PERBANDINGAN MODEL REGRESI *QUASI* POISSON, *ZERO INFLATED* POISSON DAN *ZERO INFLATED* NEGATIVE BINOMIAL PADA DATA OVERDISPERSI

(Studi Kasus: Data Angka Kematian Ibu Kabupaten Lampung Timur Tahun 2013)

(Skripsi)

Oleh

ANISA FITRIYANI

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023

PERBANDINGAN MODEL REGRESI *QUASI* POISSON, *ZERO INFLATED* POISSON DAN *ZERO INFLATED* NEGATIVE BINOMIAL PADA DATA OVERDISPERSI

(Studi Kasus: Data Angka Kematian Ibu Kabupaten Lampung Timur Tahun 2013)

Oleh

ANISA FITRIYANI

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023

ABSTRACT

COMPARISON OF QUASI POISSON REGRESSION MODELS, ZERO INFLATED POISSON AND ZERO INFLATED NEGATIVE BINOMIAL ON OVERDISPERSION DATA

(Case Study: Data on Maternal Mortality Rate in East Lampung Regency in 2013)

By

ANISA FITRIYANI

Regression analysis is a method used in analyzing the relationship between the independent variable and the dependent variable. One of the regression models that can be used to analyze the relationship between the dependent variable Y in the form of discrete data and the independent variable X in the form of continuous, discrete or mixed data is the Poisson regression model. The assumption that must be met in the case of Poisson regression is that the variance value of the response variable must be equal to its mean (equidispersion). When the variance value is greater than the average, this condition is called overdispersion.

In this study using data on maternal mortality with a response variable that has a lot of zero values zero (0), with several factors that affect the number of deaths. The results of the analysis show that the Zero Inflated Poisson (ZIP) regression is good for modeling overdispersion of data and zero inflation. This is because the AIC and RMSE values of the ZIP regression are smaller than other methods, namely Quasi Poisson and ZINB. The model obtained from the ZIP regression analysis is: $\ln(\mu) = -3.26175 + 0.173044X1 - 0.150728X2 + 0.001592X3$ and $\text{logit}(\omega) = -402.3538 + 5.3352X1 - 1.2401X2 - 0.1844X3$.

Keywords: Overdispersion, Poisson Regression, Quasi Poisson Regression, Zero Inflated Poisson Regression, Zero Inflated Negative Binomial Regression.

ABSTRAK

PERBANDINGAN MODEL REGRESI *QUASI* POISSON, *ZERO INFLATED* POISSON DAN *ZERO INFLATED* NEGATIVE BINOMIAL PADA DATA OVERDISPERSI

(Studi Kasus: Data Angka Kematian Ibu Kabupaten Lampung Timur Tahun 2013)

Oleh

ANISA FITRIYANI

Analisis regresi adalah metode yang digunakan dalam menganalisis hubungan antara vaariabe bebas dengan variabel terikat. Salah satu model regresi yang dapat digunakan untuk menganalisis hubungan antara variabel *dependent* Y yang berupa data diskrit dan variabel *independent* X berupa data kontinyu, diskrit atau campuran adalah model regresi Poisson. Asumsi yang harus dipenuhi dalam kasus regresi Poisson adalah bahwa nilai variansi dari variabel respon harus sama dengan rataannya (*equidispersi*). Ketika nilai variansi lebih besar dari rata-rata maka kondisi ini disebut *overdispersion*

Pada penelitian ini menggunakan data Angka Kematian Ibu dengan variabel respon yang memiliki banyak nilai nol (0), dengan beberapa vaktor yang mempengaruhi jumlah kematian. Hasil analisis menunjukkan bahwa regresi Zero Inflated Poisson (ZIP) baik digunakan dalam memodelkan data yang mengalami overdispersi dan *zero inflation*. Hal ini dikarenakan nilai AIC dan RMSE dari regresi ZIP lebih kecil disbanding metode lain, yaitu Quasi Poisson dan ZINB. Model yang didapat dari analisis regresi ZIP adalah: $\ln(\mu) = -3.26175 + 0.173044X1 - 0.150728X2 + 0.001592X3$ dan $\text{logit}(\omega) = -402.3538 + 5.3352X1 - 1.2401X2 - 0.1844X3$.

Kata kunci: Overdispersi, Regresi Poisson, Regresi *Quasi* Poisson, Regresi *Zero Inflated* Poisson, Regresi *Zero Inflated* Negative Binomial.

Judul Skripsi	: **PERBANDINGAN MODEL REGRESI QUASI POISSON, ZERO INFLATED POISSON DAN ZERO INFLATED NEGATIVE BINOMIAL PADA DATA OVERDISPERSI (Case Study: Data on Maternal Mortality Rate in East Lampung Regency in 2013)**
Nama Mahasiswa	: **Anisa Fitriyani**
Nomor Pokok Mahasiswa	: **1917031012**
Program Studi	: **Matematika**
Fakultas	: **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Dr. Ir. Netti Herawati, M.Sc.	**Subian Saidi, S.Si. M.Si.**
NIP.196501251990032001	NIP.198008212008121001

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

Ketua	: **Dr. Ir. Netti Herawati, M.Sc.**	....
Sekretaris	: **Subian Saidi, S.Si. M.Si.**	....
Penguji Bukan Pembimbing	: **Drs. Eri Setiawan, M.Si.**	....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T.
NIP 19740705 2000031001

Tanggal Lulus Ujian Skripsi : **2 Februari 2023**

PERNYATAAN SKRIPSI MAHASISWA

Saya yang bertanda tangan dibawah ini:

Nama Mahasiswa : **ANISA FITRIYANI**

Nomor Pokok Mahasiswa : **1917031012**

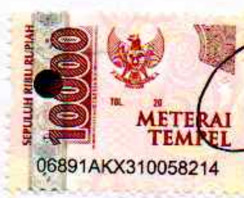
Jurusan : **Matematika**

Judul Skripsi : **PERBANDINGAN MODEL REGRESI QUASI POISSON, ZERO INFLATED POISSON DAN ZERO INFLATED NEGATIVE BINOMIAL PADA DATA OVERDISPERSI (Case Study: Data on Maternal Mortality Rate in East Lampung Regency in 2013)**

Dengan ini mengatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri, bukan hasil orang lain, dan semua hasil tulisan tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 20 Februari 2023

Yang Menyatakan



ANISA FITRIYANI
NPM. 1917031012

RIWAYAT HIDUP

Penulis bernama lengkap Anisa Fitriyani, anak satu-satunya dari pasangan Bapak Ponidin dan Ibu Yunita Ningsih yang lahir di Batangharjo pada tanggal 18 Januari 2001. Penulis menyelesaikan pendidikan sekolah dasar di SD Negeri 1 Balerejo pada tahun 2007 s.d 2013, sekolah menengah pertama di SMP Negeri 1 Batanghari pada tahun 2013 s.d 2016, dan sekolah menengah atas di Madrasah Aliyah Negeri (MAN) 1 Metro pada tahun 2016 s.d 2019.

Pada tahun 2019 penulis diterima sebagai mahasiswa S1 di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung melalui jalur SNMPTN. Selama aktif menjadi mahasiswa, penulis ikut serta dalam Himpunan Mahasiswa Jurusan Matematika (HIMATIKA) sebagai anggota Bidang Keilmuan pada Tahun 2020, dan melanjutkan sebagai Ketua Bidang Keilmuan HIMATIKA pada Tahun 2021.

Pada tahun 2022, sebagai bentuk penerapan bidang ilmu di dunia kerja, penulis melaksanakan Kerja Praktek (KP) di Badan Penyelenggara Jaminanan Sosial (BPJS) Kesehatan Cabang Metro dan sebagai bentuk pengabdian kepada masyarakat penulis melaksanakan Kuliah Kerja Nyata (KKN) di Desa Negara Saka, Kecamatan Jabung, Kabupaten Lampung Timur.

PERSEMBAHAN

Alahmdulillahirabbil'alamin,
Puji dan syukur tiada hentinya terhanturkan kepada Allah SWT
Kupersembahkan karya ini kepada:

Diri Sendiri
Terimakasih telah bertahan dari segala tekanan sampai dititik ini. Yang pundaknya setiap malam menyokong setiap titik beban. Dan matanya yang lelah karna cahaya yang terlalu terang.

Ayah dan Ibu
Orang tuaku tercinta, Bapak Ponidin dan Ibu Yunita Ningsih yang selalu memberikan doa, dukungan dan kasih sayang.

Dosen
Dosen-dosen pembimbing dan pembahas yang sangat berjasa dalam membimbing dan memberikan masukan yang membangun serta menyampaikan ilmu kepada saya.

Sahabat-sahabatku
Para sahabat tersayang yang terus saling mendukung, menolong, serta memberikan warna dalam hidupku.

Almamater kebanggaan, Universitas Lampung.

SANWACANA

Puji syukur kepada Tuhan yang Maha Esa, yang telah memberikan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Perbandingan Model Regresi *Quasi* Poisson, *Zero Inflated* Poisson Dan *Zero Inflated* Negative Binomial Pada Data Overdispersi (Studi Kasus: Data Angka Kematian Ibu Kabupaten Lampung Timur Tahun 2013)”.

Dalam penyelesaian skripsi ini, penulis menyadari adanya keterbatasan pengetahuan dan kemampuan yang dimiliki. Sehingga, penulis banyak mendapatkan arahan dan bantuan dari berbagai pihak. Oleh karena itu penulis mengucapkan terima kasih kepada :

1. Ibu Dr. Ir. Netti Herawati, M.Sc., selaku dosen pembimbing I yang senantiasa memberikan arahan, bantuan, dan saran kepada penulis dalam menyelesaikan skripsi ini.
2. Bapak Subian Saidi, S.Si., M.Si., selaku dosen pembimbing II yang telah memberikan bimbingan, saran dan kritik salam proses menyelesaikan skripsi ini.
3. Bapak Drs. Eri Setiawan, M.Si. selaku dosen Pembahas skripsi yang telah memberikan evaluasi dan saran bagi perbaikan skripsi penulis.
4. Bapak Dr. Aang Nuryaman, S.Si. M.Si., selaku ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Bapak Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Seluruh dosen, staf, dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Univeristas Lampung.

