ESTIMASI MODEL REGRESI NONPARAMETRIK *SPLINE* POISSON PADA TINGKAT KEMATIAN BAYI DI SULAWESI SELATAN

SKRIPSI

NOVILIA JAO

H051171302

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

NOVEMBER 2020

ESTIMASI MODEL REGRESI NONPARAMETRIK *SPLINE* POISSON PADA TINGKAT KEMATIAN BAYI DI SULAWESI SELATAN

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin Makassar

NOVILIA JAO

H051171302

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

NOVEMBER 2020

LEMBAR PERNYATAAN KEONTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sesungguh-sungguhnya bahwa skripsi yang saya buat dengan judul :

ESTIMASI MODEL REGRESI NONPARAMETRIK *SPLINE* POISSON PADA TINGKAT KEMATIAN BAYI DI SULAWESI SELATAN

adalah benar hasil kerja saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 5 November 2020

METERAI TEMPEL
7B6B8AHF73798397
6000
ENAM RIBU RUPIAH

NOVILIA JAO

NIM : H051171302

ESTIMASI MODEL REGRESI NONPARAMETRIK *SPLINE* POISSON PADA TINGKAT KEMATIAN BAYI DI SULAWESI SELATAN

disetujui oleh:

Pembimbing Utama	Pembimbing Pertama
Dr. Anna Islamiyati, S.Si., M.Si.	Dr. Nurtiti Sunusi, S.Si., M.Si.
NIP. 197708082005012002	NIP. 197201171997032002

Pada tanggal: 5 November 2020

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Novilia Jao

NIM : H051171302

Program Studi : Statistika

Judul Skripsi : Estimasi Model Regresi Nonparametrik *Spline* Poisson Pada Tingkat Kematian Bayi di Sulawesi Selatan

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

Tanda Tangan

1. Ketua : **Dr. Anna Islamiyati, S.Si., M.Si.** (........................)

2. Sekretaris : **Dr. Nurtiti Sunusi, S.Si., M.Si.** (........................)

3. Anggota : **Dr. Dr. Georgina Maria Tinungki, M.Si.** (........................)

4. Anggota : **Dr. Erna Tri Herdiani, S.Si., M.Si.** (........................)

KEMENTERIAN RISET, TEKNOLOGI DAN PENDIDIKAN TINGGI UNIVERSITAS HASANUDDIN DEPARTEMEN STATISTIKA FAKULTAS MIPA

Ditetapkan di: Makassar

Tanggal : 5 November 2020

KATA PENGANTAR

Puji syukur penulis panjatkan kepada **Tuhan Yang Maha Esa** atas berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan tugas akhir ini tepat waktu. Dalam menyelesaikan tugas akhir ini, ada banyak pihak yang terlibat dalam memberikan dukungan dan motivasi untuk segera menyelesaikan tugas akhir ini. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada semua pihak yang telah membantu baik secara langsung maupun tidak langsung.

Pada kesempatan ini penulis mengucapkan terima kasih yang sebesar-besarnya kepada keluarga terutama **papa, mama**, dan **ce Yuli** atas semua kasih sayang, kesabaran, perhatian, dan dukungan yang diberikan. Demikian pula penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada :

1. Ibu **Dr. Nurtiti Sunusi, S.Si., M.Si.** selaku Ketua Departemen Statistika FMIPA UNHAS dan Bapak **Andi Kresna Jaya, S.Si., M.Si.** selaku Sekretaris Departemen Statistika dan **segenap dosen pengajar**, terima kasih telah meberikan ilmu kepada penulis serta **staf Departemen Statistika** terima kasih telah dengan sabar selalu membantu dalam pengurusan persuratan.
2. Ibu **Dr. Anna Islamiyati, S.Si., M.Si.** dan Ibu **Dr. Nurtiti Sunusi, S.Si., M.Si.** selaku pembimbing utama dan pembimbing pertama terima kasih telah membimbing, memberikan motivasi, dan saran kepada penulis hingga tugas akhir ini dapat diselesaikan dengan baik.
3. Ibu **Dr. Dr. Georgina Maria Tinungki, M.Si** selaku penguji dan pembimbing akademik penulis serta Ibu **Dr. Erna Tri Herdiani, S.Si., M.Si.** selaku penguji terima kasih telah memberikan segala saran dan kritik guna membuat tugas akhir ini lebih baik.
4. Teman-teman **Statistika 2017** terima kasih untuk segala kebersamaan yang tak terlupakan selama 3 tahun, kerja sama, kenangan suka dan duka serta dukungan yang telah diberikan kepada penulis setiap kali penulis mengalami hambatan dalam menyelesaikan tugas akhir.
5. **Nurul Wahyuni** terima kasih karena selalu menemani dalam mengurus persuratan dan berkas serta menjadi tempat bertukar pikiran selama proses

penyelesaian tugas akhir. Untuk **Chika, Indry, Salsa** dan **Sri**, terima kasih untuk segala kenangan 3 tahun dan yang telah membuat hari-hari perkuliahan menjadi lebih berwarna. Untuk **Riska** terima kasih karena selalu menampung saya di kosnya untuk menyelesaikan tugas akhir ini. Untuk **Nisa, Aprilia**, dan **Dije** terima kasih untuk selalu ada setiap kali penulis butuh bantuan printer dalam mengurus persuratan. Untuk **Miftah** dan **Sakinah** terima kasih karena selalu menemani bertemu pembimbing dan memberi semangat kepada penulis.

6. Teman-teman **KKN Gelombang 104 Kecamatan Bontoala** terima kasih untuk segala dukungan dan kenangan kekeluargaannya.
7. Untuk **Selyna** terima kasih selalu ada setiap kal penulis butuh bantuan dan semangat yang diberikan kepada penulis.
8. Semua pihak yang telah memberikan bantuan dan dukungan secara langsung maupun tidak langsung dalam penyusunan tugas akhir ini yang tidak sempat penulis sebutkan satu persatu.

Penulis menyadari bahwa tugas akhir ini masih sangat jauh dari kesempurnaan sehingga segala kritikan dan saran yang membangun akan penulis terima agar tulisan ini dapat memberikan manfaat bagi semua pihak. Akhir kata semoga tugas akhir ini bermanfaat bagi para mahasiswa terutama bagi mahasiswa Statisyika Universitas Hasanuddin dan bagi perguruan tinggi lainnya.

