PERFORMA MODEL REGRESI *ZERO INFLATED NEGATIVE BINOMIAL* (ZINB) UNTUK MENGATASI OVERDISPERSI PADA DISTRIBUSI POISSON
(STUDI KASUS: JUMLAH KEMATIAN BAYI)

(Skripsi)

Oleh
REFI INDIRA FIRYAL EFFENDI

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023

ABSTRACT

PERFORMANCE MODEL ZERO INFLATED NEGATIVE BINOMIAL (ZINB) REGRESSION TO OVERCOME OVERDISPERSION IN POISSON DISTRIBUTION (STUDY CASE: NUMBER OF INFANT DEATH)

By

Refi Indira Firyal Effendi

The case of infant mortality is one measure of the state of public health. Researching the causes of infant death is one way to try and lower the number of infant deaths. Poisson regression can be used to describe the number of cases of infant mortality and the factors that influence it. However, overdispersion conditions where the variance of the dependent variable is higher than the average value are often found in Poisson regression analysis. Too many zeros in the dependent variable can cause overdispersion. One way to overcome the problem of overdispersion is the ZINB regression model. ZINB regression is a better method to use to model the number of cases of infant mortality and the factors that influence it in South Lampung Regency in 2021. In this study, data on the number of cases of infant mortality in South Lampung in 2021 has a high percentage of zero values. Therefore, the use of the ZINB regression model is suitable for modeling the data. The results of the analysis showed that Low Birth Weight Infants (LBW) (X_1), deliveries assisted by health personnel (X_2), and exclusive breastfeeding (X_4), were the variables that most influenced the number of cases of infant mortality.

Keywords: Poisson, Case of Infant Mortality, Overdispersion, Zero-Inflated, ZINB.

ABSTRAK

PERFORMA MODEL REGRESI *ZERO INFLATED NEGATIVE BINOMIAL* (ZINB) UNTUK MENGATASI OVERDISPERSI PADA DISTRIBUSI POISSON (STUDI KASUS: JUMLAH KEMATIAN BAYI)

Oleh

Refi Indira Firyal Effendi

Kasus kematian bayi merupakan salah satu tolak ukur keadaan kesehatan masyarakat. Meneliti penyebab kematian bayi adalah salah satu cara untuk mencoba dan menurunkan jumlah kematian bayi. Regresi Poisson dapat digunakan untuk menggambarkan jumlah kasus kematian bayi serta faktor-faktor yang mempengaruhinya. Namun, kondisi overdispersi di mana nilai variansi variabel terikat lebih tinggi daripada nilai rata-rata sering ditemukan dalam analisis regresi Poisson. Terlalu banyak nilai nol dalam variabel terikat dapat menyebabkan overdispersi. Salah satu cara untuk mengatasi masalah overdispersi adalah model regresi ZINB. Regresi ZINB merupakan metode yang lebih baik digunakan untuk memodelkan jumlah kasus kematian bayi dan faktor-faktor yang mempengaruhinya di Kabupaten Lampung Selatan tahun 2021. Pada penelitian ini, data kasus jumlah kematian bayi di Lampung Selatan tahun 2021 memiliki persentase nilai nol yang tinggi. Oleh karena itu, penggunaan model regresi ZINB cocok untuk memodelkan data tersebut. Hasil analisis menunjukkan bahwa Bayi Berat Lahir Rendah BBLR (X_1), persalinan ditolong tenaga kesehatan (X_2), dan pemberian ASI eksklusif (X_4) merupakan variabel yang paling berpengaruh terhadap jumlah kasus kematian bayi.

Kata Kunci: Poisson, Angka Kematian Bayi, Overdispersi, *Zero-Inflated*, ZINB.

