TUGAS AKHIR- KM184801

KAJIAN ESTIMASI PARAMETER UNTUK *REGRESI ZERO-INFLATED POISSON* (ZIP)

ALVINDA NISMA YUSNIAR
NRP 06111540000003

Dosen Pembimbing :
Dra. Laksmi Prita Wardhani, M.Si
Dr. Valeriana Lukitosari, S.Si., MT

DEPARTEMEN MATEMATIKA
Fakultas Sains dan Analitika Data
Institut Teknologi Sepuluh Nopember
Surabaya 2020

FINAL PROJECT–KM184801

STUDY OF PARAMETER ESTIMATION FOR ZERO-INFLATED POISSON (ZIP) REGRESSION

ALVINDA NISMA YUSNIAR
NRP 06111540000003

Supervisors :
Dra. Laksmi Prita Wardhani, M.Si
Dr. Valeriana Lukitosari, S.Si., MT

DEPARTMENT OF MATHEMATICS
Faculty of Science and Data Analytics
Sepuluh Nopember Institute of Technology
Surabaya 2020

LEMBAR PENGESAHAN

KAJIAN ESTIMASI PARAMETER UNTUK REGRESI *ZERO-INFLATED POISSON* (ZIP)

STUDY OF PARAMETER ESTIMATIONFOR ZERO-INFLATED POISSON (ZIP) REGRESSION

TUGAS AKHIR

Diajukan untuk memenuhi salah satu syarat
Untuk memperoleh gelar Sarjana Matematika
Pada bidang studi Matematika Terapan
Program Studi S-1 Departemen Matematika
Fakultas Sains dan Analitika Data
Institut Teknologi Sepuluh Nopember Surabaya

Oleh :
Alvinda Nisma Yusniar
NRP. 06111540000003

Menyetujui,

Dosen Pembimbing II,	Dosen Pembimbing I,
Dr. Valeriana Lukitosari, S.Si, MT	Dra. Laksmi Prita Wardhani, M.Si
NIP. 19710928 199802 2 001	NIP. 19611208 198803 2 001

Mengetahui,
Kepala Departemen Matematika
FSAD ITS

Subchan, Ph.D
NIP. 19710513 199702 1 001
Surabaya, Januari 2020

KAJIAN ESTIMASI PARAMETER UNTUK REGRESI *ZERO-INFLATED POISSON (ZIP)*

Nama	**: Alvinda Nisma Yusniar**
NRP	**: 06111540000003**
Departemen	**: Matematika FSAD-ITS**
Pembimbing	**: Dra. Laksmi Prita Wardhani, M.Si**
	Dr. Valeriana Lukitosari, S.Si, MT

ABSTRAK

Regresi Poisson merupakan salah satu analisis regresi yang sering digunakan untuk menganalisis hubungan antara beberapa variabel respon dan variabel prediktor yang berbentuk diskrit. Akan tetapi dalam beberapa kasus terdapat fenomena dimana variabel respon yang berditribusi Poisson mengandung banyak nilai nol dengan presentase di atas 50%, fenomena tersebut dapat mengakibatkan kesalahan analisis. Oleh sebab itu diperlukan metode untuk mengatasi permasalahan tersebut.

Regresi Zero-Inflated Poisson adalah salah satu metode alternative untuk mengatasi permasalahan dimana variabel respon mengandung banyak nilai nol. Setelah dilakukan pengkajian estimasi untuk Regresi Poisson dan Regresi Zero-Infalted Poisson dengan metode Maximum Likelihood Estimation dan Moment Estimation diperoleh perbedaan langkah untuk mendapatkan nilai estimasi dan jumlah parameter. Pada regresi Poisson terdapat satu parameter sedangkan pada Regresi Poisson dan Regresi Zero-Inflated Poisson terdapat dua parameter. Langkah dalam mendapatkan nilai estimasi dengan metode Maximum Likelihood Estimation pada model Regresi

Zero-Inflated Poisson membutuhkan Algoritma-EM karena hasil estimasi berbentuk implisit.

Kata-kunci :
Regresi Poisson, Regresi Zero-Inflated Poisson, Maximum Likelihood Estimation, Moment Estimation

STUDY OF PARAMETER ESTIMATION FOR ZERO-INFLATED POISSON (ZIP) REGRESSION)

Name : Alvinda Nisma Yusniar
NRP : 06111540000003
Department : Mathematics FSAD-ITS
Supervisors : Dra. Laksmi Prita Wardhani, M.Si
Dr. Valeriana Lukitosari, S.Si, MT

ABSTRACT

Poisson Regression is a regression analysis that is often used to analyze the relationship between several response variables and predictor variables that are discrete. However, in some cases there are phenomena where the response variable with Poisson distribution contains a lot of zero values with a percentage above 50%, this phenomenon can lead to analysis errors. Therefore we need a method to overcome these problems.

Zero-Inflated Poisson Regression is an alternative method to solve problems where the response variable contains many zero values. After estimating the Poisson Regression and Zero-Inflated Poisson Regression with the Maximum Likelihood Estimation and Moment Estimation methods, different steps are obtained to obtain the estimated value and the number of parameters. In Poisson regression there are one parameter while in Poisson Regression and Zero-Inflated Poisson Regression there are two parameters. The step in getting the estimated value using the Maximum Likelihood Estimation method in the Zero-Inflated Poisson Regression model requires an EM-algorithm because the estimation results are in the form of implicit.

Keywords :
Poisson Regression, Zero-Inflated Poisson Regression, Maximum Likelihood Estimation, Moment Estimation

KATA PENGANTAR

Alhamdulillaahirobbil'aalamiin, segala puji dan syukur penulis panjatkan ke hadirat Allah SWT yang telah memberikan limpahan rahmat, taufik serta hidayah-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul

"KAJIAN ESTIMASI PARAMETER UNTUK REGRESI *ZERO-INFLATED POISSON (ZIP)*"

sebagai salah satu syarat kelulusan Program Sarjana Departemen Matematika FSAD Institut Teknologi Sepuluh Nopember (ITS) Surabaya.

Tugas Akhir ini dapat diselesaikan dengan baik berkat bantuan dan dukungan dari berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih dan penghargaan kepada:

1. Orang tua penulis (Bapak Isdijono dan Ibu Sumini) serta kakak kandung penulis (Elmida Nismawandani) yang selalu mendoakan dan mendukung penulis selama menempuh perkuliahan di Departemen Matematika ITS.
2. Bapak Subchan, S.Si, M.Sc, Ph.D selaku Kepala Departemen Matematika FSAD-ITS yang telah memberikan dukungan dan motivasi selama perkuliahan hingga terselesaikannya Tugas Akhir ini.
3. Ibu Dr. Dwi Ratna Sulistyaningrum, S.Si, MT selaku Sekretaris Bidang Akademik Departemen Matematika FSAD-ITS yang telah membantu penulis dalam hal menyelesaikan kebutuhan administrasi selama periode pengambilan Tugas Akhir.

4. Ibu Dra. Laksmi Prita Wardhani, M.Si dan ibu Dr. Valeriana Lukitosari, S.Si, MT selaku dosen pembimbing yang telah memberikan motivasi dan pengarahan dalam penyelesaian Tugas Akhir ini.
5. Ibu Endah Rokhmati Merdika Putri, S.Si, MT, Ph.D, bapak Drs. Sentot Didik Surjanto, M.Si dan bapak Drs. Soetrisno, MI.Komp. selaku dosen penguji atas saran yang telah diberikan untuk Tugas Akhir ini.
6. Bapak Dr. Dieky Adzkiya, S.Si, M.Si selaku dosen wali yang telah memberikan arahan akademik selama penulis kuliah di Departemen Matematika FSAD-ITS.
7. Bapak dan Ibu dosen serta para staf Departemen Matematika FSAD-ITS yang tidak dapat penulis sebutkan satu-persatu.
8. Ibu Dewi Novita Sari yang mau direpotkan mengajarkan ilmu mengenai Regresi *Zero-Inflated Poisson* kepada penulis.
9. Sahabat penulis yaitu Qori' Afiata Fiddina, Rudat Ilaina dan Nirwana Fatria Kridayati yang selalu menemani, mendengarkan keluh resah penulis dan selalu mendukung penulis selama menjalani perkuliahan di Departemen Matematika FSAD-ITS.
10. Teman-teman mahasiswa Matematika ITS khususnya DOHMAIn yang telah memberikan masa-masa berkesan bagi penulis selama menjalani perkuliahan di Departemen Matematika FSAD-ITS.

11. Semua pihak yang tidak bisa penulis sebutkan satu-persatu, terimakasih sudah memberikan motivasi dan dukungan kepada penulis.

Penulis menyadari bahwa dalam penyusunan Tugas Akhir ini masih mempunyai banyak kekurangan. Oleh karena itu, kritik dan saran yang bersifat membangun sangat diharapkan oleh penulis. Semoga Tugas Akhir ini dapat bermanfaat bagi penulis dan semua pihak yang berkepentingan.

Surabaya, Januari 2020

Penulis

DAFTAR ISI

Halaman

Halaman

DAFTAR TABEL

Halaman

DAFTAR LAMPIRAN

DAFTAR GAMBAR

BAB I
PENDAHULUAN

Bab ini menjelaskan hal-hal yang menjadi latar belakang munculnya permasalahan yang dibahas dalam Tugas Akhir. Selanjutnya permasalahan tersebut disusun kedalam suatu rumusan masalah dan dari rumusan masalah digunakan untuk menentukan tujuan. Penjabaran batasan masalah dilakukan untuk memperoleh manfaat. Adapun sistematika penulisan Tugas Akhir ini diuraikan pada bagian bab ini.

1.1 Latar Belakang Masalah

Model statistik sering kali diselesaikan dengan distribusi Normal, tetapi untuk data *count* distribusi Normal tidak sesuai. Data *count* berkaitan dengan model jumlah kejadian yang mengacu pada suatu peristiwa yang terjadi, misalnya jumlah kasus penyakit, jumlah bencana alam, hasil pemungutan suara pada pemilihan umum, dan sebagainya. Jumlah kejadian merupakan realisasi dari variable acak *nonnegative integer-valued*. Sebuah model univariate dari jumlah kejadian menentukan distribusi probabilitas dari data *count* dengan beberapa parameter kejadian. Akan tetapi estimasi dan interface dalam model tersebut tidak diketahui paramaternya pada distribusi probabilitas dari data *count*. Spesifikasinya tidak melibatkan variabel lain dan jumlah kejadian diasumsikan independen [1]. Sehingga bermunculan analisis hubungan pada model statistik untuk data count seperti Regresi Poisson, Regresi *Binomial Negative*, Regresi *Zero-Inflated Poisson* (ZIP), dan masih banyak lagi.

Data count dengan banyak nilai *zero* di samping nilai-nilai *non-zero* yang besar sudah umum di berbagai data. Fenomena ini dapat ditangani oleh campuran dua komponen dimana salah satu komponen yang diambil untuk turunan distribusi, memiliki massa satu di nol. Komponen lainnya adalah distribusi non-degenerate seperti Poisson, bentuk Binomial atau lainnya, negatif tergantung pada situasi [2]. Permasalah "*excess zero*" sering terjadi pada aplikasi data count dibidang ekonomi kesehatan dan ilmu sosial. Proporsi dengan *zero count* pada sampel pengamatan sering lebih besar dari yang diperkirakan oleh model *standard count* [3]. Regresi *Zero-Inflated Poisson* (ZIP) adalah model untuk data count dengan *excess zero* [4]. Estimasi parameter model Regresi *Zero-Inflated Poisson* (ZIP) dapat menggunakan metode *Maximum Likelihood Estimation* (MLE), metode *Moment Estimation* (MME), metode *Bayesian Estimation* dan lain-lain.

Penelitian mengenai estimasi model *Zero-Inflated Poisson* (ZIP) pernah dilakukan oleh Lestari pada tahun 2008 [5]. Pada penelitiannya, Lestari menaksir nilai parameter untuk model Regresi *Zero-Inflated Poisson* menggunakan metode *Maximum Likelihood Estimation* (MLE) dan uji hipotesis dengan *Likelihood Ratio Test*. Lestari juga melakukan pemodelan data pekerja seks komersial di Klinik Reproduksi Putat Jaya Surabaya dengan banyaknya PSK yang menderita PMS *Trikomoniasis Vaginalis* di setiap rumah sebagai variabel respon. Selanjutnya, membandingkan antara pemodelan Regresi *Zero-Inflated Poisson* dan pemodelan Regresi Poisson dengan hasil yang menunjukkan bahwa pemodelan Regresi *Zero-Inflated Poisson* pada data lebih baik jika terdapat sebanyak 108 atau lebih respon yang bernilai nol dari 403 orang. Pada tahun 2015, Nusantara melakukan analisis jumlah penyakit Tetanus Neonatorum di Provinsi Jawa timur tahun 2012 dengan menggunakan *Geographically Weighted Zero-Inflated Poisson Regression* [6]. GWZIPR merupakan salah satu pengembangan dari Regresi *Zero-Inflated Poisson* yang menggunakan data koordinat geografis selain data utama. Nusantara juga menguji kesamaan model GWZIPR dan model Regresi *Zero-Inflated Poisson* dengan pendekatan distribusi F dan memberikan kesimpulan bahwa terdapat perbedaan anatara kedua model tersebut.

Penelitian tugas akhir ini mengkaji mengenai beberapa metode estimasi parameter untuk model Regresi *Zero-Inflated Poisson* (ZIP) yaitu metode *Maximum Likelihood Estimation* (MLE) dan metode *Moment Estimation* (MME) serta menerapkannya pada data jumlah kejadian meninggal kasus penyakit difteri di Indonesia tahun 2017 untuk 34 provinsi.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dituliskan, maka rumusan masalah pada Tugas Akhir ini adalah sebagai berikut:

1. Bagaimana menentukan estimasi parameter untuk model Regresi *Zero-Inflated Poisson* (ZIP) menggunakan metode *Maximum Likelihood Estimation* (MLE) dan metode *Moment Estimation* (MME)?
2. Bagaimana penerapan estimasi parameter untuk model Regresi *Zero-Inflated Poisson* (ZIP) pada data jumlah kejadian meninggal kasus penyakit difteri di Indonesia tahun 2017?

1.3 Tujuan

Berdasarkan rumusan masalah tersebut, tujuan penulisan Tugas Akhir ini adalah:

1. Menentukan estimasi parameter untuk model Regresi *Zero-Inflated Poisson* (ZIP) menggunakan metode *Maximum Likelihood Estimation* (MLE) dan metode *Moment Estimation* (MME).

2. Menerapkan estimasi parameter untuk model Regresi *Zero-Inflated Poisson* (ZIP) pada data jumlah kejadian meninggal kasus penyakit difteri di Indonesia tahun 2017.

1.4 Batasan Masalah

Ruang lingkup permasalahan yang dikaji dalam Tugas Akhir ini adalah:

1. Mendapatkan persamaan persamaan estimasi parameter untuk model model Regresi *Zero-Inflated Poisson* (ZIP) dengan metode *Maximum Likelihood Estimation* (MLE) dan metode *Moment Estimation* (MME).
2. Menggunakan data jumlah kejadian meninggal kasus penyakit difteri di Indonesia tahun 2017 sebagai variabel respon. Data yang digunakan adalah “Data penyakit Difteri pada Profil Kesehatan tahun 2017”. Data untuk variabel prediktor yaitu presentase Universal Child Immunisation, presentase penduduk miskin, presentase rumah tangga yang memiliki akses terhadap sanitasi layak, dan presentase gizi buruk.

1.5 Manfaat

Manfaat setelah diselesaikan Tugas Akhir ini adalah diperolehnya kajian mengenai estimasi parameter pada model Regresi *Zero-Inflated Poisson* sehingga dapat digunakan sebagai rujukan untuk kajian sejenisnya.

1.6 Sistematika Penulisan

Tugas Akhir ini secara garis besar memberikan gambaran mengenai keseluruhan isi Tugas Akhir ini, dikemukakan sistematika penulisan dalam Tugas Akhir ini yang terdiri dari lima bab yaitu sebagai berikut:

BAB I PENDAHULUAN

Bab ini menjelaskan tentang latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat dan sistematika penulisan hasil Tugas Akhir.

BAB II TINJAUAN PUSTAKA

Bab ini membahas mengenai penelitian terdahulu serta teori-teori yang mendukung dalam proses penyelesaian permasalahan dalam Tugas Akhir ini.

BAB III METODE PENELITIAN

Bab ini menjelaskan tentang tahapan-tahapan dalam proses menyelesaikan masalah dan mencapai tujuan Tugas Akhir.

BAB IV ANALISIS DAN PEMBAHASAN

Bab ini membahas mengenai kajian teoritis mengenai model Regresi Poisson dan Regresi *Zero-Inflated Poisson*, serta kajian terapan menggunakan jumlah kejadian meninggal kasus penyakit difteri di Indonesia tahun 2017.

BAB V PENUTUP

Bab ini menjelaskan kesimpulan yang diperoleh dari pembahasan dan saran untuk pengembangan penelitian selanjutnya.

BAB II
TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian mengenai estimasi model *Zero-Inflated Poisson* (ZIP) pernah dilakukan oleh Lestari pada tahun 2008 [5]. Pada penelitiannya, Lestari menaksir nilai parameter untuk model Regresi *Zero-Inflated Poisson* menggunkan metode *Maximum Likelihood Estimation* (MLE) dan uji hipotesis dengan *Likelihood Ratio Test*. Lestari jugamelakukan pemodelan data pekerja seks komersial di Klinik Reproduksi Putat Jaya Surabaya dengan banyaknya PSK yang menderita PMS *Trikomoniasis Vaginalis* di setiap rumah sebagai variabel respon. Selanjutnya, membandingkan antara pemodelan Regresi *Zero-Inflated Poisson* dan pemodelan Regresi Poisson dengan hasil yang menunjukkan bahwa pemodelan Regresi *Zero-Inflated Poisson* pada data lebih baik jika terdapat sebanyak 108 atau lebih respon yang bernilai nol dari 403 orang. Pada tahun 2015, Nusantara melakukan analisis jumlah penyakit Tetanus Neonatorum di Provinsi Jawa timur tahun 2012 dengan menggunakan *Geographically Weighted Zero-Inflated Poisson Regression* [6]. GWZIPR merupakan salah satu pengembangan dari Regresi *Zero-Inflated Poisson* yang menggunakan data koordinat geografis selain data utama. Nusantara juga menguji kesamaan model GWZIPR dan model Regresi *Zero-Inflated Poisson* dengan pendekatan distribusi F dan memberikan kesimpulan bahwa terdapat perbedaan anatara kedua model tersebut.

2.2 Distribusi Poisson

Distribusi Poisson merupakan bentuk distribusi dari percobaan yang menghasilkan nilai numerik bagi suatu peubah acak Y, yaitu banyaknya hasil percobaan yang terjadi dengan peluang yang sangat kecil dan bergantung pada interval waktu tertentu dengan hasil pengamatan berupa peubah diskrit [7].

Bilangan Y yang menyatakan banyaknya hasil percobaan dalam suatu percobaan Poisson disebut peubah acak Poisson dan sebaran peluangnya disebut sebaran Poisson. Nilai peluang dari percobaan Poisson bergantung pada μ, yaitu rata-rata dari banyaknya hasil percobaan yang terjadi selama selang waktu atau daerah yang diberikan, dapat dinotasikan dengan $f(y;\mu)$.

Definisi 2.1 Suatu peubah acak Y dikatakan berdistribusi Poisson jika memiliki *probability density function* (pdf) sebagai berikut [8]:

$$f(y;\mu) = \frac{e^{-\mu}(\mu)^y}{y!}, \qquad y = 0,1,2,\dots\infty \tag{2.1}$$

dimana $0 < \mu < \infty$ adalah parameter, $f(y;\mu)$ adalah peluang kejadian, y adalah banyaknya jumlah kejadian yang terjadi dalam suatu interval dan e adalah konstanta euler (e = 2.71828…).. Peubah acak Y berdistribusi Poisson dengan parameter μ dapat dinotasikan dengan $Y \sim POI(\mu)$.

Teorema 2.1 Jika peubah acak Y berdistribusi Poisson, $Y \sim POI(\mu)$ maka :

$$\begin{aligned} E(Y) &= \mu \\ Var(Y) &= \mu \\ M_Y(t) &= e^{\mu(e^t-1)} \end{aligned} \tag{2.2}$$

dimana $E(Y)$ adalah nilai ekspektasi (nilai harapan), $Var(Y)$ adalah varians dan $M_Y(t)$ adalah *Moment Generating Function* (MGF) dari peubah acak Y. Teorema 2.1 dapat dibuktikan pada Lampiran A.

2.3 Distribusi Binomial

Distribus Binomial adalah bentuk distribusi dari percobaan Binomial, yaitu suatu percobaan Bernouli yang diulang sebanyak n kali pengulangan dan setiap pengulangan saling bebas secara statistik terhadap pengulangan berikutnya. Sama halnya dengan percobaan Bernouli, pada setiap pengulangan percobaan Binomial memiliki salah satu peluang antara sukses atau gagal. Peluang sukses dari percobaan ini dapat dinotasikan dengan p, dan jika gagal dinotasikan dengan $q = 1 - p$. Ketika $n = 1$, distribusi Binomial adalah distribusi Bernouli. Distribusi Binomial digunakan untuk data diskrit yang dihasilkan dari percobaan Bernouli [8].

Definisi 2.2 Suatu peubah acak Y dikatakan distribusi Binomial dengan parameter p dan n jika memiliki *probability density function* (pdf) sebagai berikut :

$$f(y; p) = \binom{n}{y} p^y (1-p)^{n-y}, \qquad y = 0,1,2, \dots, n \qquad (2.3)$$

dimana $f(y)$ adalah peluang kejadian, y adalah banyaknya jumlah sukses dalam n kali perulangan, p adalah peluang sukses dengan $0 < p < 1$, dan $\binom{n}{y}$ adalah kombinasi atau koefisien binomial dengan,

$$\binom{n}{y} = \frac{n!}{y!\,(n-y)!}$$

Peubah acak Y berdistribusi Binomial dengan parameter p dan n dapat dinotasikan dengan $Y \sim BIN(n, p)$.

Teorema 2.2 Jika Y adalah peubah acak yang berdistribusi Binomial dengan parameter p dan n, $Y \sim BIN(n, p)$ maka :

$$E(Y) = np$$

$$Var(Y) = np(1-p)$$

$$M_Y(t) = [(1-p) + pe^t]^n \qquad (2.4)$$

dimana $E(Y)$ adalah nilai ekspektasi (nilai harapan), $Var(Y)$ adalah varians dan $M_Y(t)$ adalah *Moment Generating Function* (MGF) dari peubah acak Y. Teorema 2.2 dapat dibuktikan pada Lampiran B.

2.4 Distribusi *Zero-Inflated Poisson*

Distribusi *Zero-Inflated Poisson* merupakan pengembangan dari model ditribusi Poisson yang digunakan untuk mengidentifikasi peubah acak Y_i yang banyak memiliki nilai 0. Suatu peubah acak Y_i saling bebas dengan $i = 1, 2, 3, \dots, n$ dan [4]

$$Y_i \sim \begin{cases} 0 & \text{dengan } P(Y_i) = \pi_i \\ POI(\mu_i) & \text{dengan } P(Y_i) = 1 - \pi_i \end{cases} \tag{2.5}$$

dikatakan berdistribusi *Zero-Inflated Poisson* jika memiliki *probability density function* (pdf) sebagai berikut :

$$P(Y_i = y_i) = \begin{cases} \pi_i + (1 - \pi_i)e^{-\mu_i} & , \ y_i = 0 \\ (1 - \pi_i)\dfrac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} & , y_i > 0 \end{cases} \tag{2.6}$$

dengan parameter $\mu = (\mu_1, \dots, \mu_n)$ dan $\pi = (\pi_1, \dots, \pi_n)$.

Nilai ekspektasi dan varians dari Y_i dapat ditulis sebagai berikut,

$$\begin{aligned} E(Y_i) &= \mu_i(1 - \pi_i) \\ Var(Y_i) &= \mu_i(1 - \pi_i)(1 + \mu_i\pi_i) \end{aligned} \tag{2.7}$$

dengan pembuktian pada Lampiran C.

