = \sum_{m=0}^d \hat{\beta}_m(x_i)x_i^m = \sum_{k=1}^n r_k(x_i)y_k$, dengan r_k tidak bergantung pada y_j , dan $j = 1, \dots, n$.

3. Mengambil B , fungsi bobot bisquare yang didefinisikan oleh

$$B(x) = (1 - x^2)^2, \text{ untuk } |x| < 1 \\ = 0, \text{ untuk } |x| \geq 1$$

Misalkan $e_i = y_i - \hat{y}_i$ dan s adalah median dari $|e_i|$, dengan e_i adalah residual atau sisa dari nilai pengepasan dari tahap 2. Maka, pendefinisian *robustness weight* dinyatakan oleh $\delta_k = B(e_k/6s)$.

4. Menghitung \hat{y}_i baru untuk setiap i oleh pengepasan polinomial berderajat d menggunakan bobot kuadrat terkecil seperti pada tahap 2 dengan bobot $w_k(x_i)$ yang diganti dengan *robustness weight* $\delta_k w_k(x_i)$ pada titik (x_k, y_k) .
5. Mengulangi langkah 3 dan 4 selama t kali, dengan t merupakan jumlah iterasi pada prosedur pencocokan *robust*.

2.4.4 Distribusi untuk Data Cacahan yang Dapat Dipakai pada GAMLSS

Distribusi yang dapat dianalisis pada GAMLSS meliputi distribusi keluarga eksponensial dan distribusi *non*-Keluarga Eksponensial. Untuk data berupa cacahan atau diskrit, ada beberapa distribusi yang berpotensi dapat dipakai, diantaranya seperti berikut:

a. Distribusi Poisson

Fungsi peluang dari sebaran Poisson dengan parameter μ adalah:

$$f(y | \lambda) = \frac{e^{-\lambda} \lambda^y}{y!} \text{ untuk } y = 0, 1, 2, \dots, \lambda > 0.$$

dengan μ adalah mean dan y adalah variabel respon yang berbentuk cacahan. Hubungan antara varian dan meannya adalah sama, yang dinamakan equidispersi, atau $E(y) = \mu = \sigma$. Faktanya, banyak ditemukan data yang tidak memenuhi asumsi ini (Hilbe, 2011).

b. Distribusi Binomial Negatif I

Dalam distribusi Binomial Negatif tipe I, terdapat 2 parameter distribusi yaitu μ dan σ . Adapun fungsi kepadatan peluang untuk distribusi ini adalah sebagai berikut:

$$f(y | \mu, \sigma) = \frac{\Gamma(y+1/\sigma)}{\Gamma(y+1)\Gamma(1/\sigma)} \left[\frac{(\mu\sigma)^y}{(\mu\sigma+1)} \right]^{y+\frac{1}{\sigma}} \text{ untuk } y = 0, 1, 2, \dots, \infty,$$

dengan mean $E(y) = \mu$ dan varian $\sigma = \mu(1 + \alpha)$, sehingga saat $\alpha > 0$, varian akan melebihi mean dan terjadi overdispersi. Menurut Cameron dan Trivedi (1986) dalam Hilbe (2011) nilai μ pada formula diatas diklasifikasikan

sebagai persamaan linier, sehingga Binomial Negatif 1 (NB1) dapat disebut juga sebagai Binomial Negatif Linier.

c. Distribusi Binomial Negatif II

Distribusi Binomial Negatif II (NBII) atau Binomial Negatif Kuadratik, juga mempunyai 2 parameter distribusi, yaitu μ dan σ . Stasinopoulos dan Rigby (2014) menyatakan fungsi kepadatan peluang distribusi Binomial Negatif II adalah:

$$f(y | \mu, \sigma) = \frac{\Gamma(y+(\mu/\sigma))\sigma^y}{\Gamma(\mu/\sigma)\Gamma(y+1)(1+\sigma)^{y+(\mu/\sigma)}} \text{ untuk } y = 0,1,2, \dots, \infty, \mu > 0 \text{ dan } \alpha > 0, \text{ dengan mean } E(y) = \mu \text{ dan varian } \sigma = \mu + \alpha\mu^2, \text{ maka akan terjadi overdispersi karena varian lebih besar dari meannya.}$$

d. Distribusi *Zero-Inflated* Poisson (ZIP)

Lambert (1992) menjelaskan bahwa *Zero-Inflated* Poisson (ZIP) adalah model regresi yang digunakan untuk data cacah dengan banyak peristiwa nol. Jika Y_i merupakan peubah acak independen berdistribusi *Zero-Inflated* Poisson, maka penelitian nol dapat didefinisikan sebagai berikut.

$$Y_i \sim \begin{cases} 0, & \text{dengan peluang } \omega \\ \text{Poisson}(\lambda_i), & \text{dengan peluang } (1 - \omega) \end{cases}$$

Mean dan varian masing-masing didefinisikan sebagai berikut.

$$E(Y_i) = (1 - \omega)\lambda_i = \mu_i \text{ dan } \text{var}(Y_i) = \mu_i + \left(\frac{\omega}{1-\omega}\right)\mu_i^2$$

Berdasarkan mean dan variannya, dapat dilihat bahwa distribusi dari Y_i menunjukkan adanya overdispersi saat $\omega > 0$, karena varian $>$ mean.

Dari keempat distribusi diatas, 3 diantaranya dapat digunakan untuk menganalisis data dengan kejadian overdispersi.

2.5 Akaike's Information Criterion (AIC)

AIC adalah salah satu kriteria yang dijadikan patokan memilih model yang baik dengan menghitung pertimbangan besarnya maksimum likelihood dan banyaknya variabel yang digunakan dalam model. Besarnya AIC dihitung melalui rumus berikut:

$$AIC = -2l(\hat{\theta}) + 2q$$

dengan $l(\hat{\theta})$ adalah nilai likelihood dari model yang dihadapi dan q adalah banyaknya parameter dalam model. (Tirta, 2009)

2.6 Data Cacahan

Dalam pengumpulan data, khususnya data kuantitatif, data cacahan adalah data yang diperoleh dengan cara mencacah, membilang, atau menghitung banyak objek. Datanya berupa bilangan cacah.

Hilbe (2011) menyatakan bahwa overdispersi pada model Poisson terjadi ketika varian dari respon lebih besar dari nilai mean, sehingga dapat dinyatakan dengan $var(Y) > E(Y)$. Jika terjadi overdispersi dan tetap menggunakan regresi Poisson, maka dugaan dari parameter koefisien regresinya konsisten namun tidak efisien. Hal ini berdampak pada nilai standar error yang menjadi *underestimate*, sehingga kesimpulan menjadi tidak valid.

Menurut Hardin dan Hilbe (2007), overdispersi pada regresi Poisson dapat dideteksi dengan melihat nilai dari *Pearson Chi-Square* dan *residual deviance* yang dibagi dengan derajat bebasnya. Apabila kedua nilai ini secara signifikan lebih besar dari satu, maka terjadi overdispersi pada data. Overdispersi bisa disebabkan oleh hal-hal berikut:

1. *Missing value* pada variabel bebas
2. Data mengandung pencilan
3. Kesalahan memilih *link function*.

BAB 3 METODE PENELITIAN

Pada bab ini akan dipaparkan beberapa uraian mengenai metode penelitian sebagai konsep dalam melakukan penelitian ini. Adapun metode penelitian ini meliputi data dan analisisnya dengan menggunakan paket *gamlss* dan *ggplot2*.

3.1 Data

Data yang digunakan adalah data sekunder yang diambil dari *website* resmi Dinas Kesehatan Republik Indonesia berupa laporan tentang jumlah kematian balita karena pneumonia dan profil kesehatan per provinsi di seluruh Indonesia (kecuali Provinsi Bengkulu dan Jawa Barat) selama tahun 2013. Data ini meliputi jumlah kematian balita karena pneumonia, jumlah kasus gizi buruk, persentase cakupan imunisasi dasar pada bayi, rumah tangga perilaku hidup bersih dan sehat, serta banyaknya puskesmas yang terdapat di masing – masing provinsi (lihat lampiran A).

Data tersebut dapat diklasifikasikan ke dalam 2 variabel, yaitu variabel respon dan variabel eksplanatori (penjelas), yaitu:

Tabel 3.1 Variabel Prediktor (x) yang Digunakan

| Variabel | Jenis Variabel | Jenis data |
|---|----------------|------------|
| Jumlah kematian balita (Y) | Respon | Cacahan |
| Kasus gizi buruk (x_1) | Prediktor | Cacahan |
| Cakupan imunisasi dasar pada bayi(x_2) | Prediktor | Persentase |
| Cakupan Rumah Tangga Bersih dan Sehat (x_3) | Prediktor | Persentase |
| Jumlah Puskesmas (x_4) | Prediktor | Cacahan |

3.2 Pengolahan Data

Dalam pengolahan data, instrumen yang digunakan dalam penelitian ini adalah program R i386 3.1.2, paket *gamlss* dan paket *ggplot2*. Paket ini didapat dari *website* GAMLSS di <http://www.gamlss.com/>. Stasinopololos *et.,al* (2008) menyatakan struktur fungsi dalam GAMLSS dengan menggunakan metode pemulusan loess adalah sebagai berikut:

```
gamlss(formula = y ~ lo(~x, span=..., degree=...), family=..., data
       =..., method=...)
```

Keterangan:

formula : berisi formula yang menyatakan variabel respon dan beberapa variabel prediktor dalam pemulusan loess. Variabel yang ditulis dipisahkan oleh tanda koma (,), misalnya x_1, x_2, \dots .

span : rentang yang diperlukan, merupakan parameter utama dalam pemulusan loess. *Span default* dalam R adalah 0.5.

degree : derajat polinomial dalam pemulusan loess. Derajat default dalam R adalah 1.

data : nama data yang akan dianalisis

family : distribusi yang berkaitan dengan data. Dalam penelitian ini ada 4 distribusi yang dijadikan pilihan, yaitu:

Distribusi Poisson atau $PO()$,

Distribusi Negatif Binomial I atau $NBI()$,

Distribusi Negatif Binomial II atau $NBII()$, dan

Distribusi *Zero Inflated Poisson* atau $ZIP()$.

method : algoritma yang akan digunakan. Terdapat 3 pilihan algoritma, yaitu

Algoritma Rigby & Stasinopoulos dengan formula $RS()$

Algoritma Cole & Green dengan formula $CG()$

Algoritma Campuran dengan formula $mixed()$

3.3 Analisis Data

Menganalisis data diawali dengan menyusun data sampel. Data tersebut adalah data jumlah kematian akibat pneumonia pada balita selama tahun 2013. Ilustrasi data tersebut dilampirkan pada lampiran A.