PERFORMA MODEL REGRESI *ZERO INFLATED NEGATIVE BINOMIAL* (ZINB) UNTUK MENGATASI OVERDISPERSI PADA DISTRIBUSI POISSON
(STUDI KASUS: JUMLAH KEMATIAN BAYI)

Oleh

Refi Indira Firyal Effendi

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

Judul Skripsi	: **PERFORMA MODEL REGRESI *ZERO INFLATED NEGATIVE BINOMIAL* (ZINB) UNTUK MENGATASI OVERDISPERSI PADA DISTRIBUSI POISSON (STUDI KASUS: JUMLAH KEMATIAN BAYI)**
Nama Mahasiswa	: **Refi Indira Firyal Effendi**
Nomor Pokok Mahasiswa	: **1817031089**
Program Studi	: **Matematika**
Fakultas	: **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Ir. Netti Herawati, M.Sc., Ph.D.
NIP. 196501251990032001

Agus Sutrisno S.Si., M.Si.
NIP. 197008311999031001

2. Ketua Jurusan Matematika

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

MENGESAHKAN

1. Tim Penguji

   Ketua : **Ir. Netti Herawati, M.Sc., Ph.D.** ....

   Sekretaris : **Agus Sutrisno S.Si., M.Si.** ....

   Penguji
   Bukan Pembimbing : **Dr. Khoirin Nisa S.Si., 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

Yang bertanda tangan di bawah ini:

Nama Mahasiswa	: **Refi Indira Firyal Effendi**
Nomor Pokok Mahasiswa	: **1817031089**
Jurusan	: **Matematika**
Judul Skripsi	: **PERFORMA MODEL REGRESI *ZERO INFLATED NEGATIVE BINOMIAL* (ZINB) UNTUK MENGATASI OVERDISPERSI PADA DISTRIBUSI POISSON (STUDI: KASUS JUMLAH KEMATIAN BAYI)**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 13 Februari 2023
Penulis

SEPULUH RIBU RUPIAH 10000 METERAI TEMPEL 5AF05AKX217670756

Refi Indira Firyal Effendi

RIWAYAT HIDUP

Penulis bernama Refi Indira Firyal Effendi dilahirkan di Metro pada 13 September 2000. Penulis merupakan anak kedua dari tiga bersaudara pasangan Bapak Endang Effendi dan Ibu Yusnani.

Penulis pertama kali menempuh pendidikannya di Taman Kanak-Kanak Amarta Tani pada tahun 2005-2006 dan melanjutkan pendidikannya ke Sekolah Dasar di SD Negeri 1 Labuhan Dalam pada tahun 2006-2012. Kemudian, penulis melanjutkan jenjang pendidikannya di SMP Negeri 19 Bandar Lampung pada tahun 2012-2015. Jenjang pendidikan selanjutnya di SMA Negeri 15 Bandar Lampung pada tahun 2015-2018.

Pada tahun 2018 penulis diterima sebagai mahasiswa S1 Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung melalui jalur Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN). Selama menjadi mahasiswa penulis juga aktif dalam berorganisasi Himpunan Mahasiswa Matematika (HIMATIKA) FMIPA UNILA. Pada tahun 2021 penulis melakukan Kuliah Kerja Nyata (KKN) di Kelurahan Fajar Baru, Kecamatan Jati Agung, Lampung Selatan dan Kerja Praktik (KP) di Badan Kependudukan dan Keluarga Berencana (BKKBN) Provinsi Lampung, serta mengikuti Program Kampus Merdeka yang bernama Studi Independen di PT. Microsoft Indonesia dengan mengambil *learning track Microsoft Productivity: The Modern Workplace*. Pada tahun 2022 penulis menjadi asisten dosen pada mata kuliah Pengantar Teori Peluang.

