PEMODELAN REGRESI *ZERO INFLATED NEGATIVE BINOMIAL* PADA DATA YANG MENGALAMI OVERDISPERSI

SKRIPSI

AINUL FAJRI

H051171018

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

JANUARI 2022

PEMODELAN REGRESI *ZERO INFLATED NEGATIVE BINOMIAL* PADA DATA YANG MENGALAMI OVERDISPERSI

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

AINUL FAJRI
H051171018

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
JANUARI 2022

LEMBAR PERNYATAAN KEOTENTIKAN

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

Pemodelan Regresi *Zero Inflated Negative Binomial* pada Data yang Mengalami Overdispersi

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

Makasssar, 26 Januari 2022

Ainul Fajri

H051171018

PEMODELAN REGRESI *ZERO INFLATED NEGATIVE BINOMIAL* PADA DATA YANG MENGALAMI OVERDISPERSI

Disetujui oleh:

Pembimbing Utama	Pembimbing Pertama
Andi Kresna Jaya, S.Si., M.Si.	**Dr. La Podje Talangko., M.Si.**
NIP. 197312282000031001	**NIP. 195512191987011001**

Ketua Departemen Statistika

Dr. Nurtiti Sunusi, S.Si., M.Si.

NIP. 197201171997032002

Pada Tanggal: 26 Januari 2022

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh	:
Nama	: Ainul Fajri
NIM	: H051171018
Program Studi	: Statistika
Judul Skripsi	: Pemodelan Regresi *Zero Inflated Negative Binomial* pada Data yang Mengalami Overdispersi

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

DEWAN PENGUJI

		Tanda Tangan
1. Ketua	: Andi Kresna Jaya, S.Si., M.Si.	(…………………)
2. Sekretaris	: Dr. La Podje Talangko., M.Si.	(…………………)
3. Anggota	: Dr. Erna Tri Herdiani., S.Si., M.Si.	(…………………)
4. Anggota	: Dr. Anna Islamiyati, S.Si., M.Si.	(…………………)

Ditetapkan di : Makassar

Tanggal : 26 Januari 2022

KATA PENGANTAR

Puji syukur kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat, nikmat dan hidayah-Nya, serta shalawat dan salam semoga senantiasa tercurahkan kepada Nabi yang paling dimuliakan, pemimpin orang-orang bertakwa yakni Rasulullah *Shallallahu Alaihi Wasallam* dan kepada para keluarga serta sahabat beliau yang senantiasa kita rindukan perjumpaan dengannya. Alhamdulillah, semua kemudahan yang penulis dapatkan tidak lepas dari pertolongan Allah dan doa dari orang-orang yang tulus, akhirnya skripsi dengan judul "**Pemodelan Regresi *Zero Inflated Negative Binomial* pada Data yang Mengalami Overdispersi**" yang disusun sebagai salah satu syarat akademik untuk meraih gelar sarjana pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin ini dapat dirampungkan. Penulis berharap skripsi ini dapat memberikan tambahan pengetahuan baru bagi para pembelajar statistika.

Penulis menyadari bahwa penyelesaian skripsi ini dengan segala keterbatasan kemampuan dan pengetahuan dapat melewati segala hambatan dan masalah berkat bantuan dan dorongan dari berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang tak terhingga serta penghargaan yang setinggi-tingginya untuk Ayahanda **Sattuo** dan Ibunda tercinta **Samrah** yang tak kenal lelah mendoakan, memberikan dukungan, dan selalu melimpahkan cinta dan kasih sayangnya kepada penulis sehingga mereka menjadi motivasi terbesar penulis dalam menyelesaikan tugas akhir ini. Serta untuk keluarga besar penulis, terima kasih atas doa dan dukungannya selama ini.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada:

1. Ibu **Prof. Dr. Dwia Aries Tina Pulubuhu, M.A,** selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. Bapak **Dr. Eng. Amiruddin,** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.

3. Ibu **Dr. Nurtiti Sunusi, S.Si., M.Si.,** selaku Ketua Departemen Statistika, segenap dosen pengajar dan staf Departemen Statistika yang telah memberikan ilmu dan pengetahuan serta bantuan-bantuan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
4. Bapak **Andi Kresna Jaya, S. Si., M. Si.** selaku dosen pembimbing utama yang dengan sabar, tulus dan ikhlas meluangkan begitu banyak waktu di tengah berbagai kesibukan dan prioritasnya untuk membimbing dan memberikan masukan serta motivasi kepada penulis mulai dari awal hingga selesainya penulisan tugas akhir ini dan Alm. Bapak **Dr. La Podje Talangko., M.Si.** selaku dosen pembimbing pertama yang telah meluangkan waktunya untuk memberikan masukan dan motivasi.
5. Ibu **Dr. Erna Tri Herdiani., S.Si., M.Si.** selaku dosen penguji sekaligus pembimbing akademik penulis dan Ibu **Dr. Anna Islamiyati, S.Si., M.Si.** selaku dosen penguji terima kasih telah meluangkan waktu dan memberikan segala saran dan kritikan yang membangun dalam penyempurnaan penyusunan tugas akhir ini.
6. Teman-teman **Statistika 2017**, terkhusus **Nurul Wahyuni, S. Si.** yang merupakan sosok guru dan pembimbing bagi penulis di masa perkuliahan serta menjadi tempat bertukar pikiran selama proses penyelesaian tugas akhir, terima kasih atas ilmu, kebersamaan, suka dan dukanya.
7. Keluarga besar **DISKRIT 2017**, terima kasih untuk cerita sekaligus kenangan selama proses yang telah dilalui. Banyak pengalaman dan pelajaran berharga yang penulis dapatkan dari kalian. Tetap jaga **SATUKAN, ERATKAN, KUATKAN.**
8. Kakak-kakak, teman-teman, dan adik-adik anggota **Keluarga Mahasiswa FMIPA UNHAS**, terkhusus anggota keluarga **HIMASTAT FMIPA UNHAS** dan **HIMATIKA FMIPA UNHAS**, terima kasih atas ilmu yang mungkin tidak bisa didapatkan di proses perkuliahan. Penulis merasa bangga menjadi salah satu bagian dari keluarga ini.
9. Seluruh teman-teman **KKN Tematik Gelombang 105 Posko Kecamatan Tamalanrea 12**, yang telah menjadi sahabat sekaligus keluarga selama proses

KKN berlangsung hingga saat ini. Terima kasih atas segala canda dan tawanya dalam memotivasi penulis.