Makassar, 3 November 2020

Penulis

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIK

Sebagai sivitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Novilia Jao

Nim : H051171302

Program studi : Statistika

Departemen : Statistika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Jenis karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Noneksklusif** (***Non-exclusif Royalty-Free Right***) atas skripsi saya yang berjudul:

"Estimasi Model Regresi Nonparametrik *Spline* Poisson Pada Tingkat Kematian Bayi di Sulawesi Selatan"

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal diatas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar, pada tanggal 5 November 2020

Yang Menyatakan:

(Novilia Jao)

ABSTRAK

Analisis regresi merupakan salah satu metode statistika yang bertujuan untuk mendapatkan estimasi titik untuk suatu parameter serta mengetahui hubungan antara variabel prediktor dengan variabel respon. Apabila variabel respon berdistribusi Poisson maka analisa yang dapat digunakan adalah analisis regresi Poisson. Namun, tidak semua data memiliki pola yang beraturan sehingga regersi Poisson kurang tepat digunakan. Untuk menyelesaikan masalah tersebut digunakan regresi nonparametrik Poisson multivariabel dengan estimator *spline truncated*. Adapun estimasi parameter dilakukan dengan menggunakan metode maksimum *likelihood*. Dalam penelitian ini, estimasi parameter regresi nonparametrik Poisson multivariabel diterapkan pada data tingkat kematian bayi di Sulawesi Selatan tahun 2017. Tingkat kematian bayi dapat diukur dari jumlah kematian bayi di bawah satu tahun per kabupaten/kota. Naik turunnya jumlah kematian bayi per kabupaten/kota menyebabkan data penelitian ini dianalisis menggunakan regresi nonparametrik Poisson multivariabel *spline truncated*. Metode pemilihan titik knot optimal menggunakan metode *Generalized Cross Validation* (GCV). Model terbaik terbentuk pada model *spline* linier dengan satu titik knot. Berdasarkan estimasi parameter yang terbentuk menunjukkan bahwa variabel jumlah bayi dengan berat badan lahir rendah (x_1) dan jumlah bayi yang diberi ASI eksklusif (x_3) signifikan mempengaruhi jumlah kematian bayi.

Kata Kunci: Estimasi Parameter, *Generalized Cross Validation,* Jumlah Kematian Bayi, Regresi Nonparametrik Poisson Multivariabel, *Spline Truncated.*

ABSTRACT

Regression analysis is a statistical method that aims to obtain a point estimate for a parameter and determine the relationship between the predictor variable and the response variable. If the response variable is Poisson distributed, the analysis that can be used is Poisson regression analysis. However, not all data have an orderly pattern, so the Poisson regression is not appropriate to use. To solve this problem, a multivariable Poisson nonparametric regression with a spline truncated estimator was used. The parameter estimation is done using the maximum likelihood method. In this research, the estimation parameters of multivariable Poisson nonparametric regression was applied to data of infant mortality rates in South Sulawesi in 2017. The infant mortality rate can be measured from the number of infant deaths under one year per district/city. The fluctuation of the number of infant deaths per district/city led to the data of this research being analyzed using nonparametric Poisson multivariable spline truncated regression. The method of selecting the optimal knot point uses the Generalized Cross Validation (GCV) method. The best model is formed on a linear spline model with one knot point. Based on the estimation of the parameters formed, it shows that the variable number of babies with low birth weight (x1) and the number of infants who are exclusively breastfed (x3) significantly affect the number of infant deaths.

Keywords: Estimated Parameters, Generalized Cross Validation, Multivariable Poisson Nonparametric Regression, Total Infant Mortality, Spline Truncated.

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Pengembangan aplikasi statistika pada era globalisasi telah berkembang dengan sangat pesat sebagai suatu bidang kajian terapan. Beberapa metode aplikasi statistika berkembang sebagai bentuk respon atas berbagai masalah yang ada dalam berbagai bidang, baik bidang kesehatan, pertanian, ekonomi, dan sebagainya. Salah satu metode statistika yang paling sering digunakan dalam penarikan kesimpulan suatu penelitian adalah analisis regresi. Hal ini karena analisis regresi merupakan metode statistika yang digunakan untuk mengevaluasi hubungan antara satu atau lebih variabel prediktor X_1, X_2, ..., X_m dengan variabel respon Y (Jus'at, 2019).

Salah satu pemanfaatan analisis regresi dalam bidang kesehatan adalah mengestimasi parameter model regresi pada tingkat kematian bayi dan menganalisis faktor-faktor yang mempengaruhi tingkat kematian bayi. Tingkat kematian bayi dapat diukur dari jumlah kematian bayi di bawah satu tahun (Dinkes, 2017). Menurut Kemenkes (2013) dalam profil kesehatan Indonesia 2012, tingkat kematian bayi di Indonesia apabila dibandingkan dengan Negara ASEAN lainnya masih tergolong tinggi, yaitu 4,2 kali lebih tinggi dari Negara Malaysia, 1,2 kali lebih tinggi dari Negara Filipina, dan 2,2 kali lebih tinggi dari Negara Thailand.

Jumlah kematian bayi merupakan data diskrit sehingga tergolong dalam data yang berdistribusi Poisson. Salah satu model regresi yang dapat digunakan untuk menganalisis hubungan antara variabel prediktor dengan variabel respon yang berupa data diskrit adalah model regresi Poisson. Pemodelan regresi Poisson merupakan pemodelan dengan variabel respon berdistribusi Poisson. Oleh karena itu, jumlah kematian bayi tepat dianalisis dengan menggunakan regresi Poisson. Aulele (2012) telah memodelkan jumlah kematian bayi di Provinsi Maluku tahun 2010 dengan menggunakan regresi Poisson. Namun, dalam pemodelan regersi Poisson sering terjadi pelanggaran asumsi antara lain variansi estimasi lebih besar daripada rata-ratanya (overdispersi). Overdispersi menyebabkan keragaman variansi sehingga asumsi residual dalam model regresi tidak terpenuhi. Penanganan overdispersi dapat menggunakan regresi *generalized* Poisson ataupun regresi binomial negatif. Safrida, dkk (2013) memodelkan angka kematian bayi di Jawa Tengah tahun 2007 dengan menggunakan regresi Poisson tergeneralisasi. Model regresi Poisson tergeneralisasi mirip dengan model regresi Poisson, yaitu merupakan suatu model GLM (*Generalized Linier Model*). Akan tetapi, pada model regresi Poisson tergeneralisasi mengasumsikan bahwa komponen acaknya berdistribusi Poisson tergeneralisasi. Selanjutnya Prahutama, dkk (2017) membandingkan angka kematian bayi di Jawa Tengah dengan model regresi Poisson tergeneralisasi dan model regresi binomial negatif. Hasil penelitian tersebut menunjukkan bahwa regresi binomial negatif lebih baik dalam memodelkan angka kematian bayi dibandingkan regresi Poisson tergeneralisasi. Hal ini ditunjukkan dengan nilai AIC yang dihasilkan oleh regresi binomial negatif sebesar 375,7 sedangkan nilai AIC yang dihasilkan oleh regresi Poisson tergeneralisasi sebesar 377,3. Semua model regresi tersebut merupakan model regresi parametrik. Akan tetapi, pada data jumlah kematian bayi terdapat pola data yang tidak beraturan sehingga penggunaan regresi parametrik kurang tepat digunakan.