KATA INSPIRASI

Sesungguhnya Allah tidak merubah keadaan sesuatu kaum sehingga mereka merubah keadaan yang ada pada diri mereka sendiri.
(Q.S. Al-Ra'd: 11)

Angin tidak berhembus untuk menggoyangkan pepohonan, melainkan untuk menguji kekuatan akarnya
(Ali bin Abi Thalib)

Ketahuilah bahwa kemenangan bersama kesabaran, kelapangan bersama kesempitan, dan kesulitan bersama kemudahan
(HR. Tirmidzi)

Setiap manusia mempunyai hak untuk mempunyai masa depan yang cerah
(Penulis)

PERSEMBAHAN

Dengan mengucap puji dan syukur atas kehadirat Allah SWT, atas segala rahmat dan hidayah-Nya skripsi ini dapat terselesaikan dengan baik, serta salawat serta salam kepada junjungan kita Nabi Muhammad SAW. Dengan penuh ketulusan saya persembahkan karya tulis ini untuk:

Bapak Endang Effendi dan Ibu Yusnani

Terima kasih kepada kedua orang tua saya yang selalu memberikan dukungan dan saran dalam setiap keputusan, kasih sayang serta doa yang tak pernah putus dalam setiap langkah yang saya tempuh

Kakak Risya Delfira Cahyani Effendi dan Adik Reynisa Azzira Effendi

Terima kasih telah memberikan doa, semangat, serta dukungan selama ini

Dosen Pembimbing dan Pembahas

Terima kasih kepada bapak dan ibu dosen yang sangat berjasa, membantu, memberikan arahan, serta masukan dan ilmu yang bermanfaat

Teman-teman yang telah membantu, menemani, serta mendukung setiap langkahnya dari awal, hingga saat ini, dan seterusnya

Almamater Tercinta, Universitas Lampung

SANWACANA

Puji dan syukur penulis panjatkan kehadirat Allah SWT karena berkat segala rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi dengan judul "Performa Model Regresi *Zero Inflated Negative Binomial* (ZINB) untuk Mengatasi Overdispersi pada Distribusi Poisson (Studi kasus: Jumlah Kematian Bayi)".

Dalam menyusun laporan ini penulis banyak mendapatkan bantuan. Oleh karena itu, penulis mengucapkan terimakasih sebesar-besarnya kepada:

1. Ibu Ir. Netti Herawati, M.Sc., Ph.D., selaku dosen pembimbing I yang telah bersedia membimbing, memberi saran, bantuan, motivasi, dan arahan dalam menyelesaikan skripsi ini.
2. Bapak Agus Sutrisno, S.Si., M.Si., selaku dosen pembimbing II yang telah memberikan saran serta masukan kepada penulis sehingga dapat menyelesaikan skripsi ini.
3. Ibu Dr. Khoirin Nisa, S.Si., M.Si., selaku dosen penguji yang telah memberikan kritik dan saran selama proses penyusunan skripsi.
4. Ibu Dr. Fitriani, S.Si., M.Si., selaku dosen pembimbing akademik.
5. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika.
6. Bapak Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Seluruh dosen dan staff Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
8. Ayah, bunda, kakak, adik, dan keluarga besar yang selalu mendoakan dan memberikan dukungan.

9. Jule, Dyah, dan Ristia, sahabat yang selalu ada dan memberikan dukungan.
10. Amel, Elsa, Marisa, Putri, Ajeng, Ipeh, Eja dan Kelas C 2018 yang memberi dukungan selama kuliah.
11. Sherli, Mona, Anisa, dan teman seperbimbingan yang telah membantu dan menemani selama penulisan skripsi.
12. Semua pihak yang telah membantu dalam menyelesaikan skripsi ini yang tidak bisa penulis sebutkan satu persatu.

Penulis menyadari masih banyak kekurangan dalam laporan ini. Oleh karena itu, kritik dan saran sangat diharapkan agar dapat menjadi pelajaran dan perbaikan untuk kedepannya. Semoga skripsi ini dapat memberikan manfaat baik bagi penulis maupun bagi pihak yang membutuhkan.

