

**PEMODELAN REGRESI BIVARIATE ZERO-
INFLATED POISSON UNTUK MENGATASI
OVERDISPERSI**

SKRIPSI



ANDI ISNA YUNITA

H121 16 310

PROGRAM STUDI STATISTIKA

DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

SEPTEMBER 2020



**PEMODELAN REGRESI BIVARIATE ZERO-
INFLATED POISSON UNTUK MENGATASI
OVERDISPERSI**

SKRIPSI

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

ANDI ISNA YUNITA

H12116310

**PROGRAM STUDI STATISTIKA
DEPARTEMEN STATISTIKA**

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

SEPTEMBER 2020



Optimization Software:
www.balesio.com

LEMBAR PERNYATAAN KEOTENTIKAN

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

**Pemodelan Regresi Bivariate Zero-Inflated Poisson
Untuk Mengatasi Overdispersi**

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

Makassar, 28 September 2020



Andi Isna Yunita
NIM H12116310



PEMODELAN REGRESI BIVARIATE ZERO- INFLATED POISSON UNTUK MENGATASI OVERDISPERSI

Disetujui oleh:

Pembimbing Utama



Andi Kresna Jaya, S.Si., M.Si
NIP. 19731228 200003 1001

Pembimbing Pertama



Dr. Dr. Georgina Maria Tinungki, M.Si
NIP. 19620926 198702 2001



Optimization Software:
www.balesio.com

Pada Tanggal : 28 September 2020

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Andi Isna Yunita

NIM : H12116310

Program Studi : Statistika

Judul Skripsi : Pemodelan Regresi Bivariate Zero-Inflated Poisson Untuk Mengatasi 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

1. Ketua : Andi Kresna Jaya, S.Si., M.Si

2. Sekretaris : Dr. Dr. Georgina Maria Tinungki, M.Si

3. Anggota : Dr. Erna Tri Herdiani, S.Si., M.Si

4. Anggota : Sitti Sahriman, S.Si., M.Si

Tanda Tangan



(.....)

(.....)

(.....)

kan di : Makassar

: 28 September 2020



KATA PENGANTAR

Puji syukur kepada **Allah SWT** atas segala limpahan rahmat, nikmat, dan hidayah-Nya, serta shalawat dan salam semoga selalu dilimpahkan kepada Nabi yang paling dimuliakan, pemimpin orang-orang bertakwa, **Muhammad bin Abdullah** dan kepada para keluarga serta sahabat yang berjuang bersama beliau. *Alhamdulillah*, berkat pertolongan Allah akhirnya tugas akhir dengan judul “**Pemodelan Regresi Bivariate Zero-Inflated Poisson Untuk Mengatasi Overdispersi**” yang disusun sebagai salah satu syarat akademik untuk meraih gelar Sarjana Sains pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulisan tugas akhir ini tentunya tidak lepas dari bantuan berbagai pihak baik moril maupun materil. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang tak terhingga untuk Ayahanda **Zainal Abidin** dan Ibunda tercinta **Andi Tasmawati** yang tak kenal lelah mendoakan, memberikan dukungan, dan selalu melimpahkan cinta dan kasih sayang kepada penulis sehingga mereka menjadi motivasi terbesar penulis dalam menyelesaikan tugas akhir ini. Untuk adik-adikku tercinta **Andi Amir Achmad Riyadi**, **Andi Novita Sri Ramadhani**, **Andi Devi Permatasari**, dan **Andi Desi Ratnasari** serta sepupu tercinta yang kuanggap kakak sendiri **Andi Nurhidayah** terima kasih telah memberikan semangat, motivasi, dukungan, dan doa yang diberikan kepada penulis. Dan tak terlupakan untuk seluruh keluarga besar, penulis ucapkan banyak terima kasih atas doa dan dukungan moril maupun materil selama penulis mengemban ilmu di Universitas Hasanuddin.

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

1. Ibu **Prof. Dr. Dwia Aries Tina Pulubuhu, MA**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.

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 sekaligus pembimbing akademik yang telah meluangkan waktu dan pemikirannya untuk memberikan masukan, arahan, dan motivasi kepada penulis mulai dari awal hingga selesainya penulisan tugas akhir ini dan Ibu **Dr. Dr. Georgina Maria Tinungki, 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** dan Ibu **Sitti Sahrinan, S.Si, M.Si** selaku anggota Tim Penguji. Terima kasih telah meluangkan waktu dan telah memberikan kritikan dan saran yang membangun dalam penyempurnaan penyusunan tugas akhir ini.
6. Keluarga besar **ALGORITMA 2016** yang telah penulis anggap seperti saudara sendiri yang selalu memberikan semangat dan motivasi selama di kampus. Banyak pengalaman dan pelajaran berharga yang penulis dapatkan dari kalian.
7. Kakak-kakak, teman-teman, dan adik-adik anggota **Keluarga Mahasiswa FMIPA Unhas** terkhusus anggota keluarga **Himatika FMIPA Unhas** dan **Himastat FMIPA Unhas**, terima kasih atas ilmu yang tidak didapatkan di bangku perkuliahan dan banyak belajar betapa pentingnya sebuah proses.
8. Teman-teman **Statistika 2016**, terkhusus spesial untuk **Sobat Minor**, yaitu **Ratmila, Ririn Arianti, Alimatun Najiha, Ainun Utari, Annisa AK, Asnidar, Nur Alfianingsih Raja, Nurazrawiyah Amin, Reza Ayu Lestari, Andi Tenri Riski Amalia, Risma Sari, dan Nurul Afdhaliah**. Terima kasih atas motivasi, segala cerita suka dan dukanya, serta selalu menemani penulis sehingga masa perkuliahan dapat dijalani lebih bermakna dan berwarna.
9. Spesial untuk sahabat-sahabat tercintaku **Nurul Azizah Mansyur, S.Ked; Wahidah Amir, A.Md; Nurqanaah M., S.Ked; Ayu Khumaera, S.Ak; Dwi Sinta Saputri, S.Tr.T dan Nurhashunatil Mar'ah, S.KH** yang telah mendengarkan segala curhatan, memberikan nasihat, dan memotivasi penulis hingga penulis bisa mendapatkan lebih banyak pelajaran hidup.



10. Seluruh teman-teman **KKN Unhas Gelombang 102 Kecamatan Sinjai Timur**, terkhusus kepada teman posko **Desa Salohe** yang telah menjadi sahabat sekaligus keluarga selama di lokasi KKN hingga saat ini.
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 SWT**.