2.5 *Generalized Linear Model* (GLM)

Generalized Linear Model (GLM) adalah perluasan model regresi linier dengan distribusi variabel respon yang merupakan keluarga eksponensial dan memodelkan fungsi dari mean. Suatu model dikatakan *Generalized Linear Model* (GLM), apabila memenuhi tiga komponen, yaitu [9]:

1. Komponen Acak

Komponen acak yaitu komponen dari suatu variabel respon (Y) dengan pengamatan bebas $y_1, y_2, \ldots, y_n$ dari suatu distribusi dalam keluarga eskponensial.

2. Komponen Sistematis

Komponen sistematis yaitu komponen yang menghubungkan vektor $(\mu_1, \mu_2, \ldots, \mu_n)$ dengan variabel-variabel prediktor melalui model linier.

3. Fungsi Penghubung (*link function*)

Fungsi penghubung merupakan fungsi yang menghubungkan komponen acak dan komponen sistematis, dinotasikan dengan $h(.)$. Fungsi penghubung menggambarkan hubungan antara nilai ekspektasi dari variabel respon dengan prediktor linierya, dapat ditulis berikut ini:

$$\theta_i = h(\mu_i) = \sum_{p=1}^{k} \beta_p X_{pi} \ , i = 1,2,\ldots,n \qquad (2.8)$$

dengan k adalah jumlah parameter

2.5.1. *Exponential Family Distribution*

Exponential Family Distribution adalah ditribusi yang mencangkup atau merupakan kesatuan dari distribusi-distribusi yang sering digunakan sebagai kasus khusus. Misalkan Y peubah acak saling bebas, dapat dikatakan berdistribusi yang merupakan anggota *Exponential Family* jika mempunyai bentuk *probability density function* (pdf) bersama sebagai berikut [10],

$$f(y;\theta,\phi) = \exp\left[\frac{\left(y\theta - b(\theta)\right)}{a(\phi)} + c(y,\phi)\right] \qquad (2.9)$$

dimana $a(.)$, $b(.)$ dan $c(.)$ adalah beberapa fungsi.

1. Distribusi Poisson

Distribusi Poisson dapat ditulis sebagai kasus khusus dari distribusi *Exponential Family*, dengan fungsi peluang sebagai berikut,

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

$$f(y;\mu) = \exp[y\ln(\mu) - \mu - \ln(y!)] \qquad (2.10)$$

Persamaan diatas dapat dibandingkan dengan persamaan (2.6). Perhatikan bahwa $\theta = \ln(\mu)$ berarti $\mu = \exp(\theta)$ yang disubstitusi ke persamaan (2.7) dan diperoleh,

$$f(y;\mu) = \exp[y\theta - \exp(\theta) - \ln(y!)] \qquad (2.11)$$

Persamaan (2.7) merupakan salah satu kasus khusus dari persamaan (2.6), dengan $\theta = \ln(\mu)$, $b(\theta) = \exp(\theta)$, $c(y,\phi) = -\ln(y!)$ dan $a(\phi) = 1$.

2. Distribusi Binomial

Distribusi Binomial dapat ditulis sebagai berikut,

$$f(y;p) = \binom{n}{y} p^y (1-p)^{n-y}$$

$$f(y;p) = \exp\left[y\ln\left(\frac{p}{1-p}\right) + n\ln(1-p) + \ln\binom{n}{y}\right] \qquad (2.12)$$

dimana $\theta = \ln\left(\frac{p}{1-p}\right)$ yang berarti $p = \frac{\exp(\theta)}{1+\exp(\theta)}$ disubstitusikan ke persamaan (2.9) untuk mendapatkan

$$f(y;p) = \exp\left[y\theta + n\ln\left(\frac{1}{1+\exp(\theta)}\right) + \ln\binom{n}{y}\right] \qquad (2.13)$$

Terbukti bahwa distribusi binomial merupakan distribusi *Exponential Family* dengan $\theta = \ln\left(\frac{p}{1-p}\right)$, $b(\theta) = n\ln(1+\exp(\theta))$, $c(y,\phi) = \log\binom{n}{y}$ dan $a(\phi) = 1$.

2.5.2. Fungsi Penghubung (*link function*)

Fungsi penghubung antara rata-rata atau mean ke parameter alami disebut *canonical link*. Berikut merupakan *canonical link* untuk beberapa distribusi yang merupakan anggota dari *exponential family* [10],

Poisson : $\theta = \ln(\mu)$ memiliki invers $\mu = \exp(\theta)$ pada interval $[0, \infty]$ dengan *canonical link* adalah log link.

Binomial : $\theta = \ln\left(\frac{p}{1-p}\right)$ memiliki invers $p = \frac{\exp(\theta)}{1+\exp(\theta)}$ pada interval [0,1] dengan *canonical link* adalah logit link.

Normal : $\theta = \mu$ memiliki invers $\mu = \theta$ dengan *canonical link* adalah identity link.

2.6 Model Regresi Poisson

Model Regresi Poisson merupakan suatu model regresi yang menyatakan hubungan variabel respon dengan variabel prediktornya. Model tersebut menghubungkan variabel respon dan variabel prediktor melalui parameter yang dinamakan sebagai parameter regresi. Variabel respon merupakan variabel yang dipengaruhi oleh variabel prediktor, sedangkan variabel prediktor merupakan penyebab terjadinya variabel respon. Model Regresi Poisson adalah model standar yang digunakan untuk data diskrit dan termasuk dalam model linier. Model Regresi Poisson merupakan suatu bentuk model linier umum dimana variabel responnya dimodelkan sebagai distribusi Poisson. Distribusi Poisson memberikan suatu model yang realistis untuk berbagai fenomena acak selama nilai dari variabel acak Poisson berupa bilangan bulat non-negatif.

Model Regresi Poisson termasuk ke dalam *Generalized Linear Models* (GLM). Dalam *Generalized Linear Models* (GLM) terdapat sebuah fungsi h yang menghubungkan rata-rata dari variabel responnya dengan sebuah prediktor linier [1], sebagai berikut,

$$h = \theta_i = \ln \mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} \qquad (2.14)$$

fungsi h disebut fungsi penghubung (*link function*). Fungsi hubung yang digunakan pada model Regresi Poisson adalah fungsi penghubung log karena log menjamin bahwa nilai variabel yang diharapkan dari variabel responnya akan bernilai non-negatif. Fungsi penghubung yang digunakan untuk model Regresi Poisson dapat ditulis sebagai berikut,

$$\begin{aligned} \ln E(y|x) = &\quad \ln(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} \\ \mu_i = &\quad exp(\beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki}) = exp\ (\boldsymbol{X_i^T \beta}) \end{aligned} \qquad (2.15)$$

Atau dapat ditulis secara matriks berikut

$$\mu_i = \quad exp\left(\boldsymbol{X_i^T \beta}\right) = \quad exp\left[1\ x_{1i}\ x_{2i} \ldots x_{pi}\right] \begin{bmatrix} \beta \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \qquad (2.16)$$

dengan

μ_i : rata-rata jumlah peristiwa yang terjadi pada periode waktu tertentu

$\boldsymbol{X}_i$: vektor dari variabel prediktor, $\boldsymbol{X}_i^T = [x_{1i}, \ldots, x_{ki}]$

$\boldsymbol{\beta}$: vektor dari parameter regresi Poisson $\boldsymbol{\beta}^T = [\beta_0, \ldots, \beta_k]$

Sehingga model regresi Poisson adalah

$$y_i = exp\left(\boldsymbol{X_i^T \beta}\right) \qquad (2.17)$$

2.7 Model Regresi *Zero-Inflated Poisson*

Model Regresi *Zero-Inflated Poisson* pertama kali diperkenalkan oleh Diane Lambert pada tahun 1992. Model Regresi *Zero-Inflated Poisson* merupakan pengembangan dari model Regresi Poisson dimana variabel respon berdistribusi *Zero-Inflated Poisson*. Model Regresi *Zero-Inflated Poisson* menyatakan hubungan antara variabel respon dan variabel prediktornya [4].

Fungsi penghubung yang digunakan pada model Regresi *Zero-Inflated Poisson* adalah fungsi penghubung log dan logit, yaitu,

$$\ln(\mu) = \boldsymbol{X\beta} \tag{2.18}$$

dan

$$\text{logit}(\pi) = \ln\left(\frac{\pi}{1-\pi}\right) = \boldsymbol{X\gamma} \tag{2.19}$$

dengan $\boldsymbol{X}_i$ adalah matriks dari variabel prediktor, sedangkan $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ adalah vektor dari parameter yang akan ditaksir. Sehingga model Regresi *Zero-Inflated Poisson* adalah

$$\mu_i = exp(\boldsymbol{X_i'\beta}) \tag{2.20}$$

dan

$$\pi_i = \frac{\exp(\boldsymbol{X}_i^T\boldsymbol{\gamma})}{(1 + \exp(\boldsymbol{X}_i^T\boldsymbol{\gamma}))} \tag{2.21}$$

2.8 Metode Estimasi Parameter

Pada suatu data populasi biasanya nilai parameternya tidak diketahui, oleh karena itu untuk mengetahuinya dilakukan estimasi atau pendugaan terhadap parameter tersebut melalui data sampel. Berikut merupakan metode estimasi yang digunakan dalam penelitian tugas akhir ini.

2.8.1 Metode *Maximum Likelihood Estimation*

Metode *Maximum Likelihood Estimation* merupakan sebuah metode perkiraan untuk menemukan parameter-parameter yang memaksimumkan kemungkinan bersama (*joint likelihood*) dari sebuah data. Atau bisa dikatakan *Maximum Likelihood Estimation* adalah suatu metode yang memaksimumkan fungsi *likelihood*. *Maximum Likelihood Estimation* memiliki prinsip yaitu memilih $\hat{\theta}$ sebagai estimator titik untuk θ yang memaksimumkan $L(\theta)$, dimana $L(\theta)$ adalah fungsi *likelihood*. [11].

Definisi 2.3 Jika diberikan $Y_1, Y_2, \ldots, Y_n$ adalah sampel acak dari populasi Y yang memiliki *joint density* $f_\theta(Y_1, Y_2, \ldots, Y_n) = f(Y_1, Y_2, \ldots, Y_n|\theta)$, maka $Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n$ adalah nilai yang diamati sehingga fungsi *likelihood* dari θ sebagai berikut,

$$L(\theta) = \prod_{i=1}^{n} f(y_i;\ \theta) \tag{2.22}$$

Selanjutnya, untuk memaksimumkan fungsi *likelihood* digunakan fungsi *ln likelihood*,

$$\ln L(\theta) = l(\theta|Y_1, Y_2, \ldots, Y_n) = \sum_{i=1}^{n} \ln f(y_n|\theta) \tag{2.23}$$

Estimator *likelihood* diperoleh dari turunan fungsi *loglikelihood* $\ln L(\theta)$ terhadap parameternya, yaitu $\frac{d}{d\theta} \ln L(\theta) = 0$.

2.8.2 Metode *Moment Estimation*

Metode *Moment Estimation* merupakan metode yang digunakan untuk mendapatkan estimasi parameter populasi dengan menyamakan momen-momen populasi (teoritis) dengan momen-momen sampel. Jika diberikan $Y_1, Y_2, \ldots, Y_n$ adalah sampel acak dari populasi Y dengan *probability density function* (pdf) $f(y;\ \theta_1, \theta_2, \ldots, \theta_k)$, dimana $\theta_1, \theta_2, \ldots, \theta_k$ adalah parameter k yang tidak diketahui, dengan [8]

$$E(Y^k) = \int_{-\infty}^{\infty} y^k f(y;\ \theta_1, \theta_2, \ldots, \theta_k) dx \tag{2.24}$$

maka momen sampel ke- k dapat didefinisikan sebagi berikut,

$$M_k = \frac{1}{n}\sum_{i=1}^{n} Y_i^k \tag{2.25}$$

dan momen pupulasi ke- k dapat didefinisikan sebagai berikut,

$$\begin{aligned} E(Y) &= M_1 \\ E(Y^2) &= M_2 \\ &\vdots \\ E(Y^k) &= M_k \end{aligned} \tag{2.26}$$

Pada penelitian ini, kedua metode estimasi yang digunakan yaitu metode *Maximum Likelihood Estimation* dan metode *Moment Estimation* memiliki perbedaan salah satunya dalam langkah untuk mendapatkan nilai estimasi. Metode *Maximum Likelihood Estimation* membangun likelihood dari fungsi $f(y)$ kemuadian diubah ke dalam bentuk fungsi *loglikelihood.* Untuk estimasi parameter diperoleh dari turunan pertama fungsi *loglikelihood* terhadap parameternya yang disama dengankan dengan nol. Jika persamaan turunan pertama tidak analitik maka diselesaikan secara numerik. Sedangkan untuk metode *Moment Estimation*, estimasi parameter diperoleh dengan menyamakan momen sampel dan momen populasi diperoleh persamaan estimasi parameter yang analitik.

Metode *Maximum Likelihood Estimation* merupakan metode yang sering digunakan dan dianggap menjadi metode terbaik dalam mencari estimasi parameter karena estimasi parameter yang dihasilkan memiliki kemungkinan lebih besar dengan nilai yang diperkirakan. Namun, jika pada persamaan turunan pertama tidak analitik maka sulit untuk diselesaikan secara manual sehingga membutuhkan komputer untuk menyelesaikan secara numerik.

Metode *Moment Estimation* memiliki prosedur yang paling mudah dan cepat dalam memperoleh estimasi parameter atau bisa dikatakan metode yang cukup sederhana dan menghasilkan estimasi parameter yang konsisten. Namun, estimasi parameter yang dihasilkan dengan metode *Moment Estimation* seringkali bias dan pada beberapa kasus nilai estimasi parameter jauh dari yang dihasilkan.

2.9 Metode Newton-Rapshon

Metode Newton-Rapshon merupakan salah satu metode yang sering digunakan dalam optimasi statistika. Metode ini adalah sebuah aturan perulangan untuk menghasilkan nilai $\widehat{\boldsymbol{\beta}}$ dari $\boldsymbol{\beta}$ dengan memaksimumkan fungsi $l(\boldsymbol{\beta})$. Misalkan $\boldsymbol{\beta}^{(m)}$ adalah penaksiran ke- m dari $\widehat{\boldsymbol{\beta}}$, dimana $m = 0, 1, 2, \ldots$.. Metode Newton Rapshon membutuhkan nilai taksiran awal $\boldsymbol{\beta}^{(0)}$ sebagai nilai untuk memaksimumkan fungsi tersebut. Pada setiap langkah ke- m dalam proses perulangan fungsi $l(\boldsymbol{\beta})$ ditaksir dengan syarat sampai urutan kedua dalam perluasan deret Taylor terhadap $l(\boldsymbol{\beta})$, sebagai berikut, [12]

$$\begin{aligned} R^{(m)}(\boldsymbol{\beta}) = \quad & l\left(\boldsymbol{\beta}^{(m)}\right) + \boldsymbol{g}^{(m)}\left(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)}\right) \\ & + \frac{1}{2}\left(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)}\right)' \boldsymbol{H}^{(m)}\left(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)}\right) \\ & + \left(\left\|\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)}\right\|\right) \end{aligned} \tag{2.27}$$

Dalam hal ini $\boldsymbol{H}$ adalah matriks Hessian $k \times k$ yang merupakan turunan kedua dari $l(\boldsymbol{\beta})$ dengan memiliki elemen $\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_c \partial \boldsymbol{\beta}_j}$ dengan $c, j = 0,1,2 \ldots., k$ dan $\boldsymbol{g}$ adalah vektor gradien yang merupakan turunan pertama dari $l(\beta)$ dengan elemen $\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k}$. Sehingga dapat dikatakan bahwa $\boldsymbol{H}^{(m)}$ dan $\boldsymbol{g}^{(m)}$ adalah H dan g yang dihitung pada saat $\boldsymbol{\beta} = \boldsymbol{\beta}^{(m)}$. Penaksiran nilai $\hat{\beta}$ diperoleh dari nilai maksimum $R^{(m)}(\boldsymbol{\beta})$ dengan menyelesaikan

$\frac{\partial R^{(m)} l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \boldsymbol{\beta}^{(m)} + \boldsymbol{H}^{(m)}\left(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)}\right) = 0.$ Berikut untuk menghasilkan nilai penaksiran selanjutnya dari $\hat{\beta}$,

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} - \left(\boldsymbol{H}^{(m)}\right)^{-1} \boldsymbol{g}^{(m)} \tag{2.28}$$

2.10 Algoritma-EM

Algoritma-EM (*Expectation-Maximization*) adalah algoritma yang digunakan untuk menduga suatu parameter dalam suatu fungsi dengan menggunakan *Maximum Likelihood Estimation* (MLE), suatu fungsi tersebut berada pada kondisi dimana mengandung data yang tidak lengkap (*missing data* atau *incomplete data*). Pada Algoritma-EM terdiri dari dua tahapan dalam setiap iterasinya, yaitu tahap *Expectation* (E-Step) dan tahap *Maximization* (M-Step). Tahap *Expectation* (E-Step) adalah tahap perhitungan ekspektasi dan fungsi *ln likelihood* dengan memperhatikan data yang tidak lengkap. Sedangkan tahap *Maximization* (M-Step) merupakan tahap perhitungan untuk mecari penaksir parameter yang memaksimumkan fungsi *ln likelihood* dari hasil tahap *Expectation* sebelumnya. Parameter-parameter yang dihasilkan dari M-Step akan digunakan kembali untuk E-Step selanjutnya. Langkah ini akan terus diulang hingga tercapai suatu nilai yang konvergen. [13]

Berikut merupakan langkah-langkah Algoritma-EM untuk penaksiran parameter.

1. Menentukan inisialisasi parameter $\hat{\theta}_k;\ k = 0$
2. Langkah *Expectation* :
   Menghitung *complete data likelihood* dengan cara substitusi $\hat{\theta}_k$ pada fungsi $Q\left(\hat{\theta}_k\right) = E\left[L\left(\hat{\theta}_k \middle| x, y\right) \middle| x\right]$
3. Langkah *Maximization*:

Dilakukan dengan mengacu pada kondisi $\partial Q(\hat{\theta}_k)/\partial\hat{\theta}_k = 0$ untuk mendapatkan inisialisasi yang baru yaitu $\hat{\theta}^{(k+1)}$

4. Langkah E dan M dilakukan secara iteratif sampai diperoleh nilai $\left|\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)}\right| < \varepsilon$.

2.11 Multikolinieritas

Adanya kolerasi antar variabel prediktor dalam model regresi linear atau yang biasa disebut dengan multikolinearitas, akan menyebabkan error yang besar pada pendugaan parameter regresi. Untuk itu perlu dilakukan uji multikolinearitas yang dapat diketahui melalui nilai koefisien *Pearson* (r_{ij}) antar variabel prediktor yang lebih besar dari 0.95. Selain itu adanya kasus mutikolinearitas dapat juga diketahui melalui *Variance Inflation Factors* (VIF) yang bernilai lebih besar dari 10, dengan nilai VIF yang dinyatakan sebagai berikut [14]:

$$VIF_k = \frac{1}{1 - R_k^2} \tag{2.29}$$

dengan

$$R_k^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - \widehat{x_\iota})^2}{\sum_{i=1}^{n}(x_i - \overline{x_\iota})^2} \tag{2.30}$$

dan R_k^2 adalah koefisien determinasi yang diperoleh dari hasil regresi antara x_k dengan variabel pediktor lainnya, $\bar{x}$ adalah *mean* dari variabel prediktor sebagai variabel respon, $\widehat{x_\iota}$ adalah nilai prediksi variabel prediktor ke-i, dan x_i adalah nilai observasi variabel prediktor ke-i dengan $i = 1, 2, 3, \dots, n$ dan $k = 1, 2, 3, 4$. Jika terdapat kasus multikolinearitas maka untuk mengatasinya menggunakan *Principal Component Analiysis* (PCA), yaitu dengan membentuk komponen-komponen utama sebagai variabel prediktor baru yang merupakan kombinasi linear dari variabel-variabel prediktor sebelumnya.

2.12 Uji Parameter Serentak

Untuk mengetahui ada atau tidaknya pengaruh variabel prediktor terhadap variabel respon, dilakukan pengujian parameter serentak menggunakan satistik uji *Likelihood Ratio Test* (LRT) dengan hipotesa sebagai berikut [1].

H_0 : $\beta_1 = \beta_2 = \ldots = \beta_k = 0$, yaitu tidak ada variabel yang berpengaruh secara signifikan

H_1 : $\beta_j \neq 0$, yaitu paling sedikit ada satu variabel yang berpengaruh secara signifikan

dimana $i = 1, 2, 3, \ldots, k$. Berikut merupakan notasi dari *Likelihood Ratio*

$$\Lambda = \frac{L(\widehat{\omega})}{L(\widehat{\Omega})} \tag{2.31}$$

dengan:

$L(\widehat{\omega})$: nilai *likelihood* pada model dibawah kondisi H_0, tanpa melibatkan variabel prediktor

$L(\widehat{\Omega})$ **:** nilai *likelihood* dengan melibatkan variabel prediktor

Namun, karena regresi Poisson merupakan salah satu keluarga eksponensial maka satistik uji *Likelihood Ratio* juga dapat ditulis dalam bentuk berikut:

$$G = -2 \ ln \ \Lambda = -2 \ ln \ \frac{L(\widehat{\omega})}{L(\widehat{\Omega})} = -2 \ ln \left[L(\widehat{\omega}) - L(\widehat{\Omega})\right] \tag{2.32}$$

Pengujian parameter ini memiliki kriteria yaitu tolak H_0 jika $G > \chi^2_{(\alpha,k)}$ yang menyatakan bahwa paling sedikit ada satu parameter yang berpengaruh secara signifikan terhadap model, dimana α adalah tingkat signifikasi dan k adalah banyakknya parameter.

2.13 Uji Parameter Parsial

Uji Parameter parsial digunakan untuk menguji apakah masing-masing variabel prediktor berpengaruh signifikan terhadap variabel respon. Hipotesis yang digunakan untuk uji parameter parsial adalah sebagai berikut [1]..

H_0 : $\beta_j = 0,\ j = 1,2,3,\dots,k$ (pengaruh variabel ke- j tidak signifikan)

H_1 : $\beta_j \neq 0$ (pengaruh variabel ke- j signifikan)

Statistika uji yang digunakan adalah uji Z,

$$\mathrm{Z_{hit}} = \frac{\widehat{\beta_j}}{\sqrt{Var(\boldsymbol{\beta})}}$$

Pengujian parameter ini memiliki kriteria yaitu tolak H_0 jika $|\mathrm{Z_{hit}}| \geq |\mathrm{Z}_\alpha|$ yang menyatakan bahwa variabel prediktor yang diuji berpengaruh terhadap variabel respon, dimana α adalah tingkat signifikasi dan k adalah banyakknya parameter.

2.14 Pemilihan Model Terbaik

Salah satu pemilihan model terbaik yang bisa digunakan adalah *Akaike's Information Criteria* (AIC) dengan rumus sebagai berikut [15]:

$$\mathrm{AIC} = -2\ln L(\hat{\beta},\hat{\gamma}) + 2k \qquad (2.33)$$

dengan k adalah jumlah parameter yang terdapat pada model terbaik dengan menggunakan Regresi Poisson dan Regresi *Zero-Inflated Poisson* (ZIP) adalah dengan nilai AIC terkecil.

BAB III
METODE PENELITIAN

Bab ini menjelaskan langkah-langkah sistematis yang digunakan dalam mengerjakan Tugas Akhir. Metode penelitian dalam Tugas Akhir terdiri dalam dua tahapan utama, yaitu kajian teoritis dan kajian terapan untuk model Regresi Poisson dan model Regresi *Zero-Inflated Poisson*. Rincian dari ketiga tahapan dalam menggerjakan Tugas Akhir, antara lain:

3.1 Kajian Teoritis tentang Model Regresi Poisson dan Model Regresi *Zero-Inflated Poisson*

Pengkajian model Regresi Poisson dan model *Zero-Inflated Poisson* dilakukan dengan penaksiran nilai estimasi parameter untuk masing-masing model menggunakan metode *Maximum Likelihood Estimation* (MLE) dan metode *Moment Estimation* (MME).

3.1.1 Kajian Estimasi Parameter Model Regresi Poisson

Penaksiran estimasi untuk model Regresi Poisson dilakukan dengan dua metode sebagai berikut.