7. Ayah dan ibu yang senantiasa tidak pernah putus doa, dukungan dan perhatian kepada penulis sehingga dapat menyelesaikan skripsi ini.
8. Sahabat-sahabat terbaik yang tidak lelah memberikan semangat, doa dan waktunya untuk membantu penulis dalam menyelesaikan skripsi ini.
9. Seluruh pihak yang telah membantu dan terlibat dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa dalam penulisan skripsi ini masih jauh dari kata sempurna, oleh karena itu kritik dan saran sangat penulis harapkan. Semoga skripsi ini dapat bermanfaat bagi penulis dan bagi pembaca.

Bandar Lampung, 20 Februari 2023
Penulis

ANISA FITRIYANI

DAFTAR ISI

Halaman

DAFTAR TABEL

I. PENDAHULUAN

1.1. Latar Belakang dan Masalah

Analisis regresi berkaitan dengan pengaruh satu atau lebih variabel penjelas terhadap variabel respon (Gujarati, 1991). Banyak ditemukan variabel yang saling berhubungan dalam kehidupan sehari-hari, sehingga dapat dibangun suatu model untuk mengetahui hubungan antar variabel tersebut. Salah satu model regresi yang dapat digunakan untuk menganalisis hubungan antara variabel *dependent* Y yang berupa data diskrit dan variabel *independent* X berupa data kontinyu, diskrit atau campuran adalah model regresi Poisson. Asumsi yang harus dipenuhi dalam kasus regresi Poisson adalah bahwa nilai variansi dari variabel respon harus sama dengan rataannya (*equidispersi*). Namun dalam analisis data sering dijumpai data yang nilai variansinya lebih kecil atau lebih besar dari rataannnya. Keadaan ini dikenal dengan *underdispersion* yaitu variansi lebih kecil dari rata-rata, dan *overdispersion* yaitu variansi lebih besar dari rata-rata. Menurut Agresti (1990), kondisi *overdispersion* dapat dilihat dari nilai taksiran dispersi yaitu nilai *Pearson chi square/df* dan *Deviance/df* yang nilai keduanya lebih besar dari 1. Menurut Riner Winkelman (2008), salah satu penyebab terjadinya *overdispersion* adalah terlalu banyak nilai nol (*excess zero*) pada variabel respon.

Ada beberapa penelitian membahas terkait masalah data yang mengalami *overdispersion* seperti penelitian Pratiwi (2018), tentang data Banyaknya Kematian Bayi di Kabupaten Pasuruan Tahun 2013. Diketahui bahwa regresi *Quasi* Poisson dapat digunakan untuk membentuk model dari data tersebut yang menyebar Poisson. Kemudian penelitian Kartiningrum (2013), tentang data Kematian Ibu di Jawa Timur Tahun 2010 dengan kesimpulan bahwa regresi *Zero-Inflated* Poisson (ZIP) dapat digunakan untuk membentuk model dari data tersebut yang menyebar Poisson. Dalam penelitian lain yang dilakukan oleh Sekarmini, dkk. (2013), peneliti menggunakan regresi *Zero-Inflated* Negative Binomial (ZINB) pada data kematian yang menyebar Poisson. Ketiga penelitian tersebut, regresi *Quasi* Poisson, ZIP dan regresi ZINB dapat digunakan untuk data kematian yang mengalami *overdispersion* serta data yang banyak terdapat nilai 0 pada variabel respon.

Berdasarkan hal tersebut, maka akan dilakukan perbandingan model regresi *Quasi* Poisson, ZIP dan ZINB pada data Angka Kematian Ibu di Kabupaten Lampung Timur Tahun 2013 yang mengalami overdispersi, menggunakan software *R*.

1.2. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah

1. Membentuk model regresi *Quasi* Poisson, ZIP dan ZINB pada kasus Angka Kematian Ibu di Kabupaten Lampung Timur Tahun 2013.
2. Menentukan model terbaik antara regresi *Quasi* Poisson, ZIP dan ZINB pada kasus Angka Kematian Ibu di Kabupaten Lampung Timur Tahun 2013.
3. Melihat pengaruh K1, K4, dan Komplikasi Kebidanan terhadap Jumlah Kematian Ibu di Kabupaten Lampung Timur Tahun 2013.

1.3. Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah

1. Mengetahui apakah regresi *Quasi* Poisson, ZIP dan ZINB dapat digunakan untuk memodelkan kasus Angka Kematian Ibu di Kabupaten Lampung Timur Tahun 2013.
2. Mengetahui model terbaik antara regresi *Quasi* Poisson, ZIP dan ZINB pada kasus Angka Kematian Ibu di Kabupaten Lampung Timur Tahun 2013.
3. Mengetahui pengaruh K1, K4, dan Komplikasi Kebidanan terhadap Jumlah Kematian Ibu di Kabupaten Lampung Timur Tahun 2013

II. TINJAUAN PUSTAKA

2.1. Sebaran Poisson

Percobaan yang menghasilkan nilai-nilai bagi suatu peubah acak *Y*, yaitu banyaknya hasil percobaan yang terjadi selama suatu selang waktu tertentu atau di suatu daerah tertentu, sering disebut percobaan poisson. Menurut Walpole (1995), percobaan poisson memiliki cici-ciri berikut:

1. Banyaknya hasil percobaan yang terjadi dalam suatu selang waktu atau suatu daerah tertentu, tidak bergantung pada banyaknya hasil percobaan yang terjadi pada aselang waktu atau daerah lain.
2. Peluang terjadinya satu hasil percobaan selama satu selang waktu yang singkat sekali atau dalam suatu daerah yang kecil, sebanding dengan panjang selang waktu tersebut atau besarnya daerah tersebut, dan tidak bergantung pada banyaknya hasil percobaan yang terjadi di luar selang waktu atau daerah tersebut.
3. Peluang bahwa lebih dari satu hasil percobaan akan terjadi dalam selang waktu yang singkat tersebut atau dalam daerah yang kecil tersebut, dapat diabaikan.

Bilangan *Y* yang menyatakan banyaknya hasil percobaan dalam suatu percobaan Poisson disebut peubah acak Poisson dan sebaran peluangnya disebut sebaran Poisson. Sebaran peluang bagi peubah acak Poisson *Y*, yang menyatakan banyaknya hasil percobaan yang terjadi selama suatu selang waktu tertentu. . Menurut Walpole (1995), sebaran Poisson memiliki fungsi peluang sebagai berikut:

$$p(y;\mu) = \frac{e^{-\mu}\mu^{y}}{y!} \ldots. \quad (2.1)$$

dengan:

y = 0,1,2,…..

μ = rata-rata banyaknya hasil percobaan yang terjadi selama selang waktu atau dalam daerah yang dinyatakan

e = 2.71828…

Sebaran Poisson dapat digunakan untuk menghampiri sebaran Binomial apabila *n* besar dan peluang sukses *p* kecil atau dekat dengan nol (Walpole, 1995).

Misalkan *Y* variabel random berdistribusi Poisson dengan parameter μ maka rata rata dan variansi *Y* adalah μ.

Bukti

$$E(Y) = \sum_{y=0}^{\infty} y\,.\,p(y;\mu)$$

$$= \sum_{y=0}^{1} y\,.\frac{\mu^{y}e^{-\mu}}{y(y-1)!} + \sum_{y=1}^{\infty} y\,.\frac{\mu^{y}e^{-\mu}}{y(y-1)!}$$

$$= \sum_{y=0}^{1} 0\,.\frac{\mu^{0}e^{-\mu}}{0!} + \sum_{y=1}^{\infty} y\,.\frac{\mu^{y}e^{-\mu}}{y(y-1)!}$$

$$= \sum_{y=1}^{\infty} y\,.\frac{\mu^{y}e^{-\mu}}{y(y-1)!}$$

$$= \sum_{y=1}^{\infty} \frac{\mu^{y}e^{-\mu}}{(y-1)!}$$

$$= \sum_{y=1}^{\infty} \frac{\mu.\mu^{(y-1)}e^{-\mu}}{(y-1)!}$$

$$= \mu \sum_{z=0}^{\infty} \frac{\mu^{(z)}e^{-\mu}}{z!} \text{ (misalkan } z = y-1, y = 1 \text{ } maka \text{ } z = 0)$$

$$= \mu \sum_{z=0}^{\infty} P(z, \mu)$$

$$= \mu . 1$$

$$E(Y) = \mu$$

Sedangkan variansi dari Y yang berdistribusi Poisson dengan parameter μ adalah:

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

$$= E(Y^2 - Y + Y) - [E(Y)]^2$$

$$= E[Y(Y-1) + Y] - [E(Y)]^2$$

$$= E[Y(Y-1)] + E(Y) - [E(Y)]^2$$

$$= \sum_{y=0}^{\infty} y(y-1). \frac{\mu^y e^{-\mu}}{y!} + E(Y) - [E(Y)]^2$$

$$= \sum_{y=0}^{\infty} y(y-1). \frac{\mu^y e^{-\mu}}{y(y-1)(y-2)!} + E(Y) - [E(Y)]^2$$

$$= \sum_{y=0}^{\infty} \frac{\mu^y e^{-\mu}}{(y-2)!} + E(Y) - [E(Y)]^2$$

$$= \sum_{y=2}^{\infty} \frac{\mu^2 \mu^y e^{-\mu}}{(y-2)!} + E(Y) - [E(Y)]^2$$

$$= \mu^2 \sum_{y=2}^{\infty} \frac{\mu^y e^{-\mu}}{(y-2)!} + E(Y) - [E(Y)]^2$$

$= \mu^2 \sum_{z=0}^{\infty} \frac{\mu^z e^{-\mu}}{z!} + E(Y) - [E(Y)]^2$ (misalkan $z = y - 2, y = 2 \; maka \; z = 0$)

$$= \mu^2 + \mu + \mu^2$$

$$Var(Y) = \mu$$

Distribusi Poisson merupakan distribusi diskrit. Untuk melihat nilai μ yang besar akan lebih mendekati distribusi normal. Untuk kasus yang jarang terjadi maka nilai μ akan kecil. Distribusi Poisson adalah suatu distribusi yang paling sederhana dalam memodelkan data cacah, tetapi bukan satu satunya

2.2. Regresi Poisson

Regresi Poisson dapat digunakan untuk menganalisis data bertipe *count* (jumlahan) seperti misalnya data banyaknya kejadian yang terjadi dalam kurun waktu dan atau wilayah tertentu. Model regresi Poisson termasuk dalam model regresi non linier dan merupakan model standar untuk data diskrit (Cameron dan Trivedi, 1998).