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, 28 September 2020



Penulis



PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama : Andi Isna Yunita
NIM : H12116310
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 Bivariate Zero-Inflated Poisson Untuk Mengatasi 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 28 September 2020

Yang menyatakan



Andi Isna Yunita

ABSTRAK

Regresi *Poisson* merupakan model regresi non-linier dengan variabel responnya berupa data diskrit dan berdistribusi *Poisson*. Salah satu asumsi yang harus dipenuhi dalam regresi *Poisson* adalah asumsi *equidispersi*, yaitu keadaan dengan mean dan variansi dari variabel responnya bernilai sama. Namun dalam aplikasinya asumsi tersebut kadang dilanggar karena nilai variansinya lebih besar daripada meannya yang disebut *overdispersi*. Salah satu penyebab terjadinya *overdispersi* adalah lebih banyaknya pengamatan bernilai nol sehingga dapat digunakan model regresi *Zero-Inflated Poisson* (ZIP). Sedangkan untuk memodelkan sepasang data diskrit yang saling berkorelasi dan *overdispersi*, maka dapat digunakan model regresi *Bivariate Zero-Inflated Poisson* (BZIP). Model regresi BZIP merupakan model dengan variabel respon yang berdistribusi campuran antara distribusi *Bivariate Poisson* dan probabilitas titik di (0,0). Parameter model regresi BZIP diestimasi menggunakan metode *maximum likelihood estimation* (MLE) dengan algoritma *expectation maximization* (EM). Penelitian ini diaplikasikan pada data jumlah kematian ibu dan bayi di Kota Makassar tahun 2017. Hasil yang diperoleh adalah nilai AIC dari model regresi BZIP, yaitu 170,976 lebih kecil dibandingkan dengan model regresi *Bivariate Poisson*, yaitu 198,120. Hal ini menunjukkan bahwa model regresi BZIP lebih baik digunakan pada data yang mengalami *overdispersi*.

Kata Kunci: Kematian Ibu, Kematian Bayi, *Overdispersi*, Regresi BZIP, MLE



ABSTRACT

Poisson regression is a non-linear regression model with response variable in form of discrete data and Poisson distribution. One of assumptions that must be met in Poisson regression is assumption of equidispersion, which is a condition where mean and variance of response variable are the same. However in its application, this assumption is sometimes violated because the variance is greater than the mean which is called overdispersion. One of the causes overdispersion is more zero-value observations so the Zero-Inflated Poisson (ZIP) regression model can be used. As for modeling a pair of discrete data is correlated and overdispersi, it can be used the Bivariate Zero-Inflated Poisson (BZIP) regression model. The BZIP regression model is a model with response variables with mixed distributions between Bivariate Poisson distribution and a point probability at (0,0). Parameters of the BZIP regression model are estimated using maximum likelihood estimation (MLE) with expectation maximization (EM) algorithm. This research was applied to data on number of maternal and infant mortality in the city of Makassar in 2017. The result obtained is the AIC value of the BZIP regression model is 170,976 smaller than the Bivariate Poisson regression model is 198,120. This shows that the BZIP regression model is better used for data with overdispersion.

Keywords: Maternal Mortality, Infant Mortality, Overdispersion, BZIP Regression, MLE



DAFTAR ISI

HALAMAN JUDUL..... i

HALAMAN PERNYATAAN KEOTENTIKAN..... ii

HALAMAN PERSETUJUAN PEMBIMBING iii

HALAMAN PENGESAHAN..... iv

KATA PENGANTARv

PERSETUJUAN PUBLIKASI KARYA ILMIAH..... viii

ABSTRAK ix

DAFTAR ISI..... xi

DAFTAR TABEL..... xiii

DAFTAR LAMPIRAN..... xiv

BAB I PENDAHULUAN.....1

 1.1 Latar Belakang1

 1.2 Rumusan Masalah.....3

 1.3 Batasan Masalah3

 1.4 Tujuan Penelitian4

 1.5 Manfaat Penelitian4

BAB II TINJAUAN PUSTAKA.....5

 2.1 Analisis Regresi5

 2.2 Regresi Poisson.....6

 2.3 Regresi Bivariate Poisson8

 2.4 Equidispersi.....10

 2.5 Regresi Zero-Inflated Poisson.....10

 2.6 Regresi Bivariate Zero-Inflated Poisson.....12

 2.7 Metode Maximum Likelihood Estimation.....13

 2.8 Algoritma Expectation Maximization.....14

 2.9 Uji Serentak Parameter Model.....14

 2.10 Uji Parsial Parameter Model.....15

 2.10.1 Uji Parsial Parameter β 15

 2.10.2 Uji Parsial Parameter γ 16

 Pemilihan Model Terbaik16

 Kematian Ibu dan Bayi17



BAB III METODOLOGI PENELITIAN20

3.1 Sumber Data.....20

3.2 Identifikasi Variabel.....20

3.3 Metode Analisis Data.....21

BAB IV HASIL DAN PEMBAHASAN25

4.1 Uji Asumsi Model Regresi Bivariate Zero-Inflated Poisson25

4.1.1 Uji Kecocokan Distribusi Poisson25

4.1.2 Uji Overdispersi26

4.1.3 Uji Korelasi26

4.2 Estimasi Parameter Regresi Bivariate Zero-Inflated Poisson27

4.3 Pemodelan Regresi Bivariate Zero-Inflated Poisson pada Data Kematian Ibu dan Kematian Bayi di Kota Makassar Tahun 201737

4.3.1 Uji Serentak Parameter Model38

4.3.2 Uji Parsial Parameter β 40

4.3.3 Uji Parsial Parameter γ 41

4.4 Pemilihan Model Terbaik43

BAB V PENUTUP.....44

5.1 Kesimpulan44

5.2 Saran44

DAFTAR PUSTAKA45



DAFTAR TABEL

Tabel 4.1 Uji Kecocokan Distribusi Poisson.....	25
Tabel 4.2 Uji Overdispersi	26
Tabel 4.3 Uji Korelasi Pearson.....	26
Tabel 4.4 Estimasi Parameter λ_0	37
Tabel 4.5 Estimasi Parameter β_1, β_2 , dan γ	38
Tabel 4.6 Uji Parsial Parameter β_1 dan β_2	41
Tabel 4.7 Uji Parsial Parameter γ	42
Tabel 4.8 Nilai AIC model regresi <i>Bivariate</i> Poisson dan BZIP	43



DAFTAR LAMPIRAN

Lampiran 1.	Data Variabel Respon, Jumlah Kematian Ibu dan Bayi di Puskesmas Kota Makassar Provinsi Sulawesi Selatan Tahun 2017.....	48
Lampiran 2.	Data Variabel Prediktor, Faktor-faktor yang Mempengaruhi Jumlah Kematian Ibu dan Bayi dalam Persen (%).....	50
Lampiran 3.	Output SPSS untuk Uji Kecocokan Distribusi Poisson.....	52
Lampiran 4.	Output SPSS untuk Uji Overdispersi.....	53
Lampiran 5.	Output SPSS untuk Uji Korelasi Pearson.....	54
Lampiran 6.	Sintaks MATLAB untuk Estimasi Parameter.....	55
Lampiran 7.	Sintaks MATLAB untuk Uji Serentak Parameter Model.....	60
Lampiran 8.	Sintaks MATLAB untuk Uji Parsial Parameter Model.....	61
Lampiran 9.	Sintaks MATLAB untuk Nilai AIC Model Regresi <i>Bivariate</i> Poisson dan BZIP	62