Adapun analisis data riil menggunakan program R dengan paket *gamlss* langkah-langkahnya sebagai berikut:

1. Menginputkan data pengamatan dalam program R.
2. Melihat sebaran data dengan menggunakan *scatterplot* untuk mengetahui distribusi awal yang cocok dengan data.
3. Menentukan variabel yang akan di-*smoothing* dengan melihat *scatterplotmatrix* hubungan antara variabel respon dan variabel eksplanatori.
4. Menganalisis data riil menggunakan GAMLSS dengan algoritma RS dan metode pemulusan loess.
5. Melihat diagnostik data.
6. Menentukan model terbaik. Ada beberapa langkah dalam menentukan model terbaik, yaitu:
 - a. Menentukan distribusi terbaik untuk data cacahan dengan membandingkan nilai AIC.
 - b. Menentukan derajat polinomial dan *span* yang cocok dengan data pneumonia berdasarkan pemulusan loess.
 - c. Menentukan nilai estimasi parameter dari model terbaik.
7. Melakukan uji hipotesis. Dalam hal ini yang akan diuji adalah apakah ada pengaruh dari variabel prediktor terhadap variabel respon, dengan:

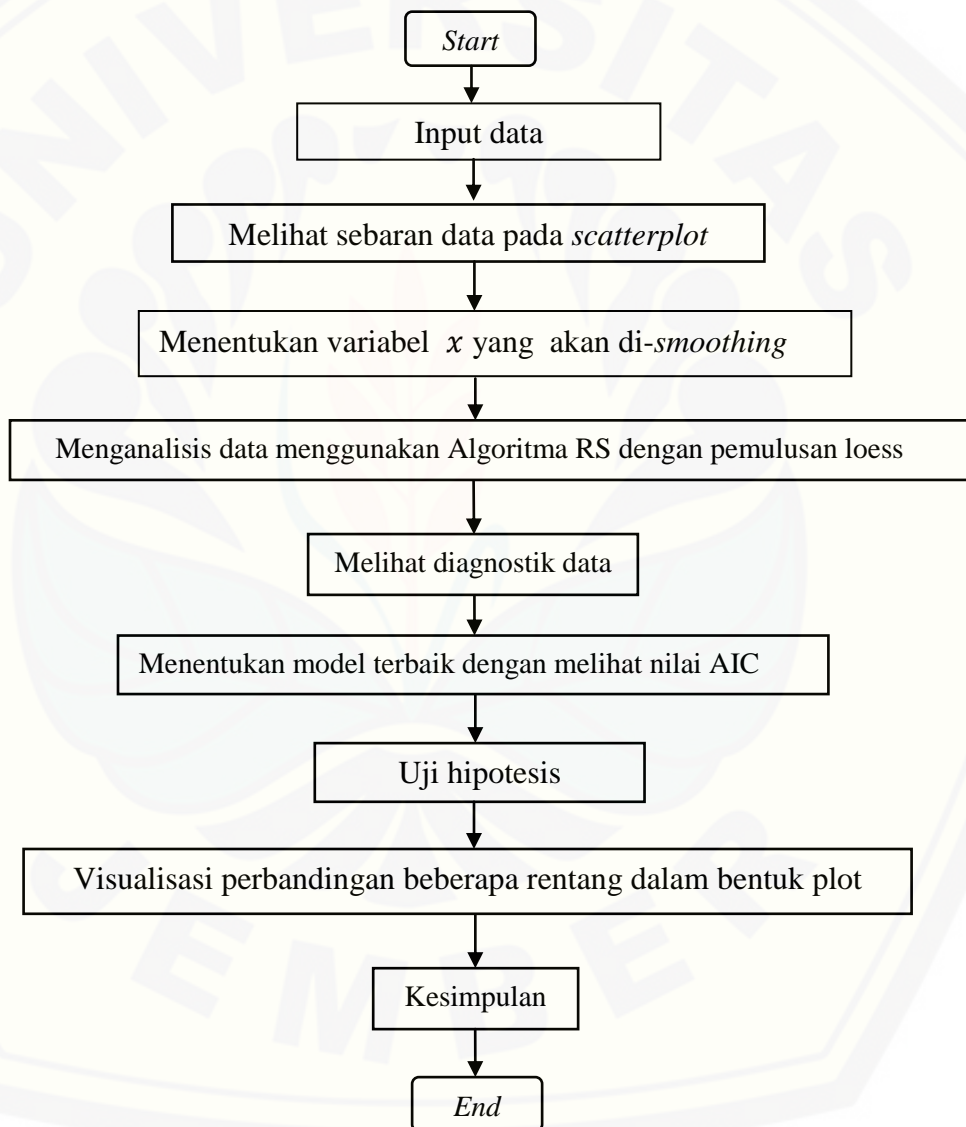
H_0 : tidak ada pengaruh banyaknya kasus gizi buruk, cakupan imunisasi yang diberikan, jumlah rumah tangga yang sehat dan bersih, serta jumlah puskesmas pada jumlah kematian bayi akibat pneumonia ($\forall_j \beta_j = 0$ untuk $j = 0,1,2,3$)

H_1 : Ada pengaruh (minimal satu) banyaknya kasus gizi buruk, cakupan imunisasi yang diberikan, jumlah rumah tangga yang sehat dan bersih, serta

jumlah puskesmas pada jumlah kematian bayi akibat pneumonia ($\exists_j \beta_j \neq 0$ untuk $j = 0,1,2,3$)

8. Memvisualisasikan perbandingan beberapa pilihan rentang dalam bentuk plot.
9. Menarik kesimpulan berdasarkan uji hipotesis dan analisis data.

Secara skematis, langkah-langkah yang akan dilakukan dalam penelitian ini dapat digambarkan sebagai berikut:



Gambar 3.1 Skema Metode Penelitian

BAB 4 HASIL DAN PEMBAHASAN

Bab ini akan menjelaskan tentang hasil dan analisis dari penerapan GAMLSS dengan menggunakan algoritma RS dan metode pemulusan loess pada data diskrit. Analisis data pada penelitian ini menggunakan program R 3.1.2, dengan mengaktifkan paket *gamlss* dan *ggplot2*.

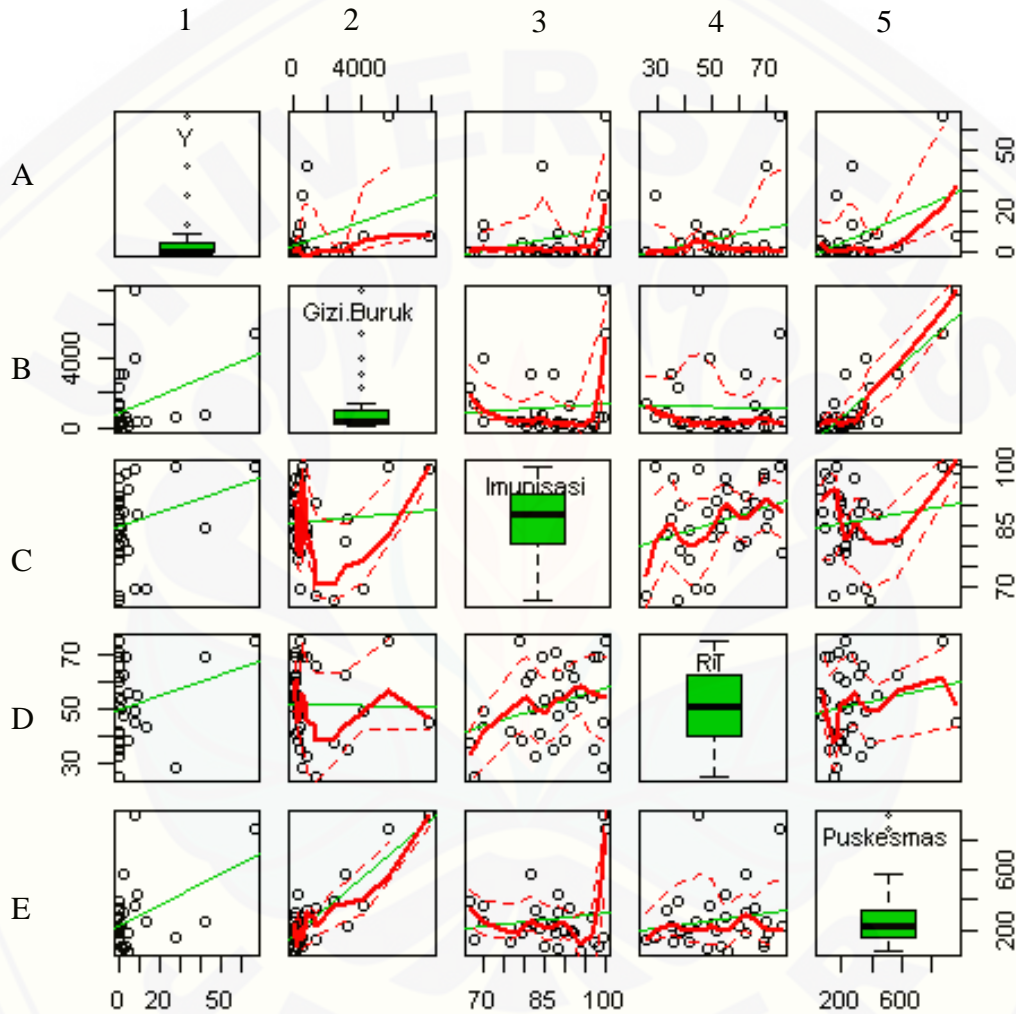
Tahap analisis data dibagi menjadi beberapa tahap, yaitu eksplorasi data pneumonia dengan variabel respon jumlah kematian balita, termasuk menentukan variabel eksplanatori yang akan di-*smoothing*. Selanjutnya data akan dianalisis dengan menggunakan GAMLSS untuk menentukan distribusi awal yang dapat menggambarkan data, kemudian menganalisis data dengan pemulusan loess (menentukan *degree* dan *span* terbaik) dan mendapatkan nilai estimasi parameter dari model terbaik.

4.1 Eksplorasi Data

Langkah pertama setelah mendapatkan data adalah melihat sebaran data. Hal ini dilakukan untuk mendapatkan karakteristik dari data dan dapat menentukan pilihan distribusi yang dapat diimplementasikan pada data tersebut. Data yang digunakan merupakan data diskrit yang menunjukkan jumlah kematian akibat pneumonia pada masing-masing provinsi di Indonesia (kecuali Provinsi Jawa Barat dan Bengkulu) serta beberapa variabel yang diduga berhubungan dengan hal tersebut. Rincian untuk data tersebut dapat dilihat pada Lampiran A.

Pada penelitian ini model yang digunakan adalah model semiparametrik yang merupakan gabungan dari model parametrik dan *non* parametrik, sehingga sebelum dianalisis lebih jauh perlu dilakukan eksplorasi data dengan menggunakan *scatterplot smoother*. Hasil *scatterplot smoother* akan memberikan gambaran hubungan

kelinieran antara variabel respon (Y) dengan masing-masing variabel eksplanatori, variabel mana yang tidak linier sehingga variabel yang tidak linier tersebut yang akan diestimasi dengan menggunakan metode pemulusan loess. Perintah *scatterplot matrix* dalam R (Lampiran D1) menghasilkan beberapa visualisasi gambar sebagai berikut.



Gambar 4.1 *Scatterplotmatrix* Hubungan Antara Variabel Respon dan Variabel Eksplanatori

Pada Gambar 4.1, ditunjukkan bahwa *scatterplot* antara variabel Y dengan variabel gizi buruk (A2), hubungan antara Y dengan variabel RT atau rumah tangga