Untuk mengatasi masalah tersebut, jumlah kematian bayi dimodelkan dengan menggunakan model regresi nonparametrik. Dalam regresi nonparametrik, data diharapkan mencari sendiri bentuk estimasinya tanpa dipengaruhi oleh faktor subjektivitas dari perancang penelitian. Salah satu metode regresi nonparametrik yang dapat mengatasi pola data ysang tidak beraturan adalah dengan menggunakan regresi nonparametrik *spline truncated*. Regresi nonparametrik *spline truncated* dapat memodelkan data pada pola data yang berubah-ubah pada sub-sub interval tertentu. Pratiwi (2017) memodelkan data angka kematian bayi dan status gizi buruk balita dengan menggunakan regresi *spline truncated* birespon. Alexander dan Alkema (2018) meneliti data kematian 195 negara pada tahun 1990-2015 menggunakan model regresi *spline* hirarki bayesian. Terdapat beberapa negara dengan jumlah kematian bayi yang lebih rendah maupun lebih tinggi dari biasanya sehingga membuat pola data menjadi tidak beraturan sehingga peneliti menggunakan metode regresi *spline* dalam memodelkan data kematian. Selanjutnya Kilinc dan Asfha (2019) memodelkan data kematian menggunakan regresi *penalized spline* dengan membandingkan pemodelan data kematian menggunakan GLM dan GAM (*Generalized Additive Model*). Pemodelan menggunakan GAM menghasilkan model yang lebih baik dibandingkan GLM, yang ditunjukkan dengan nilai AIC yang dihasilkan GLM sebesar 5137,492 sedangkan AIC yang dihasilkan model GAM sebesar 4834,453.

Dalam penelitian ini, penulis memodelkan jumlah kematian bayi di Provinsi Sulawesi Selatan yang merupakan data berdistribusi Poisson menggunakan regresi nonparametrik estimator *spline truncated* dengan melibatkan beberapa variabel prediktor, yaitu jumlah bayi dengan berat lahir rendah, jumlah bayi yang mendapatkan vitamin A, jumlah bayi yang diberi ASI eksklusif, jumlah bayi yang diberi imunisasi lengkap, dan jumlah ibu bersalin yang ditolong dengan tenaga medis.

Oleh sebab itu, penulis melakukan penelitian yang dituangkan dalam skripsi berjudul,

"Estimasi Model Regresi Nonparametrik *Spline* Poisson Pada Tingkat Kematian Bayi di Sulawesi Selatan"

1.2 Rumusan Masalah

Berdasarkan latar belakang, rumusan masalah yang dikaji pada penelitian ini, yaitu:

1. Bagaimana estimasi parameter model regresi nonparametrik melalui estimator *spline* Poisson pada tingkat kematian bayi di Sulawesi Selatan ?
2. Bagaimana hubungan variabel prediktor terhadap tingkat kematian bayi di Sulawesi Selatan dengan menggunakan model regresi nonparametrik *spline* Poisson ?

1.3 Batasan Masalah

Batasan masalah diperlukan agar pembahasan masalah dalam penelitian tidak terlalu luas. Batasan masalah untuk penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data sekunder jumlah kematian bayi di Provinsi Sulawesi Selatan tahun 2017 yang dirinci berdasarkan 24 kabupaten/kota.

2. Metode analisis yang digunakan adalah regresi nonparametrik dengan estimator *spline truncated*.
3. Pemilihan titik knot didasari pada nilai *Generalized Cross Validation* (GCV) minimum.
4. Jumlah titik knot yang dicobakan adalah 1-2 titik knot dengan orde dibatasi pada penggunaan 1-2 orde.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan dari penelitian ini diuraikan sebagai berikut:

1. Memperoleh estimator model regresi nonparametrik *spline* Poisson pada tingkat kematian bayi di Sulawesi Selatan.
2. Memperoleh hubungan variabel prediktor terhadap tingkat kematian bayi di Sulawesi Selatan menggunakan regresi nonparametrik *spline* Poisson.

1.5 Manfaat Penelitian

Adapun manfaat dari proposal penelitian ini, yaitu:

1. Memberikan gambaran bagaimana hubungan jumlah bayi dengan berat lahir rendah, jumlah bayi yang mendapatkan vitamin A, jumlah bayi yang diberi ASI eksklusif, jumlah bayi yang diberi imunisasi lengkap, dan jumlah ibu bersalin yang ditolong dengan tenaga medis terhadap tingkat kematian bayi di Sulawesi Selatan dalam bahasa matematika (model matematika).
2. Memberikan gagasan objektif dari sudut pandang statistika kepada para tenaga medis dalam memprediksi tingkat kematian bayi di Sulawesi Selatan.
3. Memberikan gambaran ke pemerintah untuk menggalakkan sosialisasi pemberian imuniasasi pada bayi kepada masyarakat luas.
4. Memberikan gambaran kepada masyarakat luas mengenai aplikasi dari ilmu statistika dalam kehidupan nyata khususnya dalam bidang kesehatan.

BAB II
TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi digunakan untuk mengetahui pola hubungan antara variabel respon dan variabel prediktor. Jika variabel respon adalah y_i dan variabel prediktor adalah x_i dengan $i = 1, 2, \cdots, n,$ maka pasangan data x_i, y_i akan memiliki model hubungan fungsional

$$y_i = f\,(x_i) + \varepsilon_i \;, \; i = 1, 2, ..., n \tag{2.1}$$

dengan:

$f(x_i)$ = kurva regresi

ε_i = error acak yang diasumsikan identik, independen, dan berdistribusi normal dengan mean nol dan varian σ^2 (Eubank, 1998).