10. Kak **Muhammad Fadil** dan Kak **Irfan Taufik** yang telah memberikan banyak ilmu, petunjuk, saran dan motivasi.
11. Serta kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih untuk semuanya. Semoga segala dukungan dan partisipasi yang diberikan kepada penulis bernilai ibadah di sisi Allah *Subhanahu Wa Ta'ala*.

Penulis berharap tugas akhir ini dapat memberikan tambahan pengetahuan baru bagi para pembelajar statistika. Penulis menyadari bahwa dalam penulisan tugas akhir ini masih banyak kekurangan. Oleh karena itu, dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga apa yang telah kita lakukan hari ini dapat membuat kita selangkah lebih maju dari hari-hari sebelumnya dan mudah-mudahan tugas akhir ini bermanfaat bagi orang-orang yang berkepentingan. Aamiin….

Makassar, 26 Januari 2022

Ainul Fajri

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama	: Ainul Fajri
NIM	: H051171018
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 Non-eksklusif** (***Non-exclusive Royalty- Free Right***) atas karya ilmiah saya yang berjudul:

"Pemodelan Regresi *Zero Inflated Negative Binomial pada Data yang Mengalami Overdispersi*"

beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir 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 26 Januari 2022

Yang menyatakan,

Ainul Fajri

ABSTRAK

Regresi Poisson merupakan model regresi non linear dengan variabel responnya berupa data diskrit dan berdistribusi Poisson. Analisis data menggunakan regresi Poisson harus memenuhi asumsi seperti nilai varians dan nilai rata-rata dari variabel respon memiliki nilai yang sama. Akan tetapi, pada penerapannya sering terjadi overdispersi, yaitu nilai varians lebih besar dari nilai rata-ratanya. Overdispersi pada regresi Poisson dapat terjadi karena banyaknya pengamatan yang bernilai nol pada variabel respon (*excess zeros*). Data yang terdapat *excess zeros* dan overdispersi lebih sesuai menggunakan regresi *Zero-Inflated Negative Binomial* (ZINB). Model regresi *Zero Inflated Negative Binomial* (ZINB) merupakan model yang dibentuk dari distribusi campuran Poisson gamma. Parameter model regresi *Zero Inflated Negative Binomial* (ZINB) diestimasi menggunakan metode *Maximum Likelihood Estimation* (MLE) dengan algoritma *Expectation Maximization* (EM). Penelitian ini diaplikasikan pada data jumlah kematian neonatal di Kota Makassar tahun 2018. Hasil pengujian parameter model regresi *Zero Inflated Negative Binomial* (ZINB) menunjukkan bahwa variabel prediktor yang berpengaruh signifikan secara parsial adalah jumlah bayi yang baru lahir dengan Berat Badan Lahir Rendah (BBLR).

Kata Kunci: Overdispersi, Kematian Neonatal, Regresi ZINB, MLE

ABSTRACT

Poisson regression is a non-linear regression model with the response variable in the form of discrete data and Poisson distribution. Data analysis using Poisson regression must meet assumptions such as the variance value and the average value of the response variables have the same value. However, in its application, overdispersion often occurs, namely the variance value is greater than the average value. Overdispersion in Poisson regression can occur because of the number of zero observations on the response variable (zero excess). Data with zero excess and overdispersion are more suitable using Zero Inflated Negative Binomial (ZINB) regression. The Zero Inflated Negative Binomial (ZINB) regression model is a model formed from the mixed distribution of Poisson gamma. The parameters of the Zero Inflated Negative Binomial (ZINB) regression model were estimated using the Maximum Likelihood Estimation (MLE) method with the Expectation Maximization (EM) algorithm. This study was applied to data on the number of neonatal deaths in Makassar City in 2018. The results of testing the parameters of the Zero Inflated Negative Binomial (ZINB) regression model showed that the predictor variable that had a partially significant effect was the number of newborns with low birth weight (LBW).

Kata Kunci: Overdispersion, Neonatal Mortality, ZINB Regression, MLE

DAFTAR ISI

HALAMAN SAMPUL i
HALAMAN JUDUL ii
HALAMAN PERNYATAAN KEOTENTIKAN iii
HALAMAN PERSETUJUAN PEMBIMBING iv
HALAMAN PENGESAHAN v
KATA PENGANTAR vi
PERSETUJUAN PUBLIKASI KARYA ILMIAH ix
ABSTRAK x
ABSTRACT xi
DAFTAR ISI xii
DAFTAR TABEL xiv
DAFTAR LAMPIRAN xv
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 2
1.3 Batasan Masalah 3
1.4 Tujuan Penelitian 3
1.5 Manfaat Penelitian 3
BAB II TINJAUAN PUSTAKA 4
2.1 *Generalized Linear Model* (GLM) 4
2.2 Regresi Poisson 4
2.2.1 Overdispersi 6
2.2.2 *Excess Zeros* 6
2.3 Regresi Binomial Negatif 7
2.4 Regresi *Zero Inflated Negative Binomial* 9
2.5 Metode *Maximum Likelihood Estimation* 10
2.6 Algoritma *Expectation Maximization* 11
2.7 Uji Serentak Parameter Model 12

2.8 Uji Parsial Parameter Model 13
2.8.1 Uji Parsial Parameter $\boldsymbol{\beta}$ 13
2.8.2 Uji Parsial Parameter $\boldsymbol{\gamma}$ 13
2.9 Pemilihan Model Terbaik 14
2.10 Jumlah Kematian Neonatal 14
BAB III METODE PENELITIAN 16
3.1 Sumber Data 16
3.2 Identifikasi Variabel 16
3.3 Metode Analisis 16
BAB IV HASIL DAN PEMBAHASAN 19
4.1 Estimasi Parameter Regresi *Zero Inflated Negative Binomial* (ZINB) .. 19
4.2 Penerapan Model Regresi *Zero Inflated Negative Binomial* (ZINB) pada Data Jumlah Kematian Neonatal di Kota Makassar Tahun 2018 27
4.2.1 Uji Kecocokan Distribusi Poisson 27
4.2.2 Uji Overdispersi 27
4.2.3 Pemeriksaan *Excess Zeros* pada Variabel Respon 28
4.2.4 Pemodelan Regresi *Zero Inflated Negative Binomial* (ZINB) pada Data Jumlah Kematian Neonatal di Kota Makassar Tahun 2018 28
4.2.5 Uji Serentak Parameter Model 29
4.2.6 Uji Parsial Parameter $\boldsymbol{\beta}$ 30
4.2.7 Uji Parsial Parameter $\boldsymbol{\gamma}$ 31
4.3 Pemilihan Model Terbaik 32
BAB V PENUTUP 33
5.1 Kesimpulan 33
5.2 Saran 33
DAFTAR PUSTAKA 34
LAMPIRAN 37