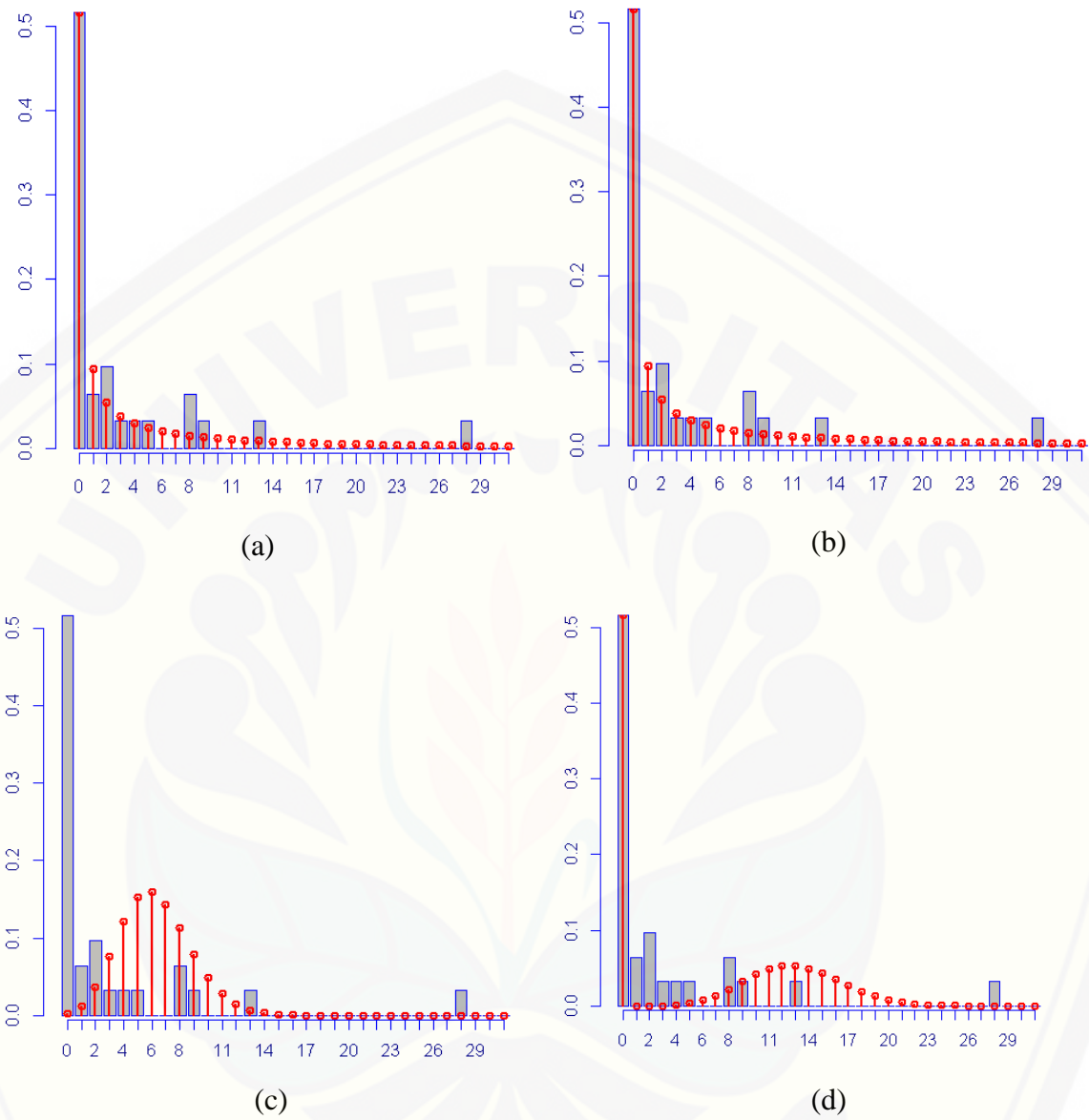
bersih dan sehat (A4) dan puskesmas (A5) menunjukkan hubungan yang lebih cocok dengan pendekatan linier. Sedangkan hubungan antara variabel Y dan variabel Imunisasi (A3) lebih cocok menggunakan pendekatan *non*-linier. Dalam hal ini variabel gizi buruk, RT, dan puskesmas adalah variabel bebas (X) dan Imunisasi adalah variabel acak (Z), sehingga dapat disimpulkan bahwa variabel gizi buruk, RT, dan puskesmas dimodelkan secara parametrik, sedangkan variabel Imunisasi secara *non*-parametrik, yang kemudian akan diestimasi menggunakan pemulusan loess.

4.2 Pengepasan Model pada Data Menggunakan GAMLSS dengan Algoritma RS dan Pemulusan Loess.

Untuk menentukan model terbaik dilakukan dengan cara membandingkan nilai AIC-nya. Adapun langkah pertama dalam tahap ini dimulai dengan menentukan distribusi yang terbaik untuk data tersebut dengan variabel respon jumlah kematian balita akibat pneumonia (Y).

4.2.1 Menentukan Distribusi Terbaik

Pada sub-bab sebelumnya telah dipaparkan tentang beberapa distribusi yang akan diaplikasikan pada data pneumonia ini, yaitu Distribusi Poisson, Distribusi Negatif Binomial I, Distribusi Negatif Binomial II, dan Distribusi *Zero-Inflated* Poisson. Dalam menentukan distribusi terbaik dilakukan secara intuitif yaitu dengan melihat histogram masing-masing pilihan distribusi (skrip R terdapat pada lampiran D2), yaitu sebagai berikut:



Gambar 4.2 Histogram Distribusi (a) Negatif Binomial I (b) Negatif Binomial II
(c) Poisson (d) *Zero-Inflated* Poisson

Berdasarkan histogram diatas maka dapat disimpulkan bahwa Negatif Binomial I dan Negatif Binomial II merupakan distribusi yang cocok dengan data tersebut. Namun, untuk memilih distribusi yang terbaik maka perlu dilakukan analisis

secara eksak dengan menentukan nilai AIC terkecil. Formula yang dipakai adalah sebagai berikut:

```
>m1<-gamlss(Y~Gizi.Buruk+Imunisasi+RT+Puskesmas,
  data=pneumonia1,family=NBI, method=RS())
>m2<-gamlss(Y~Gizi.Buruk+Imunisasi+RT+Puskesmas,
  data=pneumonia1,family=NBII, method=RS())
```

Adapun nilai AIC dari kedua model diatas terangkum dalam tabel berikut.

Tabel 4.1 Nilai AIC Distribusi Model

| Model | Distribusi | AIC |
|-----------|----------------------------|-----------------|
| m1 | Negatif Binomial I | 152,1758 |
| m2 | Negatif Binomial II | 149,4690 |

Berdasarkan tabel diatas dapat diambil kesimpulan bahwa model dengan distribusi Negatif Binomial II mempunyai nilai AIC yang paling kecil, yaitu 149,4690, sehingga untuk data pneumonia dengan variabel respon jumlah kematian balita lebih cocok dianalisis dengan menggunakan Distribusi Negatif Binomial II.

4.2.2 Menganalisis Data Menggunakan Pemulusan Loess

Pada subbab sebelum sudah ditentukan bahwa variabel respon mengikuti Distribusi Negatif Binomial II dan variabel imunisasi akan diestimasi dengan metode pemulusan *loess*. Terdapat 2 parameter pemulusan dalam loess, yaitu derajat polinomial (*degree*) dan rentang (*span*) yang harus ditentukan dengan melihat nilai AIC yang terkecil. Berikut ulasannya.

4.2.2.a Memilih derajat polinomial.

Ada dua pilihan derajat polinomial dalam pemulusan loess, yaitu derajat 1 (linier) dan polinomial berderajat 2 (kuadrat). Berikut analisisnya

```
>md1<-
gamlss (Y~Gizi.Buruk+lo (~Imunisasi, degree=1) +RT+Puskesmas,
data=pneumonia1, family=NBII, method=RS ())
>md2<-
gamlss (Y~Gizi.Buruk+lo (~Imunisasi, degree=2) +RT+Puskesmas,
data=pneumonia1, family=NBII, method=RS ())
```

dari kedua model diatas akan dibandingkan nilai AIC dari masing-masing model. Nilai AIC dari tiap model disajikan pada Tabel 4.2 (skrip lebih lengkap dari *output* program diatas dapat dilihat pada lampiran D3.

Tabel 4.2 Tabel *Degree* Terbaik

| Model | Derajat polinomial | AIC |
|------------|--------------------|-----------------|
| md1 | 1 | 150,5939 |
| md2 | 2 | 146,6413 |

Berdasarkan tabel diatas dapat ditarik kesimpulan bahwa polinomial berderajat 2 menjadi derajat polinomial yang cocok untuk data tersebut.

4.2.2.b Memilih span yang cocok

Rentang (*span*) sebagai parameter pemulusan diperlukan untuk mengetahui berapa banyak pemulusan yang diperlukan untuk memodelkan variabel tersebut dengan pemulusan loess, oleh karena itu diperlukan *span* yang tepat. *Span* yang terbaik ditentukan mulai rentang 0,2 hingga 0,8 (Cleveland,1979). Beberapa model akan diuji dengan beberapa pilihan span yaitu 0.4, 0.5, 0.6, 0.7, dan 0.8. Hasil skrip dirangkum dalam tabel berikut (skrip tercantum pada lampiran D4).

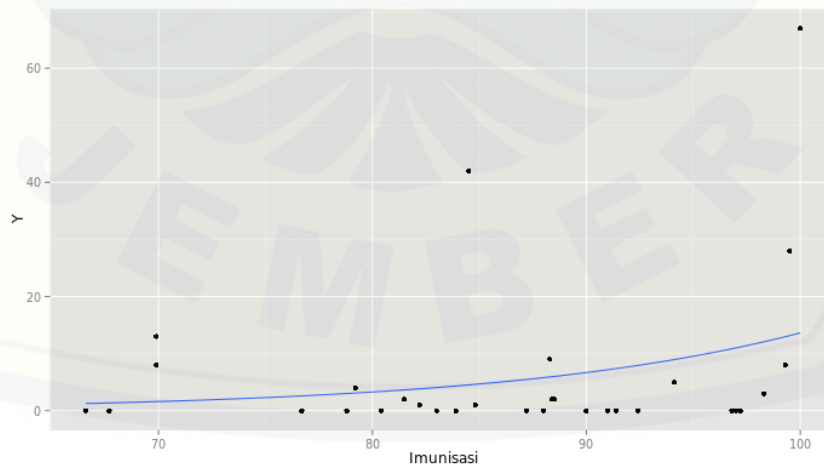
Tabel 4.3 Tabel Nilai AIC *Span* terbaik

| Model | Span | AIC |
|------------|------------|-----------------|
| ms4 | 0,4 | 291,9743 |
| ms5 | 0,5 | 145,2962 |
| ms6 | 0,6 | 146,0660 |
| ms7 | 0,7 | 146,0316 |
| ms8r | 0,8 | 146,8004 |

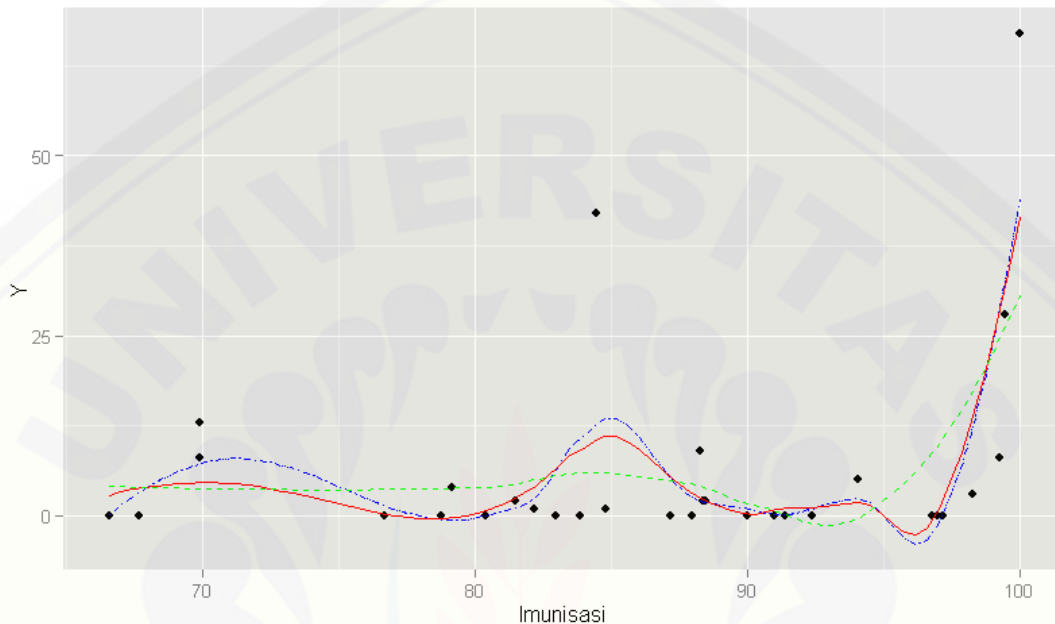
Berdasarkan tabel 4.3 diatas diperoleh bahwa rentang 0,5 merupakan nilai *span* terbaik untuk data tersebut dengan nilai AIC sebesar 145,2962.. Untuk selanjut akan dibandingkan nilai AIC model sebelum dan sesudah dilakukan *smoothing* serta melakukan visualisasi dari keempat pilihan *span* tersebut pada data.

4.3 Memvisualisasikan Perbandingan Rentang (*Span*) dalam Bentuk Plot

Menggambar plot data sebelum dan sesudah dilakukan *smoothing* dilakukan dengan mengujicoba beberapa pilihan rentang seperti pada sub-bab sebelumnya. Dalam hal ini perlukan pula paket *ggplot2*. Untuk data sebelum dilakukan *smoothing*, cara memvisualisasikannya adalah dengan *fitting* distribusi yang terbaik pada data, yaitu seperti pada Gambar 4.3.