BAB I

PENDAHULUAN

1.1 Latar Belakang

Angka kematian ibu dan bayi merupakan salah satu indikator yang paling menonjol untuk menilai derajat kesehatan masyarakat di suatu daerah. Namun, sampai saat ini angka kematian ibu dan bayi di Sulawesi Selatan masih tergolong tinggi. Bahkan Sulawesi Selatan termasuk provinsi dengan jumlah kematian ibu dan bayi terbanyak di Indonesia sehingga masalah Kesehatan Ibu dan Anak (KIA) masih menjadi masalah yang perlu diperhatikan. Dinas Kesehatan Sulawesi Selatan menyatakan bahwa setiap minggunya, 2 ibu dan 16 bayi baru lahir meninggal. Sepanjang tahun 2017, terdapat 115 kematian ibu dan 1.059 kematian bayi di Sulawesi Selatan. Kota Makassar menjadi salah satu dari 5 kabupaten/kota di Sulawesi Selatan dengan tingkat kematian ibu dan bayi terbanyak (Dinkes, 2018).

Menurut *World Health Organization* (WHO), kematian ibu dan bayi merupakan dua hal yang saling berkaitan karena selama masa kandungan, gizi yang diperoleh janin disalurkan dari tubuh ibu melalui plasenta sehingga kondisi ibu selama masa kehamilan akan berpengaruh pada janin dan bayi yang dilahirkannya kelak. Peran ibu juga sangat berpengaruh dalam merawat bayi mulai saat ia dilahirkan hingga berumur satu tahun (Arkandi & Winahju, 2015). Adapun upaya yang dilakukan oleh pemerintah dalam mengantisipasi angka kematian ibu dan bayi, antara lain melalui peningkatan pelayanan kesehatan ibu hamil, bersalin, dan nifas serta peningkatan pelayanan kesehatan bayi (Dinkes, 2018). Oleh karena itu, agar upaya tersebut tepat guna, maka diperlukan adanya penelitian tentang jumlah kematian ibu dan bayi dengan melibatkan faktor-faktor yang mempengaruhi keduanya secara bersamaan.

Jumlah kematian ibu dan bayi merupakan data hitung yang mengikuti distribusi *Poisson* dan mempunyai keterkaitan atau korelasi satu sama lain. Regresi *Poisson* merupakan metode regresi yang digunakan untuk memodelkan data hitung (diskrit) yang memiliki korelasi (Karlis & Ntzoufras, 2005). Penelitian sebelumnya dilakukan oleh Arkandi dan Winahju (2015) menggunakan



model regresi *Bivariate Poisson* pada data kematian ibu dan bayi yang menghasilkan bahwa model tersebut baik untuk memodelkan sepasang data hitung (diskrit) yang saling berkorelasi. Namun, model yang dihasilkan memiliki nilai variansi yang lebih besar daripada nilai meannya atau data mengalami *overdispersi*. Dengan demikian, model yang diperoleh tidak memenuhi asumsi distribusi *Poisson* dengan nilai mean dan variansi yang sama.

Penelitian lainnya yang telah menggunakan model regresi *Bivariate Poisson* dilakukan oleh Ilmi (2015) pada data jumlah kasus malaria dan filariasis yang menghasilkan model regresi dengan tiga parameter, yaitu λ_1 sebagai kasus malaria, λ_2 sebagai kasus filariasis, dan λ_0 sebagai kasus malaria dan filariasis. Model regresi tersebut dibangun dengan tiga buah nilai λ_0 yang berbeda, yaitu λ_0 sama dengan nol, λ_0 adalah suatu konstanta, dan λ_0 adalah suatu persamaan. Selain itu, Armawati (2016) juga memodelkan data kasus HIV dan AIDS dengan tiga buah nilai λ_0 yang berbeda. Sedangkan penelitian yang dilakukan oleh Kurniawan (2018) hanya memodelkan data jumlah kematian ibu dan bayi dengan λ_0 adalah suatu konstanta. Penelitian-penelitian tersebut menghasilkan model yang baik untuk memodelkan data hitung yang berpasangan namun tidak dapat mengatasi masalah *overdispersi*. Oleh karena itu, model regresi *Bivariate Zero-Inflated Poisson* (BZIP) digunakan untuk mengatasi masalah respon bivariat yang bernilai nol lebih banyak dan sekaligus mengatasi *overdispersi* secara bersamaan.

Model regresi BZIP merupakan model dengan variabel respon yang berdistribusi campuran antara distribusi *Bivariate Poisson* dan probabilitas titik di (0,0) (Wang dkk., 2003). Untuk mengestimasi parameter, metode *maximum likelihood estimation* (MLE) dapat digunakan pada model regresi yang datanya mengikuti distribusi tertentu. Metode MLE dilakukan dengan memaksimumkan fungsi *likelihood*. Namun pada umumnya, maksimum suatu fungsi *likelihood* tidak bisa diselesaikan secara analitik. Oleh karena jika diperoleh bentuk implisit atau non-linier, maka dapat diselesaikan dengan algoritma *Newton-Raphson* (NR), *scoring*, atau *Expectation Maximization* (EM) untuk mendapatkan solusi terbaiknya (Purba, 2018).



Banyaknya pengamatan bernilai nol merupakan salah satu penyebab terjadinya *overdispersi* pada data model regresi BZIP sehingga algoritma EM cocok digunakan untuk mengestimasi parameter yang menyertakan pengamatan yang bernilai nol (Kusuma dkk., 2013). Algoritma EM terdiri dari dua tahap, yaitu tahap *Expectation (E-Step)* untuk mencari nilai ekspektasi dari fungsi *likelihood* dan tahap *Maximization (M-Step)* untuk memaksimumkan fungsi yang telah didefinisikan pada tahap ekspektasi sehingga didapatkan estimator parameter yang konvergen. Algoritma EM juga lebih mudah diterapkan ketika masalah optimasi memiliki banyak parameter dibandingkan dengan metode iterasi yang lainnya (Purba, 2018).

Berdasarkan uraian tersebut, peneliti tertarik untuk membahas tentang pemodelan regresi BZIP untuk mengatasi *overdispersi* pada data kematian ibu dan bayi di Kota Makassar tahun 2017. Namun, peneliti membatasi model BZIP dengan λ_0 adalah suatu konstanta.

1.2 Rumusan Masalah

Permasalahan yang akan dikaji dalam penelitian ini adalah sebagai berikut.

1. Bagaimana mengestimasi parameter model regresi BZIP dengan menggunakan metode MLE?
2. Bagaimana memodelkan data jumlah kematian ibu dan kematian bayi di Kota Makassar tahun 2017 yang mengalami *overdispersi* dengan menggunakan model regresi BZIP?