DAFTAR TABEL

Tabel 4.1 Uji Kecocokan Distribusi Poisson 27
Tabel 4.2 Uji Overdispersi 27
Tabel 4.3 Pemeriksaan excess zeros pada variabel respon 28
Tabel 4.4 Estimasi Parameter $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ 29
Tabel 4.5 Uji Parsial Parameter $\boldsymbol{\beta}$ 30
Tabel 4.6 Uji Parsial Parameter $\boldsymbol{\gamma}$ 31
Tabel 4.7 Nilai AIC model regresi binomial negatif dan ZINB 32

DAFTAR LAMPIRAN

Lampiran 1. Data Jumlah Kematian Neonatal dan Faktor-Faktor yang Mempengaruhi di Kota Makassar Provinsi Sulawesi Selatan Tahun 2018 38
Lampiran 2. Hasil Output untuk Uji Kecocokan Distribusi Poisson 40
Lampiran 3. Hasil Output untuk Uji Overdispersi 41
Lampiran 4. Hasil Output untuk Estimasi Parameter 42
Lampiran 5. Hasil Output untuk Uji Serentak Parameter Model 43
Lampiran 6. Hasil Output untuk Uji Parsial Parameter Model 44
Lampiran 7. Hasil Output untuk Nilai AIC Model Regresi Binomial Negatif dan ZINB 45

BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis yang digunakan untuk memodelkan hubungan antara variabel respon dengan satu atau lebih variabel prediktor adalah analisis regresi. Hubungan antara variabel respon dan variabel prediktor dinyatakan dalam persamaan regresi yang dapat berbentuk hubungan linear atau non linear. Persamaan regresi linear digunakan untuk menganalisis variabel respon yang berupa data kontinu dan mengikuti distribusi normal, namun banyak ditemukan variabel respon yang tidak berdistribusi normal. Untuk mengatasi hal tersebut dikembangkan *Generalized Linear Model* (GLM). *Generalized Linear Model* digunakan sebagai perluasan model regresi umum dengan variabel responnya berdistribusi keluarga eksponensial, meliputi distribusi normal, binomial, Poisson, binomial negatif, eksponensial, gamma, dan invers normal (Myers dkk., 2010).

Jika variabel respon yang digunakan merupakan data diskrit dan berdistribusi Poisson, maka dapat digunakan model regresi Poisson untuk pembentukan model regresi. Analisis data menggunakan regresi Poisson harus memenuhi asumsi seperti nilai varians dan nilai rata-rata dari variabel respon memiliki nilai yang sama atau equidispersi (Myers dkk., 2010). Akan tetapi, pada penerapannya sering terjadi overdispersi, yaitu nilai varians lebih besar dari nilai rata-ratanya. Data yang mengandung overdispersi dan dianalisis dengan regresi Poisson menghasilkan galat baku yang lebih kecil dari nilai sesungguhnya (*underestimate*). Hal ini menyebabkan kesimpulan yang diperoleh menjadi tidak valid (McCullagh dan Nelder, 1989). Overdispersi pada regresi Poisson dapat terjadi karena banyaknya pengamatan yang bernilai nol pada variabel respon (*excess zeros*).

Lambert (1992) mengatakan bahwa data yang terdapat kasus *excess zeros* dapat dimodelkan dengan menggunakan regresi *Zero-Inflated Poisson* (ZIP). Regresi ini hanya mampu untuk menangani data *excess zeros,* tidak untuk kasus overdispersi. Oleh karena itu Ismail dan Zamani (2013) menyarankan bahwa data yang terdapat *excess zeros* dan overdispersi lebih sesuai menggunakan regresi *Zero-Inflated Negative Binomial* (ZINB) dari pada *Zero-Inflated Poisson* (ZIP).

Menurut Hilbe (2011), model regresi *Zero Inflated Negative Binomial* (ZINB) merupakan model yang dibentuk dari distribusi campuran Poisson gamma. Penggunaan model regresi *Zero Inflated Negative Binomial* (ZINB) telah dilakukan oleh beberapa peneliti, diantaranya Ariawan dkk. (2012) menggunakan regresi *Zero Inflated Negative Binomial* (ZINB) untuk mengatasi overdispersi akibat *excess zeros* pada data asuransi mobil PT. Asuransi Sinar Mas Cabang Semarang tahun 2010. Astuti dan Zain (2015) memodelkan data penderita tetanus neonatorum di Provinsi Jawa Timur menggunakan regresi *Zero Inflated Negative Binomial* (ZINB). Dewanti dan Susilawati (2016) menggunakan regresi *Zero Inflated Negative Binomial* (ZINB) untuk mengatasi masalah overdispersi pada data angka kematian ibu di Provinsi Bali. Yulian (2018) menggunakan regresi *Zero Inflated Negative Binomial* (ZINB) dalam memodelkan frekuensi bepergian penduduk Kabupaten Tapanuli Selatan tahun 2016 yang melanggar asumsi equidispersi.

Penelitian ini menerapkan model regresi *Zero Inflated Negative Binomial* (ZINB) pada kasus jumlah kematian neonatal di Kota Makassar tahun 2018. Jumlah kematian neonatal adalah banyaknya kematian bayi yang terjadi saat bayi baru lahir hingga bayi berusia di bawah 28 hari, yang mana data tersebut merupakan data jumlahan sehingga tergolong ke dalam data diskrit. Kematian neonatal umumnya disebabkan oleh faktor yang dibawa anak sejak lahir yang diperoleh dari orang tua pada saat konsepsi atau selama kehamilan (Dinkes, 2019). Penelitian terkait jumlah kematian neonatal telah dilakukan sebelumnya. Seperti Ratmila (2020) yang menerapkan peta kendali atribut menggunakan *Zero Inflated Generalized Poisson* (ZIGP) pada data jumlah kematian neonatal di Kota Makassar tahun 2016 & 2017.