Menurut Myers (1990), model regresi Poisson adalah:

$$y = \mu_i + \varepsilon_i \tag{2.2}$$

$$\mu_i = exp^{(x_i * \beta)} \tag{2.3}$$

dengan:

i = 1,2,..,n

y = Jumlah kejadian

μ_{i} = Rata-rata jumlah kejadian yang terjadi dalam selang waktu tertentu

ε_i = Sisaan ke-i

β = Koefisien regresi poisson

Menurut Agresti (2002), regresi Poisson merupakan salah satu kasus khusus *Generalized Linear Model* (GLM) yang digunakan untuk membentuk membentuk model di mana peubah responnya tidak mengharuskan mengikuti sebaran normal. Dalam model GLM, terdapat sebuah fungsi *g* yang linier yang menghubungkan nilai rata-rata dari peubah respon dengan peubah prediktor yang terdapat pada persamaan berikut:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} \tag{2.4}$$

2.3. Penduga Parameter Regresi Poisson

Dalam pendugaan parameter regresi Poisson dilakukan penaksiran pada β menggunakan metode *Maximum Likelihood Estimator* (Myers, 1990). Jika diberikan sebuah sampel berisi n buah pasangan pengamatan yang saling bebas, yaitu $\{(X_{1i}, X_{2i}, X_{3i}, \dots, X_{pi}, Y_{1i}); i = 1,2,3, \dots, n\}$ dengan X_i dan Y_i berturut-turut adalah pengamatan ke-i dari variabel X dan Y dan asumsi untuk setiap $X_1 = x_1, X_2 = x_2, X_3 = x_3, \dots, X_p = x_p$, distribusi dari Y_i adalah Poisson dan $E(Y_i|X_i = x_i) = \mu_i(x_i)$, maka fungsi probabilitas bersyarat dari Y_i oleh $x_{1i}, x_{2i}, x_{3i}, \dots, x_{pi}$ adalah:

$$P(y_i|x_{1i}, x_{2i}, x_{3i}, \dots, x_{pi};\ \mu_i(x_{1i}, x_{2i}, x_{3i}, \dots, x_{pi}))$$
$$= \frac{[\mu_i(x_{1i}, x_{2i}, x_{3i}, \dots, x_{pi})]^{y_i} e^{[-\mu_i(x_{1i}, x_{2i}, x_{3i}, \dots, x_{pi})]}}{y_i!}$$

karena $\mu_i(x_{1i}, x_{2i}, x_{3i}, \dots, x_{pi}) = e^{(\beta_0 + \sum_{j=1}^{p} \beta_j X_{ji})}$ maka diperoleh:

$$P(y_i|x_{1i}, x_{2i}, x_{3i}, \dots, x_{pi};\ \mu_i(x_{1i}, x_{2i}, x_{3i}, \dots, x_{pi}))$$
$$= \frac{[e^{(\beta_0 + \sum_{j=1}^{p} \beta_j X_{ji})}]^{y_i} e^{[-e^{(\beta_0 + \sum_{j=1}^{p} \beta_j X_{ji})}]}}{y_i!}$$

Fungsi *likelihood* diperoleh dengan mengalikan semua fungsi probabbilitas bersyarat dari $y_i$oleh x_i sehingga:

$$L(\beta) = \prod_{i=1}^{n} P(y_i|x_{1i}, x_{2i}, x_{3i}, \dots, x_{pi}; \beta)$$

$$= \prod_{i=1}^{n} \frac{[e^{(\beta_0 + \sum_{j=1}^{p} \beta_j X_{ji})}]^{y_i} e^{[-e^{(\beta_0 + \sum_{j=1}^{p} \beta_j X_{ji})}]}}{y_i!}$$

$$= \frac{[e^{(\sum_{i=1}^{n} \beta_0 + \sum_{j=1}^{p} \beta_j X_{ji})}]^{y_i} e^{[-e^{(\sum_{i=1}^{n} \beta_0 + \sum_{j=1}^{p} \beta_j X_{ji})}]}}{\prod_{i=1}^{n} y_i!} \qquad (2.5)$$

Agar persamaan (2.5) mudah diselesaikan maka bentuk menjadi fungsi *ln likelihood.*

$$\ln L(\beta) = \ln\left([e^{(\sum_{i=1}^{n} \beta_0 + \sum_{j=1}^{p} \beta_j X_{ji})}]^{y_i} + e^{[-e^{(\sum_{i=1}^{n} \beta_0 + \sum_{j=1}^{p} \beta_j X_{ji})}]} - \prod_{i=1}^{n} y_i! \right)$$
$$= \left(e^{(\sum_{i=1}^{n} y_i(\beta_0 + \sum_{j=1}^{p} \beta_j X_{ji})} \right) - e^{\left(\sum_{i=1}^{n} \beta_0 + \sum_{j=1}^{p} \beta_j X_{ji}\right)} - ln \prod_{i=1}^{n} y_i!$$

Turunan pertama dari $lnL(\beta)$ terhadap β_0 adalah:

$$\frac{\partial lnL(\beta)}{\partial(\beta_0)} = \sum_{i=1}^{n} \left\{ y_i - e^{\beta_0 + \sum_{j=1}^{p} \beta_j X_{ji}} \right\} = 0$$
$$\frac{\partial lnL(\beta)}{\partial(\beta_1)} = \sum_{i=1}^{n} \left\{ y_i x_{1i} - x_{1i} e^{\beta_0 + \sum_{j=1}^{p} \beta_j X_{ji}} \right\} = 0$$
$$= \sum_{i=1}^{n} \left\{ x_{1i}(y_i - e^{\beta_0 + \sum_{j=1}^{p} \beta_j X_{ji}}) \right\} = 0$$
$$\ldots \tag{2.6}$$
$$\frac{\partial lnL(\beta)}{\partial(\beta_p)} = \sum_{i=1}^{n} \left\{ y_i x_{pi} - x_{pi} e^{\beta_0 + \sum_{j=1}^{p} \beta_j X_{ji}} \right\} = 0$$
$$= \sum_{i=1}^{n} \left\{ x_{pi}(y_i - e^{\beta_0 + \sum_{j=1}^{p} \beta_j X_{ji}}) \right\} = 0$$

Bentuk vektor dari persamaan (2.6) yaitu:

$$\begin{bmatrix} \frac{\partial lnL(\beta)}{\partial(\beta_0)} \\ \frac{\partial lnL(\beta)}{\partial(\beta_1)} \\ \ldots \\ \frac{\partial lnL(\beta)}{\partial(\beta_p)} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} \left\{ y_i - e^{\beta_0 + \sum_{j=1}^{p} \beta_j X_{ji}} \right\} \\ \sum_{i=1}^{n} \left\{ x_{1i}(y_i - e^{\beta_0 + \sum_{j=1}^{p} \beta_j X_{ji}}) \right\} \\ \ldots \\ \sum_{i=1}^{n} \left\{ x_{pi}(y_i - e^{\beta_0 + \sum_{j=1}^{p} \beta_j X_{ji}}) \right\} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \ldots \\ 0 \end{bmatrix} \tag{2.7}$$

Karena diketahui turunan pertama masih mengandung parameter yang tidak dapat diselesaikan dengan metode linear maka dimaksimumkan dengan metode optimasi numerik *Newton Raphson*.

$$\hat{\beta}_{(i+1)} = \hat{\beta}_i - ((H(\beta))^{-1}.U_i(\beta))$$

dengan:

$\hat{\beta}_{(i+1)}$ =vektor β berukuran$(p+1) \times 1$ pada iterasi ke $(i+1)$

$U_i(\beta)$ =turunan pertama dari *ln* $L(y;\beta)$ pada iterasi ke-*i*

$H(\beta)$ = turunan kedua dari ln $L(y;\beta)$

2.4. Sebaran Binomial Negatif

Menurut Bain & Engelhardt (1992), sebaran Binomial Negatif yaitu sebagai pendekatan suatu percobaan sampai terjadi r buah sukses, dengan setiap pengulangannya saling bebas di mana peluang gagal yaitu $1 - p$ dan peluang sukses adalah p. Jika peubah acak Y menyatakan jumlah kegagalan yang dibutuhkan sampai terjadi k sukses, maka sebaran peluang peubah acak Y disebut mengikuti sebaran Binomial Negatif dengan fungsi peluang sebagai berikut:

$$P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, x = r, r+1, \ldots \qquad (2.8)$$

dengan:

r = banyak kejadian sukses

x = jumlah percobaan yang dibutuhkan sampai r sukses

p = peluang sukses

$1 - p$ = peluang gagal

2.5. Regresi Binomial Negatif

Regresi Binomial Negatif dapat digunakan untuk memodelkan data dengan permasalahan overdispersi pada data dengan sebaran Poisson (Berk & MacDonald, 2008). Berikut adalah rata-rata dan ragam sebaran Binomial Negatif.

$$E(Y_i) = \mu_i \qquad (2.9)$$

$$Var(Y_i) = \mu_i + k{\mu_i}^2 \qquad (2.10)$$

dimana k adalah parameter dispersi.