1. Estimasi parameter model Regresi Poisson dengan metode *Maximum Likelihood Estimation* (MLE). Berikut merupakan langkah-langkah dilakukan:
   a. Membuat fungsi *likelihood* berdasarkan model Regresi Poisson
   b. Membuat fungsi *ln likelihood* berdasarkan model Regresi Poisson
   c. Mendapatkan turunan pertama dan turunan kedua dari fungsi *ln likelihood* di atas
   d. Melakukan iterasi menggunakan metode Newton-Rapshon

2. Estimasi parameter model Regresi Poisson dengan metode *Moment Estimation* (MME)

3. Pengujian parameter untuk model Regresi Poisson dengan hipotesis-hipotesis berikut ini:

   a. Uji parameter serentak menggunakan uji *Likelihood Ratio Test* (LRT)

      $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$

      H_1 : paling sedikit ada satu $\beta_i \neq 0$

      dengan $i = 1, 2, 3, \ldots, k$

   Berikut ini merupakan langkah-langkah dalam pengujian hipotesis:

   - Membuat himpunan parameter dibawah populasi (Ω):

   $$\Omega = \{\beta\}$$

   - Membuat himpunan parameter di bawah $H_0(\omega)$:

   $$\omega = \{\beta_0\}$$

   - Membuat fungsi likelihood di bawah populasi $\left(L(\Omega)\right)$:

   $$L(\Omega) = \prod_{i=1}^{n} f(y_i; \boldsymbol{\beta})$$

   - Membuat fungsi likelihood di bawah H_0 $\left(L(\omega)\right)$:

   $$L(\omega) = \prod_{i=1}^{n} f(y_i; \beta_0)$$

   - Menentukan $L\left(\hat{\Omega}\right) = \max_{\Omega} L(\Omega)$
   - Menentukan $L(\hat{\omega}) = \max_{\omega} L(\omega)$
   - Membuat ratio $\frac{L(\hat{\omega})}{L(\hat{\Omega})}$
   - Tolak H_0 jika $\frac{L(\hat{\omega})}{L(\hat{\Omega})} < \mu_0 < 1$, dimana $0 < \mu_0 < 1$
   - Membuat $G = -2 \ln \left[\frac{L(\hat{\omega})}{L(\hat{\Omega})}\right]$

   Tolak H_0 apabila $G_{\text{hitung}} > \chi^2_{(\alpha,k)}$

b. Uji parsial parameter $\boldsymbol{\beta}$ menggunakan uji Z
$H_0 : \beta_i = 0$
$H_1 : \beta_i \neq 0$
dengan $i = 1, 2, 3, \dots, k$
Menghitung nilai $\mathrm{Z_{hit}}$ untuk setiap parameter β_i
Terjadi tolak H_0 jika $|\mathrm{Z_{hit}}| > |\mathrm{Z_\alpha}|$

3.1.2 Kajian Estimasi Parameter Model Regresi *Zero-Inflated Poisson*

Penaksiran estimasi parameter model Regresi *Zero-Inflated Poisson* menggunakan dua metode, yaitu:

1. Estimasi parameter untuk model Regresi *Zero-Inflated Poisson* dengan metode *Maximum Likelihood Estimation* (MLE). Berikut merupakan langkah-langkah dilakukan:
   c. Membuat fungsi *likelihood* berdasarkan model Regresi *Zero-Inflated Poisson*
   d. Membuat fungsi *ln likelihood* berdasarkan model Regresi *Zero-Inflated Poisson*
   e. Mendefinisikan variabel Z_i sebagai berikut:
   f. Membuat fungsi *likelihood* dengan variabel Z_i, $i = 1,2, \dots, n$
   g. Membuat fungsi *ln likelihood* dengan variabel Z_i, $i = 1,2, \dots, n$
   h. Menaksir parameter β dan γ dengan memaksimumkan fungsi *ln likelihood* yang diperoleh di atas menggunakan Algoritma-EM.
   i. Melakukan iterasi menggunakan metode Newton-Rapshon

2. Estimasi parameter model Regresi Regresi *Zero-Inflated Poisson* dengan metode *Moment Estimation* (MME)

3. Pengujian parameter untuk model Regresi *Zero-Inflated Poisson* dengan hipotesis-hipotesis berikut ini:
   a. Uji parameter serentak menggunakan uji *Likelihood Ratio Test* (LRT)
      $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = \gamma_1 = \gamma_2 = \ldots = \gamma_k = 0$
      H_1 : paling sedikit ada satu $\beta_i \neq 0$ atau $\gamma_i \neq 0$
      dengan $i = 1, 2, 3, \ldots, k$

   Berikut ini merupakan langkah-langkah dalam pengujian hipotesis:

   - Membuat himpunan parameter dibawah populasi (Ω):

   $$\Omega = \{\beta, \gamma\}$$

   - Membuat himpunan parameter di bawah $H_0(\omega)$:

   $$\omega = \{\beta_0, \gamma_0\}$$

   - Membuat fungsi likelihood di bawah populasi $\left(L(\Omega)\right)$:

   $$L(\Omega) = \prod_{i=1}^{n} f(y_i; \boldsymbol{\beta}, \boldsymbol{\gamma})$$

   - Membuat fungsi likelihood di bawah H_0 $\left(L(\omega)\right)$:

   $$L(\omega) = \prod_{i=1}^{n} f(y_i; \beta_0, \gamma_0)$$

   - Menentukan $L(\hat{\Omega}) = \max_{\Omega} L(\Omega)$
   - Menentukan $L(\hat{\omega}) = \max_{\omega} L(\omega)$
   - Membuat ratio $\frac{L(\hat{\omega})}{L(\hat{\Omega})}$
   - Tolak H_0 jika $\frac{L(\hat{\omega})}{L(\hat{\Omega})} < \mu_0 < 1$, dimana $0 < \mu_0 < 1$
   - Membuat $G = -2 \ln \left[\frac{L(\hat{\omega})}{L(\hat{\Omega})}\right]$

   Tolak H_0 apabila $G_{\text{hitung}} > \chi^2_{(\alpha,k)}$

b. Uji parsial parameter $\boldsymbol{\beta}$ menggunakan uji Z
$H_0 : \beta_i = 0$
$H_1 : \beta_i \neq 0$
dengan $i = 1, 2, 3, \dots, k$
Menghitung nilai $\mathrm{Z_{hit}}$ untuk setiap parameter β_i
Terjadi tolak H_0 jika $|\mathrm{Z_{hit}}| > |\mathrm{Z}_\alpha|$

c. Uji parsial parameter $\boldsymbol{\gamma}$ menggunakan uji Z
$H_0 : \gamma_i = 0$
$H_1 : \gamma_i \neq 0$
dengan $i = 1, 2, 3, \dots, k$
Menghitung nilai $\mathrm{Z_{hit}}$ untuk setiap parameter γ_i
Terjadi tolak H_0 jika $|\mathrm{Z_{hit}}| > |\mathrm{Z}_\alpha|$

3.2 Kajian Terapan Menggunakan Data Jumlah Meninggal Kasus Difteri di Indonesia Tahun 2017

Data yang digunakan pada penelitian ini adalah data sekunder yang berasal dari data Profil Kesehatan Indonesia tahun 2017 yang dipublikasikan oleh Kementrian Kesehatan Republik Indonesia. Penelitian ini menggunakan dua variabel penelitian yaitu variabel respon (Y) dan variabel prediktor (X) yang disajikan pada Tabel 3.1.

Tabel 3.1 Variabel Penelitian

Variabel	Keterangan
Y	Jumlah Meninggal Kasus Difteri
X_1	Presentase Universal Child Immunisation
X_2	Presentase Penduduk Miskin
X_3	Presentase Rumah Tangga yang Memiliki Akses terhadap Sanitasi Layak
X_4	Presentase Gizi Buruk

3.2.1 Pemodelan Model Regresi Poisson untuk Data Jumlah Kejadian Meninggal Kasus Penyakit Difteri di Indonesia Tahun 2017

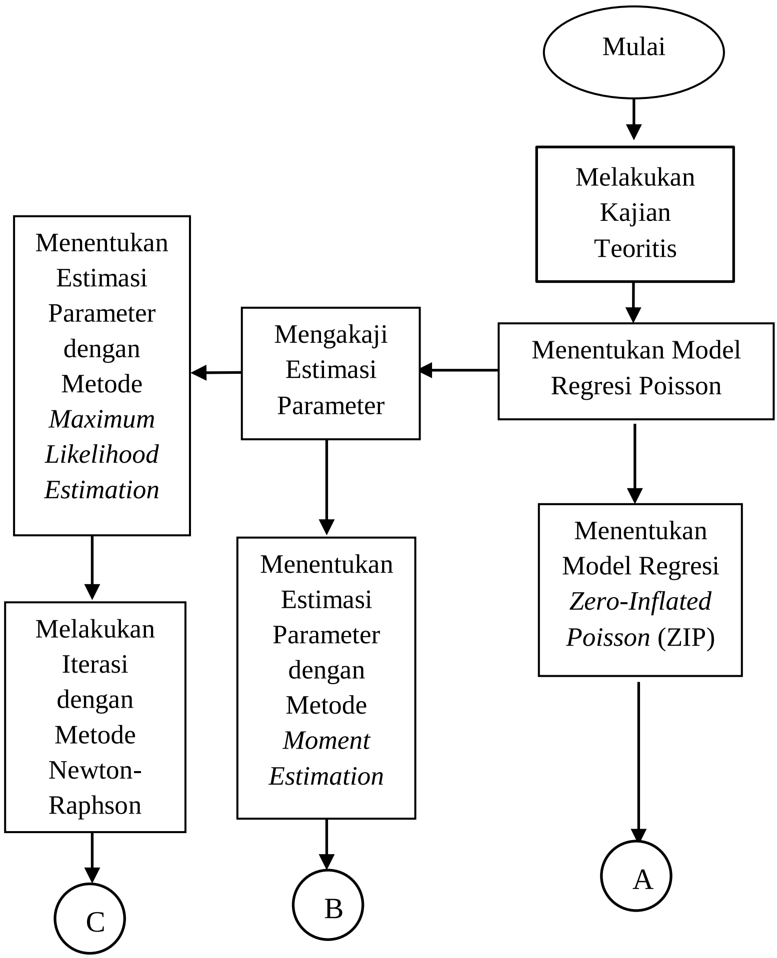
1. Mencari nilai estimasi parameter model Regresi Poisson dengan menggunakan metode *Maximum Likelihood Estimation* (MLE).
   - Mencari nilai awal parameter $\boldsymbol{\beta}$
   - Membentuk vektor gradien dan matriks *Hessian*
   - Melakukan iterasi menggunakan metode Newton-Raphson sampai parameter konvergen
   - Membentuk model Regresi Poisson
2. Mencari nilai estimasi parameter dan membentuk model Regresi Poisson dengan menggunakan metode *Moment Estimation* (MME).
3. Melakukan uji serentak parameter dan uji parsial parameter model Regresi Poisson.

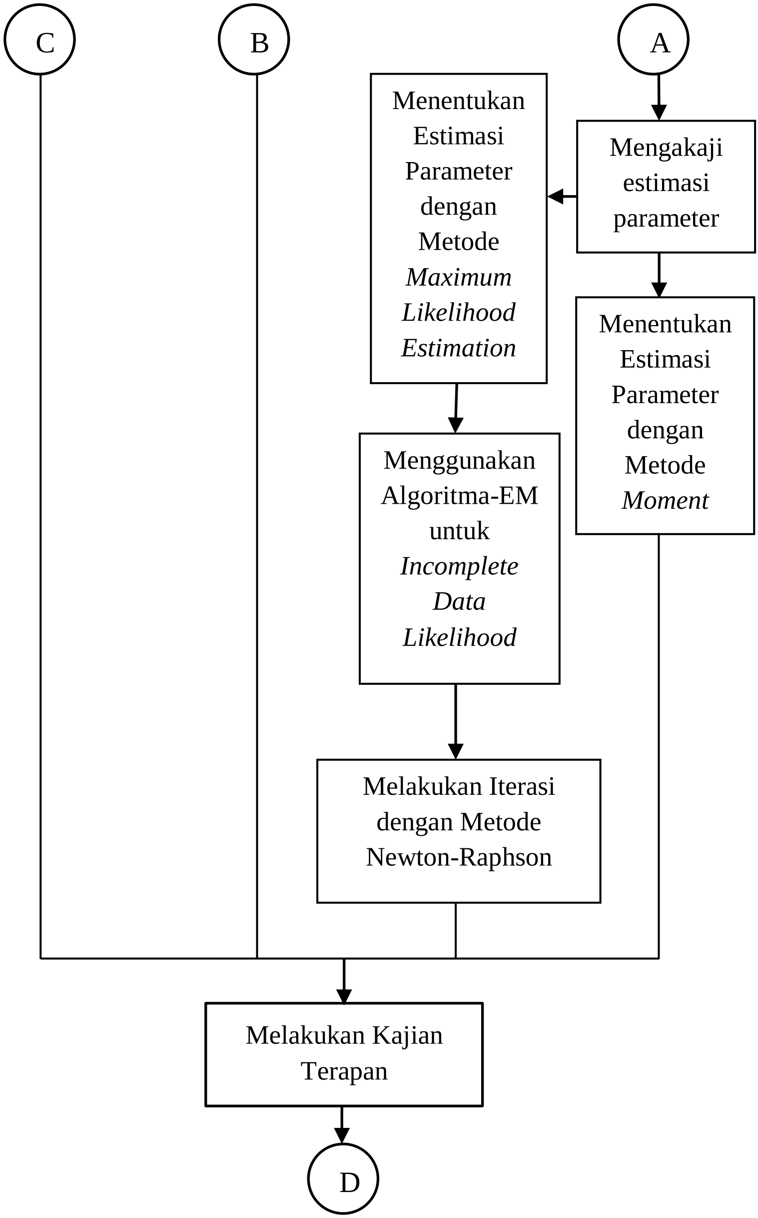
3.2.2 Pemodelan Model Regresi *Zero-Inflated Poisson* untuk Data Jumlah Kejadian Meninggal Kasus Penyakit Difteri di Indonesia Tahun 2017

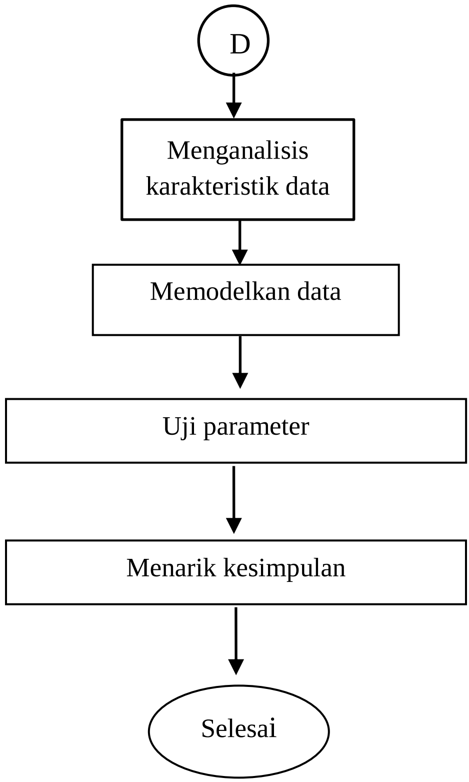
1. Mencari nilai estimasi parameter model Regresi *Zero-Inflated Poisson* dengan menggunakan metode *Maximum Likelihood Estimation* (MLE).
   - Mencari nilai awal parameter $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$
   - Membentuk vektor gradien dan matriks *Hessian* untuk masing-masing parameter
   - Melakukan iterasi menggunakan metode Newton-Raphson sampai kedua parameter konvergen
   - Membentuk model Regresi *Zero-Inflated Poisson*

2. Mencari nilai estimasi parameter dan membentuk model Regresi Poisson dengan menggunakan metode *Moment Estimation* (MME).
3. Melakukan uji serentak parameter dan uji parsial parameter model Regresi *Zero-Inflated Poisson.*

3.3 Diagram Alir




C
B
A
Mengakaji estimasi parameter
Menentukan Estimasi Parameter dengan Metode *Maximum Likelihood Estimation*
Menentukan Estimasi Parameter dengan Metode *Moment*
Menggunakan Algoritma-EM untuk *Incomplete Data Likelihood*
Melakukan Iterasi dengan Metode Newton-Raphson
Melakukan Kajian Terapan
D



Gambar 3.1 Diagram Alir

BAB IV
ANALISIS DAN PEMBAHASAN

Bab ini membahas mengenai kajian dalam penaksiran estimasi parameter dari model Regresi Poisson dan model Regresi *Zero-Inflated Poisson*. Penaksiran parameter dilakukan dengan menggunakan dua metode, yaitu metode *Maximum Likelihood Estimation* (MLE), dan metode *Moment Estimation*. Setelah didapatkan estimasi parameter, dilakukan penerapan estimasi parameter model Regresi Poisson dan model Regresi *Zero-Inflated Poisson* pada data.

4.1 Kajian Teoritis untuk Model Regresi Poisson

Pada sub bab ini membahas tentang analisis estimasi parameter untuk model Regresi Poisson. Telah dijelaskan pada bab 2 bahwa peubah acak Y dikatakan berdistribusi Poisson jika memiliki *probability density function* (pdf) seperti pada persamaan (2.1),

$$f(y|\mu) = \frac{e^{-\mu}(\mu)^{y}}{y!}, \qquad y = 0, 1, 2, \dots, \infty \tag{4.1}$$

Model Regresi Poisson merupakan model regresi dengan variabel respon yang berdistribusi Poisson. Model Regresi Poisson didefinisikan seperti persamaan (2.16) adalah,

$$\ln(\mu_i) = \boldsymbol{X}_i^T \boldsymbol{\beta} \text{ atau } \mu_i = \exp(\boldsymbol{X}_i^T \boldsymbol{\beta}) \tag{4.2}$$

dengan fungsi linknya adalah fungsi log link.

Berikut dilakukan analisis estimasi parameter model Regresi Poisson dengan menggunakan dua metode estimasi, yaitu

a. Metode *Maximum Likelihood Estimation* (MLE)
b. Metode *Moment Estimation* (MME)

4.1.1 Estimasi Parameter Model Regresi Poisson dengan Metode *Maximum Likelihood Estimation* (MLE)

Estimasi parameter menggunakan metode *Maximum Likelihood Estimation* (MLE) dilakukan dengan membangun *likelihood* dari fungsi Regresi Poisson. Adapun langkah-langkah dalam mengestimasi parameter dengan metode *Maximum Likelihood Estimation* (MLE).

1. Ambil sampel acak $y_1, y_2, \ldots, y_n$ dengan y_i berdistribusi Poisson atau $y_i \sim P(\mu_i)$, maka bentuk model Regresi Poisson diperoleh dari substitusi persamaan (4.2) ke persamaan (4.1) sebagai berikut:

$$f(y_1, y_2, \ldots, y_n, \mu) = \frac{e^{-\mu}(\mu)^y}{y!}$$

$$f(y_1, y_2, \ldots, y_n, \boldsymbol{\beta}) = \frac{e^{-e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}}\left(e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}\right)^y}{y!} \tag{4.3}$$

2. Setelah diperoleh bentuk model Regresi Poisson, kemudian persamaan (4.3) diubah ke bentuk fungsi *likelihood* model Regresi Poisson,

$$L(y_1, y_2, \ldots, y_n, \boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{e^{-e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}}\left(e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}\right)^y}{y!}$$

$$L(y_1, y_2, \ldots, y_n, \boldsymbol{\beta}) = \frac{\prod_{i=1}^{n} e^{-e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}} \prod_{i=1}^{n}\left(e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}\right)^{y_i}}{\prod_{i=1}^{n} y_i!} \tag{4.4}$$

3. Bentuk fungsi *likelihood* model Regresi Poisson pada persamaan (4.4) diubah menjadi fungsi *ln likelihood* sebagai berikut,

$$\ln L(y_i, \boldsymbol{\beta}) = \ln\left(\frac{\prod_{i=1}^{n} e^{-e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}} \prod_{i=1}^{n}\left(e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}\right)^{y_i}}{\prod_{i=1}^{n} y_i!}\right)$$

$$\ln L(y_i, \boldsymbol{\beta}) = -\sum_{i=1}^{n} e^{\boldsymbol{X}_i^T \boldsymbol{\beta}} + \sum_{i=1}^{n} y_i \ln\left(e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}\right) - \sum_{i=1}^{n} \ln(y_i!)$$

$$\ln L(y_i, \boldsymbol{\beta}) = -\sum_{i=1}^{n} e^{\boldsymbol{X}_i^T \boldsymbol{\beta}} + \sum_{i=1}^{n} y_i \boldsymbol{X}_i^T \boldsymbol{\beta} - \sum_{i=1}^{n} \ln(y_i!)$$

$$\ln L(y_i, \boldsymbol{\beta}) = \sum_{i=1}^{n} \left(-e^{\boldsymbol{X}_i^T \boldsymbol{\beta}} + y_i \boldsymbol{X}_i^T \boldsymbol{\beta} + \ln(y_i!)\right) \tag{4.5}$$

4. Mendapatkan $\widehat{\boldsymbol{\beta}}$ yang memaksimalkan fungsi *ln likelihood* dengan mencari turunan pertama dari persamaan (4.5) dan menyamakan dengan nol,

$$\frac{\partial \ln L(y, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0$$

$$\frac{\partial\left(-\sum_{i=1}^{n} e^{\boldsymbol{X}_i^T \boldsymbol{\beta}} + \sum_{i=1}^{n} y_i \boldsymbol{X}_i^T \boldsymbol{\beta} - \sum_{i=1}^{n} \ln(y_i!)\right)}{\partial \beta} = 0$$

$$\sum_{i=1}^{n} \left(-e^{\boldsymbol{X}_i^T \boldsymbol{\beta}} + y_i\right)(\boldsymbol{X}_i^T) = 0 \tag{4.6}$$

dengan $\widehat{\boldsymbol{\beta}}$ merupakan nilai taksiran dari $\boldsymbol{\beta}$.

Hasil penurunan pertama fungsi *ln likelihood* model Regresi Poisson pada persamaan (4.6) tidak menunjukkan adanya solusi analitik untuk parameter $\widehat{\boldsymbol{\beta}}$. Oleh karena itu perlu menggunakan prosedur iteratif numerik, yaitu dengan menggunakan metode Newton-Raphson.

Metode Newton-Raphson merupakan salah satu metode yang umum digunakan untuk menghitung taksiran nilai $\boldsymbol{\beta}$. Berikut adalah rumus iterasi dari metode Newton-Raphson pada persamaan (2.28):

$$\widehat{\boldsymbol{\beta}}^{(m+1)} = \widehat{\boldsymbol{\beta}}^{(m)} - \boldsymbol{H}^{-1} \boldsymbol{g}\left(\widehat{\boldsymbol{\beta}}^{(m)}\right) \tag{4.7}$$

dengan

$\widehat{\boldsymbol{\beta}}^{(m)}$: vektor parameter regresi pada iterasi ke-m, $m = 0,1,2,\ldots$

$\widehat{\boldsymbol{\beta}}^{(m+1)}$: vektor parameter regresi pada iterasi ke $m+1$

$\boldsymbol{g}\left(\widehat{\boldsymbol{\beta}}^{(m)}\right)$: vektor gradient yang dievaluasi pada $\widehat{\boldsymbol{\beta}}^{(m)}$

$\boldsymbol{H}$: matriks Hessian yang dievaluasi pada $\widehat{\boldsymbol{\beta}}^{(m)}$

dan

$$\boldsymbol{g}\left(\widehat{\boldsymbol{\beta}}^{(m)}\right) = \left.\frac{\partial L(y,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\widehat{\beta}^{(m)}} \quad (4.8)$$

$$\boldsymbol{H} = \left.\frac{\partial^2 L(y,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T}\right|_{\widehat{\beta}^{(m)}} \quad (4.9)$$

Algoritma dari metode Newton-Raphson untuk mendapatkan estimasi parameter model Regresi Poisson adalah sebagai berikut.