Berdasarkan uraian tersebut, penulis tertarik untuk membahas tentang pemodelan regresi *Zero Inflated Negative Binomial* (ZINB) untuk mengatasi masalah overdispersi pada data jumlah kematian neonatal di Kota Makassar tahun 2018.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka permasalahan yang akan dibahas dalam penelitian ini adalah sebagai berikut:

1. Bagaimana mengestimasi parameter model regresi *Zero Inflated Negative Binomial* (ZINB) menggunakan metode *Maximum Likelihood Estimation* (MLE)?
2. Bagaimana memodelkan data jumlah kematian neonatal di Kota Makassar tahun 2018 dengan menggunakan model regresi *Zero Inflated Negative Binomial* (ZINB)?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. Data yang digunakan adalah jumlah kematian neonatal di Kota Makassar tahun 2018.
2. Estimasi parameter model regresi *Zero Inflated Negative Binomial* (ZINB) dilakukan menggunakan metode *Maximum Likelihood Estimation* (MLE) dengan algoritma *Expectation Maximization* (EM).

1.4 Tujuan Penelitian

Tujuan dalam penelitian ini adalah sebagai berikut:

1. Mengestimasi parameter model regresi *Zero Inflated Negative Binomial* (ZINB) menggunakan metode *Maximum Likelihood Estimation* (MLE).
2. Memodelkan data jumlah kematian neonatal di Kota Makassar tahun 2018 dengan menggunakan model regresi *Zero Inflated Negative Binomial* (ZINB).

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan wawasan keilmuan yang lebih khusus kepada penulis tentang pemodelan regresi *Zero Inflated Negative Binomial* (ZINB) yang mengalami masalah overdispersi.
2. Memberikan informasi kepada pemerintah untuk menetapkan kebijakan dalam rangka mengatasi jumlah kematian neonatal di Kota Makassar.

BAB II
TINJAUAN PUSTAKA

2.1 *Generalized Linear Model* (GLM)

Generalized Linear Model (GLM) merupakan perluasan dari model regresi ketika variabel respon mengikuti distribusi keluarga eksponensial yaitu normal, binomial, Poisson, geometrik, binomial negatif, eksponensial, dan gamma. McCullagh dan Neldar (1989) menyatakan terdapat tiga komponen utama yang harus ada dalam suatu *Generalized Linear Model*, yaitu:

1. Komponen acak, yaitu suatu komponen yang mengidentifikasi distribusi dari variabel respon (Y) berasal dari keluarga eksponensial. Bentuk keluarga distribusi eksponensial dapat dinyatakan dalam fungsi peluang berikut:
$$\begin{aligned} f(y) &= s(y)t(\theta)e^{a(y)b(\theta)} \\ &= \exp\left[a(y)b(\theta) + c(\theta) + d(y)\right] \end{aligned} \tag{2.1}$$
dengan $a(.)$, $b(.)$, $s(.)$, $t(.)$ adalah fungsi yang diketahui, $s(y) = \exp d(y)$, $t(\theta) = \exp c(\theta)$ dan θ adalah parameter dispersi.
2. Komponen sistematik, meliputi variabel-variabel prediktor dari model untuk menjelaskan variabel-variabel yang berhubungan dalam sebuah model linear.
$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_1 x_{1i} + \cdots + \beta_1 x_{pi} \tag{2.2}$$
3. Fungsi penghubung (*link function*), yaitu fungsi yang menghubungkan rata-rata dari variabel respon (Y) dengan variabel-variabel prediktor melalui persamaan linear. Misalkan $\lambda_i = E(Y_i), i = 1,2,\dots,n$, fungsi untuk menghubungkan λ_i dengan η_i adalah $g(.)$ sehingga $g(\lambda_i) = \eta_i$. Fungsi penghubung $g(.)$ menghubungkan $E(Y_i)$ dengan variabel prediktor, yaitu:
$$g(\lambda_i) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ji}\,, \;\; i = 1,2,\dots,n \tag{2.3}$$

2.2 Regresi Poisson

Regresi Poisson didapatkan dari data yang berdistribusi Poisson, yaitu suatu distribusi untuk peristiwa yang probabilitas kejadiannya kecil, yang mana kejadiannya tergantung pada interval waktu tertentu atau di suatu daerah tertentu dengan hasil pengamatan berupa variabel diskrit dan antar variabel saling independen (Walpole, 1992). Menurut Cameron dan Trivedi (1998), suatu variabel acak Y yang bertipe diskrit akan mengikuti distribusi Poisson dengan fungsi massa peluang sebagai berikut:

$$f(y) = \frac{e^{-\lambda}\lambda^{y}}{y!}, y = 0,1,2, \dots ; \lambda > 0 \tag{2.4}$$

dan $E(Y) = \lambda, Var(Y) = \lambda$. Distribusi Poisson memiliki fungsi peluang yang dapat dinyatakan sebagai berikut:

$$\begin{aligned} f(y|\lambda) &= \frac{e^{-\lambda}\lambda^{y}}{y!} \\ &= \exp\left\{\ln\left(\frac{e^{-\lambda}\lambda^{y}}{y!}\right)\right\} \\ &= \exp\left\{\ln\left(e^{-\lambda}\lambda^{y}\right) - \ln(y!)\right\} \\ &= \exp\left\{\ln\left(\lambda^{y}e^{-\lambda}\right) - \ln(y!)\right\} \\ &= \exp\{(y \ln \lambda - \lambda) - \ln y!\} \end{aligned}$$

dengan $a(y) = y$ $\quad c(\theta) = -\lambda$

$b(\theta) = \ln \lambda$ $\quad d(y) = -\ln y!$

Distribusi Poisson merupakan anggota dari distribusi keluarga eksponensial. Hal ini dibuktikan dari fungsi massa peluang distribusi Poisson memenuhi Persamaan (2.1).

Untuk mengetahui bahwa suatu data merupakan distribusi tertentu dapat digunakan uji *Kolmogorov-Smirnov* (Ruliana, 2015). Adapun hipotesisnya adalah sebagai berikut:

H_0 : variabel respon berdistribusi Poisson

H_1 : variabel respon tidak berdistribusi Poisson

Statistik uji:

$$D_{hitung} = maks|F_n(Y) - F_0(Y)| \tag{2.5}$$

dengan:

$F_0(Y)$ = fungsi distribusi kumulatif yang dihipotesiskan

$F_n(Y)$ = fungsi distribusi kumulatif yang diamati

Kriteria pengujiannya yaitu tolak H_0 apabila $D_{hitung} > D_{tabel}\,(\alpha)$ atau nilai signifikansi $< \alpha$.