Analisis regresi memiliki dua tujuan utama, yaitu memberikan cara megeksplorasi hubungan antara variabel respon dan variabel prediktor serta membuat prediksi (Silverman, 1985).

Model regresi umumnya dibagi ke dalam tiga bentuk, yaitu model regresi parametrik, model regresi nonparametrik, dan model regresi semiparametrik. Bentuk model regresi tersebut tergantung pada kurva $f(x_i)$. Apabila bentuk kurva $f(x_i)$ diketahui, maka pendekatan model regresi yang digunakan adalah pendekatan model regresi parametrik (Islamiyati & Budiantara, 2007). Apabila bentuk kurva $f(x_i)$ tidak diketahui atau tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola datanya, maka pendekatan model regresi yang digunakan adalah pendekatan model regresi nonparametrik (Lestari & Budiantara, 2010). Dalam beberapa kasus, sebagian bentuk pola data diketahui, sedangkan untuk sebagian yang lain bentuk pola datanya tidak diketahui. Pada kasus ini, pendekatan model regresi yang disarankan adalah pendekatan model regresi semiparametrik (Wahba, 1990).

2.2 Model Regresi Poisson

Salah satu model regresi yang dapat menggambarkan hubungan antara variabel respon *Y* yang berupa data diskrit dengan variabel prediktor *X* adalah regresi Poisson, yang mana model regresi Poisson termasuk dalam model linier tergeneralisasi. Jika μ_i adalah rata-rata jumlah kejadian dalam periode t_i dan diasumsikan μ_i tidak berubah dari titik data ke titik data secara bebas maka μ_i dapat dimodelkan sebagai fungsi dari *k* variabel prediktor. Dalam *Generalized Linier Model* (GLM), terdapat sebuah fungsi *g* yang menghubungkan rata-rata dari variabel responnya dengan sebuah prediktor linier, yaitu:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_m x_{im} \qquad (2.2)$$

dengan *g* disebut fungsi penghubung (*link function*), *i* = 1, 2, ..., *n* dan *j* = 1, 2, .., *m*.

Pada model regresi Poisson, fungsi penghubung yang digunakan adalah fungsi penghubung log karena fungsi log menjamin bahwa nilai variabel yang diharapkan dari variabel responnya akan bernilai nonnegatif. Berikut ini adalah fungsi penghubung yang digunakan untuk model regresi Poisson,

$$\ln E(y\,|\,x) = \ln(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{21} + \cdots + \beta_m x_{mi}$$

$$\mu_i = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_m x_{mi}) \qquad (2.3)$$

sehingga model regresi Poisson dapat ditulis sebagai berikut (Myres, 1990),

$$y_i = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) + \varepsilon_i \;,\; i = 1, 2, ..., n \qquad (2.4)$$

$\boldsymbol{\beta}$ ditaksir dengan $\hat{\boldsymbol{\beta}}$ sehingga,

$$\hat{y}_i = \exp(\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}) + \varepsilon_i \;,\; i = 1, 2, ..., n \qquad (2.5)$$

2.3 Model Regresi Nonparametrik

Regresi nonparametrik merupakan salah satu pendekatan yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dan variabel respon yang tidak diketahui kurva regresinya. Pendekatan nonparametrik memiliki fleksibelitas yang tinggi karena tidak mengasumsikan bentuk kurva regresi. Dalam hal ini, kurva diharapkan dapat mengestimasi sendiri bentuk kurva berdasarkan sebaran data dengan memerhatikan sifat kemulusan (*smooth*) kurva tersebut. Secara umum model regresi nonparametrik adalah,

$$y_i = f\,(x_i) + \varepsilon_i \;\; , i = 1, 2, ..., n \qquad (2.6)$$

dengan:

y_i = variabel respon

x_i = variabel prediktor

ε_i = error acak yang diasumsikan identik, independen, dan berdistribusi normal dengan mean nol dan varian σ^2

$f\,(x_i)$ = fungsi yang tidak diketahui, hanya diasumsikan termuat dalam suatu ruang fungsi tertentu, dengan pemilihan ruang fungsi tersebut biasanya termotivasi oleh sifat kemulusan (*smooth*) yang dimiliki oleh fungsi $f\,(x_i)$ tersebut (Herawati, 2011).

2.4 *Generalized Additive Model* (GAM)

Stone (1985) memperluas model sederhana pada persamaan (2.6) untuk variabel prediktor lebih dari satu dan dikenal dengan model aditif. Model aditif didefinisikan sebagai:

$$y_i = \sum_{j=1}^{m} f_j\left(x_{ji}\right) + \varepsilon_i \, , \;\; i = 1, 2, ..., n \qquad (2.7)$$

Hastie dan Tibshirani (1990) menjelaskan bahwa *Generalized Additive Model* (GAM) merupakan perluasan atau generalisasi dari model aditif pada persamaan (2.7). GAM memperluas model aditif dengan cara menghubungkan nilai harapan dari sebuah respon terhadap prediktor aditif melalui *link function* sehingga memungkinkan distribusi dari variabel respon berasal dari keluarga distribusi eksponensial. Model GAM dapat dituliskan sebagai berikut,

$$g(\mu_i) = \eta_i = f_0 + \sum_{j=1}^{m} f_j\left(x_{ji}\right) , \; i = 1, 2, ..., n \qquad (2.8)$$

dengan fungsi *g* merupakan *link function* yang bentuk fungsinya berbeda-beda untuk setiap distribusi dari anggota keluarga eksponensial.

2.5 Keluarga Eksponensial

Berdasarkan Collett (2002), suatu variabel acak *Y* dengan fungsi kepadatan peluang *f* dan parameter θ dan ϕ, dikatakan menjadi anggota distribusi keluarga eksponensial jika *f* dapat dinyatakan sebagai berikut,

$$f(y) = exp\left\{\frac{1}{\phi}\left[y\theta + c(\theta)\right] + d(\phi, y)\right\} \tag{2.9}$$

dengan $c(\theta)$ hanya merupakan fungsi dari θ sedangkan $d(\phi, y)$ merupakan fungsi dari ϕ dan *y*.