1. Menentukan nilai taksiran awal parameter $\widehat{\boldsymbol{\beta}}$, yaitu $\widehat{\boldsymbol{\beta}}^{(0)}$ dengan metode Ordinary Least Square (OLS) :

$$\widehat{\boldsymbol{\beta}}^{(\mathbf{0})} = \left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\right)^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}^{*} \quad (4.10)$$

dengan

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix}; \boldsymbol{Y}^{*} = \begin{bmatrix} \ln y_1 \\ \ln y_2 \\ \vdots \\ \ln y_n \end{bmatrix} \text{ dan } \widehat{\boldsymbol{\beta}}^{(0)} = \begin{bmatrix} \hat{\beta}_0^{(0)} \\ \hat{\beta}_1^{(0)} \\ \vdots \\ \hat{\beta}_k^{(0)} \end{bmatrix}$$

2. Membentuk vektor gradien g berukuran $(k+1)\times 1$ untuk nilai $\widehat{\boldsymbol{\beta}}^{(\mathbf{0})}$ yaitu:

$$\boldsymbol{g}\left(\widehat{\boldsymbol{\beta}}^{0}\right) = \begin{bmatrix} \dfrac{\partial \ln L(Y,\boldsymbol{\beta})}{\partial \beta_0} \\ \dfrac{\partial \ln L(Y,\boldsymbol{\beta})}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial \ln L(Y,\boldsymbol{\beta})}{\partial \beta_k} \end{bmatrix}_{\beta=\widehat{\beta}^{(0)}} \quad (4.11)$$

dengan k adalah jumlah parameter yang ditaksir.

3. Membentuk matrik Hessian H berukuran$(k+1)\times(k+1)$, yang bersifat simetris. Matriks Hessian H disebut juga sebagai matriks informasi.

$$\boldsymbol{H}\left(\widehat{\boldsymbol{\beta}}^{(0)}\right)=\begin{bmatrix}\frac{\partial^2 \ln L(Y,\boldsymbol{\beta})}{\partial\beta_0^2} & \cdots & \frac{\partial^2 \ln L(Y,\boldsymbol{\beta})}{\partial\beta_0\partial\beta_k}\\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln L(Y,\boldsymbol{\beta})}{\partial\beta_k\partial\beta_0} & \cdots & \frac{\partial^2 \ln L(Y,\boldsymbol{\beta})}{\partial\beta_k^2}\end{bmatrix}_{\beta=\widehat{\beta}^{(0)}} \quad (4.12)$$

4. Melakukan iterasi dengan rumus pada persamaan (4.7) untuk mendapatkan $\widehat{\boldsymbol{\beta}}^{(m+1)}$

$$\widehat{\boldsymbol{\beta}}^{(m+1)}=\widehat{\boldsymbol{\beta}}^{(m)}-\boldsymbol{H}^{-1}\boldsymbol{g}\left(\widehat{\boldsymbol{\beta}}^{(m)}\right), \qquad m=0,1,2,\dots,n$$

5. Iterasi berhenti ketika parameter konvergen dimana memenuhi kondisi,

$$\left\|\widehat{\boldsymbol{\beta}}^{(m+1)}-\widehat{\boldsymbol{\beta}}^{(m)}\right\|\leq\boldsymbol{\varepsilon} \quad (4.13)$$

dimana $\boldsymbol{\varepsilon}$ adalah nilai toleransi error yang diterapkan, berupa nilai yang sangat kecil dan telah ditetapkan sebelumnya. Atau memenuhi kondisi dimana nilai parameter mendekati ke satu titik konstan.

Jika belum didapatkan penaksir parameter yang konvergen, maka dilajutkan kembali ke langkah 4. Kondisi pada persamaan (4.13) menunjukkan bahwa $\widehat{\boldsymbol{\beta}}^{(m+1)}$ konvergen ke $\widehat{\boldsymbol{\beta}}$. Apabila telah memenuhi kondisi pada persamaan (4.13) maka $\widehat{\boldsymbol{\beta}}^{(m+1)}$ merupakan solusi dari persamaan (4.7). Sehingga nilai taksiran parameter untuk Model Regresi *Poisson* pada persamaan (4.2) adalah

$$\widehat{\boldsymbol{\beta}}^{(m+1)}=\begin{pmatrix}\hat{\beta}_0^{(m+1)}\\ \hat{\beta}_1^{(m+1)}\\ \vdots \\ \hat{\beta}_k^{(m+1)}\end{pmatrix}, \quad m=0,1,2,\dots,k \quad (4.14)$$

4.1.2 Estimasi Parameter Model Regresi Poisson dengan Metode *Moment Estimation*

Dalam menaksir estimasi parameter model Regresi Poisson dengan berdasarkan yang sudah dijelaskan pada persamaan (2.2) bahwa ekspektasi dan varians dari Y yaitu:

$$\begin{aligned} E(Y) &= \mu \\ Var(Y) &= \mu \end{aligned} \tag{4.15}$$

Estimasi parameter dengan menggunakan metode *Moment Estimation* dilakukan dengan menyamakan ekspektasi populasi dengan moment ke-k. Ambil n sampel random $y_1, y_2, \ldots, y_n$ sehingga memenuhi metode *Moment Estimation*, sebagai berikut.

$$\begin{aligned} E(Y) = M_1 &= \frac{1}{n}\sum_{i=1}^{n} y_i \\ \mu &= M_1 \\ \mu &= \frac{1}{n}\sum_{i=1}^{n} y_i \\ \mu &= \bar{y} \end{aligned} \tag{4.16}$$

Setelah diperoleh hasil menyamaan ekspektasi populasi dengan moment ke-k, selanjutnya mensubstitusi ke persamaan (2.16) adalah:

$$\begin{aligned} e^{\boldsymbol{X}_i^T\boldsymbol{\beta}} &= \bar{y} \\ \boldsymbol{X}_i^T\boldsymbol{\beta} &= \ln(\bar{y}) \\ \boldsymbol{X}_i\boldsymbol{X}_i^T\boldsymbol{\beta} &= \boldsymbol{X}_i \ln(\bar{y}) \\ (\boldsymbol{X}_i\boldsymbol{X}_i^T)(\boldsymbol{X}_i\boldsymbol{X}_i^T)^{-1}\boldsymbol{\beta} &= \boldsymbol{X}_i\left(\boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1} \ln(\bar{y}) \\ \boldsymbol{\beta} &= \boldsymbol{X}_i\left(\boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1} \ln(\bar{y}) \\ \hat{\boldsymbol{\beta}} &= \boldsymbol{X}_i(\boldsymbol{X}_i\boldsymbol{X}_i^T)^{-1} \ln(\bar{y}) \end{aligned} \tag{4.17}$$

Untuk mendapatkan nilai parameter $\hat{\boldsymbol{\beta}}$ dengan memperoleh nilai matriks $\left(\boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1}$ dimana harus memenuhi syarat invers $\boldsymbol{X}_i\boldsymbol{X}_i^T$ yaitu, matriks $\boldsymbol{X}_i\boldsymbol{X}_i^T$ merupakan matriks persegi dengan determinan tidak sama dengan nol.

4.1.3 Pengujian Parameter Model Regresi *Poisson*

Pengujian parameter model Regresi Poisson dilakukan dengan menguji hipotesis-hipotesis menggunakan uji *Likelihood Ratio Test* dan uji Z.

1. Pengujian Serentak (Parameter $\boldsymbol{\beta}$)

Parameter yang diuji pada pengujian serentak ini mencakup semua parameter $\boldsymbol{\beta}$ secara bersama.

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$

H_1 : paling sedikit ada satu $\beta_j \neq 0$

$j = 1, 2, 3, \ldots, k$

Statistika uji yang digunakan adalah ukuran statistika *likelihood ratio* (devians) yang dibentuk dengan menentukan himpunan parameter di bawah populasi (Ω) yaitu $\Omega = \{\beta\}$ dan himpunan parameter jika H_0 benar (ω) yaitu $\omega = \{\beta_0,\}$. Berikut merupakan fungsi *likelihood* parameter di bawah populasi $\left(L(\Omega)\right)$:

$$L(\Omega) = \prod_{i=1}^{n} f(y_i; \boldsymbol{\beta})$$

$$L(\Omega) = \prod_{i=1}^{n} \left(\frac{\exp(-\exp(\boldsymbol{X}_i^T \boldsymbol{\beta}))(\exp(\boldsymbol{X}_i^T \boldsymbol{\beta}))^{y_i}}{y_i!} \right) \tag{4.18}$$

dan fungsi *likelihood* jika H_0 benar $\left(L(\omega)\right)$:

$$L(\omega) = \prod_{i=1}^{n} f(y_i; \boldsymbol{\beta}_0)$$

$$L(\omega) = \prod_{i=1}^{n} \left(\frac{\exp(-\exp(\boldsymbol{\beta}_0))(\exp(\boldsymbol{\beta}_0))^{y_i}}{y_i!} \right) \tag{4.19}$$

Sehingga diperoleh

$$L(\hat{\Omega}) = \max_{\Omega} L(\Omega)$$

$$L(\hat{\Omega}) = \prod_{i=1}^{n} \left(\frac{\exp\left(-\exp\left(\boldsymbol{X}_i^T \hat{\boldsymbol{\beta}}\right)\right)\left(\exp\left(\boldsymbol{X}_i^T \hat{\boldsymbol{\beta}}\right)\right)^{y_i}}{y_i!} \right) \tag{4.20}$$

dan

$$L(\widehat{\omega}) = \max_{\omega} L(\omega)$$

$$L(\widehat{\omega}) = \prod_{i=1}^{n}\left(\frac{\exp\left(-\exp\left(\widehat{\boldsymbol{\beta}}_0\right)\right)\left(\exp\left(\widehat{\boldsymbol{\beta}}_0\right)\right)^{y_i}}{y_i!}\right) \tag{4.22}$$

Nilai $\widehat{\boldsymbol{\beta}}$ merupakan hasil estimasi parameter yang diperoleh dari subbab 4.1 yang disubstitusikan pada fungsi $L(\widehat{\Omega})$ dan $L(\widehat{\omega})$. Kedua fungsi $L(\widehat{\Omega})$ dan $L(\widehat{\omega})$ dibandingkan dalam bentuk berikut:

$$\frac{L(\widehat{\omega})}{L(\widehat{\Omega})} = \frac{\prod_{i=1}^{n}\left(\frac{\exp\left(-\exp\left(\hat{\beta}_0\right)\right)\left(\exp\left(\hat{\beta}_0\right)\right)^{y_i}}{y_i!}\right)}{\prod_{i=1}^{n}\left(\frac{\exp\left(-\exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}\right)\right)\left(\exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}\right)\right)^{y_i}}{y_i!}\right)}$$

$$\frac{L(\widehat{\omega})}{L(\widehat{\Omega})} = \frac{\prod_{i=1}^{n}\left(\exp\left(-\exp\left(\hat{\beta}_0\right)\right)\left(\exp\left(\hat{\beta}_0\right)\right)^{y_i}\right)}{\prod_{i=1}^{n}\left(\exp\left(-\exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}\right)\right)\left(\exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}\right)\right)^{y_i}\right)} \tag{4.23}$$

Daerah penolakan yaitu Tolak H_0 jika $\frac{L(\widehat{\omega})}{L(\widehat{\Omega})} < \mu_0 < 1$, dimana $0 < \mu_0 < 1$. Kemudian dibandingkan dalam bentuk *Likelihood Ratio,*

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

$$G = 2\left(\sum_{i=1}^{n}\left(-exp(\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}) + y_i(\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}})\right)\right) - 2\left(\sum_{i=1}^{n}\left(-exp(\hat{\beta}_0) + y_i\hat{\beta}_0\right)\right) \tag{4.24}$$

Terjadi Tolak H_0 apabila $G_{\text{hitung}} > \chi^2_{(\alpha,k)}$

2. Pengujian Parsial (Parameter $\boldsymbol{\beta}$)

Parameter yang diuji pada pengujian ini mencakup semua parameter $\boldsymbol{\beta}$ secara parsial.

$H_0 : \beta_j = 0$
$H_1 : \beta_j \neq 0$
$j = 1, 2, 3, \dots, k$

Statistika uji yang digunakan adalah uji Z dengan mendapatkan $\mathrm{Z_{hit}}$ dari masing-masing parameter $\hat{\beta}_j$ yang nilainya diperoleh dari subbab 4.1

$$\mathrm{Z_{hit}} = \frac{\widehat{\beta_J}}{\sqrt{Var(\boldsymbol{\beta})}} \tag{4.25}$$

Terjadi tolak H_0 jika $|\mathrm{Z_{hit}}| > |\mathrm{Z}_\alpha|$

4.2 Kajian Teoristis untuk Regresi *Zero-Inflated Poisson* (ZIP)

Penaksiran parameter Model Regresi *Zero-Inflated Poisson* dilakukan seperti pada model Regresi *Poisson* dengan menggunakan dua metode, yaitu:

a. Metode *Maximum Likelihood Estimation* (MLE)
b. Metode *Moment Estimation* (MME)

Model Regresi *Zero-Inflated Poisson* adalah model regresi dengan variabel respon *Y* yang berdistribusi pada persamaan (2.5),

$$y_i = \begin{cases} 0 & \text{dengan probabilitas } \pi_i \\ > 0 & \text{dengan probabilitas } 1 - \pi_i \end{cases} \tag{4.26}$$

Pada dasarnya variabel respon *Y* distribusi *Zero-Inflated Poisson* berasal dari distribusi Poisson dan *zero inflation* atau dari distribusi Poisson dan distribusi Binomial/Bernouli. Jika $y_i > 0$ maka dapat diasumsikan berdistribusi *Poisson*, sehingga *probability density function* (pdf) untuk $y_i > 0$ yaitu,

$$P(Y_i > 0) = (1 - \pi_i) \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \tag{4.27}$$

Sedangkan jika $y_i = 0$ maka distribusi yang digunakan adalah distribusi Poisson atau distribusi Bernoulli/Binomial. Sehingga *probability density function* (pdf) untuk $y_i = 0$ sebagai berikut,

$$
\begin{aligned}
P(Y_i = y_i) = &\quad \pi_i + ({\pi_i}^{y_i}(1-\pi_i)^{1-y_i})\left(\frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}\right) \\
P(Y_i = 0) = &\quad \pi_i + ({\pi_i}^{0}(1-\pi_i)^{1-0})\left(\frac{e^{-\mu_i}\mu_i^{0}}{0!}\right) \\
P(Y_i = 0) = &\quad \pi_i + (1-\pi_i)e^{-\mu_i}.
\end{aligned}
\tag{4.28}
$$

Bedasarkan persamaan (4.27) dan persamaan (4.28), *probability density function* (pdf) dari distribusi *Zero-Inflated Poisson* adalah

$$
P(Y_i = y_i) = \begin{cases} \pi_i + (1-\pi_i)e^{-\mu_i} & , \; y_i = 0 \\ (1-\pi_i)\dfrac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} & , \; y_i > 0 \end{cases} \tag{4.29}
$$

Selanjutnya dalam mengestimasi parameter perlu menetukan persamaan π_i dan persamaan $(1-\pi_i)$ yang diperoleh model Regresi *Zero-Inflated Poisson* pada persamaan (2.18) dan persamaan (2.19), yaitu

$$
\ln(\mu) = \boldsymbol{X\beta} \tag{4.30}
$$

dan

$$
logit(\pi) = \ln\left(\frac{\pi}{1-\pi}\right) = \boldsymbol{X\gamma} \tag{4.31}
$$

Model Regresi *Zero-Inflated Poisson* merupakan *canonical link* dari distribusi Poisson dan distribusi Binomial. *Canonical link* merupakan *link function* yang menghubungkan mean ke parameter. Berikut merupakan penjabaran dalam menentukan persamaan π_i dan persamaan $(1-\pi_i)$,

$$
logit(\pi) = \quad \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \boldsymbol{X}_i^T\boldsymbol{\gamma}
$$

$$\begin{aligned}
\frac{\pi_i}{1-\pi_i} &= e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}} \\
\pi_i &= (1-\pi_i)e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}} \\
\pi_i &= e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}} - \pi_i e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}} \\
\pi_i + \pi_i e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}} &= e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}} \\
\pi_i\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right) &= e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}} \\
\pi_i &= \frac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)} \qquad (4.32)
\end{aligned}$$

dan

$$\begin{aligned}
(1-\pi_i) &= 1 - \frac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)} \\
(1-\pi_i) &= \frac{1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}} - e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)} \\
(1-\pi_i) &= \frac{1}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)} \qquad (4.33)
\end{aligned}$$

Setelah diperoleh persamaan π_i dan persamaan $(1-\pi_i)$, selanjutnya mensubstitusikan persamaan (4.2), (4.32) dan (4.33) ke dalam persamaan (4.29), diperoleh:

Untuk $y_i = 0$

$$\begin{aligned}
P(Y_i = 0) &= \pi_i + (1-\pi_i)e^{-\mu_i} \\
P(Y_i = 0) &= \frac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)} + \left(\frac{1}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)}\right)\left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right) \\
P(Y_i = 0) &= \frac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}} + \left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right)}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)} \qquad (4.34)
\end{aligned}$$

Untuk $y_i > 0$

$$P(Y_i > 0) = (1-\pi_i)\frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}$$

$$P(Y_i > 0) = \left(\frac{1}{\left(1 + e^{\boldsymbol{X}_i^T \boldsymbol{\gamma}}\right)}\right)\left(\frac{\left(\exp\left(-e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}\right)\right)\left(e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}\right)^{y_i}}{y_i!}\right) \quad (4.35)$$

Sehingga dapat ditulis kembali dalam bentuk gabungan berikut ini,

$$P(Y_i = y_i) = \begin{cases} \dfrac{e^{\boldsymbol{X}_i^T \boldsymbol{\gamma}} + \left(\exp\left(-e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}\right)\right)}{\left(1 + e^{\boldsymbol{X}_i^T \boldsymbol{\gamma}}\right)} & , \ y_i = 0 \\ \left(\dfrac{1}{\left(1 + e^{\boldsymbol{X}_i^T \boldsymbol{\gamma}}\right)}\right)\left(\dfrac{\left(\exp\left(-e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}\right)\right)\left(e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}\right)^{y_i}}{y_i!}\right) & , \ y_i > 0 \end{cases} \quad (4.36)$$

Setelah diperoleh *probability density function* (pdf) untuk model distribusi *Zero-Inflated Poisson*, selanjutnya persamaan (4.36) dilakukan penaksiran estimasi parameter menggunakan dua metode yang telah disebutkan sebelumnya, yaitu metode *Maximum Likelihood Estimation* (MLE), dan metode *Moment Estimation* (MME) sebagai berikut.

4.2.1 Estimasi Parameter Model Regresi *Zero-Inflated Poisson* (ZIP) dengan Metode *Maximum Likelihood Estimation* (MLE)

Untuk mendapatkan estimasi parameter model Regresi *Zero-Inflated Poisson* menggunakan metode *Maximum Likelihood Estimation* (MLE) dilakukan dengan membangun *likelihood* dari fungsi Regresi *Zero-Inflated Poisson.* Adapun langkah-langkah dalam penaksiran parameter model Regresi *Zero-Inflated Poisson.*

Untuk $y_i = 0$ dengan *probability density function* (pdf) seperti pada persamaan (4.34),

$$P(Y_i = 0) = \frac{e^{\boldsymbol{X}_i^T \boldsymbol{\gamma}} + \left(\exp\left(-e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}\right)\right)}{\left(1 + e^{\boldsymbol{X}_i^T \boldsymbol{\gamma}}\right)}.$$

1. Ambil n sampel acak $y_1, y_2, \ldots, y_n$ dengan $y_i = 0$, maka bentuk fungsi *likelihood* model Regresi *Zero-Inflated Poisson* adalah:

$$L_1(\boldsymbol{\gamma},\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}} + \left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right)}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)} \qquad (4.37)$$

2. Fungsi *ln likelihood* model Regresi *Zero-Inflated Poisson* adalah $L_1(\boldsymbol{\gamma},\boldsymbol{\beta})$.

$$\ln L_1(\boldsymbol{\gamma},\boldsymbol{\beta}) = \ln \prod_{i=1}^{n} \frac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}} + \left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right)}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)}$$

$$\ln L_1(\boldsymbol{\gamma},\boldsymbol{\beta}) = \sum_{i=1}^{n} \ln\left(e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}} + \left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right)\right) - \sum_{i=1}^{n} \ln\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right) \qquad (4.38)$$

Untuk $y_i > 0$ dengan *probability density function* (pdf) seperti pada persamaan (4.35),

$$P(Y_i > 0) = \left(\frac{1}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)}\right)\left(\frac{\left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right)\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i}}{y_i!}\right)$$

1. Ambil n sampel acak $y_1, y_2, \ldots, y_n$ dengan $y_i > 0$, maka bentuk fungsi *likelihood* model Regresi *Zero-Inflated Poisson* adalah:

$$L_2(\boldsymbol{\gamma},\boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{1}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)}\right) \frac{\left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right)\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i}}{y_i!}$$

$$L_2(\boldsymbol{\gamma},\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{\left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right)\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i}}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)y_i!} \qquad (4.39)$$

2. Fungsi *ln likelihood* model Regresi *Zero-Inflated Poisson* adalah $L_1(\boldsymbol{\gamma},\boldsymbol{\beta})$.

$$\ln L_2(\boldsymbol{\gamma},\boldsymbol{\beta}) = \ln\left(\prod_{i=1}^{n} \frac{\left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right)\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i}}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)y_i!}\right)$$

$$\ln L_2(\boldsymbol{\gamma},\boldsymbol{\beta}) = \sum_{i=1}^{n} \ln\left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right) + \sum_{i=1}^{n} \ln\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i} - \sum_{i=1}^{n} \ln\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right) - \sum_{i=1}^{n} \ln y_i!$$

$$\ln L_2(\boldsymbol{\gamma},\boldsymbol{\beta}) = -\sum_{i=1}^{n} e^{\boldsymbol{X}_i^T\boldsymbol{\beta}} + \sum_{i=1}^{n} \ln\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i} - \sum_{i=1}^{n} \ln\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right) - \sum_{i=1}^{n} \ln y_i!$$

$$\ln L_2(\boldsymbol{\gamma},\boldsymbol{\beta}) = -\sum_{i=1}^{n} e^{\boldsymbol{X}_i^T\boldsymbol{\beta}} + \sum_{i=1}^{n} \ln e^{\boldsymbol{X}_i^T\boldsymbol{\beta} y_i} - \sum_{i=1}^{n} \ln\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right) - \sum_{i=1}^{n} \ln y_i!$$

$$\ln L_2(\boldsymbol{\gamma},\boldsymbol{\beta}) = -\sum_{i=1}^{n} e^{\boldsymbol{X}_i^T\boldsymbol{\beta}} + \sum_{i=1}^{n} \boldsymbol{X}_i^T\boldsymbol{\beta} y_i - \sum_{i=1}^{n} \ln\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right) - \sum_{i=1}^{n} \ln y_i! \tag{4.40}$$

Dari persamaan (4.38) dan persamaan (4.40) diperoleh persamaan untuk $\ln L(\boldsymbol{\gamma},\boldsymbol{\beta})$, yaitu:

$$\ln L(\boldsymbol{\gamma},\boldsymbol{\beta}) = \begin{cases} \displaystyle\sum_{i=1}^{n} \ln\left(e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}} + \left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right)\right) - \sum_{i=1}^{n} \ln\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right) & y_i = 0 \\ \displaystyle -\sum_{i=1}^{n} e^{\boldsymbol{X}_i^T\boldsymbol{\beta}} + \sum_{i=1}^{n} \boldsymbol{X}_i^T\boldsymbol{\beta} y_i - \sum_{i=1}^{n} \ln\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right) - \sum_{i=1}^{n} \ln y_i! & y_i > 0 \end{cases} \tag{4.41}$$

Persamaan (4.41) juga dapat ditulis seperti pada persamaan (4.42) berikut,

$$\ln L(\boldsymbol{\gamma},\boldsymbol{\beta}) = \sum_{\substack{i=1 \\ y_i=0}}^{n} \ln\left(e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}} + \left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right)\right) - \sum_{\substack{i=1 \\ y_i=0}}^{n} \ln\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right) - \sum_{\substack{i=1 \\ y_i>0}}^{n} e^{\boldsymbol{X}_i^T\boldsymbol{\beta}} + \sum_{\substack{i=1 \\ y_i>0}}^{n} \boldsymbol{X}_i^T\boldsymbol{\beta} y_i - \sum_{\substack{i=1 \\ y_i>0}}^{n} \ln\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right) - \sum_{\substack{i=1 \\ y_i>0}}^{n} \ln y_i! \tag{4.42}$$

Setelah didapatkan persamaan fungsi *ln likelihood* dari model Regresi *Zero-Inflated Poisson*, muncul asumsi bahwa penjumlahan fungsi eksponensial $\sum_{\substack{i=0 \\ y_i=0}}^{n} \ln\left(\exp(\boldsymbol{X}_i^T\boldsymbol{\gamma}) + (\exp(-\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})))\right)$ dari fungsi *ln likelihood* pada persamaan (4.42) mempersulit perhitungan, karena tidak diketahui nilai nol pada $\sum_{\substack{i=0 \\ y_i=0}}^{n} \ln\left(\exp\left(\boldsymbol{X}_i^T\boldsymbol{\gamma}\right) + \left(\exp\left(-\exp\left(\boldsymbol{X}_i^T\boldsymbol{\beta}\right)\right)\right)\right)$ mana yang berasal dari *zero state* dan mana yang berasal *Poisson state*. Sehingga persamaan (4.42) tidak dapat diselesaikan dengan metode analitik. Maka persamaan (4.42) disebut juga *incomplete data likelihood*.