Menurut Green (2008), fungsi regresi Binomial Negatif mempunyai fungsi peluang sebagai berikut:

$$P(y_i, \mu_i, k) = \frac{\Gamma(y_i + 1/k)}{\Gamma(1/k) y_i!} \left(\frac{1}{1 + k\mu_i}\right)^{1/k} \left(\frac{k\mu_i}{1 + k\mu_i}\right)^{y_i}, y_i = 1,2, \ldots \quad (2.11)$$

dengan:

y_i = Jumlah kejadian

μ_i = Rata-rata jumlah kejadian yang terjadi dalam selang waktu tertentu

2.6. Regresi *Quasi* Poisson

Menurut Ver Hoef & Boveng (2007), regresi *Quasi* Poisson sering digunakan untuk menangani masalah overdispersi karena dapat memperbaiki nilai *standard error* pada regresi Poisson. Nilai standard error pada regresi *Quasi* Poisson disesuaikan dengan nilai parameter dispersinya sehingga menghasilkan *standard error* yang lebih besar dari regresi Poisson. Model regresi *Quasi* Poisson sama seperti model regresi Poisson seperti pada persamaan (2.2), dengan rata-rata dan ragam sebagai berikut:

$$E(y_i) = \mu_i \quad (2.12)$$

$$var(y_i) = \mu_i k \quad (2.13)$$

Pada regresi Poisson diasumsikan peubah respon mengikuti sebaran keluarga Exponensial. Menurut Agesti (2002), fungsi peluang sebaran keluarga Exponensial yang melibatkan k sebagai parameter dispersi yang terdapat pada persamaan (2.14).

$$f(y_i; \theta_i, k) = exp\left(\frac{y_i k_i - bk_i}{a(k)} + c(y_i, k)\right) \quad (2.14)$$

Fungsi *Likelihood* dari persamaan (2.14) terdapat pada persamaan berikut:

$$L(\beta) = \prod_{i=1}^{n} f(y_i; \theta_i, k) = \prod_{i=1}^{n} exp\left(\frac{y_i k_i - bk_i}{a(k)} + c(y_i, k)\right) \quad (2.14)$$

Fungsi ln *Likelihood* persamaan (2.15) sebagai berikut:

$$\ln L(\beta) = \sum_{i=1}^{n} \ln f(y_i; \theta_i, k) = \sum_{i=1}^{n} \frac{y_i k_i - bk_i}{a(k)} + \sum_{i=1}^{n} c(y_i, k) \quad (2.16)$$

2.7. Penduga Parameter Regresi *Quasi* Poisson

Berdasarkan dalil rantai, penduga parameter β pada regresi *Quasi* Poisson menggunakan persamaan berikut:

$$\frac{\partial lnL(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} \left(\frac{\partial lnL(\beta)}{\partial \theta_i} \times \frac{\partial \theta_i}{\partial \mu_i} \times \frac{\partial \mu_i}{\partial \eta_i} \times \frac{\partial \eta_i}{\partial \beta_j} \right) \tag{2.17}$$

Penduga maksimum *likelihood* $\hat{\beta}$ dapat diduga dengan fungsi *log likelihood* pada persamaan (2.17) diturunkan secara parsial terhadap θ_i, sehingga diperoleh persamaan berikut:

$$\frac{\partial lnL(\beta)}{\partial \beta_j} = \frac{y_i - b'\theta_i}{a(k)} = \frac{y_i - \mu_i}{a(k)} \tag{2.18}$$

Menurunkan μ_i terhadap θ_i diperoleh persamaan (2.19).

$$\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i) = \frac{var(y_i)}{a(k)} \tag{2.19}$$

Kemudian pada persamaan (2.4) yang menggunakan fungsi g linear yang menghubungkan nilai rata-rata dan peubah respon dengan peubah prediktor, diturunkan terhadap β_j, sehingga diperoleh persaman (2.20).

$$\frac{\partial \eta_i}{\partial \beta_i} = x_{ij} \tag{2.20}$$

Sehingga diperoleh persamaan dalil rantai seperti persamaan berikut:

$$\frac{\partial lnL(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} \left(\frac{(y_i - \mu_i) x_{ij}}{var(y_i)} \frac{\partial \mu_i}{\partial \eta_i} \right) \tag{2.21}$$

$$i = 1,2, \dots, n$$
$$j = 1,2, \dots, n$$

Persamaan (2.21) merupakan fungsi *score* bagi *Quasi likelihood*. Untuk mendapatkan solusi dari persamaan tersebut, digunakan metode iterasi dengan prosedur *Iteratively Reweighted Least Square* (IRLS).

$$\hat{\beta}^{(t+1)} = (X^T W^{(t)} X)^{-1} X^T W^{(t)} z^{(t)}) \quad (2.22)$$

dengan:

$\hat{\beta}^{(t+1)}$ = vektor β berukuran$(p+1) \times 1$ pada iterasi ke $(t+1)$

X^T = matriks kebalikan peubah prediktor berukuran $(p+1) \times n$

$W^{(t)}$ = diagonal matriks pembobot dengan elemen w_i pada iterasi ke-(t),

dimana $w_i = \frac{{\mu_i}^2}{k\mu_i} = \frac{\mu_i}{k}$

$z^{(t)}$ = vektor kolom dengan elemen ke-i, dimana

$$z_i = \sum_{j=1}^{p} {\beta_j}^t x_{ij} + (y_i - \mu_i)\left(\frac{1}{\mu_i}\right)$$

T = langkah dalam iterasi $(t = 1,2,\dots,m)$

Melakukan iterasi berdasarkan persamaan (2.22) hingga didapatkan penduga bagi β yang konvergen, yaitu ketika $|\widehat{\beta}^{(t+1)} - \hat{\beta}^{(t)}| < 10^{-5}$.

2.8. *Zero Inflation*

Menurut Famoye & Singh (2006), *Zero Inflation* adalah banyaknya nilai nol yang lebih dari 50% pada peubah respon. Pada variabel respon pada data diskrit mungkin ditemukan data bernilai kosong/nol. Akan tetapi dalam banyak kasus, kosong memiliki arti penting dalam penelitian. Jika nilai nol memiliki aratai penting dalam data diskrit maka harus dimasukkan dalam analisis. Dalam penelitian ini ditemukan kondisi dimana terlalu banyak nilai nol. Banyaknya nilai nol yang lebih dari 50% (*zero inflation*) berakibat penggunaan regresi Poisson menjadi tidak tepat untuk memodelkan data penelitian.

2.9. Regresi *Zero Inflated* Poisson (ZIP)

Penyebab terjadinya overdispersi salah satunya adalah banyaknya nilai nol yang ada pada suatu penelitian daripada yang diduga untuk model regresi Poisson. Regresi ZIP adalah model campuran untuk data cacah dengan banyaknya nilai nol pada peubah responnya. Model ini merupakan kombinasi dari sebaran kejadian yang bernilai nol (Cameron & Trivedi, 1998).

Nilai nol pada peubah respon penelitian diduga muncul dalam dua cara yang sesuai untuk keadaan (*state*) yang terpisah. Keadaan yang pertama disebut dengan *zero state* yang terjadi ketika peluang ω dan menghasilkan penelitian yang bernilai nol, sementara itu untuk keadaan kedua disebut *Poisson State* terjadi dengan peluang (1 - ω) dan berdistribusi Poisson dengan rata-rata μ (Lambert, 1992).

Fungsi sebaran ZIP sebagai berikut:

$$P(Y = y_i) = \begin{cases} \omega_i + (1 - \omega_i)e^{-\mu_i}, \text{ untuk } y_i = 0 \\ \frac{(1-\omega_i)e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \text{ untuk } y_i > 0 \end{cases} \tag{2.23}$$

dengan $Y_i \sim ZIP(\mu, \omega)$ dan μ adalah parameter dari sebaran Poisson sedangkan ω adalah peluang kejadian bernilai nol.

Memodelkan ω dengan model *ln* sebagai berikut:

$$\omega = \frac{\exp(x_i^T \gamma)}{1+\exp(x_i^T \gamma)} \tag{2.24}$$

dengan:

x_i = Vektor (1 × p) dari matriks penjelas ke-i

γ = Vektor (p × 1) dari parameter tambahan

Menurut Jansakul & Hinde (2002), menyatakan model penghubung μ yaitu:

$$\ln(\mu) = \boldsymbol{X}_i^T \boldsymbol{\beta} \text{ dan logit } (\omega) = \ln\left[\frac{\omega}{1-\omega}\right] = \boldsymbol{X}_i^T \boldsymbol{\gamma} \tag{2.25}$$

dengan:

X = Matriks peubah penjelas

β = Matriks berukuran $(p + 1) \times 1$ dari parameter yang akan diduga

γ = Matriks berukuran $(q + 1) \times 1$ dari parameter yang akan diduga

ω = Peluang kejadian bernilai nol

Menurut Jansakul & Hinde (2002), rata-rata dan varian regresi ZIP adalah:

$$E(Y_i) = \mu_i$$

$$Var(Y_i) = \mu_i + \frac{\omega}{1-\omega}{\mu_i}^2 \tag{2.26}$$

dan dapat dilihat bahwa distribusi dari ZIP menunjukkan adanya overdispersi jika $\omega > 0$, dikarenakan varian lebih besar daripada rata-rata.

2.10. Penduga Parameter Regresi ZIP

Pendugaan parameter regresi ZIP dapat dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Dilakukan turunan parsial fungsi kemungkinan pada parameter yang diduga.

$$\mu_i \quad = \exp(x_i^\tau \beta)$$

$$\omega_i \quad = \frac{e^{x_i^\tau \gamma}}{1+e^{x_i^\tau \gamma}}$$

$$(1-\omega_i) = \frac{1}{1+x_i^\tau \gamma} \tag{2.27}$$

Persamaan (2.27) disubtitusikan dalam persamaan (2.23) dan didapat.

$$P(Y = y_i) = \begin{cases} \dfrac{\exp(x_i^T \gamma) + \exp\left(-e^{x_i^T \beta}\right)}{1 + x_i^\tau \gamma}, untuk\ y_i = 0 \\ \left(\dfrac{1}{1 + x_i^\tau \gamma}\right)\dfrac{(\exp\left(-e^{x_i^T \beta}\right))(-e^{x_i^T \beta})^{y_i}}{y_i!}, \\ \text{untuk } y_i > 0 \end{cases} \tag{2.28}$$

dengan β dan γ adalah parameter yanga akan diduga.