Gambar 4.3 Plot Data Tanpa di-*smoothing*

Setelah itu, dilakukan visualisasikan data setelah dilakukan *smoothing* dengan menguji beberapa nilai rentang, seperti pada Gambar 4.4:



Gambar 4.4 Visualisasi *Smoothing* dengan Beberapa Pilihan *Span*

Nilai span terkecil yaitu 0.4 ditunjukkan oleh kurva *dotdash* (- · ·) berwarna biru. Nilai span terbaik yaitu 0.5 ditunjukkan oleh kurva *solid* (—) berwarna merah, sedangkan model dengan nilai span terbesar yaitu 0.8 ditunjukkan oleh kurva *dashed* (- - -) berwarna hijau (skrip R tertera pada lampiran D5).

4.4 Meng-*update* Distribusi Model Terbaik Setelah Dilakukan *Smoothing*.

Pada subbab sebelumnya telah ditentukan distribusi dasar sebelum melakukan *smoothing*. Melakukan *update* dilakukan untuk melihat apakah distribusi awal masih sesuai atau tidak setelah dilakukan pemulusan, sehingga perlu dilakukan pengepasan ulang dengan menggunakan beberapa pilihan distribusi seperti pada sub-bab 4.1.

Adapun nilai AIC dari masing-masing model yang baru tercantum pada Tabel 4.4

Tabel 4.4 Tabel *Update* Distribusi

| Model | Distribusi baru | AIC |
|----------------|------------------------------|-----------------|
| ms5PO | Poisson | 226,1339 |
| ms5NBI | Negatif Binomial I | 151,6802 |
| ms5NBII | Negatif Binomial II | 145,2962 |
| ms5ZIPr | <i>Zero-Inflated Poisson</i> | 172,8780 |

Berdasarkan tabel diatas dapat dilihat bahwa Distribusi Negatif Binomial II masih sesuai dengan data tersebut, meskipun sudah melewati tahap *smoothing* dengan nilai AIC 145,2962. Nilai ini masih lebih kecil daripada nilai AIC model sebelum dilakukan *smoothing* yaitu 149,4690. Oleh sebab itu model dengan pemulusan loess dianggap lebih baik daripada tidak dilakukan *smoothing* (skrip pada lampiran D6).

4.5 Menentukan Variabel yang Berpengaruh pada Data Pneumonia

Setelah mendapatkan model yang terbaik maka kemudian akan ditentukan variabel yang berpengaruh pada jumlah kematian balita (variabel Y).

```
>ms5a<-gamlss (Y~Gizi.Buruk+lo (~Imunisasi, span=.5,
degree=2), data=pneumonia1, family=NBII, method=RS ())
>ms5b<-gamlss (Y~lo (~Imunisasi, span=.5, degree=2),
data=pneumonia1, family=NBII, method=RS ())
>ms5c<-gamlss (Y~RT+lo (~Imunisasi, span=.5, degree=2),
data=pneumonia1, family=NBII, method=RS ())
>ms5d<-gamlss (Y~Puskesmas+lo (~Imunisasi, span=.5, degree=
2), data=pneumonia1, family=NBII, method=RS ())
```

Jika pada model ms5 merupakan model dengan meregresikan variabel Y dengan semua variabel eksplanatori, maka pada keempat model diatas merupakan model dengan hubungan antara variabel respon dengan masing-masing (satu per satu) variabel eksplanatori. Dengan membandingkan keempat model diatas dengan model ms5 yang terdapat pada subbab sebelumnya akan diperoleh model terbaik untuk data pneumonia tersebut. Nilai AIC masing-masing model disajikan pada tabel berikut (*output* lebih lengkap pada lampiran D7).

Tabel 4.5 Nilai AIC Model GAMLSS dengan Variasi Variabel

| Model | AIC |
|-------------|-----------------|
| ms5a | 141,9148 |
| ms5b | 140,9062 |
| ms5c | 142,0194 |
| ms5d | 143,0460 |
| ms5 | 145,2962 |

Berdasarkan tabel diatas model terbaik adalah model ms5b dengan AIC sebesar 140,9062, dengan variabel yang paling berpengaruh terhadap jumlah kematian balita akibat pneumonia adalah persentase imunisasi dasar yang diberikan.

4.6 Menentukan Model Terbaik dengan Menggunakan `stepGAIC()`

Formula `stepGAIC` menyediakan beberapa kombinasi variabel untuk masing-masing model yang akan diuji. Adapun langkah-langkahnya adalah sebagai berikut:

1. Meng-*input*-kan model yang akan diuji

```
ms5<-gamlss(Y~Gizi.Buruk+lo(~Imunisasi,span=0.5,
degree=2)+RT+Puskesmas,data=pneumonia1,family=NBII,
method=RS())
```

2. Memasukkan fungsi `gamlss.scope()`

Fungsi ini digunakan untuk mendefinisikan *range* yang akan diuji secara bertahap.

```
>go<-gamlss.scope(model.frame(pneumonia1),
  smoother="lo", arg="span=.5")
```

3. Mulai `stepGAIC()`

`stepGAIC` adalah fungsi untuk mengujicobakan setiap kemungkinan dari kombinasi dari tiap variabel pada data tersebut. Dalam fungsi ini terdapat pilihan argumen `additive=TRUE` yang memungkinkan pada proses ini dapat dilakukan pemulusan pada semua *trial*.

```
step1<-stepGAIC(ms5, go, additive=TRUE)
```

Berdasarkan proses diatas diperoleh model terbaik adalah model sebagai berikut: $Y \sim \text{loess}(\text{Imunisasi}, \text{span}=0,5, \text{degree}=2)$ dengan nilai AIC sebesar 140,8898 yaitu pada *trial* ke-13. Kemudian akan dilakukan `stepGAIC()` kembali untuk memastikan bahwa model diatas masih distribusi yang terbaik untuk data pneumonia. Adapun proses dan tahapnya sama, namun input yang diuji merupakan model pada subbab 4.5, yaitu sebagai berikut:

```
>ms5b<-
  gamlss(Y~lo(~Imunisasi, span=.5, degree=2), data=pneumonia1, family=NBII, method=RS())
>go<-gamlss.scope(model.frame(pneumonia1),
  smoother="lo", arg="span=.5")
>step2<-stepGAIC(ms5b, go, additive=TRUE)
```

dari step ke-2 ini didapat bahwa model `ms5b` masih merupakan model terbaik untuk data pneumonia dengan variabel respon jumlah kematian balita (*output* program dapat dilihat pada lampiran D9). Adapun ringkasan modelnya adalah sebagai berikut:

untuk parameter μ :

Mu link function: log

Mu Coefficients:

| Estimate | Std.Error | t value | Pr(> t) |
|----------|-----------|---------|-------------------------|
| -15,84 | 0,2634 | -60,13 | $2,041 \times 10^{-26}$ |

untuk parameter σ :

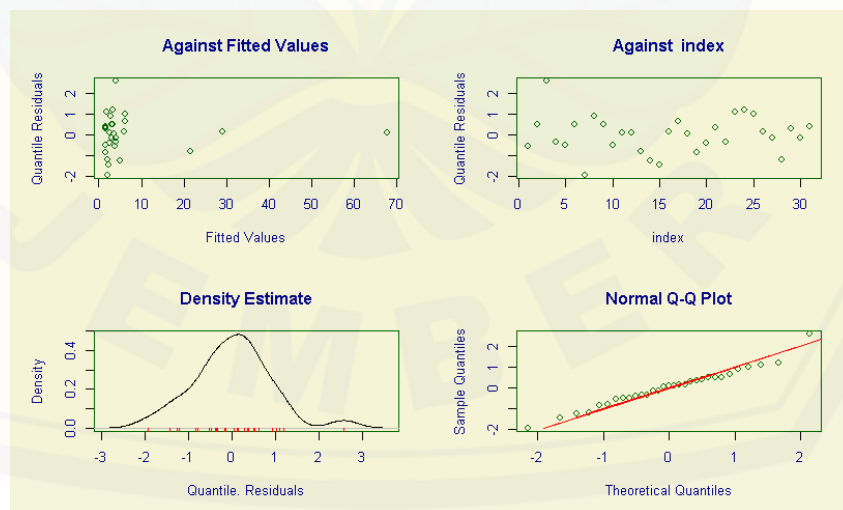
Sigma link function: log

Sigma Coefficients:

| Estimate | Std. Error | t value | Pr(> t) |
|----------|------------|---------|------------------------|
| 2,495 | 0,4151 | 6,010 | $4,279 \times 10^{-6}$ |

dengan nilai AIC model sebesar 140,9062. Plot residual untuk model diatas adalah sebagai berikut:

mean = 0,0858855
 varian = 0,6805388
 koefisien skewness = 0,5003565
 koefisien kurtosis = 4,095777



Gambar 4.5 Plot Residual Model ms5b

Suatu model data dikatakan mengikuti sebaran normal baku saat nilai *skewness*-nya 0 dan koefisien kurtosisnya bernilai 3 (lihat subbab 2.3). Jika dilihat pada ringkasan model diatas maka model ms5b dapat didekati dengan sebaran normal. Berdasarkan gambar 4.5, plot residual dari model ms5b terlihat bahwa sebaran normal Q-Q plotnya mendekati garis lurus. Hal tersebut juga menunjukkan bahwa model data tersebut dapat didekati dengan sebaran normal.

4.7 Menentukan Nilai Estimasi Parameter dari Model Terbaik

Berdasarkan pembahasan pada subbab sebelumnya dapat diketahui bahwa model terbaik untuk data pneumonia dengan menggunakan GAMLSS dengan metode pemulusan loess dan algoritma RS adalah ms5b, dengan rincian formula yaitu:

```
>ms5b<-
gamlss(Y~lo(~Imunisasi, span=.5, degree=2), data=pneumonia1,
family=NBII, method=RS())
```

Summary model diatas untuk nilai estimasi dan *p-value* dari masing-masing komponen diberikan pada Tabel 4.6:

Tabel 4.6 Nilai Estimasi, nilai t dan *p-value*

| | Estimasi μ | <i>t</i> value | Pr(> t) |
|--------------------------------------|----------------|----------------|-------------------------|
| loess(Imunisasi, span=0.5, degree=2) | -15,84 | 60,13 | $2,041 \times 10^{-26}$ |

Berdasarkan tabel diatas, dengan nilai *p-value* < 0,05 menunjukkan bahwa variabel imunisasi berpengaruh secara signifikan terhadap jumlah kematian balita akibat penyakit pneumonia.