Selain itu, fungsi *ln likelihood* dari model Regresi *Zero-Inflated Poisson* terdiri dari dua kondisi yaitu $y_i = 0$ dan $y_i > 0$, serta telah diketahui juga bahwa variabel respon y_i pada penelitian ini terdiri dalam dua kondisi yaitu *zero state* dan mana yang berasal *Poisson state*. Oleh karena itu, untuk menggambarkan kondisi pada model Regresi *Zero-Inflated Poisson* secara terperinci, maka didefinisikan kembali variabel *Y* dengan suatu varibel indikator Z_i, yaitu:

$$Z_i = \begin{cases} 0 & \text{, jika } y_i \text{ dari } \textit{Poisson state} \\ 1 & \text{, jika } y_i \text{ dari } \textit{zero state} \end{cases} \qquad (4.43)$$

Berdasarkan persamaan (4.43) dapat diasumsikan bahwa apabila variabel respon $y_i > 0$, maka z_i bernilai 0. Akan tetapi, ketika variabel respon $y_i = 0$, maka z_i dapat bernilai 0 atau 1. Jadi untuk variabel respon $y_i = 0$, nilai z_i belum bisa ditentukan. Sehingga z_i dianggap sebagai *incomplete data.* Permasalahan penaksiran parameter dengan *incomplete data* dapat diselesaikan dengan menggunakan Algoritma-EM (*Expectation-Maximization*).

4.2.2 Estimasi Parameter Fungsi Likelihood dengan Algoritma-EM (*Expectation-Maximization*)

Algoritma-EM (*Expectation-Maximization*) merupakan metode optimasi yang digunakan untuk memaksimumkan fungsi *likelihood* yang mengandung data tidak lengkap (*missing*). Pada Algoritma-EM terdapat dua tahapan yaitu tahap *Expectation* dan tahap *Maximization*. Tahap *Expectation* adalah tahap perhitungan ekspektasi dan fungsi *ln likelihood* dengan memperhatikan data yang tidak lengkap. Sedangkan tahap *Maximization* merupakan tahap perhitungan untuk mecari penaksir parameter yang memaksimumkan fungsi *ln likelihood* dari hasil tahap *Expectation* sebelumnya.

Sebelum masuk kedalam tahapan Algoritma-EM untuk mengestimasi parameter dari variabel indikator Z_i, ditentukan terlebih dahulu distribusi dari variabel Z_i, yaitu:

$$P(Z_i = z_i) = \begin{cases} \pi_i & , \text{ jika } z_i = 0 \\ 1 - \pi_i & , \text{ jika } z_i > 0 \end{cases} \tag{4.44}$$

Persamaan (4.44) dapat diartikan sebagai berikut, ketika $z_i = 0$ maka peluangnya sama dengan peluang dari y_i yang berdistribusi Poisson dengan parameter π_i, yaitu sebesar $(1 - \pi_i)$, sehingga dapat dikatakan bahwa variabel Z_i berdistribusi Binomial $(1, \pi_i)$ dengan $E(Z_i) = \pi_i$ dan $Var(Z_i) = \pi_i(1 - \pi_i)$. Maka bentuk distribusi gabungan antara Y_i dan Z_i sebagai berikut.

$$\begin{aligned} f(\pi; \mu|y, z) &= f(z)f(y|z) \\ &= f(z|1, \pi)f(y|z, \mu) \\ &= (\pi_i)^{z_i}(1 - \pi_i)^{1-z_i}\left(\frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}\right)^{1-z_i} \end{aligned} \tag{4.45}$$

Selanjutnya, persamaan (4.2), (4.32) dan (4.33) yang diperoleh sebelumnya disubstitusi ke dalam hasil persamaan

(4.45) yang merupakan persamaan fungsi distribusi gabungan dari Y_i dan Z_i.

$$f(\boldsymbol{\beta};\boldsymbol{\gamma}|y,z) = \left(\frac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}}{\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)}\right)^{z_i}\left(\frac{1}{\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)}\right)^{1-z_i}\left(\frac{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i}}{y_i!}\right)^{1-z_i}$$

$$f(\boldsymbol{\beta};\boldsymbol{\gamma}|y,z) = \left(\frac{1}{\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)}\right)^{z_i}\left(e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)^{z_i}\left(\frac{1}{\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)}\right)\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)^{z_i}\left(\frac{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i}}{y_i!}\right)^{1-z_i}$$

$$f(\boldsymbol{\beta};\boldsymbol{\gamma}|y,z) = \left(e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)^{z_i}\left(\frac{1}{\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)}\right)\left(\frac{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i}}{y_i!}\right)^{1-z_i} \tag{4.46}$$

Setelah diperoleh hasil substitusi fungsi gabungan Y_i dan Z_i, persamaan (4.46) dibentuk kedalam fungsi *likelihood*:

$$L(\boldsymbol{\beta};\boldsymbol{\gamma}|y,z) = \prod_{i=1}^{n}\left(e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)^{z_i}\left(\frac{1}{\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)}\right)\left(\frac{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i}}{y_i!}\right)^{1-z_i}$$

$$L(\boldsymbol{\beta};\boldsymbol{\gamma}|y,z) = \prod_{i=1}^{n}\left(e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)^{z_i}\left(\frac{1}{\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)}\right)\left(\frac{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i}}{y_i!}\right)\left(\frac{y_i!}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i}}\right)^{z_i}$$

$$L(\boldsymbol{\beta};\boldsymbol{\gamma}|y,z) = \prod_{i=1}^{n}\left(\frac{1}{\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)}\right)\left(\frac{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i}}{y_i!}\right)\left(\frac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}(y_i!)}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i}}\right)^{z_i} \tag{4.47}$$

Hasil dari fungsi *likelihood* yaitu persamaan $f(\beta;\gamma|y,z)$ kemudian diubah dalam bentuk fungsi *ln likelihood*.

$$\ln L(\boldsymbol{\beta};\boldsymbol{\gamma}|y,z) = \ln\left(\prod_{i=1}^{n}\left(\frac{1}{\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)}\right)\left(\frac{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i}}{y_i!}\right)\left(\frac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}(y_i!)}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i}}\right)^{z_i}\right)$$

$$\begin{aligned}\ln L(\boldsymbol{\beta};\boldsymbol{\gamma}|y,z) = {} & -\sum_{i=1}^{n}\ln\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)+\sum_{i=1}^{n}\ln\left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right) \\ & +\sum_{i=1}^{n}\ln\left(\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i}\right)-\sum_{i=1}^{n}\ln y_i!+\sum_{i=1}^{n}\ln\left(e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)^{z_i} \\ & +\sum_{i=1}^{n}\ln(y_i!)^{z_i}-\sum_{i=1}^{n}\ln\left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right)^{z_i} \\ & -\sum_{i=1}^{n}\ln\left(\left(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)^{y_i}\right)^{z_i}\end{aligned}$$

$$\begin{aligned}\ln L(\boldsymbol{\beta};\boldsymbol{\gamma}|y,z) = {} & -\sum_{i=1}^{n}\ln\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)-\sum_{i=1}^{n}e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}+\sum_{i=1}^{n}y_i\boldsymbol{X}_i^T\boldsymbol{\beta} \\ & -\sum_{i=1}^{n}\ln y_i!+\sum_{i=1}^{n}z_i\boldsymbol{X}_i^T\boldsymbol{\gamma}+\sum_{i=1}^{n}\ln(y_i!)^{z_i}+\sum_{i=1}^{n}z_i\,e^{\boldsymbol{X}_i^T\boldsymbol{\beta}} \\ & -\sum_{i=1}^{n}z_iy_i\boldsymbol{X}_i^T\boldsymbol{\beta}\end{aligned}$$

$$\ln L(\boldsymbol{\beta};\boldsymbol{\gamma}|y,z) = -\sum_{i=1}^{n}\ln\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)+\sum_{i=1}^{n}z_i\boldsymbol{X}_i^T\boldsymbol{\gamma}-\sum_{i=1}^{n}e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}+\sum_{i=1}^{n}y_i\boldsymbol{X}_i^T\boldsymbol{\beta}$$
$$+\sum_{i=1}^{n}z_ie^{\boldsymbol{X}_i^T\boldsymbol{\beta}}-\sum_{i=1}^{n}z_iy_i\boldsymbol{X}_i^T\boldsymbol{\beta}-\sum_{i=1}^{n}\ln y_i!+\sum_{i=1}^{n}\ln(y_i!)^{z_i}$$

$$\ln L(\boldsymbol{\beta};\boldsymbol{\gamma}|y,z) = \sum_{i=1}^{n}\left(z_i\boldsymbol{X}_i^T\boldsymbol{\gamma}-\ln\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)\right)$$
$$+\sum_{i=1}^{n}\left(y_i\boldsymbol{X}_i^T\boldsymbol{\beta}-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}+z_ie^{\boldsymbol{X}_i^T\boldsymbol{\beta}}-z_iy_i\boldsymbol{X}_i^T\boldsymbol{\beta}\right)$$
$$-\sum_{i=1}^{n}(\ln y_i!-\ln(y_i!)^{z_i})$$

(4.48)

Persamaaan (4.48) merupakan hasil *ln likelihood* dari fungsi gabungan Y_i dan Z_i yang dapat diasumsikan sebagai *complete data likelihood*. Fungsi *ln likelihood* pada persamaan (4.48) inilah yang akan dimaksimumkan menggunakan Algoritma-EM, dimana vektor parameter β dan γ dapat diestimasi secara terpisah.

$$\ln L(\boldsymbol{\beta};\boldsymbol{\gamma}|y,z) = \ln L(\boldsymbol{\beta},y,z)+\ln L(\boldsymbol{\gamma},y,z)+\sum_{i=1}^{n}(\ln y_i!-\ln(y_i!)^{z_i})$$

dengan

$$\ln L(\boldsymbol{\beta},y,z) = \sum_{i=1}^{n}\left(y_i\boldsymbol{X}_i^T\boldsymbol{\beta}-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}+z_ie^{\boldsymbol{X}_i^T\boldsymbol{\beta}}-z_iy_i\boldsymbol{X}_i^T\boldsymbol{\beta}\right)$$

(4.49)

dan

$$\ln L(\boldsymbol{\gamma},y,z) = \sum_{i=1}^{n}\left(z_i\boldsymbol{X}_i^T\boldsymbol{\gamma}-\ln\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)\right) \quad (4.50)$$

Sedangkan untuk $\sum_{i=1}^{n}(\ln y_i!-\ln(y_i!)^{z_i})$ dapat diabaikan karena tidak mengandung vektor parameter β dan γ. Selanjutnya persamaan (4.49) dan persamaan (4.50) diestimasi dengan menggunakan Algoritma-EM. Adapun langkah-langkah dalam proses iterasi dari Algoritma-EM sebagai berikut.

1. Langkah E untuk menghitung nilai ekspektasi $Z_i^{(m)}$ dari varabel Z_i

$$Z_i^{(m)} = E\left(Z_i \middle| y_i, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\beta}^{(m)}\right)$$
$$Z_i^{(m)} = P\left(Z_i = 1 \middle| y_i, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\beta}^{(m)}\right)$$
$$Z_i^{(m)} = \begin{cases} \left(Z_i = 1 \middle| y_i = 0, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\beta}^{(m)}\right) & , y_i = 0 \\ 0 & , y_i > 0 \end{cases} \quad (4.51)$$

Selanjutnya dicari terlebih dahulu nilai ekspektasi Z_i ketika $y_i = 0$, yaitu:

$$P\left(Z_i = 1 \middle| y_i = 0, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\beta}^{(m)}\right) = \frac{P[y_i = 0|Z_i = 1]P[Z_i = 1]}{P[y_i = 0|Z_i = 1]P[Z_i = 0] + P[y_i|Z_i = 0]P[Z_i = 0]}$$
$$P\left(Z_i = 1 \middle| y_i = 0, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\beta}^{(m)}\right) = \frac{\pi_i}{\pi_i + [(1 - \pi_i)\exp(-\mu_i)]} \quad (4.52)$$

Pada langkah E ini dilakukan dahulu substitusi persamaan (4.2), (4.32) dan (4.33) ke persamaan (4.52).

$$P\left(Z_i = 1 \middle| y_i = 0, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\beta}^{(m)}\right) = \frac{\dfrac{e^{\boldsymbol{X}_i^T \boldsymbol{\gamma}^{(m)}}}{\left(1 + e^{\boldsymbol{X}_i^T \boldsymbol{\gamma}^{(m)}}\right)}}{\dfrac{e^{\boldsymbol{X}_i^T \boldsymbol{\gamma}^{(m)}}}{\left(1 + e^{\boldsymbol{X}_i^T \boldsymbol{\gamma}^{(m)}}\right)} + \left[\left(\dfrac{1}{\left(1 + e^{\boldsymbol{X}_i^T \boldsymbol{\gamma}^{(m)}}\right)}\right)\exp\left(-e^{\boldsymbol{X}_i^T \boldsymbol{\beta}^{(m)}}\right)\right]}$$

$$P\left(Z_i = 1 \middle| y_i = 0, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\beta}^{(m)}\right) = \frac{\dfrac{e^{\boldsymbol{X}_i^T \boldsymbol{\gamma}^{(m)}}}{\left(1 + e^{\boldsymbol{X}_i^T \boldsymbol{\gamma}^{(m)}}\right)}}{\dfrac{e^{\boldsymbol{X}_i^T \boldsymbol{\gamma}^{(m)}}}{\left(1 + e^{\boldsymbol{X}_i^T \boldsymbol{\gamma}^{(m)}}\right)} + \left(\dfrac{\exp\left(-e^{\boldsymbol{X}_i^T \boldsymbol{\beta}^{(m)}}\right)}{\left(1 + e^{\boldsymbol{X}_i^T \boldsymbol{\gamma}^{(m)}}\right)}\right)}$$

$$P\left(Z_i = 1 \middle| y_i = 0, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\beta}^{(m)}\right) = \frac{\dfrac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}^{(m)}}}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}^{(m)}}\right)}}{\dfrac{\left(e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}^{(m)}} + \left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}^{(m)}}\right)\right)\right)}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}^{(m)}}\right)}}$$

$$P\left(Z_i = 1 \middle| y_i = 0, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\beta}^{(m)}\right) = \frac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}^{(m)}}}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}^{(m)}}\right)}\left(\frac{1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}^{(m)}}}{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}^{(m)}} + \left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}^{(m)}}\right)\right)}\right)$$

$$P\left(Z_i = 1 \middle| y_i = 0, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\beta}^{(m)}\right) = \frac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}^{(m)}}}{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}^{(m)}} + \left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}^{(m)}}\right)\right)}$$

$$P\left(Z_i = 1 \middle| y_i = 0, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\beta}^{(m)}\right) = \frac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}^{(m)}}}{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}^{(m)}} + \left(\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\beta}^{(m)}}\right)\right)}\left(\frac{e^{-\boldsymbol{X}_i^T\boldsymbol{\gamma}^{(m)}}}{e^{-\boldsymbol{X}_i^T\boldsymbol{\gamma}^{(m)}}}\right)$$

$$P\left(Z_i = 1 \middle| y_i = 0, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\beta}^{(m)}\right) = \frac{1}{1 + \exp\left(-\boldsymbol{X}_i^T\boldsymbol{\gamma}^{(m)} - e^{\boldsymbol{X}_i^T\boldsymbol{\beta}^{(m)}}\right)} \tag{4.53}$$

Setelah diperoleh persamaan $P\left(Z_i = 1 \middle| y_i = 0, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\beta}^{(m)}\right)$, selanjutnya persamaan (4.53) disubstitusi ke dalam persamaan (4.51) untuk mendapatkan fungsi persamaan $Z_i^{(m)}$.

$$Z_i^{(m)} = \begin{cases} \dfrac{1}{1 + \exp\left(-\boldsymbol{X}_i^T\boldsymbol{\gamma}^{(m)} - e^{\boldsymbol{X}_i^T\boldsymbol{\beta}^{(m)}}\right)} & , y_i = 0 \\ 0 & , y_i > 0 \end{cases} \tag{4.54}$$

Kemudian pada langkah E ini juga dilakukan substitusi dari hasil ekspektasi ke dalam persamaan (4.49) dan persamaan (4.50).

$$\ln L\left(\boldsymbol{\beta} \middle| y, z^{(m)}\right) = \sum_{i=1}^{n}\left(y_i\boldsymbol{X}_i^T\boldsymbol{\beta} - e^{\boldsymbol{X}_i^T\boldsymbol{\beta}} + z_i e^{\boldsymbol{X}_i^T\boldsymbol{\beta}} - z_i y_i \boldsymbol{X}_i^T\boldsymbol{\beta}\right)$$

$$\ln L\left(\boldsymbol{\beta} \middle| y, z^{(m)}\right) = \sum_{i=1}^{n}\left(y_i\boldsymbol{X}_i^T\boldsymbol{\beta} - e^{\boldsymbol{X}_i^T\boldsymbol{\beta}} - z_i^{(m)}\left(y_i\boldsymbol{X}_i^T\boldsymbol{\beta} - e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right)$$

$$\ln L\left(\boldsymbol{\beta} \middle| y, z^{(m)}\right) = \sum_{i=1}^{n}\left(1 - z_i^{(m)}\right)\left(y_i\boldsymbol{X}_i^T\boldsymbol{\beta} - e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right) \tag{4.55}$$

dan

$$\ln L\left(\boldsymbol{\gamma}\middle|y, z^{(m)}\right) = \sum_{i=1}^{n}\left(z_i^{(m)}\boldsymbol{X}_i^T\boldsymbol{\gamma} - \ln\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)\right) \qquad (4.56)$$

2. Langkah M pertama, memaksimumkan persamaan (4.55), yaitu $\ln L\left(\boldsymbol{\beta}^{(m)}, y, z^{(m)}\right)$ untuk maksimalisasi parameter β dengan mencari $\beta^{(m+1)}$ dengan menggunakan metode Newton-Raphson seperti pada pembahasan sub bab 4.1.1. Berikut adalah vektor gradien $(\boldsymbol{g})$ dan matriks Hessian $(\boldsymbol{H})$ dari persamaan (4.55):

$$\frac{\partial \ln L\left(\boldsymbol{\beta}\middle|y, z^{(m)}\right)}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}\left(1 - z_i^{(m)}\right)\left(y_i\boldsymbol{X}_i^T - \boldsymbol{X}_i^T e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)$$

$$\frac{\partial \ln L\left(\boldsymbol{\beta}\middle|y, z^{(m)}\right)}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}\left(1 - z_i^{(m)}\right)(\boldsymbol{X}_i^T)\left(y_i - e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right) \qquad (4.57)$$

dan

$$\frac{\partial^2 \ln L\left(\boldsymbol{\beta}\middle|y, z^{(m)}\right)}{\partial\boldsymbol{\beta}^2} = \sum_{i=1}^{n}\left(1 - z_i^{(m)}\right)\left(-X_i^T e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\left(X_i^T e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right) \qquad (4.58)$$

3. Langkah M kedua, memaksimumkan persamaan (4.56) yaitu $\ln L\left(\boldsymbol{\gamma}\middle|y, z^{(m)}\right)$ untuk maksimalisasi parameter γ dengan mencari $\gamma^{(m+1)}$. Berikut adalah vektor gradien $(\boldsymbol{g})$ dan matriks Hessian $(\boldsymbol{H})$dari persamaan (4.56):

$$\frac{\partial \ln L\left(\boldsymbol{\gamma}\middle|y, z^{(m)}\right)}{\partial\boldsymbol{\gamma}} = \sum_{i=1}^{n}\left(z_i^{(m)}\boldsymbol{X}_i^T - \frac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)}\boldsymbol{X}_i^T\right)$$

(4.59)

dan

$$\frac{\partial^2 \ln L\left(\boldsymbol{\gamma}\middle|y, z^{(m)}\right)}{\partial\boldsymbol{\gamma}^2} = -\sum_{i=1}^{n}\left(\frac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)}\boldsymbol{X}_i^T\boldsymbol{X}_i\right)$$

(4.60)

4. Langkah-langkah untuk memaksimalisai $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ untuk mendapatkan $\widehat{\boldsymbol{\beta}}^{(m+1)}$ dan $\widehat{\boldsymbol{\gamma}}^{(m+1)}$ identik dengan metode Newton-Raphson pada persamaan (2.29).
5. Mengganti $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ dengan $\widehat{\boldsymbol{\beta}}^{(m+1)}$ dan $\widehat{\boldsymbol{\gamma}}^{(m+1)}$, kemudian melakukan kembali ke tahap ekspektasi.
6. Tahap ekspektasi dan maksimalisasi dilakukan secara terus menerus hingga memperoleh penaksir yang konvergen.