1.3 Batasan Masalah

Penelitian ini dibatasi pada permasalahan sebagai berikut.

1. Model regresi yang digunakan adalah model regresi BZIP dengan λ_0 adalah suatu konstanta.
2. Metode estimasi parameter model BZIP yang digunakan adalah metode MLE dengan algoritma EM.



1.4 Tujuan Penelitian

Tujuan dalam penelitian ini adalah sebagai berikut.

1. Memperoleh estimasi parameter model regresi BZIP dengan menggunakan metode MLE.
2. Memodelkan data jumlah kematian ibu dan bayi di Kota Makassar tahun 2017 yang mengalami *overdispersi* dengan menggunakan model regresi BZIP.

1.5 Manfaat Penelitian

Manfaat dalam penelitian ini diharapkan dapat menambah pemahaman teoritis dan praktis bagi peneliti dan pembaca tentang model regresi BZIP menggunakan metode MLE dengan algoritma EM serta pemahaman menganalisis data jumlah kematian ibu dan bayi yang mengalami *overdispersi*. Selain itu, penelitian ini diharapkan dapat menjadi bahan referensi untuk penelitian selanjutnya.



BAB II

TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi merupakan analisis untuk memodelkan dan menjelaskan hubungan antara dua variabel, yaitu variabel respon dan variabel prediktor. Bentuk umum model regresi linier adalah sebagai berikut.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \varepsilon_i \quad (2.1)$$

dengan $i = 1, 2, \dots, n$; y_i adalah variabel respon, $x_{ij} = (x_{i1}, x_{i2}, \dots, x_{ip})$ adalah variabel prediktor, $\beta_j = (\beta_1, \beta_2, \dots, \beta_p)$ adalah parameter model regresi yang tidak diketahui, dan ε_i adalah galat yang diasumsikan identik, independen, dan berdistribusi normal dengan mean nol dan variansi σ^2 . Persamaan (2.1) dapat dinyatakan dalam bentuk matriks seperti pada persamaan (2.2) berikut.

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

dengan

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Menurut Harlan (2018), asumsi-asumsi yang harus dipenuhi pada model regresi linier, yaitu:

1. Asumsi linieritas, menyatakan bahwa hubungan antara variabel respon dan variabel prediktor adalah linier.
2. Asumsi normalitas, menyatakan bahwa suku galat berdistribusi normal dengan mean nol, yaitu $E(\varepsilon_i) = 0$.
3. Asumsi homoskedastisitas, menyatakan bahwa suku galat memiliki variansi yang konstan, yaitu $Var(\varepsilon_i) = \sigma^2$.
4. Asumsi independensi dan non-autokorelasi, menyatakan bahwa suku galat

ng bebas dan tak saling berkorelasi, yaitu $cov(\varepsilon_i, \varepsilon_{i*}) = 0$.

umsi non-multikolinieritas, menyatakan bahwa tidak boleh ada korelasi
ar variabel prediktor, yaitu $cov(x_j, x_{j*}) = 0$.



2.2 Regresi Poisson

Regresi *Poisson* merupakan model regresi non-linier dengan variabel responnya berupa data hitung (diskrit) dan berdistribusi *Poisson*. Distribusi *Poisson* adalah suatu distribusi untuk peristiwa yang probabilitas kejadiannya kecil, dimana kejadian tersebut bergantung pada interval waktu tertentu atau di suatu daerah tertentu. Jika variabel acak diskrit (Y) berdistribusi *Poisson* dengan parameter $\lambda > 0$, maka fungsi probabilitasnya dinyatakan sebagai berikut.

$$f(y) = e^{-\lambda} \frac{\lambda^y}{y!}, \quad y = 0, 1, 2, \dots \quad (2.3)$$

Distribusi *Poisson* memiliki nilai mean dan variansi yang sama, yaitu $E(Y) = Var(Y) = \lambda$ (Arkandi & Winahju, 2015).

Regresi *Poisson* cocok digunakan jika variabel respon (Y) merupakan bilangan bulat non-negatif yang nilainya kecil sehingga model regresi *Poisson* dapat dituliskan sebagai berikut.

$$y_i = \lambda_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.4)$$

dengan y_i adalah jumlah kejadian yang terjadi, λ_i adalah rata-rata jumlah kejadian pada waktu tertentu, dan ε_i adalah galat yang diasumsikan identik, independen, dan berdistribusi normal dengan mean nol dan variansi σ^2 (Cahyandari, 2014).

Regresi *Poisson* menggunakan *generalized linear model* (GLM) agar modelnya dapat digunakan dalam data pengamatan, dimana variabel responnya tidak mengharuskan berdistribusi normal. Dalam GLM, terdapat sebuah fungsi g linier yang menghubungkan rata-rata dari variabel respon dengan variabel prediktor, yaitu:

$$g(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$$

$$g(\lambda_i) = \mathbf{X}_i \boldsymbol{\beta}$$

dengan

fungsi g menunjukkan fungsi penghubung

λ_i menunjukkan rata-rata dari variabel respon (y_i) yang berdistribusi *Poisson*.

$\mathbf{X}_i = [1 \quad x_{i1} \quad \dots \quad x_{ip}]$ menunjukkan vektor yang berukuran $1 \times (p + 1)$ dari prediktor.

$\boldsymbol{\beta} = [\beta_0 \quad \beta_1 \quad \dots \quad \beta_p]^T$ menunjukkan vektor yang berukuran $(p + 1) \times 1$ dari regresi yang tidak diketahui.



Oleh karena itu, hubungan antara rata-rata dari variabel respon dan variabel prediktor adalah sebagai berikut

$$\lambda_i = g^{-1}(X_i\boldsymbol{\beta})$$

(Cahyandari, 2014).

Regresi *Poisson* biasanya menggunakan fungsi penghubung logaritma natural, karena rata-rata dari variabel responnya akan berbentuk fungsi eksponensial dan menjamin nilai variabel yang diestimasi dari variabel responnya akan bernilai non-negatif. Fungsi penghubung tersebut berbentuk seperti berikut.

$$g(\lambda_i) = \ln \lambda_i$$

Rata-rata variabel respon dan variabel prediktor dihubungkan dengan menggunakan fungsi penghubung seperti berikut.

$$\begin{aligned} \ln \lambda_i &= X_i\boldsymbol{\beta} \\ \lambda_i &= \exp(X_i\boldsymbol{\beta}) \end{aligned} \quad (2.5)$$

Oleh karena itu, fungsi probabilitas distribusi *Poisson* dapat dituliskan berikut.

$$f(y_i; \boldsymbol{\beta}) = \exp\{-[\lambda_i(X_i; \boldsymbol{\beta})]\} \frac{[\lambda_i(X_i; \boldsymbol{\beta})]^{y_i}}{y_i!}$$

dengan $\lambda(X_i; \boldsymbol{\beta})$ adalah mean dari distribusi *Poisson* dan vektor $\boldsymbol{\beta}$ menunjukkan parameter regresi yang tidak diketahui. Oleh karena itu, model regresi *Poisson* dapat dituliskan sebagai berikut.

$$\begin{aligned} y_i &= \lambda_i + \varepsilon_i \\ y_i &= \exp(X_i\boldsymbol{\beta}) + \varepsilon_i \end{aligned} \quad (2.6)$$

dengan

$X_i = [1 \ x_{i1} \ \dots \ x_{ip}]$ menunjukkan vektor yang berukuran $1 \times (p + 1)$ dari variabel prediktor yang digunakan untuk memodelkan λ_i .