Estimasi dari hasil *summary* diatas didapatkan hasil sebagai berikut:

$$\log(\hat{\mu}_Y) = \hat{\gamma}_1 \text{ loess(Imunisasi, span=0,5, degree=2)}$$

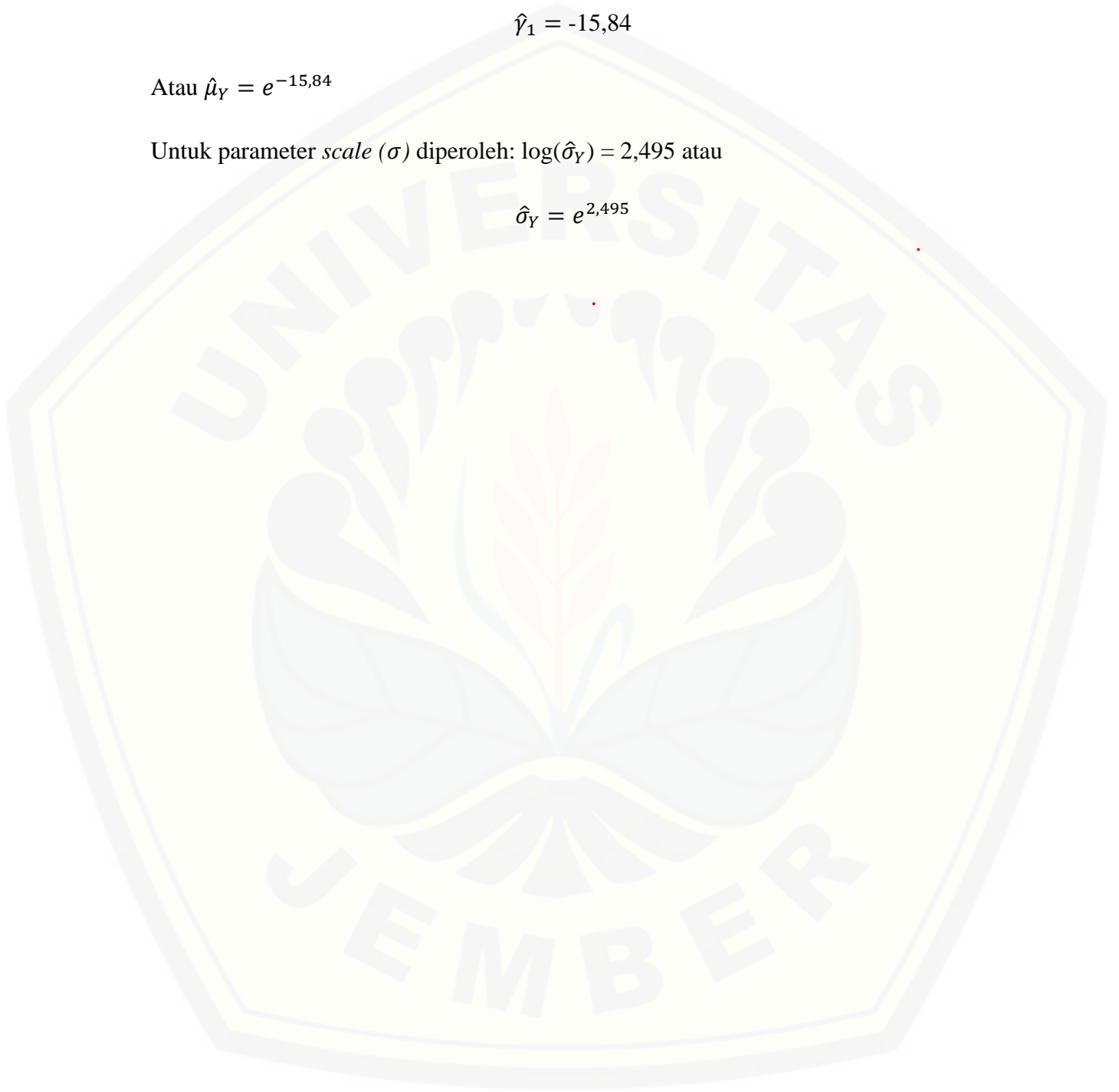
dengan

$$\hat{\gamma}_1 = -15,84$$

$$\text{Atau } \hat{\mu}_Y = e^{-15,84}$$

Untuk parameter *scale* (σ) diperoleh: $\log(\hat{\sigma}_Y) = 2,495$ atau

$$\hat{\sigma}_Y = e^{2,495}$$



BAB 5 PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan pada bab 4, penerapan GAMLSS dengan metode pemulusan loess dan menggunakan algoritma Rigby-Stasinopoulos (RS) pada data pneumonia yang berbentuk cacahan, dapat diambil beberapa kesimpulan, yaitu sebagai berikut ini:

1. Distribusi yang paling sesuai untuk data pneumonia dengan variabel respon jumlah kematian balita tersebut adalah Negatif Binomial Tipe II dengan parameter (μ, σ)
2. Model terbaik untuk variabel jumlah kematian balita pada data pneumonia adalah:
 - a. $\log(\hat{\mu}_Y) = \hat{\gamma}_1$ loess(Imunisasi, span=0,5, degree=2) dengan $\hat{\gamma}_1 = -15,84$ sehingga parameter lokasi $\hat{\mu}_Y = e^{-15,84}$
 - b. Untuk parameter *scale* (σ) diperoleh: $\log(\hat{\sigma}_Y) = 2,495$ atau

$$\hat{\sigma}_Y = e^{2,495}$$

5.2 Saran

Untuk penelitian selanjutnya diharapkan peneliti menggunakan metode pemulusan yang lain, seperti *penalized spline (p-spline)*, dan *varying coefficient*.

DAFTAR PUSTAKA

- Bohm, G & Zech, G.2010.*Introduction to Statistics and Data Analysis for Physicists*. Jerman: Helholtz Association
- Cleveland, W.S. 1979. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of The American Statistical Association*, **74**:829-836
- Cleveland W.S, Devlin,S.J. 1988. An Approach to Regression Analysis by Local Fitting. *Journal of The American Statistical Association*, **83**:596-610
- Gilchrist, R., Rigby, R.,Sedgwick,J., Stasinopoulos,M., Voudouris,V. 2011. Forecasting Film Revenues using GAMLSS. *Proceeding of The 26th International Workshop on Statistical Modelling*, **26**:263-268
- Hardin J.W, Hilbe J.M. 2007. *Generalized Linear Models and Extensions*. Texas: A Stata Press Publication.
- Hastie, T.J,Tibshirani,R.J.1990. *Generalized Additive Models*. Chapman and Hall: London
- Hilbe, J.M. 2011. *Negative Binomial Regression Second Edition*. New York: Cambridge University Press
- Lambert, D. 1992. Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing. *Technometrics*, **34**(1): 1-14.
- Nelder, J.A.& Wedderburn, R. W. M. 1972. Generalized Linear Models, *J. R Statist. Soc. A.*,**135**: 370-384
- Ramachandran,K.M & Tsokos,C.P.2009. *Mathematical Statistic with Applications*, United States of America: Academic Press.
- Rigby,R.A & Stasinopoulos D.M.2005 Generalized Additive Models for Location, and Shape. *Applied Statistic*,**54**.507-554

Stasinopoulos, D. M. & Rigby, R.A. 2007. Generalized Additive Models for Location, Scale and Shape (GAMLSS) in R. *J.R.Statisc.* [23]:7

Stasinopoulos,D.M., Rigby,R.B.,Akantziliotou,C.2008. Instruction on How to Use The GAMLSS Package in R Second Edition. *STORM Research Centre*, London Metropolitan University, London. <http://www.gamlss.com/>

Stasinopoulos, D. M. & Rigby, R.A. 2014. Package 'gamlss.dist'. <http://cran.r-project.org/web/packages/gamlss.dist/gamlss.dist.pdf>

Tirta, IM. 2009. *Analisis Regresi dengan R (ANRER)*. Jember: UPT Penerbitan Universitas Jember

Wandi, W. F. 2012. *Aplikasi GAMLSS dengan Pemulusan Cubic-Spline dan Algoritma Rigby & Stasinopoulos pada Data Agroklimatologi*. Skripsi Jurusan Matematika FMIPA Universitas Jember, Tidak Diterbitkan.

**Lampiran A. Ilustrasi Data Jumlah Kematian Balita Karena Pneumonia di
Indonesia selama tahun 2013**

| No | Provinsi | Y | x_1 | x_2 | x_3 | x_4 |
|----|---------------------------|----|-------|-------|-------|-------|
| 1 | Aceh | 0 | 807 | 83 | 33,07 | 334 |
| 2 | Sumatera Utara | 2 | 3088 | 81,5 | 62,71 | 570 |
| 3 | Sumatera Barat | 42 | 747 | 84,5 | 69,55 | 262 |
| 4 | Riau | 0 | 119 | 83,9 | 41,49 | 207 |
| 5 | Jambi | 0 | 184 | 92,4 | 62,07 | 176 |
| 6 | Sumatera Selatan | 2 | 201 | 88,5 | 62,61 | 319 |
| 7 | Lampung | 0 | 175 | 96,8 | 54,77 | 280 |
| 8 | Kepulauan Bangka Belitung | 5 | 114 | 94,1 | 55,73 | 60 |
| 9 | Kep Riau | 2 | 223 | 88,4 | 47,89 | 70 |
| 10 | DKI Jakarta | 0 | 1254 | 91 | 65,66 | 340 |
| 11 | Jawa Tengah | 67 | 5540 | 100 | 75,14 | 873 |
| 12 | DI Yogyakarta | 0 | 397 | 97,2 | 35,51 | 121 |
| 13 | Jawa Timur | 8 | 7965 | 99,3 | 45,02 | 960 |
| 14 | Banten | 0 | 3102 | 87,2 | 35,34 | 230 |
| 15 | Bali | 0 | 96 | 97 | 69,39 | 120 |
| 16 | NTB | 28 | 591 | 99,5 | 28,94 | 158 |
| 17 | NTT | 8 | 4038 | 69,9 | 49,02 | 362 |
| 18 | Kalimantan Barat | 1 | 310 | 82,2 | 50,02 | 237 |
| 19 | Kalimantan Tengah | 0 | 63 | 91,4 | 51,28 | 194 |
| 20 | Kalimantan Selatan | 0 | 132 | 80,4 | 60,09 | 228 |
| 21 | Kalimantan Timur | 0 | 392 | 78,8 | 75,26 | 222 |
| 22 | Sulawesi Utara | 0 | 75 | 88 | 70,7 | 183 |
| 23 | Sulawesi Tengah | 4 | 310 | 79,2 | 38,23 | 183 |
| 24 | Sulawesi Selatan | 9 | 317 | 88,3 | 54,87 | 440 |
| 25 | Sulawesi Tenggara | 13 | 340 | 69,9 | 43,8 | 264 |
| 26 | Gorontalo | 3 | 587 | 98,3 | 69,37 | 91 |
| 27 | Sulawesi Barat | 1 | 246 | 84,8 | 53,63 | 92 |
| 28 | Maluku | 0 | 223 | 90 | 38,39 | 190 |
| 29 | Maluku Utara | 0 | 328 | 76,7 | 41,75 | 125 |
| 30 | Papua Barat | 0 | 1379 | 67,7 | 25,5 | 143 |
| 31 | Papua | 0 | 2379 | 66,6 | 37,44 | 391 |

Keterangan:

Y : Jumlah kematian balita karena pneumonia (jiwa)

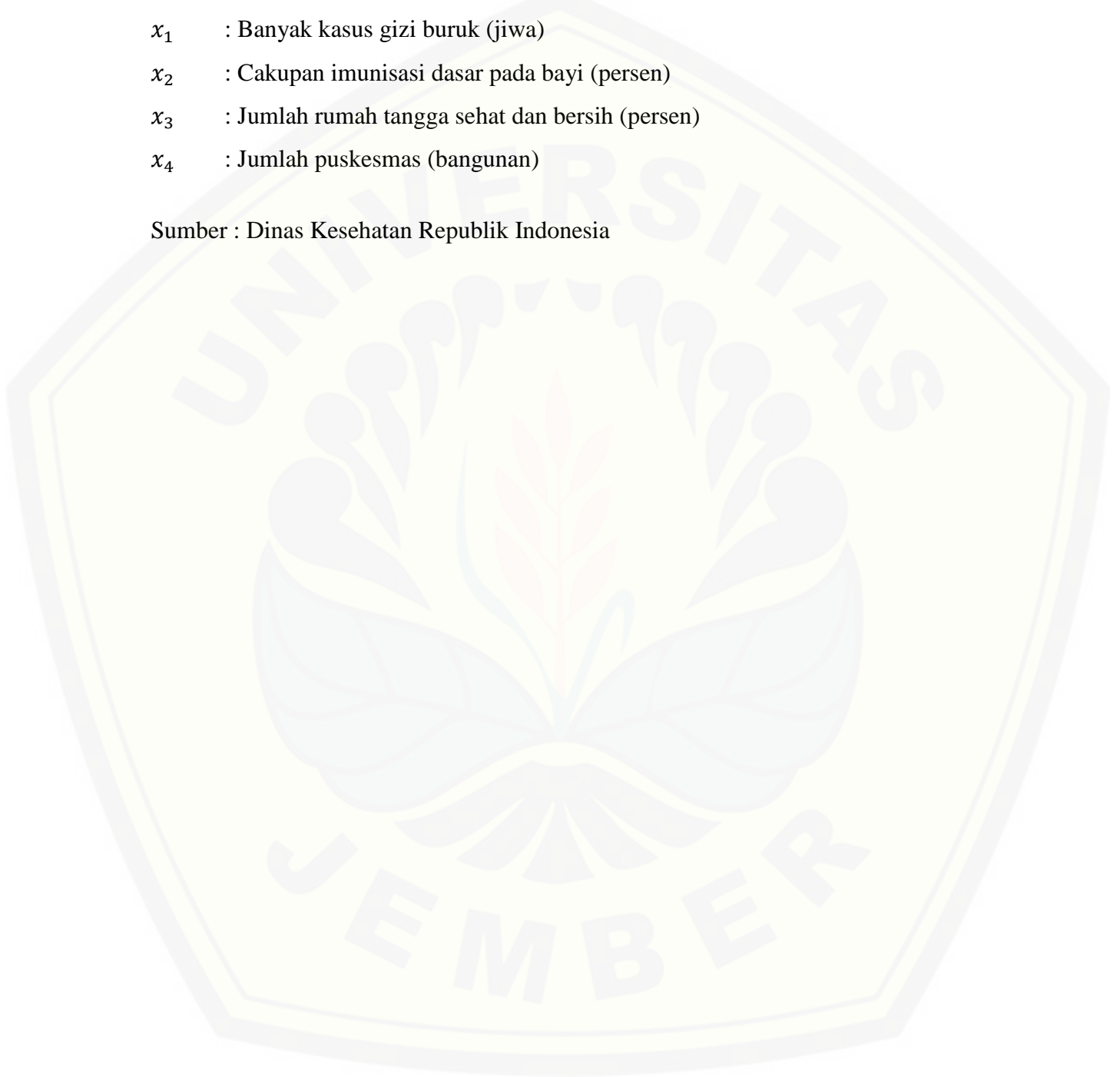
x_1 : Banyak kasus gizi buruk (jiwa)