2.2.1 Overdispersi

Asumsi penting pada model regresi Poisson adalah equidispersi yaitu kesamaan varians dengan nilai rata-rata. Kondisi ini sering tidak terpenuhi pada data diskrit. Pada umumnya ditemukan data diskrit yang memiliki varians lebih besar dari nilai rata-rata (overdispersi) atau varians lebih kecil dari nilai rata-rata (underdispersi). Model regresi Poisson yang diterapkan pada data yang overdispersi mengakibatkan simpangan baku dari parameter dugaan menjadi berbias ke bawah (*underestimate*) dan signifikansi dari pengaruh variabel prediktor menjadi berbias ke atas (*overstate*).

Pemeriksaan overdispersi dapat dilakukan menggunakan nilai *deviance* dibagi dengan derajat bebas dan mempunyai nilai lebih besar dari 1, sedangkan underdispersi dideteksi menggunakan nilai *deviance* dibagi dengan derajat bebas dan mempunyai nilai kurang dari 1. Menurut Agresti (2002), nilai *deviance* dapat dinyatakan seperti pada persamaan (2.7).

$$\phi = \frac{D}{db} \tag{2.6}$$

$$D = 2\ln\left[\frac{L(y_i, y_i)}{L(\hat{\lambda}_i, y_i)}\right]$$

$$D = 2\sum_{i=1}^{n} y_i \ln\left(\frac{y_i}{\hat{\lambda}_i}\right) \tag{2.7}$$

dengan, $db = n - c$ $(n > c)$, n merupakan banyaknya pengamatan, $c = p + 1$ merupakan banyaknya parameter, dan D adalah nilai *deviance*. Jika nilai statistik uji overdispersi (ϕ) lebih dari 1 maka dapat disimpulkan data mengalami overdispersi

2.2.2 *Excess Zeros*

Salah satu permasalahan pada regresi Poisson adalah nilai nol yang berlebih (*excess zeros*). *Excess zeros* dapat dilihat pada proporsi variabel respon yang bernilai nol lebih besar dari data diskrit lainnya. Menurut Winkleman (2008), banyaknya proporsi data yang bernilai nol dari proporsi data lainnya ($> 50\%$) dapat berakibat pada ketepatan (presisi) dalam pengambilan keputusan. Selain itu regresi Poisson juga menjadi tidak tepat digunakan karena *excess zeros* merupakan salah satu penyebab terjadinya overdispersi (Hinde dan Demetrio, 1998).

2.3 Regresi Binomial Negatif

Regresi binomial negatif merupakan salah satu model terapan dari GLM. Sebagai penerapan dari GLM maka distribusi binomial negatif memiliki tiga komponen yaitu:

1. Komponen Acak

   Pada model binomial negatif variabel respon diasumsikan berdistribusi binomial negatif yang dihasilkan dari distribusi campuran Poisson gamma (Simarmata dan Ispriyanti, 2011).

   Misalkan:

   $Y \sim \text{Poisson}(\lambda_i)$

   $\lambda_i \sim \text{Gamma}(\alpha, \beta)$

   Distribusi peluang campuran Poisson-gamma dapat diperoleh dengan cara:

$$P(Y = y_i) = \int_0^\infty \text{Poisson}(Y|\lambda_i)\text{Gamma}(\lambda_i|\alpha, \beta)\; d\lambda_i$$

$$= \int_0^\infty \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda_i^{\alpha-1} e^{-\frac{\lambda_i}{\beta}}\; d\lambda_i$$

$$= \frac{\beta^{-\alpha}}{\Gamma(\alpha)y_i!} \int_0^\infty e^{-\lambda_i\left(1+\frac{1}{\beta}\right)} \lambda_i^{y_i+\alpha-1}\; d\lambda_i$$

   Misalkan $v = \left(1 + \frac{1}{\beta}\right)\lambda$ maka $dv = \left(1 + \frac{1}{\beta}\right) d\lambda$, dan untuk $\lambda = 0 \to v = 0$,

   $\lambda = \infty \to v = \infty$

$$P(Y = y_i) = \frac{\beta^{-\alpha}}{\Gamma(\alpha)y_i!} \int_0^\infty e^{-v} \left(\frac{\beta v}{1+\beta}\right)^{y_i+\alpha-1} \left(\frac{\beta v}{1+\beta}\right) dv$$

$$= \frac{\beta^{-\alpha}}{\Gamma(\alpha)y_i!} \left(\frac{\beta}{1+\beta}\right)^{y_i+\alpha} \int_0^\infty e^{-v} v^{y_i+\alpha-1}\; dv$$

$$= \frac{\beta^{-\alpha}}{\Gamma(\alpha)y_i!} \left(\frac{\beta}{1+\beta}\right)^{\alpha} \left(\frac{\beta}{1+\beta}\right)^{y_i} \Gamma(y+\alpha)$$

$$P(Y_i = y_i) = \frac{\Gamma(y+\alpha)}{\Gamma(\alpha)y_i!} \left(\frac{1}{1+\beta}\right)^{\alpha} \left(\frac{\beta}{1+\beta}\right)^{y_i}, \text{dengan y} = 0,1,2, \ldots$$

$$= \frac{\Gamma(y+\alpha)}{\Gamma(\alpha)y_i!} \left(\frac{1}{1+\beta}\right)^{\alpha} \left(1 - \frac{1}{1+\beta}\right)^{y}$$

$P(Y|\alpha,\beta)$ merupakan fungsi massa peluang binomial negatif yang dihasilkan dari distribusi campuran Poisson gamma. Nilai rata-rata dan variansi campuran Poisson gamma adalah: $E[Y] = \alpha\beta$ dan $V[Y] = \alpha\beta + \alpha\beta^2$

Untuk membentuk suatu model regresi pada distribusi binomial negatif, maka nilai parameter dari distribusi campuran Poisson gamma dinyatakan dalam bentuk $\lambda = \alpha\beta$ dan $\kappa = \frac{1}{\alpha}$ sehingga diperoleh rata-rata dan variansi dalam bentuk:

$E[Y] = \lambda$ dan $V[Y] = \lambda + \kappa\lambda^2$

Kemudian fungsi massa peluang binomial negatif menjadi:

$$f(y_i|\lambda_i,\kappa) = \frac{\Gamma\left(y_i+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y_i!}\left(\frac{1}{1+\kappa\lambda_i}\right)^{\frac{1}{\kappa}}\left(\frac{\kappa\lambda_i}{1+\kappa\lambda_i}\right)^{y_i} \text{ dengan } y_i = 0,1,2,\ldots,n \qquad (2.8)$$