$\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \dots \ \beta_p]^T$ menunjukkan parameter regresi yang berukuran $(p + 1) \times 1$. ε_i menunjukkan galat yang diasumsikan identik, independen, dan berdistribusi normal dengan mean nol dan variansi σ^2 .

(Cahyandari, 2014)



regresi *Poisson* mengasumsikan bahwa variabel respon (Y) merupakan diskrit dan harus memenuhi asumsi *equidispersi*. *Equidispersi* yaitu nilai rata-rata dengan nilai variansi, yaitu $E(Y) = Var(Y) = \lambda$ (Kurniawan, 2017).

2.3 Regresi Bivariate Poisson

Regresi *Bivariate Poisson* adalah sebuah metode yang digunakan untuk memodelkan sepasang data hitung (diskrit) yang berdistribusi *Poisson* dan saling berkorelasi (Karlis & Ntzoufras, 2005). Karena terdapat dua variabel respon yang saling berkorelasi, maka distribusinya akan mengikuti distribusi *Bivariate Poisson*.

Distribusi *Bivariate Poisson* terjadi saat variabel acak Z_0, Z_1 , dan Z_2 masing-masing berdistribusi *Poisson* dengan parameter λ_0, λ_1 , dan λ_2 . Terdapat variabel acak Y_1 dan Y_2 yang terbentuk dari variabel Z_0, Z_1 , dan Z_2 yang saling bebas (independen) seperti berikut.

$$Y_1 = Z_1 + Z_0 \quad \text{dan} \quad Y_2 = Z_2 + Z_0$$

Oleh karena itu, variabel acak Y_1 dan Y_2 secara bersama-sama berdistribusi *Bivariate Poisson* dengan fungsi probabilitasnya berbentuk seperti berikut.

$$f_{BP}(y_1, y_2) = \exp(-\lambda) \sum_{r=0}^{\min\{y_1, y_2\}} \frac{\lambda_1^{y_1-r} \lambda_2^{y_2-r} \lambda_0^r}{(y_1 - r)! (y_2 - r)! r!} \quad (2.7)$$

dengan $\lambda = \lambda_1 + \lambda_2 + \lambda_0$ (Wang Y. , 2017).

Menurut Kawamura (1973), nilai ekpektasi dan kovariansi dari variabel acak Y_1 dan Y_2 adalah sebagai berikut.

$$E(Y_1) = E(Z_1 + Z_0) = E(Z_1) + E(Z_0) = \lambda_1 + \lambda_0$$

$$E(Y_2) = E(Z_2 + Z_0) = E(Z_2) + E(Z_0) = \lambda_2 + \lambda_0$$

Setelah diketahui nilai ekpektasi dari masing-masing variabel acak Y_1 dan Y_2 , maka dapat diketahui pula nilai $E(Y_1 Y_2)$ sebagai berikut.

$$E(Y_1 Y_2) = E[(Z_1 + Z_0)(Z_2 + Z_0)]$$

$$E(Y_1 Y_2) = E(Z_1 Z_0) + E(Z_1 Z_2) + E(Z_0 Z_2) + E(Z_0^2)$$

$$E(Y_1 Y_2) = (\lambda_1 + \lambda_0)(\lambda_2 + \lambda_0) + \lambda_0$$

Kemudian diperoleh nilai variansi, yaitu :

$$Var(Y_1) = Var(Z_1 + Z_0) = Var(Z_1) + Var(Z_0) = \lambda_1 + \lambda_0$$

$$Var(Y_2) = Var(Z_2 + Z_0) = Var(Z_2) + Var(Z_0) = \lambda_2 + \lambda_0$$



Sedangkan nilai kovariansinya, yaitu :

$$\begin{aligned} Cov(Y_1, Y_2) &= E[(Y_1 - E(Y_1))(Y_2 - E(Y_2))] \\ Cov(Y_1, Y_2) &= E[Y_1Y_2 - Y_1E(Y_2) - Y_2E(Y_1) + E(Y_1)E(Y_2)] \\ Cov(Y_1, Y_2) &= E(Y_1Y_2) - E(Y_1)E(Y_2) \\ Cov(Y_1, Y_2) &= ((\lambda_1 + \lambda_0)(\lambda_2 + \lambda_0) + \lambda_0) - ((\lambda_1 + \lambda_0)(\lambda_2 + \lambda_0)) \\ Cov(Y_1, Y_2) &= \lambda_0 \end{aligned}$$

Dengan demikian, koefisien korelasi untuk Y_1 dan Y_2 , yaitu :

$$\rho_{Y_1Y_2} = \frac{Cov(Y_1, Y_2)}{\sqrt{Var(Y_1)Var(Y_2)}} = \frac{\lambda_0}{\sqrt{(\lambda_1 + \lambda_0)(\lambda_2 + \lambda_0)}}$$

(Ilmi, 2015).

Regresi *Bivariate Poisson* mendapatkan model dari bentuk berikut.

$$\begin{aligned} (Y_1, Y_2) &\sim BP(\lambda_1, \lambda_2, \lambda_0) \\ \lambda_{ik}^* &= \lambda_{ik} + \lambda_0 = \exp(\mathbf{X}_i\boldsymbol{\beta}_k), \quad k = 1, 2 \end{aligned} \tag{2.8}$$

dengan

λ_{ik}^* menunjukkan rata-rata dari variabel respon (y_{ik}) yang berdistribusi *Poisson*.

$\mathbf{X}_i = [1 \quad x_{i1} \quad \dots \quad x_{ip}]$ menunjukkan vektor yang berukuran $1 \times (p + 1)$ dari variabel prediktor yang digunakan untuk memodelkan λ_{ik}^* .

$\boldsymbol{\beta}_k = [\beta_{0k} \quad \beta_{1k} \quad \dots \quad \beta_{pk}]^T$ menunjukkan vektor yang berukuran $(p + 1) \times 1$ dari parameter regresi yang tidak diketahui.