x_2 : Cakupan imunisasi dasar pada bayi (persen)

x_3 : Jumlah rumah tangga sehat dan bersih (persen)

x_4 : Jumlah puskesmas (bangunan)

Sumber : Dinas Kesehatan Republik Indonesia



Lampiran B. Tabel Daftar Distribusi Diskrit yang Dapat Diimplementasikan dalam Paket GAMLSS

Tabel berikut merupakan daftar distribusi diskrit yang dapat diimplementasikan dalam paket gamlss beserta fungsi *link default*-nya. *R Name* adalah penamaan distribusi tersebut dalam program R

| Distribusi | R name | μ | σ | ν |
|---|----------|-------|----------|----------|
| Beta Binomial | BB() | logit | log | - |
| Binomial | BI() | logit | - | - |
| Delaporte | DEL() | log | log | logit |
| Negative Binomial Type I | NBI() | log | log | - |
| Negative Binomial Type II | NBII() | log | log | - |
| Poisson | PO() | log | - | - |
| Poisson Inverse Gaussian | PIG() | log | log | - |
| Sichel | SIG() | log | log | Identity |
| Sichel (μ the mean) | SICHEL() | log | log | Identity |
| Zero Inflated Poisson | ZIP() | log | logit | - |
| Zero Inflated Poisson (μ the mean) | ZIP2() | log | logit | - |

Lampiran C. Pencocokan *Generalized Additive Model for Location, Scale and Shape* (GAMLSS) Secara Umum

Formula berikut digunakan untuk memcocokkan model objek menggunakan GAMLSS. Berdasarkan formula didapatkan hasil estimasi dari parameter mean atau lokasi (μ), parameter varian atau skala (σ) dan parameter ukuran yaitu *skewness*(ν) dan kurtosis (τ). Adapun formulanya adalah sebagai berikut.

```
gamlss(formula=formula(data),sigma.formula=~1,nu.formula=~1,tau.formula=~1,family=...,method=...,data=sys.parent(),weight=NULL,contrast=NULL,sigma.start=NULL,mu.start=NULL,nu.start=NULL,tau.start=NULL,mu.fix=FALSE,sigma.fix=FALSE,nu.fix=FALSE,tau.fix=FALSE,control=gamlss.control(...),i.control=glimcontrol(...))
```

Keterangan:

formula : rumus dari model. Contoh penulisannya $y \sim x$

mu.formula : rumus model untuk parameter lokasi (μ)

sigma.formula : rumus model untuk parameter skala (σ)

nu.formula : rumus model untuk parameter *skewness*(ν)

tau.formula : rumus model untuk parameter kurtosis (τ)

family : distribusi yang masuk ke dalam *gamlss*. Dalam penelitian ini menggunakan 4 pilihan distribusi yaitu Distribusi Poisson, Negatif Binomial I, Negatif Binomial II, dan Zero Inflated Poisson.

method: algoritma yang digunakan. Terdapat 3 pilihan algoritma, yaitu

Algoritma Rigby & Stasinopoulos atau ditulis RS(),

Algoritma Cole & Green atau ditulis RCG(),

Algoritma campuran atau *mixed* yang ditulis mixed()

data : data yang akan dianalisis

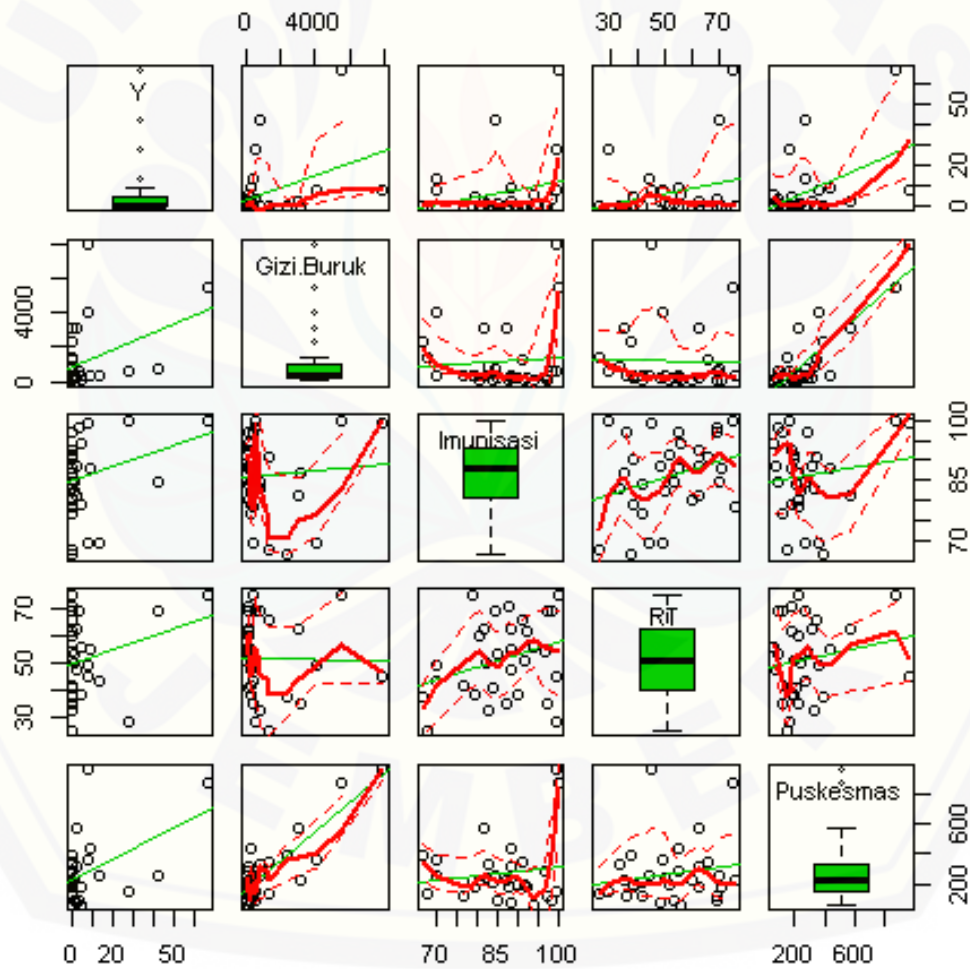
- `weights` : suatu vector dari bobot (*weights*) yang digunakan dalam proses pencocokan
- `contrast` : daftar kontras yang digunakan untuk memunculkan variabel yang hilang dalam formula.
- `start.from` : pencocokan model GAMLSS dengan nilai pencocokan yang akan digunakan untuk memulai nilai awal pada pencocokan model.
- `mu.start` : berupa vektor atau skalar untuk inisialisasi nilai untuk parameter `mu`.
Contoh penulisan `mu.start=2`
- `sigma.start` : berupa vektor atau skalar untuk inisialisasi nilai untuk parameter `sigma`. Contoh penulisan `mu.start=5`
- `nu.start` : berupa vektor atau skalar untuk inisialisasi nilai untuk parameter `nu`.
Contoh penulisan `nu.start=8`
- `tau.start` : berupa vektor atau skalar untuk inisialisasi nilai untuk parameter `mu`.
Contoh penulisan `tau.start=2`
- `mu.fix` : digunakan ketika parameter `mu` ditetapkan memiliki nilai tetap dalam pencocokan model.
- `sigma.fix` : digunakan ketika parameter `sigma` ditetapkan memiliki nilai tetap dalam pencocokan model.
- `nu.fix` : digunakan ketika parameter `nu` ditetapkan memiliki nilai tetap dalam pencocokan model.
- `tau.fix` : digunakan ketika parameter `tau` ditetapkan memiliki nilai tetap dalam pencocokan model.
- `Control` : menjaga parameter untuk tetap dalam outer *cycle*
- `i.control` : menjaga parameter untuk tetap dalam inner *cycle*

Lampiran D. PROGRAM

Berikut program analisis data pneumonia dengan menggunakan formula GAMLSS dengan metode pemulusan loess dan algoritma Rigby-Stasinopoulos (RS)

D1. Eksplorasi Data

```
>data(pneumonia1)
>scatterplotMatrix(~Y+Gizi.Buruk+Imunisasi+RT+Puskesmas,r
eg.line=lm,diagonal='boxplot',data=pneumonia1)
```



D2. Analisis Distribusi Terbaik

```

#melihat Distribusi Terbaik dengan histogram
>histDist(Y, family=NBII, xlim=c(0, 31), data=pneumonia1)
>histDist(Y, family=ZIP, xlim=c(0, 31), data=pneumonia1)
>histDist(Y, family=NBI, xlim=c(0, 31), data=pneumonia1)
>histDist(Y, family=PO, xlim=c(0, 31), data=pneumonia1)

>m1<-gamlss(Y~Gizi.Buruk+Imunisasi+RT+Puskesmas,
  data=pneumonia1, family=PO, method=RS())
GAMLSS-RS iteration 1: Global Deviance = 429.2455
GAMLSS-RS iteration 2: Global Deviance = 429.2455

>m2<-gamlss(Y~Gizi.Buruk+Imunisasi+RT+Puskesmas,
  data=pneumonia1, family=NBI)
GAMLSS-RS iteration 1: Global Deviance = 140.1758
GAMLSS-RS iteration 2: Global Deviance = 140.1758

>m3<-gamlss(Y~Gizi.Buruk+Imunisasi+RT+Puskesmas,
  data=pneumonia1, family=NBII)
GAMLSS-RS iteration 1: Global Deviance = 137.7802
GAMLSS-RS iteration 2: Global Deviance = 137.5988
GAMLSS-RS iteration 3: Global Deviance = 137.5197
GAMLSS-RS iteration 4: Global Deviance = 137.4888
GAMLSS-RS iteration 5: Global Deviance = 137.4767
GAMLSS-RS iteration 6: Global Deviance = 137.4716
GAMLSS-RS iteration 7: Global Deviance = 137.4698
GAMLSS-RS iteration 8: Global Deviance = 137.469

>m4<-gamlss(Y~Gizi.Buruk+Imunisasi+RT+Puskesmas,
  data=pneumonia1, family=ZIP)
GAMLSS-RS iteration 1: Global Deviance = 280.3798
GAMLSS-RS iteration 2: Global Deviance = 279.9752
GAMLSS-RS iteration 3: Global Deviance = 279.9743

> GAIC(m1, m2, m3, m4)
      df      AIC
m3    6 149.4690
m2    6 152.1758
m4    6 291.9743
m1    5 439.2455