Pada saat $\kappa \to 0$ maka distribusi binomial negatif memiliki varians $V(Y) \to \lambda$. Distribusi binomial negatif akan mendekati suatu distribusi Poisson yang mengasumsikan rata-rata dan varians sama yaitu $E(Y) = V(Y) = \lambda$. Fungsi distribusi keluarga eksponensial dari distribusi binomial negatif (Greene, 2008) adalah sebagai berikut:

$$\begin{aligned}
f(y_i|\lambda_i,\kappa) &= \frac{\Gamma\left(y_i+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y_i!}\left(\frac{1}{1+\kappa\lambda_i}\right)^{\frac{1}{\kappa}}\left(\frac{\kappa\lambda_i}{1+\kappa\lambda_i}\right)^{y_i} \\
&= \exp\left\{\ln\left(\frac{\Gamma\left(y_i+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y_i!}\left(\frac{1}{1+\kappa\lambda_i}\right)^{\frac{1}{\kappa}}\left(\frac{\kappa\lambda_i}{1+\kappa\lambda_i}\right)^{y_i}\right)\right\} \\
&= \exp\left\{\ln\left(\frac{\Gamma\left(y_i+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y_i!}\right) + \ln\left(\frac{1}{1+\kappa\lambda_i}\right)^{\frac{1}{\kappa}} + \ln\left(\frac{\kappa\lambda_i}{1+\kappa\lambda_i}\right)^{y_i}\right\} \\
&= \exp\left\{y_i \ln\left(\frac{\kappa\lambda}{1+\kappa\lambda}\right) + \frac{1}{\kappa}\ln\left(\frac{1}{1+\kappa\lambda}\right) + \ln\left(\frac{\Gamma\left(y_i+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y_i!}\right)\right\}
\end{aligned}$$

dengan:

$a(y) = y$

$c(\theta) = -\frac{1}{\kappa}\ln\left(\frac{1}{1+\kappa\lambda}\right)$

$b(\theta) = \ln\left(\frac{\kappa\lambda}{1+\kappa\lambda}\right)$

$d(y) = \ln\left(\frac{\Gamma\left(y_i+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y_i!}\right)$

2. Komponen Sistematik

   Kontribusi variabel bebas dalam model regresi binomial negatif dinyatakan dalam bentuk kombinasi linier antara parameter $\boldsymbol{\eta}$ dengan parameter regresi yang diestimasi yaitu:

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_1 x_{1i} + \cdots + \beta_1 x_{pi} \tag{2.9}$$

   atau dituliskan dalam bentuk matriks:

$$\boldsymbol{\eta} = \boldsymbol{X\beta}$$

   dengan $\boldsymbol{\eta}$ adalah vektor berukuran $n \times 1$ dari observasi, $\boldsymbol{X}$ adalah matriks berukuran $n \times c$ yang elemennya terdiri dari variabel prediktor, $\boldsymbol{\beta}$ adalah vektor berukuran $c \times 1$ yang elemennya terdiri dari parameter regresi, dengan $c = p + 1$.

3. Fungsi Penghubung

   Nilai rata-rata dari variabel respon adalah diskrit dan bernilai positif. Maka untuk mentransformasikan nilai η_i (bilangan riil) ke rentang yang sesuai dengan rentang pada variabel respon diperlukan suatu fungsi penghubung $g(.)$ yaitu:

$$g(\lambda_i) = \ln(\lambda_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}$$

$$\lambda_i = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) \tag{2.10}$$

2.4 Regresi *Zero Inflated Negative Binomial*

Model regresi *Zero Inflated Negative Binomial* (ZINB) merupakan model yang dibentuk dari distribusi campuran Poisson gamma. Jika y_i adalah variabel acak dengan $i = 1,2, \ldots, n$ nilai nol pada observasi diduga muncul dalam dua cara yang sesuai untuk keadaan (*state*) yang terpisah. Menurut Garay dkk. (2011), regresi *Zero Inflated Negative Binomial* (ZINB) dengan keadaan pertama disebut *zero state* terjadi dengan probabilitas π_i dan menghasilkan hanya observasi bernilai nol, sementara keadaan kedua disebut *negative binomial state* terjadi dengan probabilitas $(1 - \pi_i)$ dan berdistribusi binomial negatif. Fungsi peluang model regresi *Zero Inflated Negative Binomial* (ZINB) dapat dinyatakan seperti pada persamaan (2.11).

$$P(Y_i = y_i) = \begin{cases} \pi_i + (1-\pi_i)\left(\frac{1}{1+\kappa\lambda_i}\right)^{\frac{1}{\kappa}}, & \text{untuk } y_i = 0 \\ (1-\pi_i)\frac{\Gamma\left(y_i+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y_i!}\left(\frac{1}{1+\kappa\lambda_i}\right)^{\frac{1}{\kappa}}\left(\frac{\kappa\lambda_i}{1+\kappa\lambda_i}\right)^{y_i}, & \text{untuk } y_i > 0 \end{cases} \tag{2.11}$$

dengan $0 \leq \pi_i \leq 1, \lambda_i \geq 0$, κ adalah parameter dispersi dengan $\frac{1}{\kappa} > 0$ dan $\Gamma(.)$ adalah fungsi gamma. Ketika $\pi_i = 0$, variabel acak y_i memiliki distribusi binomial negatif dengan rata-rata λ_i dan parameter dispersi κ, sehingga $Y_i \sim NB\ (\lambda_i, \kappa)$.

Pada GLM, fungsi penghubung logit digunakan Ketika parameter model regresi bernilai 0 dan 1 dan fungsi penghubung ln digunakan jika parameter diharapkan bernilai positif. Oleh karena itu fungsi penghubung yang tepat digunakan pada regresi ZINB adalah fungsi penghubung logit untuk parameter π_i dan fungsi penghubung ln untuk parameter λ_i.

Model regresi *Zero Inflated Negative Binomial* (ZINB) dapat dinyatakan seperti pada persamaan (2.12) dan (2.13).