Distribusi Poisson memiliki fungsi massa peluang yang dapat dinyatakan sebagai berikut,

$$p(y) = \frac{e^{-\mu}\mu^y}{y!}$$

$$= exp\left\{ln\left[\frac{e^{-\mu}\mu^y}{y!}\right]\right\}$$

$$= exp\left\{ln\left(e^{-\mu}\mu^y\right) - ln(y!)\right\}$$

$$= exp\left\{ln\left(\mu^y e^{-\mu}\right) - ln(y!)\right\}$$

$$= exp\left\{\left(y\, ln\, \mu - \mu\right) - ln\, y!\right\}$$

dengan ϕ = 1 $\qquad c(\theta) = -\mu = -exp(\theta)$

$\theta = ln\ \mu \qquad d(\varphi, y) = -ln\, y!$

maka fungsi massa peluang distribusi Poisson memenuhi persamaan (2.9) sehingga distribusi Poisson merupakan anggota distribusi keluarga eksponensial.

2.6 Fungsi *Spline* Polinomial *Truncated* Multiprediktor

Hardle (1990) dan Wahba (1990) menyarankan penggunaan regresi spline sebagai alternatif pendekatan nonparametrik. *Spline* merupakan potongan (*truncated*) polinomial tersegmen yang kontinu dan memiliki sifat fleksibilitas sehingga mampu mengatasi pola data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik knot, serta kurva kurva yang dihasilkan relatif mulus. Titik knot adalah titik perpaduan bersama dari potongan-potongan tersebut atau titik yang menunjukkan terjadinya perubahan-perubahan perilaku kurva pada interval-interval yang berbeda. Titik knot diambil pada selang $a < K_k < b$, dengan *a* adalah nilai minimum dan *b* adalah nilai maksimum suatu data.

Wu & Zang (2006), menuliskan secara umum fungsi *spline* berorde *p* adalah sembarang fungsi yang dapat ditulis dalam bentuk sebagai berikut:

$$f(x) = \sum_{l=0}^{p} \beta_l x_i^l + \sum_{k=1}^{r} \beta_{p+k} (x_i - K_k)_+^p \tag{2.10}$$

dengan fungsi *truncated*:

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

β_l dan β_{l+k} adalah parameter, $K = \{K_1, K_2, \ldots, K_r\}$ adalah titik knot, $f(x)$ adalah fungsi regresi, *x* adalah variabel prediktor, dan $(x - K_k)_+^p$ adalah fungsi potongan polinomial tersegmen yang kontinu.

Apabila diberikan model nonparametrik,

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

Maka bentuk umum regresi *spline* keluarga polinomial *truncated* orde ke-*p* dengan satu prediktor adalah

$$y_i = \sum_{l=0}^{p} \beta_l x_i^l + \sum_{k=1}^{r} \beta_{l+k} (x_i - K_k)_+^p + \varepsilon_i \tag{2.11}$$

dengan β adalah parameter, *p* adalah orde dengan *l* = 1, 2, ..., *p*, dan *K* adalah titik knot dengan *k* = 1, 2, 3, ..., *r*, serta *i* = 1, 2, ..., *n* adalah banyaknya pengamatan.

Sedangkan bentuk umum regresi *spline truncated* orde ke-*p* dengan multiprediktor adalah

$$y_i = \sum_{j=1}^{m} f(x_{ji}) + \varepsilon_i \tag{2.12}$$

dengan,

$$f(x_{ji}) = \beta_{j0} + \sum_{l=0}^{p} \beta_{jl} x_{ji}^l + \sum_{k=1}^{r} \beta_{j(p+k)} (x_{ji} - K_{jk})_+^p , \; j = 1, 2, ..., m \tag{2.13}$$

Maka persamaan (2.9) dapat ditulis sebagai berikut:

$$y_i = f(x_{1i}, x_{2i}, \ldots, x_{mi}) = \beta_{00} + \sum_{j=1}^{m} \left(\sum_{l=0}^{p} \beta_{jl} x_{ji}^l + \sum_{k=1}^{r} \beta_{j(p+k)} (x_{ji} - K_{jk})_+^p \right) + \varepsilon_{ji} \tag{2.14}$$

dengan, $\beta_{00} = \sum_{j=1}^{m} \beta_{j0}$

$$(x_{ji}-K_{jk})_+^p=\begin{cases}(x_{ji}-K_{jk})^p & ,x_{ji}\geq K_{jk}\\ 0 & ,x_{ji}<K_{jk}\end{cases}$$

2.6.1 Estimasi Parameter Regresi Nonparametrik *Spline Truncated*

Hidayat (2017) melakukan suatu kajian dengan kurva regresi *f* dihampiri dengan fungsi *spline* *f* dengan titik knot *K* untuk mendapatkan estimator *spline*. Dalam bentuk matriks sebagai berikut (*Spline Truncated* Uniprediktor),

$$\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}=\begin{bmatrix}f(x_1)\\f(x_2)\\\vdots\\f(x_n)\end{bmatrix}+\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{bmatrix}$$

Apabila model regresi spline disajikan dalam bentuk matriks, diperoleh:

$$\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}=\begin{bmatrix}1 & x_1 & \cdots & x_1^p & (x_1-K_1)_+^p & (x_1-K_2)_+^p & \cdots & (x_1-K_r)_+^p\\1 & x_2 & \cdots & x_2^p & (x_2-K_1)_+^p & (x_2-K_2)_+^p & \cdots & (x_2-K_r)_+^p\\\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\1 & x_n & \cdots & x_n^p & (x_n-K_1)_+^p & (x_n-K_2)_+^p & \cdots & (x_n-K_r)_+^p\end{bmatrix}\begin{bmatrix}\beta_0\\\beta_1\\\vdots\\\beta_p\\\beta_{p+1}\\\beta_{p+2}\\\vdots\\\beta_{p+r}\end{bmatrix}+\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{bmatrix}$$

atau dapat ditulis dengan,

$$\boldsymbol{y}=\boldsymbol{X}[K_1,K_2,\ldots,K_r]\boldsymbol{\beta}+\boldsymbol{\varepsilon} \qquad (2.15)$$

Maka model *spline truncated* multiprediktor dalam bentuk matriks adalah,

$$\boldsymbol{y}=\boldsymbol{X}_{\boldsymbol{j}}[K_{j1},K_{j2},\ldots,K_{mr}]\boldsymbol{\beta}_{\boldsymbol{j}}+\boldsymbol{\varepsilon} \qquad (2.16)$$