```

```

> summary(m3)
*****
Family:  c("NBII", "Negative Binomial type II")

Call:
gamlss(formula = Y ~ Gizi.Buruk + Imunisasi + RT +
Puskesmas,family = NBII, data = pneumonial)
Fitting method: RS()

-----

Mu link function:  log
Mu Coefficients:
              Estimate
(Intercept) -8.398e-01
Gizi.Buruk   7.759e-05
Imunisasi    6.294e-03
RT           2.352e-02
Puskesmas    1.925e-03
              Std. Error  t value
(Intercept)  2.3810546  -0.3527
Gizi.Buruk   0.0003049   0.2545
Imunisasi    0.0272018   0.2314
RT           0.0183670   1.2804
Puskesmas    0.0026862   0.7167
              Pr(>|t|)
(Intercept)  0.7272
Gizi.Buruk   0.8011
Imunisasi    0.8188
RT           0.2117
Puskesmas    0.4800

-----

Sigma link function:  log
Sigma Coefficients:
              Estimate  Std. Error
(Intercept)    3.102    0.3642
              t value  Pr(>|t|)
(Intercept)    8.518  1.666e-09

-----

No. of observations in the fit:  31
Degrees of Freedom for the fit:  6
  Residual Deg. of Freedom:  25
                        at cycle:  8

Global Deviance:      137.469

```

```

AIC:      149.469
SBC:      158.0729

```

D3. Analisis Derajat Polinomial (*degree*) Terbaik

```

>mdl<-gamlss(Y~Gizi.Buruk+lo(~Imunisasi,degree=1)+RT+
Puskesmas,data=pneumonia1,family=NBII,method=RS())
GAMLSS-RS iteration 1: Global Deviance = 135.9802
GAMLSS-RS iteration 2: Global Deviance = 135.6858
GAMLSS-RS iteration 3: Global Deviance = 135.5585
GAMLSS-RS iteration 4: Global Deviance = 135.5095
GAMLSS-RS iteration 5: Global Deviance = 135.4843
GAMLSS-RS iteration 6: Global Deviance = 135.4726
GAMLSS-RS iteration 7: Global Deviance = 135.4676
GAMLSS-RS iteration 8: Global Deviance = 135.4647
GAMLSS-RS iteration 9: Global Deviance = 135.4633
GAMLSS-RS iteration 10: Global Deviance = 135.4625

```

```

>md2<-gamlss(Y~Gizi.Buruk+lo(~Imunisasi,degree=2)+RT+
Puskesmas,data=pneumonia1,family=NBII,method=RS())
GAMLSS-RS iteration 1: Global Deviance = 130.6676
GAMLSS-RS iteration 2: Global Deviance = 129.7235
GAMLSS-RS iteration 3: Global Deviance = 129.1438
GAMLSS-RS iteration 4: Global Deviance = 128.8709
GAMLSS-RS iteration 5: Global Deviance = 128.7425
GAMLSS-RS iteration 6: Global Deviance = 128.7316
GAMLSS-RS iteration 7: Global Deviance = 128.5215
GAMLSS-RS iteration 8: Global Deviance = 128.6176
GAMLSS-RS iteration 9: Global Deviance = 128.6149
GAMLSS-RS iteration 10: Global Deviance = 128.6178
GAMLSS-RS iteration 11: Global Deviance = 128.6181

```

```

> GAIC(mdl,md2)
      df      AIC
md2 9.011605 146.6413
mdl 7.565701 150.5939

```

D4. Analisis *Span* Terbaik

```

>ms4<-gamlss(Y~Gizi.Buruk+lo(~Imunisasi,span=.4,degree=2)
+RT+Puskesmas,data=pneumonia1,family=NBII,method=RS())
GAMLSS-RS iteration 1: Global Deviance = 119.2654
GAMLSS-RS iteration 2: Global Deviance = 116.381
GAMLSS-RS iteration 3: Global Deviance = 114.8633
GAMLSS-RS iteration 4: Global Deviance = 114.3614

```



```
GAMLSS-RS iteration 5: Global Deviance = 114.2882
GAMLSS-RS iteration 6: Global Deviance = 114.2691
GAMLSS-RS iteration 7: Global Deviance = 114.269

>ms5<-gamlss(Y~Gizi.Buruk+lo(~Imunisasi,span=.5,degree=2)
+RT+Puskesmas,data=pneumonial,family=NBII,method=RS())
GAMLSS-RS iteration 1: Global Deviance = 125.3619
GAMLSS-RS iteration 2: Global Deviance = 123.7091
GAMLSS-RS iteration 3: Global Deviance = 122.7911
GAMLSS-RS iteration 4: Global Deviance = 122.1587
GAMLSS-RS iteration 5: Global Deviance = 122.0377
GAMLSS-RS iteration 6: Global Deviance = 121.9547
GAMLSS-RS iteration 7: Global Deviance = 121.9269
GAMLSS-RS iteration 8: Global Deviance = 121.9152
GAMLSS-RS iteration 9: Global Deviance = 121.9199
GAMLSS-RS iteration 10: Global Deviance = 121.92

>ms6<-gamlss(Y~Gizi.Buruk+lo(~Imunisasi,span=.6,degree=2)
+RT+Puskesmas,data=pneumonial,family=NBII,method=RS())
GAMLSS-RS iteration 1: Global Deviance = 127.9717
GAMLSS-RS iteration 2: Global Deviance = 126.4738
GAMLSS-RS iteration 3: Global Deviance = 125.8494
GAMLSS-RS iteration 4: Global Deviance = 125.7292
GAMLSS-RS iteration 5: Global Deviance = 125.7361
GAMLSS-RS iteration 6: Global Deviance = 125.7687
GAMLSS-RS iteration 7: Global Deviance = 125.7755
GAMLSS-RS iteration 8: Global Deviance = 125.7766
GAMLSS-RS iteration 9: Global Deviance = 125.7657
GAMLSS-RS iteration 10: Global Deviance = 125.7778
GAMLSS-RS iteration 11: Global Deviance = 125.7756
GAMLSS-RS iteration 12: Global Deviance = 125.7764

>ms7<-gamlss(Y~Gizi.Buruk+lo(~Imunisasi,span=.7,degree=2)
+RT+Puskesmas,data=pneumonial,family=NBII,method=RS())
GAMLSS-RS iteration 1: Global Deviance = 129.5327
GAMLSS-RS iteration 2: Global Deviance = 128.2778
GAMLSS-RS iteration 3: Global Deviance = 127.6289
GAMLSS-RS iteration 4: Global Deviance = 127.4088
GAMLSS-RS iteration 5: Global Deviance = 127.3195
GAMLSS-RS iteration 6: Global Deviance = 127.223
GAMLSS-RS iteration 7: Global Deviance = 127.229
GAMLSS-RS iteration 8: Global Deviance = 127.2295
```

```

>ms8<-gamlss(Y~Gizi.Buruk+lo(~Imunisasi,span=.8,degree=2)
+RT+Puskesmas,data=pneumonia1,family=NBII,method=RS())
GAMLSS-RS iteration 1: Global Deviance = 131.0431
GAMLSS-RS iteration 2: Global Deviance = 130.1949
GAMLSS-RS iteration 3: Global Deviance = 129.7175
GAMLSS-RS iteration 4: Global Deviance = 129.439
GAMLSS-RS iteration 5: Global Deviance = 129.3148
GAMLSS-RS iteration 6: Global Deviance = 129.2484
GAMLSS-RS iteration 7: Global Deviance = 129.2056
GAMLSS-RS iteration 8: Global Deviance = 129.1908
GAMLSS-RS iteration 9: Global Deviance = 129.1728
GAMLSS-RS iteration 10: Global Deviance = 129.1699
GAMLSS-RS iteration 11: Global Deviance = 129.167
GAMLSS-RS iteration 12: Global Deviance = 129.1686
GAMLSS-RS iteration 13: Global Deviance = 129.1657
GAMLSS-RS iteration 14: Global Deviance = 129.1689
GAMLSS-RS iteration 15: Global Deviance = 129.164
GAMLSS-RS iteration 16: Global Deviance = 129.1683
GAMLSS-RS iteration 17: Global Deviance = 129.1634
GAMLSS-RS iteration 18: Global Deviance = 129.1681
GAMLSS-RS iteration 19: Global Deviance = 129.1633
GAMLSS-RS iteration 20: Global Deviance = 129.168
> ms8r<-refit(ms8)
GAMLSS-RS iteration 21: Global Deviance = 129.1647
GAMLSS-RS iteration 22: Global Deviance = 129.1625
GAMLSS-RS iteration 23: Global Deviance = 129.1623
> GAIC(m4,ms5,ms6,ms7,ms8r)
      df      AIC
ms5  11.688057 145.2962
ms7   9.401005 146.0316
ms6  10.144793 146.0660
ms8r  8.819086 146.8004
m4    6.000000 291.9743

```

D5. Visualisasi Model Sesudah dilakukan *Smoothing* dengan Beberapa Pilihan *Span*.

```

names(pneumonia1)
p<-ggplot(data=pneumonia1,aes_string(x="Imunisasi",
y="Y"))+geom_point(col="black")
print(p)
p<-p+geom_smooth(method="loess",span=0.4,degree=2,lty=4,
se=FALSE,col="blue")
print(p)

```

```
p<-p+geom_smooth(method="loess",span=0.5,degree=2,lty=1,
se=FALSE,col="red")
print(p)
p<-p+geom_smooth(method="loess",span=0.8,degree=2,lty=2,
se=FALSE,col="green")
print(p)
```

D6. Meng-update Distribusi Terbaik Setelah di-Smoothing

```
> ms5PO<-update(ms5,family=PO)
GAMLSS-RS iteration 1: Global Deviance = 204.4133
GAMLSS-RS iteration 2: Global Deviance = 204.4129

> ms5NBI<-update(ms5,family=NBI)
GAMLSS-RS iteration 1: Global Deviance = 127.8337
GAMLSS-RS iteration 2: Global Deviance = 127.8185
GAMLSS-RS iteration 3: Global Deviance = 127.8403
GAMLSS-RS iteration 4: Global Deviance = 127.8394

> ms5NBII<-update(ms5,family=NBII)
GAMLSS-RS iteration 1: Global Deviance = 125.3619
GAMLSS-RS iteration 2: Global Deviance = 123.7091
GAMLSS-RS iteration 3: Global Deviance = 122.7911
GAMLSS-RS iteration 4: Global Deviance = 122.1587
GAMLSS-RS iteration 5: Global Deviance = 122.0377
GAMLSS-RS iteration 6: Global Deviance = 121.9547
GAMLSS-RS iteration 7: Global Deviance = 121.9269
GAMLSS-RS iteration 8: Global Deviance = 121.9152
GAMLSS-RS iteration 9: Global Deviance = 121.9199
GAMLSS-RS iteration 10: Global Deviance = 121.92

> ms5ZIP<-update(ms5,family=ZIP)
GAMLSS-RS iteration 1: Global Deviance = 151.9767
GAMLSS-RS iteration 2: Global Deviance = 150.4232
GAMLSS-RS iteration 3: Global Deviance = 150.3892
GAMLSS-RS iteration 4: Global Deviance = 150.4096
GAMLSS-RS iteration 5: Global Deviance = 150.4336
GAMLSS-RS iteration 6: Global Deviance = 150.4433
GAMLSS-RS iteration 7: Global Deviance = 150.4439

> GAIC(ms5PO,ms5NBI,ms5NBII,ms5ZIP)
      df      AIC
ms5NBII 11.68806 145.2962
ms5NBI  11.92036 151.6802
```

```
ms5ZIP 11.21705 172.8780
ms5PO 10.86047 226.1339
```