4.2.3 Estimasi Parameter Model Regresi *Zero-Inflated Poisson* (ZIP) dengan Metode *Moment Estimation*

Estimasi parameter dengan metode *Moment Estimation* yaitu menggunakan ekspektasi dan varians pada persamaan (2.7) sebagai berikut

$$E(Y_i) = \mu_i(1-\pi_i) = M_1 = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y} \tag{4.61}$$

dan

$$Var(Y_i) = \mu_i(1-\pi_i)(1+\mu_i\pi_i) \tag{4.62}$$

Langkah pertama untuk estimasi parameter dengan metode *Moment Estimation,* yaitu mendapatkan persamaan μ_i dari persamaan ekpektasi (4.61)

$$E(Y_i) = \mu_i(1-\pi_i) = M_1$$

$$\mu_i = \frac{M_1}{(1-\pi_i)} \tag{4.63}$$

Selanjutnya mendapatkan persamaan μ_i dari persamaan varians (4.62), seperti berikut

$$\begin{aligned}
Var(Y_i) &= \mu_i(1-\pi_i)(1+\mu_i\pi_i) \\
E(Y^2) - \big(E(Y_i)\big)^2 &= \mu_i(1-\pi_i)(1+\mu_i\pi_i) \\
E\big(Y_i^{\,2}\big) &= \mu_i(1-\pi_i)(1+\mu_i\pi_i) + \big(E(Y_i)\big)^2 \\
M_2 &= \mu_i(1-\pi_i)(1+\mu_i\pi_i) + \big(\mu_i(1-\pi_i)\big)^2 \\
M_2 &= \mu_i(1-\pi_i)[(1+\mu_i\pi_i) + \mu_i(1-\pi_i)] \\
M_2 &= \mu_i(1-\pi_i)[1+\mu_i\pi_i + \mu_i - \mu_i\pi_i]
\end{aligned}$$

$$
\begin{aligned}
M_2 &= \mu_i(1-\pi_i)[1+\mu_i] \\
M_2 &= \frac{M_1}{(1-\pi_i)}(1-\pi_i)\left[1+\mu_i\right] \\
M_2 &= M_1[1+\mu_i] \\
M_2 &= M_1 + M_1\mu_i \\
M_1\mu_i &= M_2 - M_1 \\
\mu_i &= \frac{M_2 - M_1}{M_1} \\
\mu_i &= \frac{\bar{y}^2 - \bar{y}}{\bar{y}} \qquad (4.64)
\end{aligned}
$$

Hasil persamaan μ_i dari persamaan varians dilakukan substitusi oleh persamaan (4.2) memperoleh parameter $\widehat{\boldsymbol{\beta}}$.

$$
\begin{aligned}
\mu_i &= \frac{\bar{y}^2 - \bar{y}}{\bar{y}} \\
e^{\boldsymbol{X}_i^T\boldsymbol{\beta}} &= \frac{\bar{y}^2 - \bar{y}}{\bar{y}} \\
\boldsymbol{X}_i^T\boldsymbol{\beta} &= \frac{\bar{y}^2 - \bar{y}}{\bar{y}} \\
\boldsymbol{X}_i^T\boldsymbol{\beta} &= \ln(\bar{y}^2 - \bar{y}) - \ln(\bar{y}) \\
\boldsymbol{X_i}\boldsymbol{X}_i^T\boldsymbol{\beta} &= \boldsymbol{X_i}(\ln(\bar{y}^2 - \bar{y}) - \ln(\bar{y})) \\
(\boldsymbol{X_i}\boldsymbol{X}_i^T)(\boldsymbol{X}_i\boldsymbol{X}_i^T)^{-1}\boldsymbol{\beta} &= \boldsymbol{X_i}(\boldsymbol{X}_i\boldsymbol{X}_i^T)^{-1}(\ln(\bar{y}^2 - \bar{y}) - \ln(\bar{y})) \\
\boldsymbol{\beta} &= \boldsymbol{X_i}(\boldsymbol{X}_i\boldsymbol{X}_i^T)^{-1}(\ln(\bar{y}^2 - \bar{y}) - \ln(\bar{y})) \\
\widehat{\boldsymbol{\beta}} &= \boldsymbol{X_i}(\boldsymbol{X}_i\boldsymbol{X}_i^T)^{-1}(\ln(\bar{y}^2 - \bar{y}) - \ln(\bar{y})) \qquad (4.65)
\end{aligned}
$$

Sedangkan untuk memperoleh parameter $\widehat{\boldsymbol{\gamma}}$ dilakukan dengan persamaan (4.63) menyamakan dengan persamaan (4.64), kemudian mensubstitusi persamaan (4.32) dan (4.33).

$$
\begin{aligned}
\frac{M_1}{(1-\pi_i)} &= \frac{M_2 - M_1}{M_1} \\
M_1\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right) &= \frac{M_2 - M_1}{M_1}
\end{aligned}
$$

$$
\begin{aligned}
1 + e^{X_i^T \gamma} &= \frac{M_2 - M_1}{{M_1}^2} \\
e^{X_i^T \gamma} &= \frac{M_2 - M_1}{{M_1}^2} - 1 \\
e^{X_i^T \gamma} &= \frac{\bar{y}^2 - \bar{y}}{(\bar{y})^2} - 1 \\
e^{X_i^T \gamma} &= \frac{\bar{y}^2 - \bar{y}}{(\bar{y})^2} - \frac{(\bar{y})^2}{(\bar{y})^2} \\
e^{X_i^T \gamma} &= \frac{\bar{y}^2 - \bar{y} - (\bar{y})^2}{(\bar{y})^2} \\
\boldsymbol{X}_i^T \boldsymbol{\gamma} &= \ln\left(\frac{\bar{y}^2 - \bar{y} - (\bar{y})^2}{(\bar{y})^2}\right) \\
\boldsymbol{X}_i^T \boldsymbol{\gamma} &= \ln(\bar{y}^2 - \bar{y} - (\bar{y})^2) - \ln(\bar{y})^2 \\
\boldsymbol{X}_i^T \boldsymbol{\gamma} &= \ln(\bar{y}^2 - \bar{y} - (\bar{y})^2) - 2\ln(\bar{y}) \\
\boldsymbol{X_i} \boldsymbol{X}_i^T \boldsymbol{\gamma} &= \boldsymbol{X_i}(\ln(\bar{y}^2 - \bar{y} - (\bar{y})^2) - 2\ln(\bar{y})) \\
(\boldsymbol{X_i} \boldsymbol{X}_i^T)(\boldsymbol{X}_i \boldsymbol{X}_i^T)^{-1} \boldsymbol{\gamma} &= \boldsymbol{X_i}(\boldsymbol{X}_i \boldsymbol{X}_i^T)^{-1}(\ln(\bar{y}^2 - \bar{y} - (\bar{y})^2) - 2\ln(\bar{y})) \\
\boldsymbol{\gamma} &= \boldsymbol{X_i}(\boldsymbol{X}_i \boldsymbol{X}_i^T)^{-1}(\ln(\bar{y}^2 - \bar{y} - (\bar{y})^2) - 2\ln(\bar{y})) \\
\hat{\boldsymbol{\gamma}} &= \boldsymbol{X_i}(\boldsymbol{X}_i \boldsymbol{X}_i^T)^{-1}(\ln(\bar{y}^2 - \bar{y} - (\bar{y})^2) - 2\ln(\bar{y}))
\end{aligned}
\tag{4.66}
$$

Untuk mendapatkan nilai parameter $\hat{\boldsymbol{\beta}}$ dan $\hat{\boldsymbol{\gamma}}$ dengan memperoleh nilai matriks $\left(\boldsymbol{X}_i \boldsymbol{X}_i^T\right)^{-1}$ dimana harus memenuhi syarat invers $\boldsymbol{X}_i \boldsymbol{X}_i^T$ yaitu, matriks $\boldsymbol{X}_i \boldsymbol{X}_i^T$ merupakan matriks persegi dengan determinan tidak sama dengan nol.

4.2.4 Pengujian Parameter Model Regresi Zero-Inflated Poisson

Pengujian parameter model Regresi *Zero-Inflated Poisson* dilakukan dengan menguji hipotesis-hipotesis menggunakan uji *Likelihood Ratio Test* dan uji Z.

1. Pengujian Serentak (Parameter $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$)

Parameter yang diuji pada pengujian serentak ini mencakup semua parameter $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ secara bersama.

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = \gamma_1 = \gamma_2 = \ldots = \gamma_k = 0$
H_1 : paling sedikit ada satu $\beta_j \neq 0$ atau $\gamma_j \neq 0$
$j = 1, 2, 3, \ldots, k$

Statistika uji yang digunakan adalah *likelihood ratio* yang dibentuk dengan menentukan himpunan parameter di bawah populasi (Ω) yaitu $\Omega = \{\beta, \gamma\}$ dan himpunan parameter jika H_0 benar (ω) yaitu $\omega = \{\beta_0, \gamma_0\}$. Berikut merupakan fungsi *likelihood* parameter di bawah populasi $\left(L(\Omega)\right)$:

$$L(\Omega) = \prod_{i=1}^{n} f(y_i; \boldsymbol{\beta}, \boldsymbol{\gamma})$$

$$L(\Omega) = \prod_{i=1}^{n} (\exp(\boldsymbol{X}_i^T \boldsymbol{\gamma}))^{z_i} \left(\frac{1}{(1 + \exp(\boldsymbol{X}_i^T \boldsymbol{\gamma}))}\right) \left(\frac{\exp(-\exp(\boldsymbol{X}_i^T \boldsymbol{\beta}))\,(\exp(\boldsymbol{X}_i^T \boldsymbol{\beta}))^{y_i}}{y_i!}\right)^{1-z_i} \tag{4.67}$$

dan fungsi *likelihood* jika H_0 benar $\left(L(\omega)\right)$:

$$L(\omega) = \prod_{i=1}^{n} f(y_i; \beta_0, \gamma_0)$$

$$L(\omega) = \prod_{i=1}^{n} (\exp(\gamma_0))^{z_i} \left(\frac{1}{(1 + \exp(\gamma_0))}\right) \left(\frac{\exp(-\exp(\beta_0))\,(\exp(\beta_0))^{y_i}}{y_i!}\right)^{1-z_i} \tag{4.68}$$

Sehingga diperoleh

$$L(\widehat{\Omega}) = \max_{\Omega} L(\Omega)$$

$$L(\widehat{\Omega}) = \prod_{i=1}^{n} (\exp(\boldsymbol{X}_i^T \widehat{\boldsymbol{\gamma}}))^{z_i} \left(\frac{1}{(1 + \exp(\boldsymbol{X}_i^T \widehat{\boldsymbol{\gamma}}))}\right) \left(\frac{\exp\left(-\exp\left(\boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}\right)\right)\left(\exp\left(\boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}\right)\right)^{y_i}}{y_i!}\right)^{1-z_i} \tag{4.69}$$

dan

$$L(\widehat{\omega}) = \max_{\omega} L(\omega)$$

$$L(\widehat{\omega}) = \prod_{i=1}^{n} (\exp(\hat{\gamma}_0))^{z_i} \left(\frac{1}{(1+\exp(\hat{\gamma}_0))}\right) \left(\frac{\exp\left(-\exp\left(\hat{\beta}_0\right)\right)\left(\exp\left(\hat{\beta}_0\right)\right)^{y_i}}{y_i!}\right)^{1-z_i} \quad (4.70)$$

Nilai $\widehat{\boldsymbol{\beta}}$ dan $\widehat{\boldsymbol{\gamma}}$ merupakan hasil estimasi parameter yang diperoleh dari subbab 4.2 yang disubstitusikan pada fungsi $L(\widehat{\Omega})$ dan $L(\widehat{\omega})$. Kedua fungsi $L(\widehat{\Omega})$ dan $L(\widehat{\omega})$ dibandingkan dalam bentuk berikut:

$$\frac{L(\widehat{\omega})}{L(\widehat{\Omega})} = \frac{\prod_{i=1}^{n}\left(\exp\left(\hat{\gamma}_0\right)\right)^{z_i}\left(\frac{1}{\left(1+\exp\left(\hat{\gamma}_0\right)\right)}\right)\left(\frac{\exp\left(-\exp\left(\hat{\beta}_0\right)\right)\left(\exp\left(\hat{\beta}_0\right)\right)^{y_i}}{y_i!}\right)^{1-z_i}}{\prod_{i=1}^{n}\left(\exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\gamma}}\right)\right)^{z_i}\left(\frac{1}{\left(1+\exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\gamma}}\right)\right)}\right)\left(\frac{\exp\left(-\exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}\right)\right)\left(\exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}\right)\right)^{y_i}}{y_i!}\right)^{1-z_i}}$$

$$\frac{L(\widehat{\omega})}{L(\widehat{\Omega})} = \frac{\prod_{i=1}^{n}\left(\exp\left(\hat{\gamma}_0\right)\right)^{z_i}\left(\frac{1}{\left(1+\exp\left(\hat{\gamma}_0\right)\right)}\right)\left(\exp\left(-\exp\left(\hat{\beta}_0\right)\right)\left(\exp\left(\hat{\beta}_0\right)\right)^{y_i}\right)^{1-z_i}}{\prod_{i=1}^{n}\left(\exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\gamma}}\right)\right)^{z_i}\left(\frac{1}{\left(1+\exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\gamma}}\right)\right)}\right)\left(\exp\left(-\exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}\right)\right)\left(\exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}\right)\right)^{y_i}\right)^{1-z_i}}$$

(4.71)

Daerah penolakan yaitu Tolak H_0 jika $\frac{L(\widehat{\omega})}{L(\widehat{\Omega})} < \mu_0 < 1,$ dimana $0 < \mu_0 < 1.$ Kemudian dibandingkan dalam bentuk *Likelihood Ratio,*

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

$$G = \quad 2\left(\sum_{i=1}^{n}(z_i\boldsymbol{X}_i^T\widehat{\boldsymbol{\gamma}} - \ln(1+\exp(\boldsymbol{X}_i^T\widehat{\boldsymbol{\gamma}})))\right)$$
$$+2\left(\sum_{i=1}^{n}(1-z_i)\left(y_i\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}} - \exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}\right)\right)\right)$$
$$-2\left(\sum_{i=1}^{n}(z_i\hat{\gamma}_0 - \ln(1+\exp(\hat{\gamma}_0)))\right)$$
$$-2\left(\sum_{i=1}^{n}(1-z_i)\left(y_i\hat{\beta}_0 - \exp\left(\hat{\beta}_0\right)\right)\right) \tag{4.72}$$

Terjadi Tolak H_0 apabila $G_{\text{hitung}} > \chi^2_{(\alpha,k)}$

2. Pengujian Parsial (Parameter $\boldsymbol{\beta}$)

Parameter yang diuji pada pengujian ini mencakup semua parameter $\boldsymbol{\beta}$ secara parsial.

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$

$j = 1, 2, 3, \dots, k$

Statistika uji yang digunakan adalah uji Z dengan mendapatkan $\mathrm{Z_{hit}}$ dari masing-masing parameter $\hat{\beta}_j$ yang nilainya diperoleh dari subbab 4.2

$$\mathrm{Z_{hit}} = \frac{\widehat{\beta_J}}{\sqrt{Var(\boldsymbol{\beta})}} \tag{4.73}$$

Terjadi tolak H_0 jika $|\mathrm{Z_{hit}}| \geq |\mathrm{Z_\alpha}|$

3. Pengujian Parsial (Parameter $\boldsymbol{\gamma}$)

Parameter yang diuji pada pengujian ini mencakup semua parameter $\boldsymbol{\gamma}$ secara parsial.

$H_0 : \gamma_j = 0$
$H_1 : \gamma_j \neq 0$
$j = 1, 2, 3, \dots, k$

Statistika uji yang digunakan adalah uji Z dengan mendapatkan $\mathrm{Z_{hit}}$ dari masing-masing parameter $\hat{\gamma}_j$ yang nilainya diperoleh dari subbab 4.2

$$\mathrm{Z_{hit}} = \frac{\widehat{\gamma_j}}{\sqrt{Var(\boldsymbol{\gamma})}} \tag{4.74}$$

Terjadi tolak H_0 jika $|\mathrm{Z_{hit}}| \geq |\mathrm{Z_\alpha}|$

4.3 Kajian Terapan Menggunakan Jumlah Kejadian Meninggal Kasus Difteri di Indonesia pada Tahun 2017

Pada sub bab ini membahas mengenai terapan estimasi parameter dari model Regresi Poisson dan model Regresi *Zero-Inflated Poisson* dengan menggunakan data kejadian meninggal kasus penyakit difteri di Indonesia tahun 2017 pada Lampiran D. Sebelum dilakukan analisis terapan lebih lanjut, terlebih dahulu akan dilakukan analisis statsitika deskriptif yang digunakan untuk megetahui karakteristik dari variabel respon dan variabel prediktor dalam penelitian.

4.3.1 Karakteristik Jumlah Kejadian Meninggal Kasus Difteri di Indonesia pada Tahun 2017

Pada penelitian ini dipaparkan mengenai statistika deskriptif dalam bentuk rata-rata (mean), nilai tengah (median), varians, nilai minimum, nilai maksimum, standar deviasi dari variabel respon Y yaitu jumlah kejadian meninggal kasus difteri tahun 2017 di Indonesia dan faktor-faktor yang diduga mempengaruhi jumlah kejadian kasus difteri tersebut. Berikut ini merupakan statistika deskriptif dari data Lampiran D yang disajikan pada Tabel 4.1.

Tabel 4.1 Deskripsi Data Jumlah Kejadian Menginggal Kasus Difteri di Indonesia pada Tahun 2017

Variabel	Y	X_1	X_2	X_3	X_4
Mean	1.294	81.26	10.95	65.75	19.63
Nilai Maksimum	14.000	100.00	27.76	91.13	32.60
Nilai Minimum	0.000	21.43	3.78	33.06	9.40
Median	0.000	83.91	9.38	66.27	18.15
Varians	11.305	269.82	33.49	186.46	32.04
Standar Deviasi	3.362	16.43	5.79	13.65	5.66

Tabel 4.1. menunjukkan bahwa jumlah kejadian meninggal kasus difteri di Indonesia tahun 2017 memiliki jumlah kejadian minimum yaitu tidak terjadi kejadian meninggal dan jumlah maksimum sebanyak 14 kejadian meninggal. Kejadian meninggal terbanyak sepanjang tahun 2017 terjadi pada Provinsi Jawa Barat. Rata-rata kejadian meninggal sebesar 1.294 dengan nilai varians sebesar 11.305.

DKI Jakarta dan DI Yogyakarta menjadi provinsi yang paling sering melakukan *Universal Child Immunisation* dengan presentase sebesar 100 dan Papua menjadi provinsi dengan presentase minimum, yaitu 21.43 dalam melakukan *Universal Child Immunisation*. Rata-rata dilakukannya *Universal Child Immunisation* di semua provinsi memiliki presentase sebesar 81.26 dengan varians sebesar 269.82.

Presentase maksimum penduduk miskin di Indonesia pada tahun 2017 terjadi di Provinsi Papua sebesar 27.76, sedangkan presentase minimumnya sebesar 3.78 terjadi di Provinsi

DKI Jakarta. Presentase penduduk miskin di Indonesia memiliki rata-rata dan varians sebesar 10.95 dan 33.49.

Rumah tangga yang memiliki akses terhadap sanitasi layak dengan presentase maksimum sebesar 91.13 berada di Provinsi DKI Jakarta dan presentase minimum berada di Provinsi Papua sebesar 33.06. Rata-rata dan varians untuk presentase rumah tangga yang memiliki akses terhadap sanitasi layak adalah 65.75 dan 186.46.

Presentase gizi buruk di Indonesia dengan nilai minimum sebesar 9.40 terjadi di Provinsi Bali dan Provinsi NTT menjadi provinsi dengan nilai maksimum sebesar 32.60. Sepanjang tahun 2017 presentase gizi buruk di Indonesia memiliki rata-rata 19.63 dan varians 32.04.

4.3.2 Uji Multikolinieritas

Dalam analisis regresi, adanya kasus multikolinearitas memiliki pengaruh besar terhadap hasil estimasi parameter. Oleh karena itu, sebelum dilakukan analisis regresi lebih lanjut perlu dilakukan pengecekkan untuk mengetahui ada tidaknya multikolinearitas antar variabel prediktor. Uji multikolinearitas untuk masing-masing variabel prediktor dilakukan dengan mengitung VIF (*Variance Inflation Factor*). Berdasarkan lampiran E menunjukkan nilai koefisien determinasi $\left(R_j^2\right)$ dari masing-masing variabel prediktor yang disajikan pada Tabel 4.2.

Tabel 4.2 Nilai Koefisiensi Determinasi Variabel Prediktor

Variabel	$\boldsymbol{R_j^2}$
X_1	0.5140
X_2	0.3486
X_3	0.4872
X_4	0.5350

Untuk mendapatkan nilai VIF dilakukan perhitungan dengan menggunakan persamaan (2.29), sebagai berikut.

$$VIF_{X_1} = \frac{1}{1 - R_j^2}$$

$$VIF_{X_2} = \frac{1}{1 - R_j^2}$$

$$VIF_{X_4} = \frac{1}{1 - R_j^2}$$

$$VIF_{X_4} = \frac{1}{1 - R_j^2}$$

Nilai VIF untuk masing-masing prediktor yang digunakan dalam penelitian ini disajikan kembali pada Tabel 4.3.

Tabel 4.3 Nilai VIF Veriabel Prediktor

Variabel	VIF
X_1	2.0576
X_2	1.5352
X_3	1.9501
X_4	2.1505

Jika terdapat variabel prediktor dengan nilai VIF lebih dari 10, maka mengindikasi ternjadinya multikolinearitas antar variabel prediktor. Kasus multikolinearitas dapat diselesaikan dengan mengeliminasi terhadap variabel prediktor yang mana terjadinya multikolinearitas tersebut. Berdasarkan Tabel 4.3 diatas untuk data kejadian meninggal kasus difteri dapat dikatakan bahwa semua variabel prediktor memiliki nilai VIF kurang dari 10, sehingga dapat disimpulkan tidak terjadi kasus multikolinearitas pada variabel prediktor

4.3.3 Model Regressi Poisson Jumlah Kejadian Meninggal Kasus Difteri di Indonesia pada Tahun 2017

Pada sub bab 4.1 telah diperoleh dua model regresi Poisson yaitu dengan menggunakan metode *Maximum Likelihood Estimasi* dan metode *Moment Estimation.* Kedua model tersebut diterapkan dengan menggunakan data jumlah kejadian meninggal kasus difteri di Indonesia tahun 2017.

1. Terapan model Regresi Poisson menggunakan metode *Maximum Likelihood Estimtioni* (MLE) dengan menggunakan program R pada Lampiran F

$$\hat{\boldsymbol{\beta}}^{(7)} = \begin{bmatrix} 9.40147 \\ -0.03517 \\ -0.06957 \\ -0.02214 \\ -0.23081 \end{bmatrix}$$

Dari hasil estimasi di atas, didapatkan model Regresi Posisson dengan metode *Maximum Likelihood Estimation* yaitu:

$$\mu_i = \exp(9.40147 - 0.03517x_{1i} - 0.06957x_{2i} \\ - 0.02214x_{3i} - 0.23081x_{4i}) \tag{4.75}$$

dimana:

X_1 : Presentase *Universal Child Immunisation*

X_2 : Presentase Penduduk Miskin

X_3 : Presentase Rumah Tangga yang Memiliki Akses terhadap Sanitasi Layak

X_4 : Presentase Gizi Buruk

Uji serentak parameter model Regresi Poisson menggunakan uji *Likelihood Ratio Test* dengan hipotesis sebagai berikut.

1. $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$
2. H_1 : paling sedikit ada satu $\beta_j \neq 0$

   $j = 1, 2, 3, 4$
3. $\alpha = 0.25$

4. $G_{\text{hitung}} = 2\left(\sum_{i=1}^{n}\left(-\exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}\right) + y_i\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}\right)\right)\right) - 2\left(\sum_{i=1}^{n}\left(-\exp\left(\hat{\beta}_0\right) + y_i\hat{\beta}_0\right)\right)$
5. Terjadi Tolak H_0 apabila $G_{\text{hitung}} > \chi^2_{(0.25,5)}$

Hasil output *software* R pada lampiran F menunjukkan nilai AIC model Regresi Poisson sebesar 165.71, maka diperoleh G_{hitung} sebesar -2(log lihelihood) = 155.71. Nilai χ^2 tabel dengan jumlah parameter = 5 adalah 6.626, sehingga $G_{\text{hitung}} > \chi^2_{(0.25,5)}$. Dengan demikian terjadi tolak H_0 dan dapat disimpulkan bahwa model Regresi Poisson layak digunakan.

Pengujian parsial parameter model Regresi Poisson menggunakan uji Z dengan hipotesis sebagai berikut.

1. $H_0 : \beta_j = 0$
2. $H_1 : \beta_j \neq 0$

   $j = 1, 2, 3, 4$
3. $\alpha = 0.25$
4. $\text{Z}_{\text{hit}} = \frac{\widehat{\beta}_j}{\sqrt{Var(\boldsymbol{\beta})}}$
5. Terjadi tolak H_0 jika $|\text{Z}_{\text{hit}}| > |\text{Z}_{0.25}|$

Hasil output *software* R pada lampiran F yang disajikan pada Tabel 4.4.

Tabel 4. 4 Hasil Taksiran Parameter Regresi Poisson

Parameter	Taksiran	Standar Error	Z_{hit}
$\hat{\beta}_0$	9.40147	2.35195	3.997
$\hat{\beta}_1$	-0.03517	0.01691	-2.080*
$\hat{\beta}_2$	-0.06957	0.04264	-1.632*
$\hat{\beta}_3$	-0.02214	0.01533	-1.444*
$\hat{\beta}_4$	-0.23081	0.05419	-4.259*

*) taraf signifikan $Z_{(0.25)} = 0.67$

Tabel 4.4 pada kolom $\mathrm{Z_{hit}}$ menunjukkan bahwa semua parameter pada model Regresi Poisson memenuhi $|\mathrm{Z_{hit}}| > |\mathrm{Z_{0.25}}|$, sehingga terjadi tolak H_0.

Uji parsial untuk β_1:
1. $H_0 : \beta_1 = 0$
2. $H_1 : \beta_1 \neq 0$
3. $\alpha = 0.25$
4. $\mathrm{Z_{hit}} = \frac{\widehat{\beta}_1}{\sqrt{Var(\boldsymbol{\beta})}} = -2.080$
5. $|\mathrm{Z_{hit}}| = 2.080 > |\mathrm{Z_{0.25}}| = 0.67$
6. Terjadi tolak H_0

Uji parsial untuk β_2, β_3 dan β_4 pada Lampiran J.