Selanjutnya dibuat persamaan *ln likelihood* dari persamaan (2.28) yaitu:

$$\ln L(\beta,\gamma \mid y_i) = \begin{cases} \sum_{i=1}^{n} \ln\left(e^{x_i^T\gamma} + \exp\left(-e^{x_i^T\beta}\right)\right) - \sum_{i=1}^{n} \ln\left(1 + e^{x_i^T\gamma}\right), untuk\ y_i = 0 \\ \sum_{i=1}^{n} \left(\left(x_i^T\beta\right)y_i - e^{x_i^T\beta}\right) - \sum_{i=1}^{n} \ln\left(1 + e^{x_i^T\gamma}\right) - \sum_{i=1}^{n} \ln y_i! \\ \text{untuk } y_i > 0 \end{cases}$$

(2.29)

Digunakan algoritma EM (*Expectation Maximization*) untuk memaksimalkan fungsi *ln likelihood*. Algoritma EM merupakan metode optimasi yang biasa digunakan dalam memaksimumkan fungsi *likelihood* (Hall & Shen, 2009).

Dimisalkan Y berkaitan dengan peubah indikator Z :

$$Z = \begin{cases} \text{Bernilai } 1, \text{jika } y_i \text{dari } zero\ state \\ \text{Bernilai } 0, \text{jika } y_i \text{dari } poisson\ state \end{cases}$$

Jika nilaii $y_i > 0$ maka nilai $z_i = 0$. Sedangkan jika nilai $y_i = 0$, maka nilai $z_i = 0$ atau mungkin 1. Maka nilai z_i dianggap hilang.

Langkah-langkah pendugaan parameter menggunakan algoritma EM (*Expectation Maximization*) adalah :

Penentuan sebaran peubah Z

$P(z_i = 1) = \omega_i$

$P(z_i = 0) = 1 - \omega_i$

Sehingga z_i~Binomial $(1 - \omega_i\ ), E(z_i\ )$ dan $var(z_i\ ) = \omega_i\ (1 - \omega_i\ )$

Pembentukan sebaran gabungan antara y_i dan z_i yaitu :

$$\begin{aligned} f\left(y_i, z_i \middle| \omega_i\ , \mu_i\ \right) &= f(z_i) f\left(y_i\ \middle| z_i\ \right) \\ &= f(z_i|1, \omega_i)\ f(y_i\ |z_i\ , \mu_i) \\ &= (1 - \omega_i\ )^{(1-z_i\ )}(\omega_i)^{z_i}\left(\frac{\exp(-\mu_i)\mu_i{}^{y_i}}{y_i!}\right)^{(1-z_i)} \end{aligned} \qquad (2.30)$$

Kemudian subtitusikan persamaan (2.25) ke dalam persaman (2.30) sehingga didapat:

$$\ln L(\boldsymbol{\beta},\boldsymbol{\gamma}|y,z) = \sum_{i=\frac{1}{n}}^{n}[z_i x_i^T \boldsymbol{\gamma} - \ln(1+\exp(x_i^T\gamma))] + \sum_{i=\frac{1}{n}}^{n}(1-z_i)\left(y_i x_i^T\boldsymbol{\beta} - \exp(x_i^T\boldsymbol{\beta})\right)] - \sum_{i=\frac{1}{n}}^{n}(1-z_i)\ln y_i! \qquad (2.31)$$

Persamaan (2.31) akan dimaksimumkan menggunakan algoritma EM, dimana $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ dapat diduga secara terpisah.

$$ln\, L(\beta,\gamma,y,z) = ln\, L(\beta,y,z) + ln\, L(\gamma,y,z) - \sum_{i=\frac{1}{n}}^{n}(1-z_i)\, ln\, y_i! \qquad (2.32)$$

dengan:

$$ln\, L(\gamma,y,z) = \sum_{i=\frac{1}{n}}^{n}[z_i x_i^T\gamma - ln(1+exp(x_i^T\gamma))] \qquad (2.33)$$

dan

$$ln\, L\,(\beta,y,z) = \sum_{i=\frac{1}{n}}^{n}(1-z_i)(y_i x_i^T\beta - exp(x_i^T\beta))] \qquad (2.34)$$

Tahapan pertama dalam Algoritma EM adalah dengan E-Step (tahap ekspetasi). Menggantikan peubah z_i dengan $z_i{}^{(k)}$ yang merupakan ekspektasi dari z_i.

$$z_i{}^{(k)} = E\left(z_i \middle| y_i, \boldsymbol{\gamma}^{(k)}, \boldsymbol{\beta}^{(k)}\right) = P(z_i = 1|y_i, \boldsymbol{\gamma}^{(k)}, \boldsymbol{\beta}^{(k)})$$

Untuk $y_i = 0, z_i{}^{(k)} = \frac{1}{1+\exp(-x_i^T\boldsymbol{\gamma}^{(k)} - \exp(x_i^T\boldsymbol{\beta}^{(k)}))}$

Untuk $y_i > 0, z_i{}^{(k)}$

Sehingga persamaan (2.33) dan (2.34) menjadi :

$$\ln\left(\boldsymbol{\gamma}^{(k)}, y, z^{(k)}\right) = \sum_{i=1}^{n}\left[z_i^{(k)} x_i^T \boldsymbol{\gamma}^{(k)} - \ln\left(1+\exp\left(x_i^T\boldsymbol{\gamma}^{(k)}\right)\right)\right] \qquad (2.35)$$

$$\ln L\left(\boldsymbol{\beta}^{(k)}, y, z^{(k)}\right) = \sum_{i=1}^{n}\left(1 - z_i^{(k)}\right)\left(y_i x_i^T\boldsymbol{\beta}^{(k)} - \exp\left(x_i^T\boldsymbol{\beta}^{(k)}\right)\right) \qquad (2.36)$$

Setiap tahapan selanjutnya masuk pada tahap maksimalisasi (M-step). Memaksimalkan $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ pada persamaan dengan menghitung $\boldsymbol{\beta}^{(\boldsymbol{k}+\mathbf{1})}$ dan $\boldsymbol{\gamma}^{(\boldsymbol{k}+\mathbf{1})}$ dengan metode *Newton Raphson*. Iterasi dilakukan sampai diperoleh penduga parameter yang konvergen yaitu pada saat $|\boldsymbol{\beta}^{(k+1)} - \boldsymbol{\beta}^{(k)}| < \boldsymbol{\varepsilon}$ dan $|\boldsymbol{\gamma}^{(k+1)} - \boldsymbol{\gamma}^{(k)}| < \boldsymbol{\varepsilon}$, dimana $\boldsymbol{\varepsilon} = \mathbf{10^{-6}}$**.**

2.11. Regresi *Zero Inflated* Negative Binomial (ZINB)

Menurut Berk & MacDonald (2008), regresi Binom Negatif dapat digunakan sebagai alternatif untuk memodelkan data poisson yang mengalami *overdispersion*. Model selanjutnya yang dapat digunakan untuk pengamatan yang terlalu banyak terdapat nilai nol dan mengalami overdispersi adalah *Zero Inflated* Negative Binomial. Di duga terdapat dua *state* munculnya nilai nol dalam pengamatan. Yang pertama adalah *zero state* terjadi dengan peluang p_i dan hasil penelitian bernilai nol. Tahap kedua adalah Binomial Negatif *state* dengan peluangnya $(1 - p_i)$ yang menyebar secara Binomial Negatif dengan rata-rata μ. Menurut Garay, *et al.* (2011), fungsi peluang ZINB adalah:

$$P(Y = y_i) = \begin{cases} p_i + (1 - p_i)\left(\dfrac{k}{\mu_i + k}\right)^k, \\ \text{untuk } y_i = 0 \\ (1 - p_i)\dfrac{\Gamma(k + y_i)}{\Gamma(y_i + 1)\Gamma(k)}\left(\dfrac{\mu_i}{\mu_i + k}\right)^{y_i}\left(\dfrac{k}{\mu_i + k}\right)^k \\ \text{untuk } y_i > 0 \end{cases} \tag{2.37}$$

dimana $0 \leq p_i \leq 1, \mu_i \geq 0, 1/k$ adalah parameter dispersi dan k >0. Sedangkan $\Gamma(k)$ adalah fungsi gamma. Menurut Garay, *et al.* (2011), rata-rata dan ragam regresi ZINB adalah:

$$E(y_i) = (1 - p_i)\,\mu_i \tag{2.38}$$

$$Var(y_i) = (1 - p_i)\,\mu_i(1 + \mu_i k^{-1} + p_i\mu_i) \tag{2.39}$$

Model gabungan μ dan p_i adalah:

$$\ln(\mu_i) = x_i^T\beta$$

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = z_i^T\boldsymbol{\gamma},\ i = 1,2, \dots n \qquad (2.40)$$

dengan β dan γ adalah parameter yang belum diketahui dan akan diduga.