Bandar Lampung, 13 Februari 2023
Penulis,

Refi Indira Firyal Effendi

DAFTAR ISI

Halaman

DAFTAR TABEL

Tabel Halaman

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis regresi adalah teknik statistik umum yang paling sering digunakan untuk menjelaskan hubungan antara dua atau lebih variabel. Variabel ini terdiri dari variabel bebas dan variabel terikat. Suatu variabel dikatakan bebas jika tidak dipengaruhi oleh variabel lain, sedangkan variabel dikatakan terikat jika dipengaruhi oleh variabel lainnya. Hubungan linier atau hubungan non linier antara variabel bebas dan variabel terikat dinyatakan dalam persamaan regresi. Persamaan regresi linier digunakan untuk menganalisis variabel-variabel yang berdistribusi normal, walaupun dalam prakteknya banyak ditemukan variabel-variabel yang berdistribusi tidak normal dan tidak mengikuti distribusi yang diharapkan. McCullagh dan Nelder (1989), menjelaskan bahwa pendekatan *Generalized Linear Model* (GLM) diciptakan sebagai perluasan dari model regresi klasik untuk mengatasi masalah ini. Model GLM mengandaikan bahwa variabel terikat memiliki distribusi yang termasuk ke dalam keluarga eksponensial, seperti distribusi normal, poisson, gamma, eksponensial, dan binomial negatif.

Diberbagai penelitian, sering ditemukan data yang yang memiliki distribusi poisson. Model regresi poisson merupakan salah satu model regresi yang cocok untuk menguji data yang mengikuti distribusi poisson. Terdapat beberapa asumsi untuk regresi poisson, salah satunya adalah nilai varians variabel terikat sama dengan rataannya, atau $E(Y) = \text{var}(Y) = \mu$ (equidispersi). Dalam menganalisis data, sering dijumpai data yang variansnya lebih kecil atau lebih besar dari rataannya, fenomena ini disebut sebagai underdispersi atau overdispersi. Terlalu banyak nilai nol atau yang dikenal juga sebagai *zero inflated* dalam variabel terikat, merupakan salah satu alasan terjadinya overdispersi.

Menurut Lambert (1992), regresi *zero inflated* poisson dapat memodelkan data yang mengandung *zero inflated*. Regresi zero inflated poisson (ZIP) tepat jika digunakan untuk menangani data yang mengandung *zero inflated*, kurang tepat untuk kasus overdispersi. Oleh karena itu Gustavo dan Kristina. (2022), menyarankan bahwa data yang terdapat *zero inflated* dan overdispersi lebih sesuai menggunakan regresi ZINB.

Angka Kematian Bayi (AKB) adalah banyaknya bayi yang meninggal sebelum mencapai usia yang ditentukan yaitu satu tahun untuk setiap 1.000 kelahiran hidup (UNICEF, 2020). Untuk tujuan *Sustainable Development Goals* (SDGs) yang harus dicapai, AKB digunakan untuk merepresentasikan tingkat pembangunan kesehatan suatu bangsa dan kualitas hidup masyarakat. Kematian bayi dapat terjadi karena berbagai sebab antara lain asfiksia, BBLR, tetanus, dan kesulitan menyusui (Permenkes RI, 2012). Data kasus jumlah kematian bayi di Lampung Selatan tahun 2021 memiliki persentase nilai nol yang tinggi. Oleh karena itu, penggunaan model regresi ZINB cocok untuk memodelkan data tersebut.

Penggunaan ZINB telah dilakukan oleh beberapa peneliti, diantaranya Dewanti, Susilawati, dan Srinadi (2020), telah membandingkan regresi ZIP dan regresi ZINB untuk mengatasi overdispersi. Hasil penelitian tersebut menyebutkan bahwa regresi ZINB lebih baik digunakan. Astuti (2015), juga telah melakukan penelitian pemodelan regresi ZINB pada kasus Tetanus Neonatorum di Provinsi Jawa Timur dan menyimpulkan bahwa ZINB cocok untuk memodelkan data. Amaliana, Sa’adah, dan Wayan (2019), juga telah melakukan penelitian mengenai performa proporsi ZINB pada kasus data Tetanus Neonatorum di Jawa Tengah dan menyimpulkan bahwa proporsi *zero inflated* sebesar 64,52% merupakan model terbaik pada data tetanus neonatorum di Jawa Tengah. Dari berbagai permasalahan di atas, peneliti akan melakukan pemodelan data jumlah kasus kematian bayi menggunakan regresi ZINB.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah

1. Mengaplikasikan model regresi ZINB pada data jumlah kematian bayi di kabupaten Lampung Selatan pada tahun 2021.
2. Mengetahui variabel apa saja yang berpengaruh dengan kasus jumlah kematian bayi di Lampung Selatan pada tahun 2021.
3. Mengetahui regresi terbaik untuk mengatasi data yang mengandung overdispersi.

1.3 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah;

1. Memberi pengetahuan dan wawasan terhadap pembaca mengenai metode regresi ZINB.
2. Memberi metode alternatif bagi pembaca dalam mengatasi overdispersi pada data *zero inflated.*

II. TINJAUAN PUSTAKA

2.1 Distribusi Poisson

Menurut Walpole (1995), distribusi poisson adalah distribusi untuk kejadian dengan peluang kejadian yang rendah, dimana kejadian tersebut bergantung pada interval waktu atau lokasi tertentu. Distribusi poisson memiliki ciri-ciri sebagai berikut:

1. Banyaknya percobaan yang terjadi dalam jangka waktu atau lokasi tertentu tidak berhubungan dengan jumlah hasil percobaan yang terjadi pada selang waktu atau daerah lain.
2. Kemungkinan bahwa satu hasil percobaan akan terjadi dengan cepat atau dalam waktu singkat. Hal ini tergantung pada berapa banyak hasil eksperimen yang terjadi di luar interval waktu dan wilayah tertentu, tetapi terkait dengan berapa lama interval waktu atau seberapa besar area tersebut.
3. Sangat tidak mungkin bahwa lebih dari satu hasil eksperimen akan muncul dalam periode waktu yang begitu singkat atau di wilayah yang terbatas.

Fungsi peluang dapat dituliskan sebagai persamaan berikut, jika Y adalah variabel acak poisson dengan parameter $\mu > 0$.

$$P(y,\mu) = \begin{cases} \dfrac{e^{-\mu}\mu^{y}}{y!}, & untuk\ y = 0,1,2,\dots,\infty \\ 0, & untuk\ y\ lain \end{cases} \tag{2.1}$$

dimana

μ: rata-rata banyaknya kejadian sukses

y: banyaknya unsur keberhasilan

Misalkan Y variabel berdistribusi poisson dengan parameter μ maka rata-rata dan variansi Y adalah μ.

Bukti

Rata-rata dari Y yang berdistribusi poisson dengan parameter μ.

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

$$= \sum_{y=0}^{\infty} y.\frac{e^{-\mu}\mu^y}{y!}$$

$$= e^{-\mu}\sum_{y=0}^{\infty} y.\frac{\mu^y}{y!}$$

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

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

$$= e^{-\mu}(\mu + \mu^2 + \frac{\mu^3}{2!} + \frac{\mu^4}{3!} + \cdots)$$

$$= e^{-\mu}(\mu(1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \cdots)$$

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

$$= e^0.\mu$$

$$= 1.\mu$$

$$E(Y) = \mu \tag{2.2}$$

Varians dari Y yang berdistribusi poisson dengan parameter μ.