Model untuk *negative binomial state* $\hat{\lambda}_i$

$$\ln \hat{\lambda}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij},\ i = 1,2,\dots,n \text{ dan } j = 1,2,\dots,p \tag{2.12}$$

Model untuk *zero state* π_i

$$\text{logit } \hat{\pi}_i = \ln\left(\frac{\pi}{1-\pi}\right) = \hat{\gamma}_0 + \sum_{j=i}^{p} \hat{\gamma}_j x_{ij},\ \ i = 1,2,\dots,n \text{ dan } j = 1,2,\dots,p \tag{2.13}$$

dengan:

p : jumlah variabel prediktor

n : jumlah pengamatan

$\hat{\boldsymbol{\beta}}$: parameter model regresi ZINB yang diestimasi

$\hat{\boldsymbol{\gamma}}$: parameter model regresi ZINB yang diestimasi

2.5 Metode *Maximum Likelihood Estimation*

Metode *Maximum Likelihood Estimation* (MLE) dapat digunakan untuk mengestimasi parameter suatu model yang diketahui fungsi peluangnya. Misalkan $Y_1, Y_2, \dots, Y_n$ adalah sampel acak dari populasi dengan fungsi peluang $f(y|\theta)$ yang bergantung pada $\boldsymbol{\theta} = \lambda, \boldsymbol{\beta}, \boldsymbol{\gamma}, \kappa$ dengan $\boldsymbol{\theta}$ adalah parameter yang tidak diketahui. Karena $Y_1, Y_2, \dots, Y_n$ saling bebas, maka fungsi peluang bersama dari $Y_1, Y_2, \dots, Y_n$ dapat dinyatakan sebagai berikut:

$$f(y_1, y_2, \dots, y_n|\boldsymbol{\theta}) = f(y_1|\boldsymbol{\theta}) \cdot f(y_2|\boldsymbol{\theta}) \cdot \dots \cdot f(y_n|\boldsymbol{\theta})$$

Fungsi *likelihood* didefinisikan sebagai fungsi peluang bersama dari $Y_1, Y_2, \dots, Y_n$ yang dianggap sebagai fungsi dari $\boldsymbol{\theta}$ yang dituliskan sebagai $L(\boldsymbol{\theta}|y)$, yaitu:

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

$$L(\boldsymbol{\theta}|y) = f(y_1|\boldsymbol{\theta}) \cdot f(y_2|\boldsymbol{\theta}) \cdot \dots \cdot f(y_n|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta}) \qquad (2.14)$$

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

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

$$l(\boldsymbol{\theta}|y) = \ln\left[\prod_{i=1}^{n} f(y_i|\boldsymbol{\theta})\right] = \sum_{i=1}^{n} \ln f(y_i|\boldsymbol{\theta}) \qquad (2.15)$$

Untuk memperoleh nilai $\boldsymbol{\theta}$ yang memaksimumkan fungsi $l(\boldsymbol{\theta}|y)$, maka $l(\boldsymbol{\theta}|y)$ diturunkan terhadap $\boldsymbol{\theta}$ dan kemudian menyamakannya dengan nol seperti pada persamaan (2.18) berikut (Hogg dkk., 2013).

$$l'(\boldsymbol{\theta}|y) = \frac{\partial l(\boldsymbol{\theta}|y)}{\partial \boldsymbol{\theta}} = 0 \qquad (2.16)$$

2.6 Algoritma *Expectation Maximization*

Algoritma *Expectation Maximization* (EM) pertama kali diperkenalkan oleh Dempster, Laird, dan Rubin pada tahun 1977. Algoritma EM digunakan untuk mengestimasi suatu parameter dengan memaksimumkan fungsi *likelihood* yang mengandung data hilang atau tidak lengkap (*incomplete data*). Algoritma EM terdiri dari dua tahap, yaitu tahap ekspektasi dan maksimalisasi. Misalnya, diasumsikan terdapat data pengamatan y berdistribusi tertentu yang mengandung data hilang z. Oleh karena itu, dibentuk distribusi gabungan antara y dan z, yaitu:

$$f(y, z|\boldsymbol{\theta}) = f(z) \cdot f(y/z) \qquad (2.17)$$

Tahap ekspektasi (*E-Step*) pada algoritma EM dilakukan dengan menghitung ekspektasi dari fungsi logaritma natural *likelihood* dari data hilang berdasarkan data pengamatan yang ada (tidak hilang), yang digunakan untuk mengganti keberadaan data yang dianggap hilang. Oleh karena itu, fungsi Q didefinisikan sebagai berikut:

$$Q\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(\mathbf{0})};y\right)=E\left[\ln L(\boldsymbol{\theta}^{(\mathbf{0})}|y,z)\right] \tag{2.18}$$

Pada tahap maksimalisasi (*M-Step*) dilakukan dengan mencari nilai estimator yang dapat memaksimumkan fungsi Q yang telah didefinisikan pada tahap ekspektasi. Metode yang dapat digunakan untuk memaksimumkan fungsi Q adalah metode *Newton-Raphson* yang dilakukan secara numerik. Langkah E dan M dilakukan secara iteratif sampai didapatkan estimator $\widehat{\boldsymbol{\theta}}$ yang konvergen, yaitu saat $\left\|\boldsymbol{\theta}^{(\boldsymbol{r}+\mathbf{1})}-\boldsymbol{\theta}^{\boldsymbol{r}}\right\|\leq\boldsymbol{\varepsilon}$, dengan $\boldsymbol{\varepsilon}$ adalah nilai galat yang sangat kecil (Purba, 2018).

2.7 Uji Serentak Parameter Model

Uji serentak parameter model regresi digunakan untuk mengetahui bahwa seluruh variabel prediktor tidak berpengaruh terhadap variabel respon atau terdapat minimal salah satu variabel prediktor yang berpengaruh terhadap variabel respon. Untuk melakukan uji serentak parameter model dapat dilakukan dengan statistik uji rasio *likelihood*.

Hipotesis yang digunakan adalah sebagai berikut:

H_0: $\beta_1=\beta_2=\cdots=\beta_p=\gamma_1=\gamma_2=\cdots=\gamma_p=0$ (variabel prediktor secara bersama-sama tidak mempunyai pengaruh terhadap variabel respon)

H_1: minimal ada satu $\beta_j\neq 0$ atau $\gamma_j\neq 0, j=1,2,\ldots,p$ (variabel prediktor secara bersama-sama mempunyai pengaruh terhadap variabel respon)

Statistik uji:

$$G=-2\ln\left[\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right]=-2\left[\ln L(\widehat{\omega})-\ln L\left(\widehat{\Omega}\right)\right] \tag{2.19}$$

dengan:

$L(\widehat{\omega})$ = nilai fungsi *likelihood* tanpa variabel prediktor.

$L(\widehat{\Omega})$ = nilai fungsi *likelihood* dengan variabel prediktor.

Jika $G_{hitung}>\chi^2_{\alpha;p}$ dengan p jumlah parameter, maka H_0 ditolak yang artinya variabel prediktor secara bersama-sama memiliki pengaruh terhadap variabel respon (Kurniawan, 2017).