(Kocherlakota & Kocherlakota, 2007)

Menurut Karlis dan Ntzoufras (2005), terdapat 3 model regresi *Bivariate Poisson* dengan nilai λ_0 yang berbeda, diantaranya:

- 1) Model dengan nilai λ_0 adalah suatu konstanta (tidak ada kovariat di dalam λ_0).
- 2) Model dengan nilai λ_0 merupakan fungsi dari variabel bebas (kovariat) sehingga membentuk suatu persamaan berikut

$$\lambda_0 = \exp(\beta_{00} + \beta_{10}x_1 + \beta_{20}x_2 + \dots + \beta_{p0}x_p) = \exp(\mathbf{X}\boldsymbol{\beta}_0)$$

- 3) Model dengan nilai λ_0 adalah nol, dengan tidak ada kovariat dari kedua variabel atau kedua variabel respon saling bebas, biasanya disebut model *independent Poisson*.

(Ilmi, 2018)



2.4 Equidispersi

Salah satu asumsi yang harus dipenuhi dalam regresi *Poisson* adalah asumsi *equidispersi*, yaitu keadaan dengan mean dan variansi dari variabel responnya bernilai sama. Namun dalam aplikasinya asumsi tersebut kadang dilanggar, dimana nilai variansinya lebih besar daripada nilai mean yang disebut *overdispersi* atau nilai variansinya lebih kecil daripada nilai mean yang disebut *underdispersi* (Cameron & Trivedi, 1998). *Overdispersi* pada regresi *Poisson* menghasilkan simpangan baku dari estimasi parameter jauh lebih kecil daripada nilai sebenarnya (*underestimate*) dan uji signifikansi dari variabel prediktor jauh lebih besar daripada nilai sebenarnya (*overestimate*), sehingga kesimpulan yang dihasilkan menjadi tidak valid (Ismail & Jemain, 2007).

Dalam beberapa kasus, yang ditemukan adalah variansi data yang diamati lebih besar daripada meannya atau disebut *overdispersi*. Ada beberapa hal yang dapat menyebabkan terjadinya *overdispersi*, antara lain: banyaknya pengamatan yang bernilai nol (*excess zero*), adanya sumber keragaman antar individu yang tidak teramati (*unobserved heterogeneity*), adanya pengamatan yang hilang (*data missing*), adanya pencilan (*outlier*) pada data sehingga perlunya interaksi dalam model, atau kesalahan spesifikasi fungsi penghubung (Cahyandari, 2014).

2.5 Regresi Zero-Inflated Poisson

Salah satu penyebab terjadinya *overdispersi* adalah lebih banyak pengamatan bernilai nol daripada yang diharapkan untuk model regresi *Poisson*. Salah satu metode analisis yang diusulkan untuk lebih banyak pengamatan bernilai nol adalah model regresi *Zero-Inflated Poisson* (ZIP) (Jansakul & Hinde, 2002).

Jika Y adalah variabel acak yang berdistribusi ZIP, maka nilai nol pada pengamatan diestimasi muncul dalam dua cara yang sesuai dengan keadaan (*state*) yang terpisah. Keadaan pertama disebut *zero state* terjadi dengan probabilitas π dan menghasilkan hanya pengamatan yang bernilai nol. Sedangkan keadaan kedua disebut *Poisson state* terjadi dengan probabilitas $(1 - \pi)$ dan menghasilkan pengamatan yang berdistribusi *Poisson* dengan mean λ . Proses dua keadaan ini memberikan distribusi yang terdiri dari dua komponen dengan fungsi probabilitas seperti pada Persamaan (2.9) (Jansakul & Hinde, 2002).



$$f_{ZIP}(y) = \begin{cases} \pi + (1 - \pi)e^{-\lambda}, & \text{untuk } y = 0, \quad 0 \leq \pi \leq 1 \\ \frac{(1 - \pi)e^{-\lambda}\lambda^y}{y!}, & \text{untuk } y \neq 0, \quad 0 \leq \pi \leq 1 \end{cases} \quad (2.9)$$

Lambert (1992) menyarankan model gabungan untuk λ_i dan π_i , yaitu:

$$\ln(\lambda_i) = \mathbf{X}_i\boldsymbol{\beta} \quad \text{dan} \quad \text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{X}_i\boldsymbol{\gamma} \quad (2.10)$$

dengan

λ_i menunjukkan rata-rata dari variabel respon (y_i).

π_i menunjukkan probabilitas pengamatan ke- i .

$\mathbf{X}_i = [1 \quad x_{i1} \quad \dots \quad x_{ip}]$ menunjukkan vektor berukuran $1 \times (p + 1)$ dari variabel prediktor yang digunakan untuk memodelkan λ_i .

$\boldsymbol{\beta} = [\beta_0 \quad \beta_1 \quad \dots \quad \beta_p]^T$ dan $\boldsymbol{\gamma} = [\gamma_0 \quad \gamma_1 \quad \dots \quad \gamma_p]^T$ menunjukkan vektor yang berukuran $(p + 1) \times 1$ dari parameter regresi yang tidak diketahui.

Nilai ekspektasi dan variansi dari Y sebagai berikut.

$$E(Y) = (1 - \pi)\lambda = \mu \quad \text{dan} \quad \text{Var}(Y) = \mu + \left(\frac{\pi}{\pi - 1}\right)\mu^2 \quad (2.11)$$

Overdispersi pada Y terjadi jika nilai $\pi > 0$ yang mengindikasikan bahwa terjadinya peningkatan nilai nol pada variabel respon Y . Persamaan (2.11) terlihat bahwa $\text{Var}(Y) > E(Y)$ yang mengindikasikan bahwa regresi ZIP dapat mengatasi *overdispersi* (Eminita dkk., 2019). Oleh karena itu, regresi ZIP mendapatkan model dari bentuk berikut (Wang dkk., 2003).

$$Y \sim ZIP(\lambda, \pi)$$

$$\ln(\lambda_i) = \mathbf{X}_i\boldsymbol{\beta}$$

$$\lambda_i = \exp(\mathbf{X}_i\boldsymbol{\beta}) \quad (2.12)$$

$$\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{X}_i\boldsymbol{\gamma}$$

$$\frac{\pi_i}{1 - \pi_i} = \exp(\mathbf{X}_i\boldsymbol{\gamma})$$

$$\pi_i + \pi_i \exp(\mathbf{X}_i\boldsymbol{\gamma}) = \exp(\mathbf{X}_i\boldsymbol{\gamma})$$

$$\pi_i = \frac{\exp(\mathbf{X}_i\boldsymbol{\gamma})}{1 + \exp(\mathbf{X}_i\boldsymbol{\gamma})} \quad (2.13)$$



2.6 Regresi Bivariate Zero-Inflated Poisson

Untuk data diskrit yang saling berkorelasi dan mengalami *overdispersi*, distribusi *Bivariate Zero-Inflated Poisson* (BZIP) dapat dibangun dari distribusi *Bivariate Poisson* dan probabilitas titik di (0,0) yang didefinisikan sebagai berikut.