2.12. Penduga Parameter ZINB

Maximul Likelihood Estimation digunakan untuk menduga parameter ZINB.

$$\mu_i = e^{x_i^T\beta}$$

$$\frac{p_i}{1-p_i} = e^{z_i^T\gamma}$$

$$p_i = e^{z_i^T\gamma} - p_i e^{z_i^T\gamma}$$

$$p_i\left(1 + e^{z_i^T\gamma}\right) = e^{z_i^T\gamma}$$

$$p_i = \frac{e^{z_i^T\gamma}}{1+e^{z_i^T\gamma}} \qquad (2.41)$$

Selanjutnya dilakukan substitusi persamaan (2.41) dalam persamaan (2.35) dan menghasilkan:

$$P(Y = y_i) = \begin{cases} \dfrac{1}{1+e^{z_i^T\gamma}}\left[e^{x_i^T\beta} + \left(\dfrac{k}{e^{x_i^T\beta}+k}\right)^k\right], untuk\ y_i = 0 \\ \dfrac{1}{1+e^{z_i^T\gamma}}\dfrac{\Gamma(\mathrm{k}+y_i)}{\Gamma(y_i+1)\Gamma(\mathrm{k})}\left(\dfrac{e^{x_i^T\beta}}{e^{x_i^T\beta}+k}\right)^{y_i}\left(\dfrac{k}{e^{x_i^T\beta}+k}\right)^k, \\ \text{untuk } y_i > 0 \end{cases}$$

(2.42)

Selanjutnya dari hasil substitusi persamaan (2.41) ke dalam persamaan (2.37) yang terdapat pada persamaan (2.42), dibuat persamaan ln *likelihood* dan dihasilkan:

$$\ln L(\theta \mid y_i) = \begin{cases} -\sum_{i=1}^{n} \ln\left(1+e^{z_i^T\gamma}\right) + \sum_{i=1}^{n} \ln\left[e^{z_i^T\gamma} + \left(\frac{k}{e^{x_i^T\beta}+k}\right)^{k}\right], \\ -\sum_{i=1}^{n} \ln\left(1+e^{z_i^T\gamma}\right) + \sum_{i=1}^{n} \ln[\Gamma(k+y_i)] - \sum_{i=1}^{n} \ln[\Gamma(y_i+1)] - \\ \sum_{i=1}^{n} \ln[\Gamma(k)] + \sum_{i=1}^{n} y_i \ln\left(\frac{e^{x_i^T\beta}}{e^{x_i^T\beta}+k}\right) + \sum_{i=1}^{n} k\ln\left(\frac{k}{e^{x_i^T\beta}+k}\right), \\ \text{untuk } y_i > 0 \end{cases}$$

(2.43)

Selanjutnya digunakan algoritma EM (*Ecpectation Maximization*) untuk memaksimalkan fungsi ln *likelihood*. Misalkan peubah Y berkaitan dengan peubah indikator W yaitu :

$$w_i = \begin{cases} \text{Bernilai 1, jika } y_i \text{dari } zero\ state \\ \text{Bernilai 0, jika } y_i \text{dari } poisson\ state \end{cases}$$

$$ln\, L(\theta, y_i, w_i) = \sum_{i=1}^{n} \{w_i\, z_i^T \gamma - ln(1 + z_i^T \gamma) + (1 - w_i) ln[g(y_i; \beta, \kappa)]\} \quad (2.44)$$

dimana:

$$g(y_i; \beta, \kappa) = \frac{\Gamma(k+y_i)}{\Gamma(y_i+1)\Gamma(k)} \left(\frac{\mu_i}{\mu_i+k}\right)^{y_i} \left(\frac{k}{\mu_i+k}\right)^{k} \quad (2.45)$$

dengan $\mu_i = e^{x_i^T\beta}$ dan θ adalah vektor parameter dari fungsi *likelihood* yang akan diduga.

Langkah pertama dalam Algoritma EM adalah *E-Step* (tahap ekspektasi). Dalam tahapan ini, (s) digunakan untuk menyimbolkan iterasi.

$$\widehat{w}_i^{(s)} = \begin{cases} \left[1 + e^{-z_i^T \hat{\gamma}^{(s)}} \left(\frac{\hat{k}^{(s)}}{e^{T\widehat{\boldsymbol{\beta}}^{(s)}} + \hat{k}^{(s)}}\right)^{\hat{k}^{(s)}}\right]^{-1}, \text{ jika } y_i = 0 \\ 0, \text{ jika } y_i > 0 \end{cases}$$

(2.46)

dan

$$Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(s)}) = \ln L\left((\boldsymbol{\theta} \mid \boldsymbol{y}_{i'}, \boldsymbol{w}_i) \mid \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(s)}\right)$$

$$= \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\gamma} \mid \widehat{\boldsymbol{\theta}}^{(s)}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\beta}, \boldsymbol{k} \mid \widehat{\boldsymbol{\theta}}^{(s)}) \quad (2.47)$$

dimana:

$$Q_{1i}(\boldsymbol{\gamma} \mid \widehat{\boldsymbol{\theta}}^{(s)}) \quad = \widehat{w}_i^{(s)} \boldsymbol{z}_i^T \boldsymbol{\gamma} - \ln\left(1 + e^{z_i^T \gamma}\right) \tag{2.48}$$

dan

$$Q_{2i}(\boldsymbol{\beta}, \boldsymbol{k} \mid \widehat{\boldsymbol{\theta}}^{(s)}) = \left(1 - \widehat{w}_i^{(s)}\right) \ln\left[\frac{\Gamma(k + y_i)}{\Gamma(k)\Gamma(1 + y_i)} \left(\frac{e^{x[/\beta}}{e^{x^\top \beta} + k}\right)^{y_i} \left(\frac{k}{e^{e^\top[\beta} + k}\right)^{k}\right] \tag{2.49}$$

Tahapan selanjutnya masuk pada tahap maksimalisasi (*M- Step*). Dengan $Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(s)})$ fungsi yang memaksimumkan parameter yang akan diduga. Dengan menggunakan metode iterasi *Newton Raphson* untuk memaksimalkan $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ dari hasil tahap ekspektasi (*E-Step*) dengan menghitung$\boldsymbol{\beta}^{(\boldsymbol{m}+\mathbf{1})}$ dan $\boldsymbol{\gamma}^{(\boldsymbol{m}+\mathbf{1})}$ Iterasi dilakukan hingga konvergen dengan $|\widehat{\boldsymbol{\beta}}^{(\boldsymbol{m}+\mathbf{1})} - \widehat{\boldsymbol{\beta}}^{(\boldsymbol{m})}$ dan$|\ \widehat{\boldsymbol{\gamma}}^{(\boldsymbol{m}+\mathbf{1})} - \widehat{\boldsymbol{\gamma}}^{(\boldsymbol{m})}$ kurang dari ε, dimana $\varepsilon = 10^{-6}$.

2.13. Multikolinearitas

Menurut Ghazali (2006), uji multikolinearitas bertujuan untuk menguji apakah suatu model regresi terdapat korelasi antar peubah bebas. Model regresi yang baik seharusnya tidak terjadi korelasi antar variabel bebas. Pengujian multikolinearitas dilihat dari besaran VIF (*Variance Inflation Factor*). Regresi dikatakan bebas multikolinearitas jika besar nilai VIF < 10 dirumuskan :

$$VIF = \frac{1}{1-R_j^2} \tag{2.50}$$

Dengan R_j adalah koefisien *determinasi auxiliary regression*, yaitu regresi antara X_j sebagai respon dengan $(p-1)$ peubah penjelas lain.

2.14. Overdispersi

Terdapat asumsi yang harus dipenuhi dalam regresi Poisson yaitu nilai rata-rata dari peubah respon harus bernilai sama dengan ragam peubah respon, yang disebut juga dengan equidispersi. Namun, dalam analisis data sering dijumpai data dengan ragam peubah respon lebih besar dari rata-rata peubah respon, yang biasa disebut dengan overdispersi. Fenomena overdispersi dapat ditulis $var(Y) > E(Y)$. Sebaliknya, data dengan ragam peubah respon lebih kecil dari rata-rata peubah respon disebut dengan underdispersi (McCullagh & Nelder, 1983).

Menurut Hardin & Hilbe (2007), overdispersi terjadi akibat beberapa hal yaitu:

a. Terdapat korelasi antar pengamatan
b. Terdapat pelanggaran asumsi sebaran poisson yaitu ragam lebih besar dari rata-rata
c. Terdapat *excess zeros* (nilai 0 berlebih)
d. Terdapat *outlier* pada data

Terdapat cara untuk mendeteksi overdispersi, yaitu dengan nilai $\theta_1(Deviance)$. Jika nilai *deviance*/derajat babas menghasilkan nilai lebih dari satu maka terjadi overdispersi pada data.

$$\theta_1 = \frac{D^2}{db}$$

$$D^2 = 2\sum_{i=1}^{n}\{y_i \, ln\left(\frac{y_i}{\mu_i}\right)\} \tag{2.51}$$

dimana $db = n - p - 1$ dengan p merupakan banyaknya parameter termasuk konstanta, *n* merupakan banyaknya pengamatan dan D^2 adalah nilai *deviance*.

2.15. Uji Signifikansi Parameter Secara Parsial

Menurut Hosmer & Lemeshow (2000), untuk mengetahui pengaruh peubah penjelas terhadap peubah respon diperlukan pengujian secara parsial menggunakan uji *Wald* dengan hipotesis sebagai berikut:

$H_0: \beta_j = 0$
$H_1: \beta_j \neq 0$, $j = 1, 2, 3, \ldots, p$

Statistik uji *Wald* terdapat pada persamaan (2.52).

$$W_j = \frac{\hat{\beta}_j^2}{(SE(\hat{\beta}_j^2))^2} \tag{2.52}$$

$$SE\left(\hat{\beta}_j^2\right) = \sqrt{\frac{var\left(\hat{\beta}_j\right)}{n}}$$

dengan:

p : banyaknya peubah prediktor

$\hat{\beta}_j$:penduga parameter β_j

$SE\left(\hat{\beta}_j^2\right)$: penduga *standard error* dari β_j

Jika uji *Wald* > ${\chi^2}_{\alpha,1}$ atau nilai *p-value* < taraf nyata (α) dengan α adalah taraf nyata yang diinginkan dengan keputusan hipotesis nol diterima sehingga dapat disimpulkan bahwa model yang diperoleh sesuai dan layak untuk digunakan.