Selanjutnya, estimasi parameter $\tilde{\boldsymbol{\beta}}_j$ diperoleh melalui metode kuadrat terkecil, yaitu dengan meminimumkan jumlah kuadrat error sebagai berikut:

$$\min_{\boldsymbol{\beta}\in R^{p+r}}\{\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\}=\min_{\tilde{\boldsymbol{\beta}}\in R^{p+r}}\left(\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{bmatrix}^T\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{bmatrix}\right)$$

$$\min_{\beta\in R^{p+r}}\{\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\}=\min_{\tilde{\boldsymbol{\beta}}\in R^{p+r}}\left\{\left[\boldsymbol{y}-\boldsymbol{X}_j[K_{j1},K_{j2},\ldots,K_{mr}]\boldsymbol{\beta}_j\right]^T\left[\boldsymbol{y}-\boldsymbol{X}_j[K_{j1},K_{j2},\ldots,K_{mr}]\boldsymbol{\beta}_j\right]\right\}$$

Misalkan $K_{jk} = [K_{j1}, K_{j2}, \ldots, K_{mr}]$ maka penyajian matriks diberikan sebagai berikut,

$$\sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}^{\boldsymbol{T}} \boldsymbol{\varepsilon}$$

$$= \left([\boldsymbol{y} - \boldsymbol{X}_{\boldsymbol{j}} [K_{j1}, K_{j2}, \ldots, K_{mr}] \boldsymbol{\beta}_{\boldsymbol{j}}]^T [\boldsymbol{y} - \boldsymbol{X}_{\boldsymbol{j}} [K_{j1}, K_{j2}, \ldots, K_{mr}] \boldsymbol{\beta}_{\boldsymbol{j}}] \right)$$

$$= \left([\boldsymbol{y} - \boldsymbol{X}_j [K_{jk}] \boldsymbol{\beta}_j]^T [\boldsymbol{y} - \boldsymbol{X}_j [K_{jk}] \boldsymbol{\beta}_j] \right)$$

$$= \boldsymbol{y}^{\boldsymbol{T}} \boldsymbol{y} - 2 \boldsymbol{\beta}_{\boldsymbol{j}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{j}}^{\boldsymbol{T}} [K_{jk}] \boldsymbol{y} + \boldsymbol{\beta}_{\boldsymbol{j}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{j}}^{\boldsymbol{T}} [K_{jk}] \boldsymbol{X}_{\boldsymbol{j}} [K_{jk}] \boldsymbol{\beta}_{\boldsymbol{j}} \tag{2.17}$$

Selanjutnya, persamaan (2.14) diturunkan terhadap vektor $\boldsymbol{\beta}_{\boldsymbol{j}}^T$ dan disamakan dengan nol, maka diperoleh:

$$\frac{\partial \boldsymbol{\varepsilon}^{\boldsymbol{T}} \boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta}_{\boldsymbol{j}}^T} = \frac{\partial \left(\boldsymbol{y}^{\boldsymbol{T}} \boldsymbol{y} - 2 \boldsymbol{\beta}_{\boldsymbol{j}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{j}}^{\boldsymbol{T}} [K_{jk}] \boldsymbol{y} + \boldsymbol{\beta}_{\boldsymbol{j}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{j}}^{\boldsymbol{T}} [K_{jk}] \boldsymbol{X}_{\boldsymbol{j}} [K_{jk}] \boldsymbol{\beta}_{\boldsymbol{j}} \right)}{\partial \boldsymbol{\beta}_{\boldsymbol{j}}^T}$$

$$-2 \boldsymbol{X}_{\boldsymbol{j}}^{\boldsymbol{T}} [K_{jk}] \boldsymbol{y} + 2 \boldsymbol{X}_{\boldsymbol{j}}^{\boldsymbol{T}} [K_{jk}] \boldsymbol{X}_{\boldsymbol{j}} [K_{jk}] \boldsymbol{\beta}_{\boldsymbol{j}} = 0$$

$$\boldsymbol{X}_{\boldsymbol{j}}^{\boldsymbol{T}} [K_{jk}] \boldsymbol{X}_{\boldsymbol{j}} [K_{jk}] \boldsymbol{\beta}_{\boldsymbol{j}} = \boldsymbol{X}_{\boldsymbol{j}}^{\boldsymbol{T}} [K_{jk}] \boldsymbol{y}$$

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{j}} = \left(\boldsymbol{X}_{\boldsymbol{j}}^{\boldsymbol{T}} [K_{jk}] \boldsymbol{X}_{\boldsymbol{j}} [K_{jk}] \right)^{-1} \boldsymbol{X}_{\boldsymbol{j}}^{\boldsymbol{T}} [K_{jk}] \boldsymbol{y} \tag{2.18}$$

dengan:

$$\boldsymbol{X}_{\boldsymbol{j}} \left[K_{jk} \right] = \boldsymbol{X}_{\boldsymbol{j}} [K_{j1}, K_{j2}, \ldots, K_{mr}]$$

2.6.2 Estimasi Kurva Regresi Nonparametrik *Spline Truncated*

Setelah diperoleh estimasi parameter seperti pada persamaan (2.15) maka estimasi titik kurva regresi nonparametrik *spline* dapat diperoleh dengan mensubtitusikan persamaan (2.15) ke dalam persamaan (2.16). akibatnya, estimasi untuk kurva regresi *spline* dengan titik knot K_{jk} diberikan oleh (Hidayat, 2017),

$$\hat{\boldsymbol{y}}[K_{jk}](x_{ji}) = \hat{\boldsymbol{f}}[K_{jk}](x_{ji}) = \boldsymbol{X}_{\boldsymbol{j}} [K_{jk}] \hat{\boldsymbol{\beta}}_{\boldsymbol{j}} \tag{2.19}$$

$$\hat{\boldsymbol{f}}[K_{jk}](x_{ji}) = \boldsymbol{X}_{\boldsymbol{j}} [K_{jk}] \left(\boldsymbol{X}_{\boldsymbol{j}}^{\boldsymbol{T}} [K_{jk}] \boldsymbol{X}_{\boldsymbol{j}} [K_{jk}] \right)^{-1} \boldsymbol{X}_{\boldsymbol{j}}^{\boldsymbol{T}} [K_{jk}] \boldsymbol{y}$$

Sehingga diperoleh:

$$\hat{\boldsymbol{f}}[K_{jk}](x_{ji}) = \boldsymbol{A}[K_{jk}] \boldsymbol{y} \tag{2.20}$$

dengan, $\boldsymbol{A}[K_{jk}] = \boldsymbol{X_j}[K_{jk}]\left(\boldsymbol{X_j}^{\boldsymbol{T}}[K_{jk}]\boldsymbol{X_j}[K_{jk}]\right)^{-1}\boldsymbol{X_j}^{\boldsymbol{T}}[K_{jk}]$ dan $\boldsymbol{X_j}[K_{jk}]$ adalah matriks dari model yang bergantung pada titik knot dan $K_{jk} = [K_{j1}, K_{j2}, \ldots, K_{mr}]$ merupakan titik knot.