$$(Y_1, Y_2) \sim \text{Degenerate}(0,0) \text{ dengan probabilitas } \pi \text{ disebut } \textit{zero state}$$

$$\sim \text{BP}(\lambda_1, \lambda_2, \lambda_0) \text{ dengan probabilitas } (1 - \pi) \text{ disebut } \textit{Poisson state}$$

Distribusi BZIP memiliki fungsi probabilitas seperti berikut.

$$f_{BZIP}(y_1, y_2) = \begin{cases} \pi + (1 - \pi)e^{-\lambda} , & (y_1, y_2) = (0,0) \\ (1 - \pi)e^{-\lambda} \sum_{r=0}^{\min\{y_1, y_2\}} \frac{\lambda_1^{y_1-r} \lambda_2^{y_2-r} \lambda_0^r}{(y_1 - r)! (y_2 - r)! r!}, & (y_1, y_2) \neq (0,0) \end{cases} \quad (2.14)$$

dengan $0 < \pi < 1$ dan $\lambda = \lambda_1 + \lambda_2 + \lambda_0$ (Wang dkk., 2003).

Distribusi marginal dari BZIP adalah distribusi univariat ZIP sebagai berikut.

$$f_{Y_k}(y_k) = \begin{cases} \pi + (1 - \pi) \exp(-\lambda_k - \lambda_0) , & \textit{untuk } y_k = 0 \\ (1 - \pi) \frac{(\lambda_k + \lambda_0)^{y_k}}{y_k!} \exp(-\lambda_k - \lambda_0) , & \textit{untuk } y_k \neq 0 \end{cases}$$

Nilai ekspektasi dan variansi dari distribusi BZIP adalah sebagai berikut.

$$E(Y_k) = (1 - \pi)(\lambda_k + \lambda_0) = \mu_k, \quad k = 1,2$$

$$Var(Y_k) = [1 + \pi(\lambda_k + \lambda_0)] \mu_k, \quad k = 1,2$$

$$E(Y_1 Y_2) = (1 - \pi)[(\lambda_1 + \lambda_0)(\lambda_2 + \lambda_0) + \lambda_0]$$

$$Cov(Y_1, Y_2) = (1 - \pi)[\lambda_0 + \pi(\lambda_1 + \lambda_0)(\lambda_2 + \lambda_0)]$$

Oleh karena itu, regresi BZIP mendapatkan model dari bentuk berikut.

$$(Y_1, Y_2) \sim \text{BZIP}(\lambda_1, \lambda_2, \lambda_0, \pi)$$

$$\ln(\lambda_{ik}^*) = \ln(\lambda_{ik} + \lambda_0) = \mathbf{X}_i \boldsymbol{\beta}_k, \quad k = 1,2 \quad (2.15)$$

$$\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{X}_i \boldsymbol{\gamma} \quad (2.16)$$

dengan

λ_{ik} menunjukkan rata-rata dari variabel respon (y_{ik}).

π_i menunjukkan probabilitas pengamatan ke-i.

\mathbf{X}_i menunjukkan vektor berukuran $1 \times (p + 1)$ dari variabel prediktor yang digunakan untuk memodelkan λ_{ik} .

$\boldsymbol{\beta}_k$ menunjukkan vektor yang berukuran $(p + 1) \times 1$ dari parameter regresi yang diketahui. (Wang dkk., 2003)



2.7 Metode Maximum Likelihood Estimation

Metode *maximum likelihood estimation* (MLE) dapat digunakan untuk menaksir parameter suatu model yang diketahui fungsi probabilitasnya. Misalkan Y_1, Y_2, \dots, Y_n adalah sampel acak dari populasi dengan fungsi probabilitas $f(y|\theta)$ yang bergantung pada $\theta = \lambda_0, \beta_1, \beta_2, \gamma$ dengan θ adalah parameter yang tidak diketahui. Karena Y_1, Y_2, \dots, Y_n saling bebas, maka fungsi probabilitas bersama dari Y_1, Y_2, \dots, Y_n dapat dinyatakan sebagai berikut.

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

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

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

$$L(\theta|y) = f(y_1|\theta) \cdot f(y_2|\theta) \dots f(y_n|\theta) = \prod_{i=1}^n f(y_i|\theta) \quad (2.17)$$

(Hogg, McKean, & Craig, 2013)

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

$$l(\theta|y) = \ln L(\theta|y)$$

$$l(\theta|y) = \ln \left\{ \prod_{i=1}^n f(y_i|\theta) \right\} = \sum_{i=1}^n \ln f(y_i|\theta) \quad (2.18)$$

Untuk memperoleh nilai θ yang memaksimumkan fungsi $l(\theta|y)$, maka $l(\theta|y)$ diturunkan terhadap θ dan kemudian menyamakannya dengan nol seperti pada Persamaan (2.19) berikut.

$$l'(\theta|y) = \frac{\partial l(\theta|y)}{\partial \theta} = 0 \quad (2.19)$$

(Hogg, McKean, & Craig, 2013)



2.8 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 memaksimalkan 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:

$$f(y, z | \theta) = f(z) \cdot f(y|z) \quad (2.20)$$

Tahap ekspektasi (*E-Step*) pada algoritma EM dilakukan dengan menghitung ekspektasi dari fungsi *ln-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:

$$Q(\theta | \theta^{(0)}; y) = E[\ln L(\theta^{(0)} | y, z)] \quad (2.21)$$

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

2.9 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.

$$H_0 : \beta_{11} = \dots = \beta_{p1} = \beta_{12} = \dots = \beta_{p2} = \gamma_1 = \dots = \gamma_p = 0 \quad (\text{variabel prediktor}$$

bersama-sama tidak mempunyai pengaruh terhadap variabel respon)

atau $\beta_{j1} \neq 0$ atau $\beta_{j2} \neq 0$ atau $\gamma_j \neq 0$, dengan $j = 1, 2, \dots, p$ (variabel

prediktor secara bersama-sama mempunyai pengaruh terhadap variabel

respon)



Statistik uji:

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

dengan

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

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

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

2.10 Uji Parsial Parameter Model

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

2.10.1 Uji Parsial Parameter β

Pengujian parsial parameter model $\ln(\lambda_{ik}^*) = \ln(\lambda_{ik} + \lambda_0) = \mathbf{X}_i \boldsymbol{\beta}_k$ dengan $k = 1, 2$ menggunakan hipotesis sebagai berikut.