2.16. Uji Signifikansi Secara Simultan

Untuk mengetahui signifikansi parameter regresi secara serempak atau bersama-sama dapat dilakukan dengan cara pengujian parameter simultan. Menurut Agresti (2002), statistik uji G seperti berikut:

$$G = -2\ln[L_0 - L_1] \sim {\chi^2}_{(\alpha,p)} \tag{2.53}$$

dengan:

L_0 = *Likelihood* model tanpa semua peubah penjelas

L_1 = *Likelihood* model dengan semua peubah penjelas

2.17. Uji Kebaikan Model

a. Goodness of fit

Goodness of Fit Test yaitu suatu pengujian kesesuaian model dengan data, apakah model yang digunakan sudah sesuai dengan data. *Goodness of Fit Test* memberi tahu seberapa baik model yang digunakan sesuai dengan data (Lai & Liu, 2018). Ada beberapa metode yang bisa digunakan untuk menguji uji kesesuaian model regresi, diantaranya *Pearson Chi-Square Statistic and Deviance* dan *Hosmer-Lemeshow Test*. Akan tetapi *Hosmer-Lemeshow Test* memiliki kelemahan yaitu kuasa uji akan meningkat dengan ukuran sampel n, ini berarti jumlah ukuran sampel berpengaruh terhadap kuasa uji. Kuasa ujinya akan kurang jika ukuran sampelnya kecil. Dan jika ukuran sampel sangat besar, maka kuasa uji akan sangat tinggi sehingga cenderung menolak H_0 meskipun H_0 benar. Selain itu data dari populasi yang sama, uji ini akan cenderung menerima H_0 untuk ukuran sampel kecil dan cenderung menolak H_0 untuk sampel besar, dengan kata lain ada masalah kestabilan uji akibat perbedaan ukuran sampel. Untuk itu pada penelitian ini dilakukan uji kebaikan model dengan *Pearson Chi-Square Statistic*. Statistik uji *Pearson Chi-Square* sebagai berikut:

$$\theta_2 = \frac{\chi^2}{n-p}$$

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \widehat{\mu_i})^2}{\hat{\sigma}_i^2} \quad (2.54)$$

dengan:

y_i = nilai peubah respon dari pengamatan ke-*i*

n = banyak pengamatan

p = banyak peubah penjelas

$\hat{\mu}_i$ = penduga rata-rata respon ke-*i*

$\hat{\sigma}_i^2$ = penduga bagi varian respon ke- *i*

b. Koefisien Determinasi (R^2)

Koefisien determinasi (R^2) didefinisikan sebagai besarnya keragaman peubah respon yang diterangkan/dijelaskan oleh peubah prediktor. Semakin besar koefisien determinasi maka semakin tepat dugaan dari model regresi. Dalam analisis regresi linear, koefisien determinasi didasarkan pada perhitungan jumlah kuadrat yang diperoleh melalui metode *Ordinary Least Square* (OLS), sedangkan pada GLM didasarkan pada metode *Maximum Likelihood Estimation* (MLE). Hal ini yang mendorong berkembangnya beberapa koefisien determinasi dalam GLM, antara lain R_{Dev}^2 yang didasarkan pada *deviance* dan jumlah kuadrat regresi dan juga jumlah kudrat total. Penggunaan R_{Dev}^2didasarkan pada nilai *deviance* karena lebih optimal untuk GLM seperti yang terdapat pada persamaan berikut.

$$R_{Dev}^2 = 1 - \frac{D(y;\hat{\mu})}{D(y;\bar{y})} \tag{2.55}$$

dengan:

$D(y;\hat{\mu})$ = deviance dari model penuh

$D(y;\bar{y})$ = deviance dari model yang hanya terdiri dari intersep

Dalam regresi logistik, tidak ada seperti yang terdapat dalam regresi linier. Ada beberapa langkah yang dimaksudkan untuk meniru analisis. Pada analisis regresi logistik R^2 disebut sebagai *Pseudo R Square*. Interpretasi dari R^2 dan *Pseudo R Square* tidaklah sama, tetapi penyimpangannya dapat dianggap sebagai ukuran seberapa buruk model cocok. Ada tiga jenis *Pseudo R Square*, yaitu:

1. Cox and Snell R Square

Cox and Snell didasarkan pada *log-likelihood* tetapi memperhatikan ukuran sampel.

$$R_{CS}^2 = 1 - \left[-\frac{2}{n}(L_0 - L_1)\right] \tag{2.56}$$

Nilai *Pseudo R Square Cox and Snell* tidak dapat mencapai nilai 1 (100%) seperti yang diinginkan oleh setiap peneliti.

2. *Nagelkerke R Square*

Karena nilai *Pseudo R Square Cox and Snell* tidak dapat mencapai nilai 1 (100%), maka Nagelkerke memodifikasinya sehingga dapat mencapai nilai 1.

$$R_N^2 = \frac{R_{CS}^2}{R_{max}^2}, \text{ dimana } R_{max}^2 = 1 - exp\left[-\frac{2}{n}L_0\right] \qquad (2.57)$$

3. *McFadden's*

Nilai pada McFadden ini cenderung lebih kecil dari *R Square* dan nilai 0.2-0.4 dianggap sangat memuaskan.

$$R_{MCF}^2 = 1 - \left[\frac{L_0}{L_1}\right] \qquad (2.58)$$

2.18. Kriteria Pemilihan Model Terbaik

Menurut Zucchini (2000), ada beberapa metode yang dapat digunakan untuk memilih model terbaik dalam analisis regresi antara lain *Akaike Information Criterion* (AIC), *bootstrap*, *cross validation*, dan *Bayes Information Criterion* (BIC). Kriteria yang paling banyak digunakan untuk pemilihan model terbaik adalah *Akaike Information Criterion* (AIC). Selain itu untuk mengevaluari model dilakukan juga dengan melihat nilai *Root Mean Square Error* (RMSE).

a. *Akaike's Information Criterion* (AIC)

AIC adalah metode yang dapat digunakan untuk memilih model regresi terbaik yang ditemukan oleh Akaike dan Schwarz. Metode tersebut di dasarkan pada metode *Maximum Likelihood Estimation* (MLE). Untuk menghitung nilai AIC digunakan rumus sebagai berikut:

$$AIC = -2\log(L) + 2p \qquad (2.59)$$

dengan log(L) adalah nilai *likelihood* dan p adalah jumlah parameter yang diestimasi dalam model. Model yang terbaik adalah model yang memiliki nilai AIC terkecil. Namun Burnham & Andersoon (2002), menyatakan bahwa untuk sampel berukuran kecil kriteria AIC kurang baik untuk digunakan karena akan memberikan hasil yang bias. Sehingga dilakukan peninjauan kembali dengan memberikan faktor koreksi pada AIC disebut dengan *corrected Akaike Information Criterion* (AIC_C), yang dinyatakan dalam persamaan berikut:

$$AIC_C = AIC + \frac{2p(p+1)}{n-p-1} \quad (2.60)$$

Dengan n adalah jumlah sampel pada penelitian. AIC_C digunakan jika rasio antara jumlah sampel dan jumlah parameter dalam model kecil (< 40).

Pada model regresi untuk data *count* seperti regresi binomial negatif dan *generalized poisson regression* terdapat parameter dispersi untuk mengatasi kondisi overdispersi sehingga jumlah parameter dalam model bertambah. Hal ini akan berpengaruh terhadap nilai *likelihood* model sehingga perlu dilakukan modifikasi nilai AIC yaitu dengan *Quasi Akaike's Information Criterion* (QAIC) yang didefinisikan dalam persamaan berikut:

$$QAIC = -2\left(\frac{\log(L)}{k}\right) + 2p \quad (2.61)$$

Sehingga kriteria pemilihan model untuk data *count* dengan jumlah sampel kecil dikembangkan menjadi $QAIC_C$.

$$QAIC_C = QAIC + \frac{2p(p+1)}{n-p-1} \quad (2.62)$$

dengan c adalah taksiran parameter dispersi. Model yang terbaik adalah model yang memiliki nilai $QAIC_C$ terkecil.

b. *Root Mean Square Error* (RMSE)

RMSE merupakan alat seleksi model berdasarkan pada error yang menunjukan seberapa besar perbedaan hasil estimasi dengan nilai yang akan diestimasi. Nilai ini akan digunakan untuk menentukan model mana yang terbaik.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{Y}_i)^2}{n}} \tag{2.63}$$

dengan:

n = jumlah sampel

y_i = nilai aktual

$\hat{Y}$ = nilai prediksi

2.19. Angka Kematian Ibu

Menurut Departemen Kesehatan (2009), kematian ibu didefinisikan sebagai kematian perempuan pada saat hamil atau kematian dalam kurun 42 hari sejak terminasi kehamilan tanpa memandang lamanya kehamilan, yakni kematian yang disebabkan karena kehamilannya atau penanganannya, tetapi bukan karena sebab-sebab lain seperti kecelakaan dan terjatuh. Dalam dunia pelayanan kesehatan, banyak faktor yang mempengaruhi jumlah kematian ibu hamil. Cakupan ANC (*Antenatal Care*), komplikasi kehamilan yang ditangani, dan penanganan pasca persalinan (nifas) menjadi perhatian penting untuk megetahui pengaruhnya terhadap jumlah kematian ibu (Depkes, 2009).

ANC (*Antenatal Care*) adalah pemeriksaan kehamilan untuk mengoptimalisasi kesehatan mental dan fisik ibu hamil agar sanggup menghadapi persalinan, nifas, persiapan memberikan ASI, dan pulihnya kesehatan reproduksi secara wajar. Cakupan pelayanan antenatal (*Antenatal Care*) dikenal dengan sebutan K1 dan K4.

Cakupan K1 adalah ibu hamil yang telah melakukan kunjungan ke fasilitas pelayanan kesehatan untuk mendapatkan pelayanan antenatal sesuai dengan standar untuk pertama kali pada tiga bulan pertama kehamilan. K4 adalah kontak ibu hamil dengan tenaga kesehatan yang ke-empat (atau lebih) untuk mendapatkan pelayanan antenatal sesuai standar yang ditetapkan, dengan ketentuan : satu kali pada triwulan pertama, satu kali pada triwulan kedua, dan dua kali pada triwulan ketiga (Depkes RI, 2009). Standar yang ditetapkan untuk pemeriksaan antenatal pada ibu hamil adalah pelayanan yang mencakup minimal: timbang badan dan ukur tinggi badan, ukur tekanan darah, pemberian imunisasi *Tetanus Toxoid*, ukur tinggi fundus uteri, pemberian tablet besi (90 tablet selama kehamilan), temu wicara (pemberian komunikasi interpersonal dan konseling), test laboratorium sederhana (Hb, Protein urin)

Menurut Depkes (2011), komplikasi adalah kesakitan pada ibu hamil, ibu bersalin, ibu nifas yang dapat mengancam jiwa ibu dan/atau bayi. Penanganan dengan baik terhadap kompliasi sangatlah penting untuk mengurangi jumlah kematian ibu. Cakupan komplikasi kebidanan yang ditangani adalah ibu dengan komplikasi kebidanan di satu wilayah kerja pada kurun waktu tertentu yang mendapat penanganan definitif sesuai dengan standar oleh tenaga kesehatan terlatih pada tingkat pelayanan dasar. Cakupan komplikasi kebidanan yang ditangani dibagi menjadi 3, yaitu:

1. Komplikasi dalam kehamilan, yang meliputi: Abortus, Hiperemesis Gravidarum, perdarahan per vaginam, Hipertensi dalam kehamilan (preeklampsia, eklampsia), kehamilan lewat waktu, ketuban pecah dini.
2. Komplikasi dalam persalinan, yang meliputi: Kelainan letak/presentasi janin, Partus macet/distosia, Hipertensi dalam kehamilan (preeklampsia, eklampsia), perdarahan pasca persalinan, infeksi berat/ sepsis, kontraksi dini/persalinan prematur, kehamilan ganda.
3. Komplikasi dalam Nifas, yang meliputi: Hipertensi dalam kehamilan (preeklampsia, eklampsi*a*), Infeksi nifas, perdarahan nifas.