2.6.3 Pemilihan Banyaknya Titik Knot Optimal

Menurut Eubank (1998), pemilihan K_{jk} optimal pada hakekatnya merupakan pemilihan lokasi dan banyaknya titik knot. Dalam regresi nonparametrik *spline* sangat penting untuk menentukan banyaknya titik knot optimal. Wan (2000) menyebutkan bahwa strategi untuk memilih banyaknya titik knot dengan menggunakan titik knot yang relatif sedikit. Menurut Lee (2002) alasan atau pertimbangan tersebut berkenaan dengan pemilihan model secara statistika yang mana pemilihan titik knot haruslah yang terbaik dan lebih mengarah pada kesederhanaan (*parsimony*) model. Jika diperoleh titik knot optimal maka diperoleh fungsi *spline* terbaik. Budiantara (2006) menyebutkan metode yang dapat digunakan untuk memilih jumlah titik knot optimal adalah dengan menggunakan metode *Generalized Cross Validation* (GCV).

Menurut Wu & Zang (2006), metode GCV dapat dituliskan sebagai berikut:

$$\mathrm{GCV}[K_{jk}] = \frac{\mathrm{n}^{-1}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\left(\mathrm{n}^{-1}\mathrm{tr}\left[\boldsymbol{I} - \boldsymbol{A}[K_{jk}]\right]\right)^2} \tag{2.21}$$

dengan, $\boldsymbol{A}[K_{jk}] = \boldsymbol{X_j}[K_{jk}]\left(\boldsymbol{X_j}^{\boldsymbol{T}}[K_{jk}]\boldsymbol{X_j}[K_{jk}]\right)^{-1}\boldsymbol{X_j}^{\boldsymbol{T}}[K_{jk}]$.

Pemilihan titik knot yang optimal dilakukan dengan melihat nilai GCV dari masing-masing orde dan titik knot kemudian dipilih nilai $\mathrm{GCV}[K_{jk}]$ yang paling minimum.

2.6.4 Pengujian Signifikansi Parameter Model Regresi *Spline* Poisson

Terdapat dua uji dalam mengetahui signifikansi parameter model regresi *spline* Poisson, yaitu (Agresti, 2007):

a. Uji Simultan (Uji *Likelihood Ratio*)

Uji simultan ini dilakukan untuk menguji signifikansi pengaruh dari semua variabel prediktor secara keseluruhan terhadap variabel respon.

Hipotesis:

H_0: $\beta_{11} = \beta_{12} = \cdots = \beta_{m(p+r)} = 0$

H_1: Minimal ada satu $\beta_{j(l+k)} \neq 0$, j = 1, 2, ..., m, l = 1, 2, ..., p, k = 1, 2, ..., r

Statistik uji:

$$G = -2\ln\left[\frac{L(\hat{\theta}_0)}{L(\hat{\theta})}\right] = -2\left[L(\hat{\theta}_0) - L(\hat{\theta})\right] \tag{2.22}$$

dengan, $L(\hat{\theta}_0)$ log *likelihood* untuk model yang tidak mengandung variabel bebas dan $L(\hat{\theta})$ adalah log *likelihood* untuk model yang mengandung seluruh variabel bebas.

Kriteria pengujian:

Tolak H_0 jika $G > \chi^2_{df;\alpha}$, dengan $df = \left[\sum_{j=1}^{m}(p+r)\right] - 1$.

b. Uji Individu (Uji *Wald*)

Uji *Wald* merupakan suatu pengujian yang bertujuan untuk mengetahui apakah masing-masing koefisien regresi signifikan atau tidak terhadap variabel respon dengan variabel prediktor. Dalam uji ini digunakan hipotesis sebagai berikut,

Hipotesis:

H_0: $\beta_{j(l+k)} = 0$, j = 1, 2, ..., m, l = 1, 2, ..., p, k = 1, 2, ..., r

H_1: $\beta_{j(l+k)} \neq 0$, j = 1, 2, ..., m, l = 1, 2, ..., p, k = 1, 2, ..., r

Statistik uji:

$$W_{j(l+k)} = \left[\frac{\hat{\beta}_{j(l+k)}}{SE(\hat{\beta}_{j(l+k)})}\right]^2 \tag{2.23}$$

dengan, adalah $\hat{\beta}_{j(l+k)}$ nilai estimasi parameter dan $SE(\hat{\beta}_{j(l+k)})$ adalah standar eror dari nilai estimasi parameter.

Kriteria pengujian:

Tolak H_0 jika $W_{j(l+k)} > \chi^2_{\alpha}$

2.7 Metode *Maximum Likelihood Estimation*

Metode *maksimum likelihood estimation* (MLE) dapat digunakan untuk menaksir parameter suatu model yang diketahui fungsi probabilitasnya. Misalkan Y_1, Y_2, ..., Y_n adalah sampel acak dari populasi dengan fungsi probabilitas $f(y \mid \boldsymbol{\theta})$ yang bergantung pada $\boldsymbol{\theta} = \mu, \boldsymbol{\beta}$ dengan $\boldsymbol{\theta}$ adalah parameter yang tidak diketahui. Karena Y_1, Y_2, ..., Y_n saling bebas maka fungsi probabilitas bersama dari Y_1, Y_2, ..., Y_n dapat dinyatakan sebagai berikut,

$$f(y_1, y_2, ..., y_n \mid \boldsymbol{\theta}) = f(y_1 \mid \boldsymbol{\theta}).\, f(y_2 \mid \boldsymbol{\theta}) \; ... \; f(y_n \mid \boldsymbol{\theta})$$

Fungsi *likelihood* didefinisikan sebagai fungsi probabilitas bersama dari Y_1, Y_2, ..., Y_n yang dianggap sebagai fungsi dari $\boldsymbol{\theta}$ yang dituliskan sebagai $L(\boldsymbol{\theta} \mid y)$, yaitu (Hogg, McKean, & Craig, 2013):

$$L(\boldsymbol{\theta} \mid y) = f(\boldsymbol{\theta} \mid y_1, y_2, ..., y_n)$$

$$L(\boldsymbol{\theta} \mid y) = f(y_1 \mid \boldsymbol{\theta}).\, f(y_2 \mid \boldsymbol{\theta}) \; ... \; f(y_n \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i \mid \boldsymbol{\theta}) \tag{2.24}$$