$H_0 : \beta_{jk} = 0, j = 1, 2 \dots, p, k = 1, 2$ (variabel prediktor tidak mempunyai pengaruh terhadap variabel respon)

$H_1 : \text{Ada } \beta_{jk} \neq 0, j = 1, 2 \dots, p, k = 1, 2$ (variabel prediktor mempunyai pengaruh terhadap variabel respon)

Statistik uji:

$$W(\hat{\beta}_{jk}) = \left(\frac{\hat{\beta}_{jk}}{SE(\hat{\beta}_{jk})} \right)^2$$

dengan

$\hat{\beta}_{jk}$ adalah estimasi parameter β_{jk}

$SE(\hat{\beta}_{jk})$ adalah standar error dari $\hat{\beta}_{jk}$

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



2.10.2 Uji Parsial Parameter γ

Pengujian parsial parameter model logit(π_i) = $\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{X}_i\boldsymbol{\gamma}$ dengan menggunakan hipotesis sebagai berikut.

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

$H_1 : \text{Ada } \gamma_j \neq 0, j = 1, 2, \dots, p$ (variabel prediktor mempunyai pengaruh terhadap variabel respon)

Statistik uji:

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

dengan

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

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

Jika $W(\hat{\gamma}_j) > \chi_{\alpha; dk}^2$, maka H_0 ditolak artinya variabel prediktor memiliki pengaruh yang signifikan terhadap variabel respon Y_1 dan Y_2 (Kurniawan, 2017).

2.11 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.

$$AIC = 2p - 2 \ln L(\hat{\boldsymbol{\theta}})$$

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

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



2.12 Kematian Ibu dan Bayi

Kematian ibu adalah kematian perempuan pada saat hamil atau kematian dalam kurun waktu 42 hari sejak terminasi kehamilan tanpa memandang lamanya kehamilan, yaitu kematian yang disebabkan karena kehamilannya atau penanganannya, tetapi bukan karena sebab-sebab lain seperti kecelakaan dan terjatuh. Angka Kematian Ibu (AKI) dihitung per 100.000 kelahiran hidup pada tahun tertentu. AKI berguna untuk menggambarkan tingkat kesadaran perilaku hidup sehat, status gizi dan kesehatan ibu, kondisi kesehatan lingkungan, tingkat pelayanan kesehatan terutama untuk ibu hamil, serta pelayanan kesehatan waktu ibu melahirkan dan masa nifas (Dinkes, 2018).

Menurut Dinkes (2018), berbagai pelayanan kesehatan dasar dilakukan dalam upaya penurunan AKI, diantaranya sebagai berikut :

1. Pelayanan Kesehatan Ibu Hamil

Pelayanan kesehatan ibu hamil terdiri dari kunjungan pertama (K1) dan kunjungan keempat (K4). Cakupan K1 adalah jumlah ibu hamil yang telah memperoleh pelayanan kesehatan pertama kali oleh tenaga kesehatan sesuai standar. Sedangkan cakupan K4 adalah jumlah ibu hamil yang telah mendapatkan pelayanan kesehatan sesuai standar, minimal empat kali kunjungan selama masa kehamilannya.

2. Pelayanan Kesehatan Ibu Bersalin

Pelayanan kesehatan ibu bersalin yang sesuai standar adalah persalinan yang dilakukan oleh bidan dan atau dokter spesialis kebidanan yang bekerja di fasilitas pelayanan kesehatan pemerintah maupun swasta yang memiliki Surat Tanda Register (STR) baik persalinan normal dan atau persalinan dengan komplikasi. Pencapaian upaya kesehatan ibu bersalin diukur melalui indikator persentase persalinan yang ditolong tenaga kesehatan.

3. Pelayanan Kesehatan Ibu Nifas

Pelayanan kesehatan ibu nifas adalah pelayanan kesehatan pada ibu sesuai standar yang dilakukan sekurang-kurangnya tiga kali sesuai awal yang dianjurkan pasca persalinan. Upaya untuk meningkatkan cakupan pelayanan kesehatan ibu nifas yaitu dengan melakukan kunjungan



rumah pada semua ibu bersalin selama masa nifas sebanyak 3 kali yang dibantu oleh kader kesehatan di setiap Puskesmas dan pemberian kapsul vitamin A untuk penanggulangan kekurangan vitamin A pada periode sebelum 40 hari setelah melahirkan.

4. Penanganan Komplikasi Kebidanan

Komplikasi kebidanan adalah kesakitan pada ibu hamil, ibu bersalin, ibu nifas dan atau janin dalam kandungan, baik langsung maupun tidak langsung termasuk penyakit menular dan tidak menular yang dapat mengancam jiwa ibu dan atau janin, yang tidak disebabkan oleh trauma/kecelakaan. Adapun penanganan komplikasi itu sendiri adalah penanganan terhadap komplikasi/ kegawatdaruratan yang mendapat pelayanan kesehatan sampai selesai.

5. Pelayanan Kontrasepsi

Menurut *World Health Organisation* (WHO), keluarga berencana (KB) adalah tindakan yang membantu individu atau pasangan suami istri untuk menghindari kelahiran yang tidak diinginkan, mendapatkan kelahiran yang diinginkan, mengatur interval di antara kelahiran, mengontrol waktu saat kelahiran dalam hubungan dengan umur suami dan istri, dan menentukan jumlah anak dalam keluarga. Program KB tidak hanya berfungsi untuk menurunkan laju pertumbuhan penduduk, tetapi juga mengatasi permasalahan kesehatan reproduksi di Indonesia sehingga dapat menunjang kelancaran pembangunan.

Kematian bayi adalah kematian yang terjadi saat setelah bayi lahir sampai bayi belum berusia tepat satu tahun. Angka Kematian Bayi (AKB) dihitung per 1.000 kelahiran hidup pada tahun tertentu. Adapun faktor yang menjadi penyebab kematian bayi, yaitu kematian bayi endogen atau kematian neonatal dan kematian bayi eksogen atau kematian post neo-natal (Dinkes, 2018).

Kematian neonatal adalah banyaknya kematian bayi yang terjadi pada bulan setelah dilahirkan dan umumnya disebabkan oleh faktor yang dibawa anak yang diperoleh dari orang tua pada saat konsepsi atau selama kehamilan. Kematian neonatal dapat disebabkan oleh dua faktor, yaitu faktor ibu antara lain



pelayanan kesehatan ibu hamil, infeksi ibu hamil, gizi ibu hamil, dan karakteristik dari ibu hamil (umur, paritas, dan jarak kehamilan) serta faktor janin antara lain bayi berat badan lahir rendah (BBLR), asfiksia, dan pneumonia (Dinkes, 2018).

Salah satu upaya penurunan AKB adalah meningkatkan cakupan dan kualitas pelayanan kesehatan bayi. Pelayanan kesehatan bayi bertujuan untuk meningkatkan akses bayi terhadap pelayanan kesehatan dasar, mengetahui sedini mungkin jika terdapat kelainan pada bayi sehingga cepat mendapat pertolongan, pemeliharaan kesehatan dan pencegahan penyakit. Pelayanan kesehatan bayi terdiri dari pemantauan pertumbuhan, imunisasi dasar lengkap, serta peningkatan kualitas hidup bayi dengan stimulasi tumbuh kembang (Dinkes, 2018).