Dengan demikian dapat disimpulkan bawah semua variabel signifikan terhadap model (4.75) atau bisa dikatakan semua variabel prediktor memberikan pengaruh signifikan terhadap jumlah kejadian meninggal kasus penyakit difteri di Indonesia tahun 2017.

2. Terapan menggunakan metode *Moment Estimation*, dilakukan dengan menggunakan program R pada Lampiran G. Nilai taksiran parameter yang diperoleh untuk Jumlah Kejadian Meninggal Kasus di Indonesia pada Tahun 2017 adalah

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} 34154339 \\ 286021598 \\ 35812405 \\ 230950253 \\ 64946088 \end{bmatrix}$$

Dari hasil estimasi didapatkan model Regresi Posisson dengan metode *Moment Estimation* yaitu:

$$\mu_i = \exp(34154339 + 286021598x_{1i} + 35812405x_{2i} + 230950253x_{3i} + 64946088x_{4i}) \qquad (4.76)$$

dimana:

X_1 : Presentase Universal Child Immunisation

X_2 : Presentase Penduduk Miskin

X_3 : Presentase Rumah Tangga yang Memiliki Akses terhadap Sanitasi Layak

X_4 : Presentase Gizi Buruk

Jika persamaan (4.76) diubah ke persamaan $\ln \mu_i$ dengan memasukkan data ke dalam persamaan maka akan diperoleh nilai $\ln \mu_i$ yang sangat besar. Sehingga sulit untuk melakukan uji parameter untuk melihat kelayakan dan faktor yang berpengaruh signifikan.

4.3.4 Model Regresi *Zero-Inflated Poisson* Jumlah Kejadian Meninggal Kasus Difteri di Indonesia pada Tahun 2017

Sama halnya dengan sub sub bab 4.3.3, model regresi *Zero-Inflated Poisson* pada sub bab 4.2 juga diterapkan dengan menggunakan data jumlah kejadian meninggal kasus difteri di Indonesia tahun 2017 berikut ini.

1. Terapan model Regresi *Zero-Infalted Poisson* menggunakan metode *Maximum Likelihood Estimasi* (MLE) dilakukan dengan menggunakan program R pada Lampiran H.

$$\hat{\boldsymbol{\beta}}^{(43)} = \begin{bmatrix} 9.431032 \\ -0.011949 \\ 0.003208 \\ -0.041451 \\ -0.226466 \end{bmatrix}$$

dan

$$\hat{\boldsymbol{\gamma}}^{(43)} = \begin{bmatrix} 1.41912 \\ 0.03236 \\ 0.13666 \\ -0.04912 \\ -0.05419 \end{bmatrix}$$

Dari hasil estimasi di atas, didapatkan model Regresi *Zero-Inflated Poisson* dengan metode *Maximum Likelihood Estimation* yaitu:

$$\mu_i = \exp(9.431032 - 0.011949\, x_{1i} + 0.003208 x_{2i} - 0.041451 x_{3i} - 0.226466 x_{4i})$$

$$\ln(\mu_i) = 9.431032 - 0.011949\, x_{1i} + 0.003208 x_{2i} - 0.041451 x_{3i} - 0.226466 x_{4i}$$

$$logit(\pi_i) = 1.41912 + 0.03236 x_{1i} + 0.13666 x_{2i} - 0.04912 x_{3i} - 0.05419 x_{4i} \qquad (4.77)$$

dimana:

X_1 : Presentase Universal Child Immunisation

X_2 : Presentase Penduduk Miskin

X_3 : Presentase Rumah Tangga yang Memiliki Akses terhadap Sanitasi Layak

X_4 : Presentase Gizi Buruk

Uji serentak parameter model Regresi *Zero-Inflated Poisson* menggunakan uji *Likelihood Ratio Test* dengan hipotesis sebagai berikut.

1. $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0$
2. H_1 : paling sedikit ada satu $\beta_j \neq 0$ atau $\gamma_j \neq 0$

   $j = 1, 2, 3, 4$
3. $\alpha = 0.25$
4. $$G_{\text{hitung}} = 2\left(\sum_{i=1}^{n}\left(z_i \boldsymbol{X}_i^T \widehat{\boldsymbol{\gamma}} - \ln\left(1 + \exp\left(\boldsymbol{X}_i^T \widehat{\boldsymbol{\gamma}}\right)\right)\right)\right) + 2\left(\sum_{i=1}^{n}(1 - z_i)\left(y_i \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}} - \exp\left(\boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}\right)\right)\right) - 2\left(\sum_{i=1}^{n}(z_i \hat{\gamma}_0 - \ln(1 + \exp(\hat{\gamma}_0)))\right) - 2\left(\sum_{i=1}^{n}(1 - z_i)\left(y_i \hat{\beta}_0 - \exp\left(\hat{\beta}_0\right)\right)\right)$$
5. Terjadi Tolak H_0 apabila $G_{\text{hitung}} > \chi^2_{(0.25,10)}$

Hasil output *software* R pada lampiran H menunjukan nilai devians model Regresi *Zero-Inflated Poisson* sebesar -32.01.

Maka diperoleh G_{hitung} sebesar -2(log lihelihood) = 64.02. Nilai χ^2 tabel dengan jumlah parameter =10 adalah 12.549, sehingga $G_{\text{hitung}} > \chi^2_{(0.25,10)}$. Dengan demikian terjadi tolak H_0 dan dapat disimpulkan bahwa model Regresi *Zero-Inflated Poisson* layak digunakan

Pengujian parsial parameter $\boldsymbol{\beta}$ model Regresi *Zero-Inflated Poisson* menggunakan uji Z dengan hipotesis sebagai berikut.

1. $H_0 : \beta_j = 0$
2. $H_1 : \beta_j \neq 0$
   $j = 1, 2, 3, 4$
3. $\alpha = 0.25$
4. $\text{Z}_{\text{hit}} = \frac{\widehat{\beta}_j}{\sqrt{Var(\boldsymbol{\beta})}}$
5. Terjadi tolak H_0 jika $|\text{Z}_{\text{hit}}| > |\text{Z}_{0.25}|$

Hasil output *software* R pada lampiran H yang disajikan kembali pada Tabel 4.5.

Tabel 4. 5 Hasil Taksiran Parameter Regresi *Zero-Inflated Poisson*

Parameter	Taksiran	Standar Error	Z_{hit}
$\hat{\beta}_0$	9.431032	4.842280	1.948
$\hat{\beta}_1$	-0.011949	0.038894	-0.307
$\hat{\beta}_2$	0.003208	0.081081	0.040
$\hat{\beta}_3$	-0.041451	0.021537	-1.925*
$\hat{\beta}_4$	-0.226466	0.086299	-2.624*
$\hat{\gamma}_0$	1.41912	7.04558	0.201
$\hat{\gamma}_1$	0.03236	0.004487	0.721*
$\hat{\gamma}_2$	0.13666	0.11229	1.217*
$\hat{\gamma}_3$	-0.04912	0.05009	-0.981*
$\hat{\gamma}_4$	-0.05419	0.15697	-0.345

*) taraf signifikan $Z_{(0.25)} = 0.67$

Pada kolom $\mathrm{Z_{hit}}$ menunjukkan bahwa parameter $\hat{\beta}_3$ dan $\hat{\beta}_4$ pada model Regresi *Zero-Inflated Poisson* memenuhi $|\mathrm{Z_{hit}}| > |\mathrm{Z_{0.25}}|$, sehingga terjadi tolak H_0 untuk parameter $\hat{\beta}_3$ dan $\hat{\beta}_4$.

Uji parsial untuk β_3:

1. $H_0 : \beta_3 = 0$
2. $H_1 : \beta_3 \neq 0$
3. $\alpha = 0.25$
4. $\mathrm{Z_{hit}} = \frac{\hat{\beta}_3}{\sqrt{Var(\boldsymbol{\beta})}} = -1.925$
5. $|\mathrm{Z_{hit}}| = 1.925 > |\mathrm{Z_{0.25}}| = 0.67$
6. Terjadi tolak H_0

Uji parsial untuk β_1, β_2 dan β_4 pada Lampiran K.

Variabel yang signifikan pada model (4.77) adalah X_3 dan X_4 yang menjelaskan setiap perubahan 1% presentase rumah tangga yang memiliki akses terhadap sanitasi layak melipatkan rata-rata kejadian meninggal sebesar 0.959 kali dan setiap perubahan 1% presentase gizi buruk melipatkan rata-rata kejadian meninggal sebesar 0.797 kali dengan syarat semua variabel lain konstan.

Pengujian parsial parameter $\boldsymbol{\gamma}$ model Regresi *Zero-Inflated Poisson* menggunakan uji Z dengan hipotesis sebagai berikut.

1. $H_0 : \gamma_j = 0$
2. $H_1 : \gamma_j \neq 0$
   $j = 1, 2, 3, 4$
3. $\alpha = 0.25$
4. $\mathrm{Z_{hit}} = \frac{\hat{\gamma}_j}{\sqrt{Var(\gamma)}}$
5. Terjadi tolak H_0 jika $|\mathrm{Z_{hit}}| > |\mathrm{Z_{0.25}}|$

Tabel 4.5 pada kolom Z_{hit} menunjukkan bahwa parameter $\hat{\gamma}_1$, $\hat{\gamma}_2$ dan $\hat{\gamma}_3$ pada model Regresi *Zero-Inflated Poisson* memenuhi $|Z_{\text{hit}}| > |Z_{0.25}|$, sehingga terjadi tolak H_0 untuk parameter $\hat{\gamma}_1$, $\hat{\gamma}_2$ dan $\hat{\gamma}_3$.

Uji parsial untuk γ_1:

1. $H_0 : \gamma_1 = 0$
2. $H_1 : \gamma \neq 0$
3. $\alpha = 0.25$
4. $Z_{\text{hit}} = \frac{\hat{\gamma}_1}{\sqrt{Var(\gamma)}} = 0.721$
5. $|Z_{\text{hit}}| = 0.721 > |Z_{0.25}| = 0.67$
6. Terjadi tolak H_0

Uji parsial untuk γ_2, γ_3 dan γ_4 pada Lampiran L.

Variabel yang signifikan pada model (4.77) adalah X_1, X_2 dan X_3 yang menjelaskan bahwa tidak terjadi meninggal kasus penyakit difteri di setiap provinsi di Indonesia dipengaruhi oleh presentase *Universal Child Immunisation*, presentase penduduk miskin dan presentase rumah tangga yang memiliki akses terhadap sanitasi layak.

Berdasarkan model (4.77) diketahui bahwa variabel prediktor yang mempengaruhi *poisson state* adalah X_3 (presentase rumah tangga yang memiliki akses terhadap sanitasi layak) dan X_4 (presentase gizi buruk). Sedangkan variabel prediktor yang mempengaruhi *zero state* adalah adalah X_1(presentase *Universal Child Immunisation*), X_2(presentase penduduk miskin) dan X_3(presentase rumah tangga yang memiliki akses terhadap sanitasi layak).

2. Terapan menggunakan metode *Moment Estimation*, dilakukan dengan menggunakan program R pada Lampiran I.

Nilai taksiran parameter yang diperoleh untuk Jumlah Kejadian Meninggal Kasus di Indonesia pada Tahun 2017 adalah

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} 32108603 \\ 2688894136 \\ 336673060 \\ 2171167443 \\ 610559331 \end{bmatrix}$$

dan

$$\widehat{\boldsymbol{\gamma}} = \begin{bmatrix} 25277725 \\ 211685048249 \\ 265048249 \\ 1709266936 \\ 480667155 \end{bmatrix}$$

Dari hasil estimasi didapatkan model regresi *Zero-Inflated Poisson* dengan metode *Moment Estimation*, yaitu:

$$\mu_i = \exp(32108603 + 2688894136x_{1i} + 336673060x_{2i} + 2171167443x_{3i} + 610559331x_{4i})$$

$$\ln(\mu_i) = 32108603 + 2688894136x_{1i} + 336673060x_{2i} + 2171167443x_{3i} + 610559331x_{4i}$$

$$logit(\pi_i) = 25277725 + 211685048249x_{1i} + 265048249x_{2i} + 1709266936x_{3i} + 480667155x_{4i} \quad (4.78)$$

Sama halnya dengan persamaan (4.76), pada persamaan (4.78) akan menghasilkan nilai yang besar jika data disubstitusi untuk mendapatkan nilai $\ln(\mu_i)$ dan $logit(\pi_i)$. Oleh karena itu, hasil parameter dengan metode metode *Moment Estimation* tidak bisa dilakukan uji parameter, atau bisa dikatakan nilai yang dihasilkan akan tidak terdefinisi.

4.4 *Akaike Information Criteria*

Membandingkan model-model regresi yang telah diperoleh untuk mencari model terbaik yang dapat digunakan pada data.

Pengukuran yang digunakan adalah *Akaike Information Criteria* (AIC). Pada Tabel (4.6), ditunjukkan hasil nilai AIC dari model regresi yang telah diperoleh pada persamaan (4.75), (4.76), (4.77), dan (4.78) dengan menggunakan software R.

Tabel 4. 6 Nilai AIC dari Model Regresi Poissson dan Regresi *Zero-Inlfated Poisson*

Model	Metode	AIC
Regresi Poisson	MLE	165.71
	MME	-
Regresi *Zero-Inflated Poisson*	MLE	84.02*
	MME	-

*) Model terbaik

Tabel 4.6 menunjukkan model dengan nilai AIC paling kecil adalah model Regresi *Zero-Inflated Poisson* dengan menggunakan metode *Maximum Likelihood Estimation*. Hal ini menunjukkan bahwa pemodelan menggunakan Regresi *Zero-Inflated Poisson* lebih baik jika dibandingkan dengan Regresi Poisson. Sedangkan model Regresi Poisson dan Regresi *Zero-Inflated Poisson* dengan metode *Moment Estimation* tidak memunculkan nilai AIC karena dianggap nilai dari estimasi parameter terlalu besar dan menghasilkan nilai yang tidak terdefinisi.

BAB V
PENUTUP

5.1 Kesimpulan

Kesimpulan yang dapat ditarik dari pembahasan pada penelitian ini yaitu:

1. a. Estimasi parameter $\hat{\boldsymbol{\beta}}$ dan $\hat{\boldsymbol{\gamma}}$ dari model Regresi *Zero-Inflated Poisson* (ZIP) dengan metode *Maximum Likelihood Estimation* adalah

$$\hat{\boldsymbol{\beta}}^{(m+1)} = \hat{\boldsymbol{\beta}}^{(m)} - \boldsymbol{H}^{-1}\boldsymbol{g}\left(\hat{\boldsymbol{\beta}}^{(m)}\right)$$

dengan

$$\boldsymbol{g}\left(\hat{\boldsymbol{\beta}}^{(m)}\right) = \sum_{i=1}^{n}\left(1 - z_i^{(m)}\right)\left(\boldsymbol{X}_i^T\right)\left(y_i - e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)$$

$$\boldsymbol{H} = \sum_{i=1}^{n}\left(1 - z_i^{(m)}\right)\left(-X_i^T e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\left(X_i^T e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)\right)$$

dan

$$\hat{\boldsymbol{\gamma}}^{(m+1)} = \hat{\boldsymbol{\gamma}}^{(m)} - \boldsymbol{H}^{-1}\boldsymbol{g}\left(\hat{\boldsymbol{\gamma}}^{(m)}\right)$$

dengan

$$\boldsymbol{g} = \sum_{i=1}^{n}\left(z_i^{(m)}\boldsymbol{X}_i^T - \frac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}}{\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)}\boldsymbol{X}_i^T\right)$$

$$\boldsymbol{H} = -\sum_{i=1}^{n}\left(\frac{e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}}{\left(1+e^{\boldsymbol{X}_i^T\boldsymbol{\gamma}}\right)^2}\boldsymbol{X}_i^T\boldsymbol{X}_i\right).$$

   b. Estimasi parameter $\hat{\boldsymbol{\beta}}$ dan $\hat{\boldsymbol{\gamma}}$ dari model Regresi *Zero-Inflated Poisson* (ZIP) dengan metode *Moment Estimation* adalah

$$\hat{\boldsymbol{\beta}} = \boldsymbol{X_i}\left(\boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1}(\ln(\bar{y}^2 - \bar{y}) - \ln(\bar{y}))$$

dan

$$\hat{\boldsymbol{\gamma}} = \boldsymbol{X_i}\left(\boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1}(\ln(\bar{y}^2 - \bar{y} - (\bar{y})^2) - 2\ln(\bar{y}))$$

2. a. Penerapan estimasi parameter yang diperoleh dengan metode *Maximum Likelihood Estimation* pada data jumlah kejadian meninggal kasus penyakit difteri di Indonesia tahun 2017 menghasilkan bentuk model Regresi *Zero-Inflated Poisson* (ZIP) berikut,

$$\ln(\mu_i) = 9.431032 - 0.011949\, x_{1i} + 0.003208 x_{2i} - 0.041451 x_{3i} - 0.226466 x_{4i}$$
$$logit(\pi_i) = 1.41912 + 0.03236 x_{1i} + 0.13666 x_{2i} - 0.04912 x_{3i} - 0.05419 x_{4i}$$

   b. Penerapan estimasi parameter yang diperoleh dengan metode *Moment Estimation* menghasilkan bentuk model Regresi *Zero-Inflated Poisson* (ZIP) berikut,

$$\ln(\mu_i) = 32108603 + 2688894136 x_{1i} + 336673060 x_{2i} + 2171167443 x_{3i} + 610559331 x_{4i}$$
$$logit(\pi_i) = 25277725 + 211685048249 x_{1i} + 265048249 x_{2i} + 1709266936 x_{3i} + 480667155 x_{4i}$$

   c. Pengujian parameter model Regresi *Zero-Inflated Poisson* menunjukkan bahwa hasil estimasi parameter yang diperoleh menggunakan metode *Maximum Likelihood Estimation* lebih baik untuk diterapkan pada data jumlah kejadian meninggal kasus penyakit difteri di Indonesia tahun 2017. Tidak terjadi kasus meninggal penyakit difteri di setiap provinsi di Indonesia dipengaruhi oleh presentase *Universal Child Immunisation*, presentase penduduk miskin dan presentase rumah tangga yang memiliki akses terhadap sanitasi layak.

5.2 Saran

Penelitian selanjunya disarankan untuk memodelkan kembali Regresi *Zero-Inflted Poisson* dengan metode estimasi lainnya.

DAFTAR PUSTAKA

[1] Cameron, A. Colin and Trivedi, Pravin K. *Regression Analysis of Count Data.* New York : Cambridge University Press, 1998

[2] Hall, Daniel B. and Shen, Jing. *Robust Estimation for Zero-Inflated Poisson Regression.* Scandinavian Journal of Statistics, 2009

[3] Staub, Kevin E. and Winkelmann, Rainer.*Consistent Estimation of Zero-Inflated count models.* University of Zurich : Socioeconomic Institute, Health Economics, p. DOI: 10.1002/hec.2844, 2012

[4] Lambert, Diane. *Zero-Inflated Poisson Regression, With an Appication to Defects in Manufacturing.* American Statistical Association and the American Society for Quality Control,Technometrics, pp. Vol.34, No.1.1-4, 1992

[5] Lestari, A., dkk. *Pemodelan Regresi Zero Inflated Poisson (Aplikasi Pada DataPekerja Seks Komersial di Klinik Reproduksi Putat Jaya).* Surabaya : Institut Teknologi Sepuluh Nopember, 2008.

[6] Nusantara, Ratih Kumala Puspa. *Pemodelan Jumlah Kasus Penyakit Tetanus Neonatorum di Provinsi Jawa Timur Tahun 2012 dengan Geographically Weighted Zero-Inflated Poisson Regression (GWZIPR).* Surabaya : Institut Teknologi Sepuluh Nopember, 2015.

[7] Bain, L.J dan Engelhardt, M. *Introduction to Probability and Mathematical Statistics*, Duxburry: Thomson Learning, USA. 1992

[8] Sahoo, P. *Probability and Mathematical Statistics*. USA: Department of Mathematics, University of Louisville, Louisville, 2013

[9] Agresti, A. *Caterogical Data Analysis, 2nd edition*, John Wiley & Sons, Inc., New Jersey, 2002

[10] Olsson, Ulf. *Generalized Linier Model : An Applied Approach*.Studentlitteratur, 2002.

[11] Bolstad, B.M. *Comparing some iterative method of parameter estimation for censored gamma data*. The University of Waikato, 1998

[12] Ehlers, R. *Maximum likelihood estimation prosedures for categorial data*. South Africa: University of Pretoria, 2002

[13] Dempster, A. P, Laird, N. M, Rubin, D. B. Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1, halaman 1-38, 1977

[14] James, G., dkk. *An Introduction to Statistical Learning*. New York :Springer Science + Bussiness Media, 2013

[15] Greene, W.H. *Econometric Analysis Seventh Edition*. United States: Pearson Education. 2012

[16] [Kemenkes RI]. Kementrian Kesehatan Republik Indonesia. *Profil Kesehatan Indonesia 2017*. Jakarta: Kementrian Kesehatan Republik Indonesia, 2018

LAMPIRAN

LAMPIRAN A. Pembuktian Teorema 2.1

Berikut merupakan pembuktian dari Teorema 2.1, yaitu mencari nilai *moment generating function* (MGF) dari peubah acak Y

$$M_Y(t) = E(e^{ty})$$

$$M_Y(t) = \sum_{y=0}^{n} e^{ty} f(y)$$

$$M_Y(t) = \sum_{y=0}^{n} e^{ty} \frac{e^{-\mu}(\mu)^y}{y!}$$

$$M_Y(t) = e^{-\mu} \sum_{y=0}^{n} \frac{(\mu e^t)^y}{y!}$$

$$M_Y(t) = e^{-\mu} e^{\mu e^t} \sum_{y=0}^{n} \frac{e^{-\mu e^t}(\mu e^t)^y}{y!}$$

$$M_Y(t) = e^{-\mu} e^{\mu e^t}$$

$$M_Y(t) = e^{\mu(e^t - 1)}$$

Nilai ekspektasi dari peubah acak Y atau $E(Y)$ dapat dicari sebagai berikut,

$$E(Y) = M'_y(t = 0)$$

$$E(Y) = \frac{d}{dt} e^{\mu(e^t - 1)} \Big|_{t=0}$$

$$E(Y) = \mu e^t e^{\mu(e^t - 1)} \Big|_{t=0}$$

$$E(Y) = \mu e^0 e^{\mu(e^0 - 1)}$$

$$E(Y) = \mu$$

Sama halnya dalam mencari nilai ekspektasi, untuk mencari nilai $E(Y^2)$ dapat diperoleh dari turunan kedua *moment generating function* (MGF),

$$E(Y^2) = M''_y(t = 0)$$

$$E(Y^2) = \frac{d^2}{dt^2} e^{\mu(e^t-1)}\bigg|_{t=0}$$
$$E(Y^2) = \frac{d}{dt} \mu e^t e^{\mu(e^t-1)}\bigg|_{t=0}$$
$$E(Y^2) = (\mu e^t)^2 e^{\mu(e^t-1)} + \mu e^t e^{\mu(e^t-1)}\Big|_{t=0}$$
$$E(Y^2) = (\mu e^0)^2 e^{\mu(e^0-1)} + \mu e^0 e^{\mu(e^0-1)}$$
$$E(Y^2) = \mu^2 + \mu$$

Nilai varians dari peubah acak Y atau $Var(Y)$ dapat dicari sebagai beikut:

$$Var(Y) = E\left[\left(Y - E(Y)\right)^2\right]$$
$$Var(Y) = E\left[Y^2 - 2YE(Y) + \left(E(Y)\right)^2\right]$$
$$Var(Y) = E(Y^2) - 2\,E(Y)E(Y) + \left(E(Y)\right)^2$$
$$Var(Y) = \mu^2 + \mu - 2\mu\mu + \mu^2$$
$$Var(Y) = 2\mu^2 + \mu - 2\mu^2$$
$$Var(Y) = \mu$$

atau

$$Var(Y) = E(Y^2) - \left(E(Y)\right)^2$$
$$Var(Y) = \mu^2 + \mu - \mu^2$$
$$Var(Y) = \mu$$