2.8 Uji Parsial Parameter Model

Uji parsial digunakan untuk mengetahui variabel-variabel prediktor yang berpengaruh terhadap variabel respon. Statistik uji yang digunakan untuk uji parsial, yaitu uji Wald (Kurniawan, 2017).

2.8.1 Uji Parsial Parameter $\boldsymbol{\beta}$

Parameter yang diuji pada pengujian ini mencakup seluruh parameter $\boldsymbol{\beta}$ secara parsial untuk mengetahui parameter mana saja yang memberikan pengaruh signifikan terhadap model. Hipotesis yang digunakan adalah sebagai berikut.

H_0 : $\beta_j = 0,\ j = 1,2,\dots,p$ (variabel prediktor tidak mempunyai pengaruh terhadap variabel respon)

H_1 : minimal ada satu $\beta_j \neq 0,\ j = 1,2,\dots,p$ (variabel prediktor mempunyai pengaruh terhadap variabel respon)

Statistik uji:

$$W(\hat{\beta}_j) = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2 \qquad (2.20)$$

dengan:

$\hat{\beta}_j$ = estimasi parameter β_j

$SE(\hat{\beta}_j)$ = standar error dari $\hat{\beta}_j$

Jika $W(\hat{\beta}_j) > \chi^2_{\alpha;p}$, maka H_0 ditolak artinya variabel prediktor memiliki pengaruh yang signifikan terhadap variabel respon Y (Kurniawan, 2017).

2.8.2 Uji Parsial Parameter $\boldsymbol{\gamma}$

Parameter yang diuji pada pengujian ini mencakup seluruh parameter $\boldsymbol{\gamma}$ secara parsial untuk mengetahui parameter mana saja yang memberikan pengaruh signifikan terhadap model. Hipotesis yang digunakan adalah sebagai berikut.

H_0 : $\gamma_j = 0,\quad j = 1,2,\dots,p$ (variabel prediktor tidak mempunyai pengaruh terhadap variabel respon)

H_1 : minimal ada satu $\gamma_j \neq 0,\quad j = 1,2,\dots,p$ (variabel prediktor mempunyai pengaruh terhadap variabel respon)

Statistik uji:

$$W(\hat{\gamma}_j) = \left(\frac{\hat{\gamma}_j}{SE(\hat{\gamma}_j)}\right)^2 \qquad (2.21)$$

dengan:

$\hat{\gamma}_j$ = estimasi parameter γ_j

$SE(\hat{\gamma}_j)$ = standar error dari $\hat{\gamma}_j$

Jika $W(\hat{\gamma}_j) > \chi^2_{\alpha;p}$, maka H_0 ditolak artinya variabel prediktor memiliki pengaruh yang signifikan terhadap variabel respon Y (Kurniawan, 2017).

2.9 Pemilihan Model Terbaik

Salah satu kriteria untuk menentukan model terbaik adalah *Akaike Information Criterion* (AIC). Dengan kriteria AIC, model terbaik dipilih dengan mempertimbangkan jumlah parameter dalam model. Kriteria AIC mampu menunjukkan seberapa tepat model tersebut dengan data yang dimiliki secara mutlak. Kriteria AIC didefinisikan sebagai berikut.

$$AIC = 2c - 2\ln L(\hat{\boldsymbol{\theta}}) \tag{2.22}$$

dengan $L(\hat{\boldsymbol{\theta}})$ adalah nilai fungsi *likelihood* dan $c = p + 1$ adalah banyaknya parameter.

Besarnya nilai AIC sejalan dengan nilai *deviance* dari model. Nilai *deviance* akan semakin kecil apabila rasio antara fungsi *likelihood* di bawah H_0 dengan fungsi *likelihood* di bawah populasi semakin bersar. Hal ini mengindikasi bahwa parameter yang diuji semakin mendekati nilai parameter populasi yang sebenarnya sehingga estimasi model semakin baik. Oleh karena itu, model terbaik adalah dengan AIC terkecil sekaligus dengan *deviance* terkecil pula (Ilmi, 2015).

2.10 Jumlah Kematian Neonatal

Jumlah kematian neonatal adalah banyaknya kematian bayi yang terjadi saat bayi baru lahir hingga bayi berusia di bawah 28 hari, yang mana data tersebut merupakan data jumlahan sehingga tergolong ke dalam data diskrit. Kematian neonatal umumnya disebabkan oleh faktor yang dibawa anak sejak lahir yang diperoleh dari orang tua pada saat konsepsi atau selama kehamilan (Dinkes, 2019).

Neonatal atau bayi baru lahir digolongkan menjadi 2 (dua) yaitu bayi baru lahir (BBL) normal dan bayi baru lahir (BBL) resiko tinggi. Bayi baru lahir sangat rentan terhadap infeksi baik selama proses persalinan ataupun beberapa saat setetah lahir. Upaya kesehatan yang dilakukan untuk mengurangi resiko tersebut yaitu dengan pertolongan persalinan oleh tenaga kesehatan di fasilitas pelayanan kesehatan. Hal ini menjadi penting dikarenakan 50% kematian neonatal memberi kontribusi terhadap kematian bayi.

Indikator yang menggambarkan upaya kesehatan neonatal yaitu cakupan kunjungan neonatal pertama (KN1) dan KN lengkap. Kunjungan neonatal merupakan salah satu intervensi untuk menurunkan kematian bayi baru lahir dengan melakukan Kunjungan Neonatal (KN) selama 3 (tiga) kali kunjungan yaitu Kunjungan Neonatal I (KN1) pada 6 jam sampai dengan 48 jam setelah lahir, Kunjungan Neonatal II (KN2) pada hari ke 3 sampai dengan 7 hari, dan Kunjungan Neonatal III (KN3) pada hari ke 8 sampai dengan 28 hari (Kemenkes RI, 2010).

Berbagai kegiatan yang dilaksanakan Dinas Kesehatan Kota Makassar dalam upaya penurunan Jumlah Kematian Neonatal diantaranya kampanye anak sehat, Sosialisasi Program Perencanaan Persalinan dan Pencegahan Komplikasi (P4K) bagi kader kesehatan dan tokoh masyarakat, sosialisasi penggunaan buku Kesehatan Ibu dan Anak (KIA), Gerakan Sayang Ibu (GSI), sosialisasi dan pembinaan kelas ibu hamil dan ibu balita (Dinkes, 2017).