Estimator maksimum *likelihood* $\hat{\boldsymbol{\theta}}$ adalah nilai $\boldsymbol{\theta}$ yang memaksimumkan fungsi *likelihood* $L(\boldsymbol{\theta} \mid y)$. Namun, lebih mudah dikerjakan dengan logaritma natural dari fungsi *likelihood*, yaitu $l(\boldsymbol{\theta} \mid y) = \ln L(\boldsymbol{\theta} \mid y)$. Karena fungsi logaritma natural adalah fungsi yang monoton naik maka nilai yang memaksimumkan fungsi $l(\boldsymbol{\theta} \mid y)$ sama dengan memaksimumkan fungsi $L(\boldsymbol{\theta} \mid y)$ sehingga dapat dituliskan sebagai berikut,

$$l(\boldsymbol{\theta} \mid y) = \ln L(\boldsymbol{\theta} \mid y)$$

$$l(\boldsymbol{\theta} \mid y) = \ln \left\{ \prod_{i=1}^{n} f(y_i \mid \boldsymbol{\theta}) \right\} = \sum_{i=1}^{n} \ln f(y_i \mid \boldsymbol{\theta}) \tag{2.25}$$

Untuk memperoleh nilai $\boldsymbol{\theta}$ yang memaksimumkan fungsi $l(\boldsymbol{\theta} \mid y)$ maka $l(\boldsymbol{\theta} \mid y)$ diturunkan terhadap $\boldsymbol{\theta}$ dan kemudian menyamakannya dengan 0 seperti pada persamaan (2.26) berikut,

$$l'(\boldsymbol{\theta} \mid y) = \frac{\partial l(\boldsymbol{\theta} \mid y)}{\partial \boldsymbol{\theta}} = 0 \tag{2.26}$$

2.8 Algoritma *Fisher Scoring*

Algoritma *scoring* atau biasa dikenal dengan *Fisher scoring*. Algoritma *Fisher scoring* ditemukan oleh Ronald Fisher. Algoritma *Fisher scoring* adalah sebuah bentuk dari metode *Newton* yang digunakan dalam statistik untuk menyelesaikan persamaan maksimum *likelihood*.

Algoritma *Fisher scoring* mirip dengan algoritma *Newton raphson*, perbedaannya adalah *Fisher scoring* menggunakan matriks informasi. Matriks informasi tersebut adalah negatif dari nilai ekspektasi dari matriks turunan kedua fungsi yang akan dimaksimumkan sedangkan algoritma *Newton raphson* menggunakan matriks turunan kedua dari fungsi yang akan dimaksimumkan (Ehlers, 2002). Adapun rumus perulangan *Fisher scoring* adalah sebagai berikut,

$$\hat{\boldsymbol{\beta}}^{(r+1)} = \hat{\boldsymbol{\beta}}^{(r)} + I\left(\boldsymbol{\beta}^{(r)}\right)^{-1} l'\left(\boldsymbol{\beta}^{(r)}\right) \qquad (2.27)$$

dengan $I^{(r)}$ adalah matriks informasi yang merupakan negatif dari nilai ekspektasi dari matriks turunan kedua fungsi *loglikelihood*, yaitu $I = -E\left(\frac{\partial^2 l}{\partial \boldsymbol{\beta}\, \partial \boldsymbol{\beta}^T}\right)$ dan $l'^{(r)}$ adalah turunan pertama dari fungsi *loglikelihood*.

Algoritma *Fisher scoring* menurut Smyth (2002) adalah salah satu bentuk pengembangan dari metode *Newton raphson* dengan mengganti matriks turunan kedua fungsi *loglikelihood* dengan matriks ekspektasi turunan kedua dari fungsi *loglikelihood* yang berukuran (m+1) x (m+1).

Adapun langkah-langkah dalam algoritma *Fisher scoring* adalah sebagai berikut:

a. Menentukan nilai awal $\hat{\boldsymbol{\beta}}^{(0)}$ dengan menggunakan metode OLS (*Ordinary Least Square*).
b. Menentukan $I\left(\boldsymbol{\beta}^{(0)}\right)$ dan $l'\left(\boldsymbol{\beta}^{(0)}\right)$.
c. Menghitung estimator parameter untuk r = 0, 1, 2, ... dengan menggunakan persamaan (2.27).
d. Mengulangi iterasi sampai diperoleh nilai yang konvergen, yaitu $\left|\hat{\boldsymbol{\beta}}^{(r+1)} - \hat{\boldsymbol{\beta}}^{(r)}\right| \le \varepsilon$, dengan ε adalah konstanta positif yang ditentukan.

2.9 Tingkat Kematian Bayi

Tingkat kematian bayi adalah kematian yang terjadi antara saat setelah bayi lahir sampai bayi belum berusia tepat satu tahun (Dinkes Sulsel, 2017). Banyak faktor yang dikaitkan dengan kematian bayi. Secara garis besar, dari sisi penyebabnya, kematian bayi ada dua macam, yaitu endogen dan eksogen. Kematian bayi endogen atau yang umum disebut kematian neonatal. Kematian bayi yang terjadi pada bulan pertama setelah dilahirkan dan umumnya disebabkan oleh faktor-faktor yang dibawa anak sejak lahir yang diperoleh dari orang tuanya pada saat konsepsi atau didapat selama kehamilan. Sedangkan kematian bayi eksogen atau kematian post neonatal adalah kematian bayi yang terjadi setelah usia satu bulan sampai menjelang usia satu tahun yang disebabkan oleh faktor-faktor yang berkaitan dengan pengaruh lingkungan luar.

Ada banyak faktor yang mempengaruhi tingkat kematian bayi tetapi tidak mudah untuk menentukan faktor yang paling dominan dan faktor yang kurang dominan. Namun, dalam Profil Kesehatan Indonesia dijelaskan bahwa beberapa penyebab kematian bayi dapat bermula dari masa kehamilan. Penyebab kematian bayi yang terbanyak adalah disebabkan karena pertumbuhan janin yang lambat, kekurangan gizi pada janin, kelahiran prematur dan Berat Badan Lahir Rendah (BBLR) sedangkan penyebab lainnya yang cukup banyak terjadi adalah kejadian kurangnya oksigen dalam rahim (Hipoksia Intrauterus) dan kegagalan nafas secara spontan dan teratur pada saat lahir atau beberapa saat setelah lahir (Dinkes Sulsel, 2017).