III. METODOLOGI PENELITIAN

3.1. Waktu dan tempat penelitian

Penelitian ini dilakukan di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung pada semester ganjil tahun ajaran 2022/2023.

3.2. Data Penelitian

Penggunaan data dalam penelitian ini adalah data sekunder yang didapat dari Profil Dinas Kesehatan Kabupaten Lampung Timur Provinsi Lampung. Data yang digunakan adalah Data Angka Kematian Ibu (AKI) pada setiap Puskesmas Kecamatan di Kabupaten Lampung Timur Tahun 2013. Variabel yang diamati pada penelitian ini adalah sebagai berikut:

Tabel 3.1. Variabel yang diamati dalam penelitian

Peubah	Nama Peubah	Keterangan
Y	Jumlah Kematian Ibu (Orang)	Jumlah kematian ibu yang berhubungan dengan kehamilan ataukelahiran
X1	Persentase K1 (%)	Persentase pelayanan antenatal ibuhamil sesuai standar pada triwulanpertama kehamilan

Tabel 3.1 (Lanjutan)

X2	Persentase K4 (%)	Persentase pelayanan antenatal ibu hamil sesuai standar paling sedikit 4 kali
X3	Persentase Komplikasi Kebidanan yangDitangani (%)	Persentase Komplikasi kehamilanyang mendapat penanganan sesuaistandar kebidanan oleh tenaga kesehatan terlatih pada tingkat pelayanan dasar

3.3. Metode Penelitian

Metode analisis data pada penelitian ini adalah :

1. Pemeriksaan adanya multikolinearitas menggunakan nilai VIF
2. Pemeriksaan overdispersi, penggunaan uji *Deviance*. Data overdispersi jika hasil penghitungan uji *Deviance/db* bernilai lebih dari 1, seperti pada persamaan (2.51)
3. Pemeriksaan *zero inflation*, apabila banyaknya nilai nol pada peubah respon lebih dari 50% maka mengalami *zero inflation*
4. Analisis data menggunakan regresi Poisson seperti tahap berikut:
   a. Menentukan model Regresi Poisson
   b. Uji signifikansi parameter dengan Uji *Wald*
   c. Uji kelayakan model Regresi Poisson
5. Pemeriksaan *zero inflation*, apabila banyaknya nilai nol pada peubah respon lebih dari 50% maka mengalami *zero inflation*
6. Langkah langkah analisis data dilakukan dengan *Quasi* Poisson seperti tahap berikut:
   a. Menentukan model Regresi *Quasi* Poisson
   b. Uji signifikansi parameter secara parsial dengan Uji *G*
   c. Uji signifikansi parameter secara simultan dengan Uji *Wald*
   d. Uji kelayakan model Regresi *Quasi* Poisson

7. Langkah langkah analisis data dilakukan dengan *Zero Inflated Poisson* (ZIP) seperti tahap berikut:
    a. Menentukan model Regresi ZIP
    b. Uji signifikansi parameter secara parsial dengan Uji *G*
    c. Uji signifikansi parameter secara simultan dengan Uji *Wald*
    d. Uji kelayakan model Regresi ZIP
8. Langkah langkah analisis data dilakukan dengan *Zero Inflated* Negatif Binomial (ZINB) seperti tahap berikut:
    a. Menentukan model Regresi ZINB
    b. Uji signifikansi parameter secara parsial dengan Uji *G*
    c. Uji signifikansi parameter secara simultan dengan Uji *Wald*
    d. Uji kelayakan model Regresi ZINB

e. Pemilihan model terbaik, dengan menggunakan AIC dan RMSE untuk memilih model terbaik antara *Quasi* Poisson, ZIP dan ZINB.

V. KESIMPULAN

Berdasarkan hasil penelitian Angka Kematian Ibu di Kabupaten Lampung Timur Tahun 2013 dengan peubah penjelas Persentase K1, Persentase K4, dan Persentase Komplikasi Kebidanan yang Ditangani dapat diambil kesimpulan :

1. Model regresi yang terbentuk untuk model regresi *Quasi* Poisson adalah:
   $\hat{\mu} = \exp(\ 11.933890 - 0.105148X_1 - 0.037070X_2 + 0.019935X_3\ )$

   Untuk model regresi ZIP adalah:
   $\ln(\mu) = -3.26175 + 0.173044X_1 - 0.150728X_2 + 0.001592X_3$
   dan
   $\mathrm{logit}(\omega) = -402.3538 + 5.3352X_1 - 1.2401X_2 - 0.1844X_3$
   Sedangkan model regresi ZINB adalah :
   $\ln(\mu) = 3.655531 + 0.143229X1 - 0.200539X2 + 0.008963X3$
   dan
   $\mathrm{logit}(p_i) = -270.6962 + 11.8079\mathrm{X1} - 9.7762\mathrm{X2} - 0.1568\mathrm{X3}$

   Pada regresi *Quasi* Poisson, ZIP dan ZINB dapat digunakan untuk memodelkan data Angka Kematian Ibu Kabupaten Lampung Timur Tahun 2013.

2. Dari hasil pengujian nilai QAIC dan RMSE yang telah dilakukan, model Regresi ZIP lebih tepat digunakan untuk data Angka Kematian Ibu Kabupaten Lampung Timur Tahun 2013 karena memiliki nilai QAIC dan RMSE yang terkecil.

DAFTAR PUSTAKA

Agresti, A. 1990. *Categorical Data Analysis*. John Wiley and Son, Florida.

Bain, L.J. & Engelhardt, M. 1992. *Introduction to Probability and Mathematical Statistics*. 2nd Edition. Duxbury Press, California.

Berk, R. & MacDonald. 2008. *Overdispersion and Poisson Regression*. Springer, Philadelpia.

Burnham, K. P. & Anderson, D. R. 2004. Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*. **33**(2): 261-304.

Cameron, C.A. & Trivedi, P.K. 1998. *Regression Analysis of Count Data*. Cambridge University Pr, London.

Depkes, R.I. 2009. *Sistem Kesehatan Nasional*. Departemen Kesehatan RI, Jakarta.

Famoye, F. & Singh, K.P. 2006. Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data. *Journal of Data Science*. **4**(1): 117-130.

Garay, A.M., Hashimoto, E.M., Ortega, E.M.M., & Lachos. V.H. 2011. On Estimation and Influence Diagnostics For Zero- Inflated Negative Binomial Regression Models. *Computational Statistics & Data Analysis*. **55**(3): 1304-1318.

Ghozali, I. 2006. *Analisis Multivariate dengan Program SPSS*. Edisi Ke-4. Badan Penerbit Universitas Diponegoro, Semarang.

Greene. W. 2008. Functional Forms For The Negative Binomial Model For Count Data. *Economics Letter*. **55**(1): 585- 590.

Gujrati, P. D. 1991. Geometrical properties of clusters, percolation transitions and Ising singularity. *Physics Letters A*. **156.7**(8): 410-414.

Hall, D.B. & Shen, J. 2009. Robust Estimation for Zero-Inflated Models. *Blackwell Publishing Ltd*. **40**(1): 1- 16.

Hardin, J. W. & Hilbe, J. M. 2007. *Generalized linear models and extensions*. Stata press, Texas.

Hosmer, D.W. & Lemeshow, S. 2000. *Applied Logistic Regression*. John Wiley and Sons, New York.

Jansakul, N. & Hinde. J.P. 2002. Score Tests for Zero-Inflated Models. *Computational Statistics and Data Analysis*. **40**(1): 75-96.

Kartiningrum, E.D. 2013. Pemodelan Faktor yang Mempengaruhi Kematian Ibu di Provinsi Jawa Timur Menggunakan Zero Inflated Poisson Regression. *Hospital Majapahit*. **4**(2): 153-161.

Lambert, D. 1992. Zero-Inflated Poisson Regression with an Application to Defects in Manufacturing. *Technometrics*. **34**(1): 1-14.

Lai, X. & Liu, L. 2018. A simple test procedure in standardizing the power of Hosmer–Lemeshow test in large data sets. *Journal of Statistical Computation and Simulation*. **88**(13): 2463-2472.

McCullagh, P. & Nelder, J.A. 1983. *Generalized Linear Models*. Chapman and Hall, London.

Myers, R.H. 1990. *Classical and Modern Regression with Applications*. 2nd Edition. PWS-KENT Publishing Company, Boston.

Sekarmini, N.M., Sukarsa, I.K.G., & Srinadi, I.G.A.M. 2013. Penerapan Regresi Zero Inflated Negative Binomial (ZINB) untuk Pendugaan Kematian Anak Balita. *Jurnal Statistika*. **2**(4): 11-16.

Ver Hoef, J.M. & Boveng, P.L. 2007. Quasi-Poisson Vs. Negative Binomial Regression. *Ecology*. **88**(11): 2766-2772.

Walpole, R.E. 1995. *Pengantar Statistika*. Terjemahan Sumantri. PT Gramedia Pustaka Utama, Jakarta.

Winkelmann, Rainer. 2008. *Econometric Analysis of count Data*. New York, Spinger

Zucchini, W. 2000. An introduction to model selection. *Journal of mathematical psychology*. **44**(1): 41-61.