LAMPIRAN B. Pembuktian Teorema 2.2

Berikut merupakan pembuktian dari Teorema 2.2, yaitu mencari nilai ekspektasi dari peubah acak Y atau $E(Y)$

$$E(Y) = \sum_{y=0}^{n} y f(y)$$
$$E(Y) = \sum_{y=0}^{n} y \binom{n}{y} p^y (1-p)^{n-y}$$

$$E(Y) = \sum_{y=0}^{n} y \frac{n!}{y!\,(n-y)!} p^y (1-p)^{n-y}$$
$$E(Y) = \sum_{y=0}^{n} y \frac{n(n-1)!}{y(y-1)!\,(n-y)!} pp^{y-1} (1-p)^{n-y}$$
$$E(Y) = np \sum_{y=0}^{n} \frac{(n-y)!}{(y-1)!\,(n-y)!} p^{y-1} (1-p)^{n-y}$$
$$E(Y) = np$$

Nilai varians dari peubah acak Y atau $Var(Y)$ dapat dicari sebagai berikut,

$$Var(Y) = E\left[\left(Y - E(Y)\right)^2\right]$$
$$Var(Y) = E(Y^2) - \left(E(Y)\right)^2$$

Untuk mendapatkan nilai varians, terlebih dahulu menetukan bagian yang belum diketahui yaitu $E(Y^2)$

$$E(Y^2) = E(Y^2) - E(Y) + E(Y)$$
$$E(Y^2) = E(Y^2 - Y) + E(Y)$$
$$E(Y^2) = E\big(Y(Y-1)\big) + E(Y)$$

Menyelesaikan $E\big(Y(Y-1)\big)$, sebagai berikut,

$$E\big(Y(Y-1)\big) = \sum_{y=0}^{n} y(y-1) f(y)$$
$$E\big(Y(Y-1)\big) = \sum_{y=0}^{n} y(y-1) \binom{n}{y} p^y (1-p)^{n-y}$$
$$E\big(Y(Y-1)\big) = \sum_{y=0}^{n} y(y-1) \frac{n!}{y!\,(n-y)!} p^y (1-p)^{n-y}$$
$$E\big(Y(Y-1)\big) = \sum_{y=0}^{n} y(y-1) \frac{n(n-1)(n-2)!}{y(y-1)(y-2)!\,(n-y)!} p^2 p^{y-2} (1-p)^{n-y}$$
$$E\big(Y(Y-1)\big) = n(n-1)p^2 \sum_{y=0}^{n} \frac{(n-2)!}{(y-2)!\,(n-y)!} p^{y-2} (1-p)^{n-y}$$

$E\big(Y(Y-1)\big) = \ n^2p^2 - np^2$

Selanjutnya,

$$E(Y^2) = \ n^2p^2 - np^2 + np$$
$$E(Y^2) = \ n^2p^2 + np(1-p)$$

Dengan demikian diperoleh nilai varians dari peubah acak Y, yaitu:

$$Var(Y) = \ n^2p^2 + np(1-p) - n^2p^2$$
$$Var(Y) = \ np(1-p)$$

Nilai *moment generating function* (MGF) dari peubah acak Y atau $M_Y(t)$ dapat dicari sebagai beikut:

$$M_Y(t)= \ E(e^{ty})$$
$$M_Y(t)= \ \sum_{y=0}^{n} e^{ty} f(y)$$
$$M_Y(t)= \ \sum_{y=0}^{n} e^{ty} \binom{n}{y} p^y (1-p)^{n-y}$$
$$M_Y(t)= \ \sum_{y=0}^{n} e^{ty} \frac{n!}{y!\,(n-y)!} p^y (1-p)^{n-y}$$
$$M_Y(t)= \ \sum_{y=0}^{n} \frac{n!}{y!\,(n-y)!} (pe^t)^y (1-p)^{n-y}$$

Dengan menggunakan teorema binomial newton diperoleh nilai *moment generating function* (MGF) sebagai beikut,

$$M_Y(t)= \ (pe^t + 1 - p)^{y \mp n - y}$$
$$M_Y(t)= \ (1 - p + pe^t)^n$$

LAMPIRAN C. Pembuktian Nilai Ekspektasi dan Varians Model *Zero-Inflated Poisson*

Berikut merupakan pembuktian dalam mencari nilai ekspektasi dan varians dari Y_i untuk model *Zero-Inflated Poisson*.

Misalkan $W \sim BIN\left(1, (1-\pi)\right)$ atau $W \sim Ber(1-\pi)$ dan $D \sim POI(\mu)$ merupakan variabel acak independen, jika didefinisikan

$$Y = WD$$

Maka Y berdistribusi sesuai yang dinginkan, dimana

$P(Y = 0) = P(W = 0) + P(W = 1, D = 0) = \pi + (1-\pi)e^{-\mu}$

$P(Y = k) = P(W = 1, D = k)$ untuk $k \neq 0$

Sehingga untuk mencari nilai ekspektasi dari Y_i sebagai berikut:

$$E(Y) = E(WD)$$
$$E(Y) = E(W)E(D)$$
$$E(Y) = (1-\pi)\,\mu$$

Kemudian untuk mencari $E(Y^2)$, yaitu:

$$E(Y^2) = E(W^2D^2)$$
$$E(Y^2) = E(W^2)E(D^2)$$
$$E(Y^2) = (1-\pi)(\mu^2 + \mu)$$

Sehingga diperoleh nilai varians dari Y_i berikut ini,

$$Var(Y) = E(Y^2) - \left(E(Y)\right)^2$$
$$Var(Y) = (1-\pi)(\mu^2 + \mu) - (1-\pi)^2\,\mu^2$$
$$Var(Y) = \mu(1-\pi)[(\mu + 1) - (1-\pi)\mu]$$
$$Var(Y) = \mu(1-\pi)[\mu + 1 - \mu - \pi\mu]$$
$$Var(Y) = \mu(1-\pi)(1-\pi\mu)$$

LAMPIRAN D.
Data Jumlah Kejadian Meninggal Kasus Penyakit Difteri di Indonesia Tahun 2017

i	Provinsi	Y	X_1	X_2	X_3	X_4
1	Aceh	4	65.38	15.92	63.29	23.10
2	Sumatera Utara	0	75.41	9.28	71.84	24.50
3	Sumatera Barat	0	75.71	6.75	52.77	17.10
4	Riau	1	62.57	7.41	71.68	21.00
5	Jambi	0	94.57	7.90	66.36	16.40
6	Sumatera Selatan	0	91.74	13.10	66.18	13.20
7	Bengkulu	0	83.95	15.59	45.31	14.30
8	Lampung	0	93.96	13.04	52.89	17.70
9	Kep.Bangka Belitung	2	97.67	5.30	86.33	17.80
10	Kep. Riau	0	93.27	6.13	83.56	13.00
11	DKI Jakarta	2	100.00	3.78	91.13	14.75
12	Jawa Barat	14	88.72	7.83	63.38	14.00
13	Jawa Tengah	0	99.95	12.23	70.04	14.50
14	DI Yogyakarta	0	100.00	12.36	89.40	11.30
15	Jawa Timur	12	86.18	11.20	68.83	13.10
16	Banten	8	81.04	5.59	71.93	18.50
17	Bali	0	98.60	4.14	90.51	9.40
18	NTB	0	92.52	15.05	69.52	17.70
19	NTT	0	68.28	21.38	42.71	32.60
20	Kalimantan Barat	1	70.64	7.86	49.65	25.60
21	Kalimantan Tengah	0	68.59	5.26	45.46	25.50
22	Kalimantan Selatan	0	88.79	4.65	61.12	20.50
23	Kalimantan Timur	0	81.20	6.08	76.73	15.80

LAMPIRAN D. (Lanjutan)

Data Jumlah Kejadian Meninggal Kasus Penyakit Difteri di Indonesia Tahun 2017

i	Provinsi	Y	X_1	X_2	X_3	X_4
24	Kalimantan Utara	0	51.98	6.96	64.40	20.90
25	Sulawesi Utara	0	79.40	7.90	72.83	22.80
26	Sulawesi Tengah	0	83.86	14.22	59.48	24.80
27	Sulawesi Selatan	0	96.45	9.48	73.00	17.50
28	Sulawesi Tenggara	0	85.10	11.97	69.25	27.60
29	Gorontalo	0	92.08	17.14	58.75	22.30
30	Sulawesi Barat	0	75.85	11.18	58.09	23.30
31	Maluku	0	59.95	18.29	66.59	23.40
32	Maluku Utara	0	80.77	6.44	65.30	15.90
33	Papua Barat	0	77.22	23.12	64.20	27.20
34	Papua	0	21.43	27.76	33.06	30.30

Sumber : http://www.kemkes.go.id

Keterangan :

Y : Jumlah Meninggal Kasus Difteri

X_1 : Presentase Universal Child Immunisation

X_2 : Presentase Penduduk Miskin

X_3 : Presentase Rumah Tangga yang Memiliki Akses terhadap Sanitasi Layak

X_4 : Presentase Gizi Buruk

LAMPIRAN E. Model Linier Variabel Prediktor

```
Call:
lm(formula = X1 ~ X2 + X3 + X4, data = DB)

Residuals:
    Min      1Q  Median      3Q     Max 
-31.525  -4.765   2.092   8.368  17.675 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  83.3347    20.6571   4.034 0.000348 ***
X2           -0.1124     0.4488  -0.251 0.803906    
X3            0.3794     0.2031   1.868 0.071556 .  
X4           -1.3136     0.4879  -2.693 0.011486 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.05 on 30 degrees of freedom
Multiple R-squared:  0.5104,	Adjusted R-squared:  0.4615 
F-statistic: 10.43 on 3 and 30 DF,  p-value: 7.329e-05
```

```
Call:
lm(formula = X2 ~ X1 + X3 + X4, data = DB)

Residuals:
     Min       1Q   Median       3Q      Max 
-10.4529  -2.9267  -0.5161   3.4797   9.5855 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 15.38962   10.04057   1.533    0.136
X1          -0.01857    0.07412  -0.251    0.804
X3          -0.13524    0.08364  -1.617    0.116
X4           0.30372    0.21385   1.420    0.166

Residual standard error: 4.899 on 30 degrees of freedom
Multiple R-squared:  0.3486,	Adjusted R-squared:  0.2835 
F-statistic: 5.351 on 3 and 30 DF,  p-value: 0.004504
```

LAMPIRAN E. Model Linier Variabel Prediktor (Lanjutan)

```
Call:
lm(formula = X3 ~ X1 + X2 + X4, data = DB)

Residuals:
     Min       1Q   Median       3Q      Max 
-21.6798  -5.3624   0.5536   7.8542  14.2601 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)  61.8879    18.6752   3.314  0.00241 **
X1            0.2746     0.1470   1.868  0.07156 . 
X2           -0.5927     0.3666  -1.617  0.11637   
X4           -0.6093     0.4489  -1.357  0.18485   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.26 on 30 degrees of freedom
Multiple R-squared:  0.4872,	Adjusted R-squared:  0.4359 
F-statistic:   9.5 on 3 and 30 DF,  p-value: 0.0001439
```

```
Call:
lm(formula = X4 ~ X1 + X2 + X3, data = DB)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.8370 -2.7771 -0.3561  3.1431  8.6572 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 35.64392    5.64743   6.312 5.86e-07 ***
X1          -0.14817    0.05503  -2.693   0.0115 *  
X2           0.20743    0.14605   1.420   0.1659    
X3          -0.09494    0.06996  -1.357   0.1848    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.049 on 30 degrees of freedom
Multiple R-squared:  0.535,	Adjusted R-squared:  0.4885 
F-statistic:  11.5 on 3 and 30 DF,  p-value: 3.458e-05
```

LAMPIRAN F. *Syntax* Model Regresi Poisson dengan Metode *Maximum Likelihood Estimation* (MLE)

```
Call:
glm(formula = Y ~ X1 + X2 + X3 + X4, family = "poisson", data = DB)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.9525  -1.7243  -0.9666  -0.3371   4.5894

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  9.40147    2.35195   3.997 6.41e-05 ***
X1          -0.03517    0.01691  -2.080   0.0376 *
X2          -0.06957    0.04264  -1.632   0.1028
X3          -0.02214    0.01533  -1.444   0.1488
X4          -0.23081    0.05419  -4.259 2.05e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 160.75  on 33  degrees of freedom
Residual deviance: 130.45  on 29  degrees of freedom
AIC: 165.71

Number of Fisher Scoring iterations: 7
```

LAMPIRAN G. *Syntax* Model Regresi Poisson dengan Metode *Moment Estimation*

```
mat <- as.matrix(Data_BesaraY)
d = matrix(mat, nrow = 34, ncol = 1) #matrix Y

matx<- as.matrix(Data_BesaraX)
e = matrix(matx, nrow = 34, ncol = 4) #matrix X

et = t(e) # matrix X'

lnY=log(mat) #log Y

n=34  #banyak data
matn<- as.matrix(n)
f = matrix(matn, nrow = 34, ncol = 1)
g = log(f)

s = sum(mat) #jumlah Y
matu<- as.matrix(s)
u =matrix(matu, nrow = 34, ncol = 1)
r = mat^2
t = sum(r) #jumlah Y^2
matv<- as.matrix(t)
v =matrix(matv, nrow = 34, ncol = 1)

#Regresi Poisson MME
h1=e%*%et
h2=t(h1)
h3=et%*%h2
h4=(lnY-g)
h=h3%*%h4
beta=et%*%t(e%*%et)%*%(log(u)-g)
```

beta:

```
           [,1]
[1,]    3415439
[2,]  286021598
[3,]   35812405
[4,]  230950253
[5,]   64946088
```

LAMPIRAN H. *Syntax* Model Regresi *Zero-Inflated Poisson* (ZIP) dengan Metode *Maximum Likelihood Estimation* (MLE)

```
Call:
zeroinfl(formula = Y ~ X1 + X2 + X3 + X4 | X1 + X2 + X3 + X4, data = DB, dist = "poisson")

Pearson residuals:
    Min      1Q  Median      3Q     Max 
-0.9087 -0.5240 -0.3233 -0.1908  3.0101 

Count model coefficients (poisson with log link):
             Estimate Std. Error z value Pr(>|z|)   
(Intercept)  9.431032   4.842280   1.948  0.05146 . 
X1          -0.011949   0.038894  -0.307  0.75868   
X2           0.003208   0.081081   0.040  0.96844   
X3          -0.041451   0.021537  -1.925  0.05428 . 
X4          -0.226466   0.086299  -2.624  0.00868 **

Zero-inflation model coefficients (binomial with logit link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.41912    7.04558   0.201    0.840
X1           0.03236    0.04487   0.721    0.471
X2           0.13666    0.11229   1.217    0.224
X3          -0.04912    0.05009  -0.981    0.327
X4          -0.05419    0.15697  -0.345    0.730
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Number of iterations in BFGS optimization: 43 
Log-likelihood: -32.01 on 10 Df
```

LAMPIRAN I. *Syntax* Model Regresi *Zero-Inflated Poisson* (ZIP) dengan Metode *Moment Estimation*

```
mat <- as.matrix(Data_BesaraY)
d = matrix(mat, nrow = 34, ncol = 1) #matrix Y

matx<- as.matrix(Data_BesaraX)
e = matrix(matx, nrow = 34, ncol = 4) #matrix X

et = t(e) # matrix X'

lnY=log(mat) #log Y

n=34  #banyak data
matn<- as.matrix(n)
f = matrix(matn, nrow = 34, ncol = 1)
g = log(f)

s = sum(mat) #jumlah Y
matu<- as.matrix(s)
u =matrix(matu, nrow = 34, ncol = 1)
r = mat^2
t = sum(r) #jumlah Y^2
matv<- as.matrix(t)
v =matrix(matv, nrow = 34, ncol = 1)

#Regresi ZIP MME
#parameter betta= betaz
betaz=et%*%t(e%*%et)%*%(log(v-u)-g)
#parameter gamma=gammaz
gammaz=et%*%t(e%*%et)%*%(log(v%*%n-u%*%n)-log(u)%*%2)
```

beta :

```
           [,1]
[1,]   32108603
[2,] 2688894136
[3,]  336673060
[4,] 2171167443
[5,]  610559331
```

gamma :

```
           [,1]
[1,]   25277725
[2,] 2116850940
[3,]  265048249
[4,] 1709266936
[5,]  480667155
```

LAMPIRAN J. Pengujian Parsial Parameter Model Regresi Poisson

Uji parsial untuk β_2:

1. $H_0 : \beta_2 = 0$
2. $H_1 : \beta_2 \neq 0$
3. $\alpha = 0.25$
4. $\mathrm{Z_{hit}} = \frac{\widehat{\beta}_2}{\sqrt{Var(\boldsymbol{\beta})}} = -1.632$
5. $|\mathrm{Z_{hit}}| = 1.632 > |\mathrm{Z_{0.25}}| = 0.67$
6. Terjadi tolak H_0

Uji parsial untuk β_3:

1. $H_0 : \beta_3 = 0$
2. $H_1 : \beta_3 \neq 0$
3. $\alpha = 0.25$
4. $\mathrm{Z_{hit}} = \frac{\widehat{\beta}_3}{\sqrt{Var(\boldsymbol{\beta})}} = -1.444$
5. $|\mathrm{Z_{hit}}| = 1.444 > |\mathrm{Z_{0.25}}| = 0.67$
6. Terjadi tolak H_0

Uji parsial untuk β_4:

1. $H_0 : \beta_4 = 0$
2. $H_1 : \beta_4 \neq 0$
3. $\alpha = 0.25$
4. $\mathrm{Z_{hit}} = \frac{\widehat{\beta}_4}{\sqrt{Var(\boldsymbol{\beta})}} = -4.259$
5. $|\mathrm{Z_{hit}}| = 4.259 > |\mathrm{Z_{0.25}}| = 0.67$
6. Terjadi tolak H_0

LAMPIRAN K. Pengujian Parsial Parameter $\boldsymbol{\beta}$ Model Regresi *Zero-Inflated Poisson* (ZIP)

Uji parsial untuk β_1:
1. $H_0 : \beta_1 = 0$
2. $H_1 : \beta_1 \neq 0$
3. $\alpha = 0.25$
4. $\mathrm{Z_{hit}} = \frac{\widehat{\beta_1}}{\sqrt{Var(\boldsymbol{\beta})}} = -0.307$
5. $|\mathrm{Z_{hit}}| = 0.307 < |\mathrm{Z_{0.25}}| = 0.67$
6. Terjadi terima H_0

Uji parsial untuk β_2:
1. $H_0 : \beta_2 = 0$
2. $H_1 : \beta_2 \neq 0$
3. $\alpha = 0.25$
4. $\mathrm{Z_{hit}} = \frac{\widehat{\beta_2}}{\sqrt{Var(\boldsymbol{\beta})}} = 0.040$
5. $|\mathrm{Z_{hit}}| = 0.040 < |\mathrm{Z_{0.25}}| = 0.67$
6. Terjadi terima H_0

Uji parsial untuk β_4:
1. $H_0 : \beta_4 = 0$
2. $H_1 : \beta_4 \neq 0$
3. $\alpha = 0.25$
4. $\mathrm{Z_{hit}} = \frac{\widehat{\beta_4}}{\sqrt{Var(\boldsymbol{\beta})}} = -2.624$
5. $|\mathrm{Z_{hit}}| = 2.624 > |\mathrm{Z_{0.25}}| = 0.67$
6. Terjadi tolak H_0

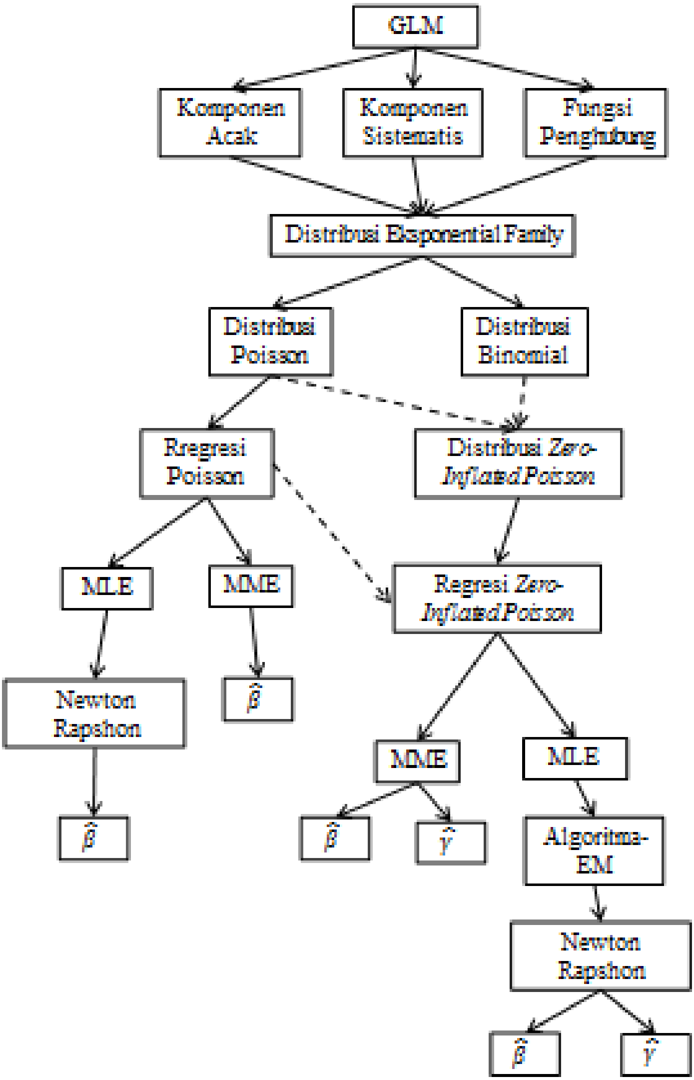
LAMPIRAN L. Pengujian Parsial Parameter γ Model Regresi *Zero-Inflated Poisson* (ZIP)

Uji parsial untuk γ_2:
1. $H_0 : \gamma_2 = 0$
2. $H_1 : \gamma_2 \neq 0$
3. $\alpha = 0.25$
4. $\mathrm{Z_{hit}} = \frac{\hat{\gamma}_2}{\sqrt{Var(\gamma)}} = 1.217$
5. $|\mathrm{Z_{hit}}| = 1.217 > |\mathrm{Z_{0.25}}| = 0.67$
6. Terjadi tolak H_0

Uji parsial untuk γ_3:
1. $H_0 : \gamma_3 = 0$
2. $H_1 : \gamma_3 \neq 0$
3. $\alpha = 0.25$
4. $\mathrm{Z_{hit}} = \frac{\hat{\gamma}_3}{\sqrt{Var(\gamma)}} = -0.981$
5. $|\mathrm{Z_{hit}}| = 0.981 > |\mathrm{Z_{0.25}}| = 0.67$
6. Terjadi tolak H_0

Uji parsial untuk γ_4:
1. $H_0 : \gamma_4 = 0$
2. $H_1 : \gamma_4 \neq 0$
3. $\alpha = 0.25$
4. $\mathrm{Z_{hit}} = \frac{\hat{\gamma}_4}{\sqrt{Var(\gamma)}} = -0.345$
5. $|\mathrm{Z_{hit}}| = 0.345 < |\mathrm{Z_{0.25}}| = 0.67$
6. Terjadi terima H_0

LAMPIRAN M. *MIND MAP*



BIODATA PENULIS

Penulis bernama Alvinda Nisma Yusniar, lahir di Ponorogo, 04 April 1996. Penulis yang merupakan anak kedua dari dua bersaudara ini telah menempuh pendidikan formal mulai dari TK Darmawanita Wagir kidul Ponorogo, SDN 1 Wagirkidul Ponorogo, SMP Terpadu Ponorogo, SMAN 2 Ponorogo. Setelah lulus dari SMAN 2 Ponorogo penulis melanjutkan pendidikan ke jenjang S1 di Departemen Matematika Institut Teknologi Sepuluh Nopember (ITS) pada tahun 2015-sekarang melalui jalur SNMPTN 2015. Penulis mengambil Bidang Minat Matematika Terapan. Selain aktif di kuliah, penulis juga aktif berorganisasi di KM ITS yaitu Himpunan Mahasiswa Matematika (HIMATIKA) ITS sebagai Kepala Sport and Art Department periode (2017-2018). Apabila ingin memberikan kritik dan saran mengenai Tugas Akhir ini dapat menghubungi melalui email penulis yaitu alvindanismay@gmail.com. Terimakasih dan semoga bermanfaat.