D7. Analisis Variabel yang Berpengaruh pada Data Pneumonia

```
>ms5a<-
gamlss(Y~Gizi.Buruk+lo(~Imunisasi,span=.5,degree=2),
data=pneumonial,family=NBII,method=RS())
GAMLSS-RS iteration 1: Global Deviance = 126.8866
GAMLSS-RS iteration 2: Global Deviance = 124.657
GAMLSS-RS iteration 3: Global Deviance = 123.6929
GAMLSS-RS iteration 4: Global Deviance = 123.1949
GAMLSS-RS iteration 5: Global Deviance = 123.2661
GAMLSS-RS iteration 6: Global Deviance = 123.0242
GAMLSS-RS iteration 7: Global Deviance = 122.8972
GAMLSS-RS iteration 8: Global Deviance = 122.8951
GAMLSS-RS iteration 9: Global Deviance = 122.8757
GAMLSS-RS iteration 10: Global Deviance = 122.8752

>ms5b<-gamlss(Y~lo(~Imunisasi,span=.5,degree=2),
data=pneumonial,family=NBII,method=RS())
GAMLSS-RS iteration 1: Global Deviance = 126.5354
GAMLSS-RS iteration 2: Global Deviance = 125.1097
GAMLSS-RS iteration 3: Global Deviance = 124.6049
GAMLSS-RS iteration 4: Global Deviance = 124.2201
GAMLSS-RS iteration 5: Global Deviance = 124.0366
GAMLSS-RS iteration 6: Global Deviance = 123.977
GAMLSS-RS iteration 7: Global Deviance = 123.966
GAMLSS-RS iteration 8: Global Deviance = 123.9799
GAMLSS-RS iteration 9: Global Deviance = 123.9909
GAMLSS-RS iteration 10: Global Deviance = 123.9955
GAMLSS-RS iteration 11: Global Deviance = 124.0102
GAMLSS-RS iteration 12: Global Deviance = 123.9879
GAMLSS-RS iteration 13: Global Deviance = 123.9944
GAMLSS-RS iteration 14: Global Deviance = 124.0071
GAMLSS-RS iteration 15: Global Deviance = 123.9883
GAMLSS-RS iteration 16: Global Deviance = 123.9911
GAMLSS-RS iteration 17: Global Deviance = 123.9927
GAMLSS-RS iteration 18: Global Deviance = 124.0044
GAMLSS-RS iteration 19: Global Deviance = 124.0139
GAMLSS-RS iteration 20: Global Deviance = 124.0511

>ms5c<-gamlss(Y~RT+lo(~Imunisasi,span=.5,degree=2),
data=pneumonial,family=NBII,method=RS())
```



```
GAMLSS-RS iteration 1: Global Deviance = 125.3901
GAMLSS-RS iteration 2: Global Deviance = 124.0459
GAMLSS-RS iteration 3: Global Deviance = 123.3323
GAMLSS-RS iteration 4: Global Deviance = 123.0601
GAMLSS-RS iteration 5: Global Deviance = 122.9764
GAMLSS-RS iteration 6: Global Deviance = 122.9615
GAMLSS-RS iteration 7: Global Deviance = 122.9618
```

```
>ms5d<-
```

```
gamlss(Y~Puskesmas+lo(~Imunisasi,span=.5,degree=2),
data=pneumonial,family=NBII,method=RS())
```

```
GAMLSS-RS iteration 1: Global Deviance = 126.8731
GAMLSS-RS iteration 2: Global Deviance = 125.8594
GAMLSS-RS iteration 3: Global Deviance = 124.8888
GAMLSS-RS iteration 4: Global Deviance = 124.4677
GAMLSS-RS iteration 5: Global Deviance = 124.2479
GAMLSS-RS iteration 6: Global Deviance = 124.2957
GAMLSS-RS iteration 7: Global Deviance = 124.1297
GAMLSS-RS iteration 8: Global Deviance = 124.0905
GAMLSS-RS iteration 9: Global Deviance = 124.2296
GAMLSS-RS iteration 10: Global Deviance = 124.172
GAMLSS-RS iteration 11: Global Deviance = 124.1932
GAMLSS-RS iteration 12: Global Deviance = 124.3477
GAMLSS-RS iteration 13: Global Deviance = 124.1448
GAMLSS-RS iteration 14: Global Deviance = 124.1243
GAMLSS-RS iteration 15: Global Deviance = 124.3507
GAMLSS-RS iteration 16: Global Deviance = 124.1381
GAMLSS-RS iteration 17: Global Deviance = 124.1153
GAMLSS-RS iteration 18: Global Deviance = 124.2772
GAMLSS-RS iteration 19: Global Deviance = 124.2282
GAMLSS-RS iteration 20: Global Deviance = 124.2214
```

```
Warning in RS() : Algorithm RS has not yet converged
```

```
>ms5dr<-refit(ms5d)
```

```
GAMLSS-RS iteration 21: Global Deviance = 124.1692
GAMLSS-RS iteration 22: Global Deviance = 124.1336
GAMLSS-RS iteration 23: Global Deviance = 124.169
GAMLSS-RS iteration 24: Global Deviance = 124.106
GAMLSS-RS iteration 25: Global Deviance = 124.0774
GAMLSS-RS iteration 26: Global Deviance = 124.0842
GAMLSS-RS iteration 27: Global Deviance = 124.1091
GAMLSS-RS iteration 28: Global Deviance = 124.1099
```



```
> GAIC(ms5,ms5a,ms5b,ms5c,ms5dr)
              df      AIC
ms5b      8.427592 140.9062
ms5a      9.519751 141.9148
ms5c      9.528795 142.0194
ms5dr     9.376194 142.8623
ms5      11.688057 145.2962
```

D8. Ringkasan Variabel Paling Berpengaruh

```
>summary(ms5b)
*****
*****
Family:  c("NBII", "Negative Binomial type II")
Call:
gamlss(formula = Y ~ lo(~Imunisasi, span = 0.5, degree =
2),family = NBII, data = pneumonial, method = RS())
Fitting method: RS()
-----
--Mu link function:  log
Mu Coefficients:
  Estimate Std. Error   t value   Pr(>|t|)
-1.584e+01  2.634e-01  -6.013e+01  2.041e-26
-----
--Sigma link function:  log
Sigma Coefficients:
  Estimate Std. Error   t value   Pr(>|t|)
 2.495e+00  4.151e-01   6.010e+00  4.279e-06
-----
--No. of observations in the fit:  31
Degrees of Freedom for the fit:  8.427592
      Residual Deg. of Freedom:  22.57241
                        at cycle:  20

Global Deviance:      124.0511
                   AIC:      140.9062
                   SBC:      152.9913
*****
```

D9. Analisis Model Terbaik dengan Menggunakan stepGAIC()

```
>ms5<-gamlss(Y~Gizi.Buruk+lo(~Imunisasi,span=.5,
degree=2)+RT+Puskesmas,data=pneumonial,family=NBII,
method=RS())
GAMLSS-RS iteration 1: Global Deviance = 125.3619
```

```
GAMLSS-RS iteration 2: Global Deviance = 123.7091
GAMLSS-RS iteration 3: Global Deviance = 122.7911
GAMLSS-RS iteration 4: Global Deviance = 122.1587
GAMLSS-RS iteration 5: Global Deviance = 122.0377
GAMLSS-RS iteration 6: Global Deviance = 121.9547
GAMLSS-RS iteration 7: Global Deviance = 121.9269
GAMLSS-RS iteration 8: Global Deviance = 121.9152
GAMLSS-RS iteration 9: Global Deviance = 121.9199
GAMLSS-RS iteration 10: Global Deviance = 121.92

go<-
gamlss.scope(model.frame(pneumonia1), smoother="lo", arg="
span=.5")
step1<-stepGAIC(ms5, go, additive=TRUE)
Distribution parameter: mu
Start: Y ~ Gizi.Buruk + lo(~Imunisasi, span = 0.5, degree
= 2) + RT + Puskesmas; AIC= 145.2962
Trial1:Y~Gizi.Buruk+lo(~Imunisasi, span=.5, degree=2)+RT+Pu
skemas, data=pneumonia1, family=NBII, method=RS();

Trial2:Y~Gizi.Buruk+lo(~Imunisasi, span=.5, degree=2), data=
pneumonia1, family=NBII, method=RS(); AIC=141.9148

Trial3:Y~lo(~Imunisasi, span=.5, degree=2)+RT, data=pneumoni
a1, family=NBII, method=RS(); AIC=142.0194

Trial4:Y~lo(~Imunisasi, span=.5, degree=2)+Puskesmas, data=p
neumonia1, family=NBII, method=RS (); AIC=142.8623

Trial5:Y~lo(~Imunisasi, span=.5, degree=2)+lo(~RT, span=.5, d
egree=2), data=pneumonia1, family=NBII, method=RS
(); AIC=149.3718

Trial6:Y~Gizi.Buruk+lo(~Imunisasi, span=.5, degree=2)+RT, da
ta=pneumonia1, family=NBII, method=RS (); AIC=143.0417

Trial7:Y~Gizi.Buruk+lo(~Imunisasi, span=.5, degree=2)+Puske
smas, data=pneumonia1, family=NBII, method=RS
(); AIC=143.7510

Trial8:Y~lo(~Imunisasi, span=.5, degree=2)+RT+Puskesmas, dat
a=pneumonia1, family=NBII, method=RS (); AIC=143.5569
```

```
Trial9:Y~Gizi.Buruk+lo(~Imunisasi,span=.5,degree=2)+RT+Pu
skesmas,data=pneumonia1,family=NBII,method=RS();
AIC=145.2962
```

```
Trial10:Y~lo(~Imunisasi,span=.5,degree=2)+lo(~RT,span=.5,
degree=2)+Puskesmas,data=pneumonia1,family=NBII,method=RS
(); AIC=150.2955
```

```
Trial11:Y~lo(~Imunisasi,span=.5,degree=2)+Gizi.Buruk+RT+P
uskesmas,data=pneumonia1,family=NBII,method=RS();
AIC=145.2962
```

```
Trial12:Y~lo(~Imunisasi,span=.5,degree=2)+Gizi.Buruk+lo(~
RT,span=.5,degree=2)+Puskesmas,data=pneumonia1,family=NBII
I,method=RS(); AIC=146.9434
```

```
Trial13:Y~lo(~Imunisasi,span=.5,degree=2),data=pneumonia1
,family=NBII,method=RS();AIC=140.8898
```

```
Trial14:Y~lo(~RT,span=.5,degree=2),data=pneumonia1,family
=NBII,method=RS();AIC=155.8262
```

```
Trial15:Y~lo(~Puskesmas,span=.5,degree=2),data=pneumonia1
,family=NBII,method=RS(); AIC=144.0185
```

```
#Step ke-2
>ms5b<-
gamlss(Y~lo(~Imunisasi,span=.5,degree=2),data=pneumonia1,
family=NBII,method=RS())
>go<-gamlss.scope(model.frame(pneumonia1), smoother="lo",
arg="span=.5")
>step2<-stepGAIC(ms5b,go,additive=TRUE)
Distribution parameter: mu
Start: AIC= 140.91
Y ~ lo(~Imunisasi, span = 0.5, degree = 2)

Df      AIC
<none>      140.91
- lo(~Imunisasi, span = 0.5, degree = 2) 6.4276 148.02
```

D10. Plotting dan Ringkasan Residual Model ms5b

```
> plot(ms5b)
```

```

*****
Summary of the Randomised Quantile Residuals
      mean      = 0.0858855
      variance   = 0.6805388
      coef. of skewness = 0.5003565
      coef. of kurtosis  = 4.095777
Filliben correlation coefficient = 0.9737896
.....

>summary(ms5b)
*****
Family:  c("NBII", "Negative Binomial type II")

Call:  gamlss(formula = Y ~ lo(~Imunisasi, span = 0.5,
      degree = 2),
      family = NBII, data = pneumonia1, method = RS())

Fitting method: RS()
-----
Mu link function:  log
Mu Coefficients:
      Estimate  Std. Error    t value    Pr(>|t|)
-1.584e+01   2.634e-01   -6.013e+01  2.041e-26
-----
Sigma link function:  log
Sigma Coefficients:
      Estimate  Std. Error    t value    Pr(>|t|)
 2.495e+00   4.151e-01    6.010e+00  4.279e-06
-----
No. of observations in the fit:  31
Degrees of Freedom for the fit:  8.427592
      Residual Deg. of Freedom:  22.57241
                        at cycle:  20

Global Deviance:  124.0511
      AIC:  140.9062
      SBC:  152.9913
*****

```