Bukti

$$E(Y^2) = \sum_{y=0}^{\infty} y^2.P(y,\mu)$$

$$= \sum_{y=0}^{\infty} y^2.P(y,\mu)$$

$$= \sum_{y=0}^{\infty} y^2 . \frac{e^{-\mu}\mu^y}{y!}$$

$$= e^{-\mu} \sum_{y=0}^{\infty} y^2 . \frac{\mu^y}{y!}$$

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

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

$$= e^{-\mu}(\mu + 2\mu^2 + 3\frac{\mu^3}{2!} + 4\frac{\mu^4}{3!} + \cdots)$$

$$= e^{-\mu}(\mu(1+\mu)(1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \cdots))$$

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

$$= e^0 . \mu(1+\mu)$$

$$= 1 + \mu(1+\mu)$$

$$E(Y^2) = \mu + \mu^2$$

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

$$= (\mu + \mu^2) - (\mu)^2$$

$$V(Y) = \mu \tag{2.3}$$

Distribusi poisson dapat digunakan untuk mendekati distribusi binomial jika n besar (n>20) dan peluang keberhasilan p kecil (0.05) atau mendekati nol. Distribusi poisson mempunyai nilai μ yang kecil karena dalam keadaan yang jarang terjadi dan untuk nilai μ yang lebih besar maka akan lebih mendekati distribusi Normal (Cameron,1998).

2.2 Regresi Poisson

Regresi poisson berasal dari data yang berdistribusi poisson, suatu distribusi yang digunakan untuk menggambarkan suatu kejadian yang relatif jarang terjadi dimana perulangannya bergantung pada periode waktu tertentu atau pada lokasi tertentu (Walpole, 1995).

Hinde dan Demetrio (1998) mengatakan bahwa bentuk umum dari persamaan regresi poisson adalah sebagai berikut:

$$\ln \mu_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij} \;; \quad i = 1,2,\dots,n \; dan \; j = 1,2,\dots,p \qquad (2.4)$$

dimana:

p: jumlah variabel bebas

n: jumlah pengamatan

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

2.3 Uji Kecocokan Distribusi Poisson

Menurut Sudjana (2005), uji kecocokan menggunakan *chi square kuadrat* dapat digunakan untuk mendeteksi distribusi poisson. Prosedur untuk melakukan pengujian kecocokan distribusi poisson adalah sebagai berikut:

Hipotesis uji:

H_0 : Data mengikuti distribusi poisson

H_1 : Data tidak mengikuti distribusi poisson

a. Taraf signifikansi: 5%
b. Statistik uji:

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{n_i} \qquad (2.5)$$

dimana:

n_i: banyaknya pengamatan

p_i: peluang masing masing pengamatan

c. Kriteria pengujian

Jika nilai *chi-square* hitung lebih besar dari *chi-square* tabel maka H_0 ditolak.

2.4 Multikolinearitas

Pengujian multikolinearitas menentukan hubungan antar variabel bebas dalam suatu model regresi. Multikolinearitas harus dihindari dalam regresi ZINB karena dapat menyebabkan hasil estimasi parameter yang tidak tepat. Menurut Hocking (1996), nilai *Variance Inflation Factor* (VIF) dapat digunakan untuk mendeteksi multikolinearitas. Jika nilai VIF lebih besar dari 10, maka model dianggap mengandung multikolinearitas.

Untuk model regresi dengan lebih dari dua variabel, nilai VIF dapat dihitung menggunakan persamaan berikut:

$$VIF = \frac{1}{1 - R^2} \tag{2.6}$$

dimana:

R^2: Koefisien determinasi dari *auxiliary regression*

2.5 Overdispersi

Masalah yang sering terjadi pada regresi poisson adalah adanya overdispersi. Menurut Hilbe (2011), data pada regresi poisson dianggap memiliki overdispersi jika nilai varians lebih besar dari nilai rata-rata. Estimasi parameter koefisien regresi tetap konsisten tetapi tidak efektif jika terjadi overdispersi pada data dan regresi poisson tetap digunakan. Akibatnya kesimpulan menjadi salah karena nilai galat baku lebih rendah dari nilai sebenarnya (*underestimate*). Jika data dianggap overdispersi, maka model poisson dapat digantikan dengan model yang lebih sesuai seperti NB atau ZINB (McCullagh dan Nelder, 1989).

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

a. Terdapat korelasi antarpengamatan.
b. Terdapat pelanggaran asumsi sebaran poisson yaitu ragam lebih besar dari rata-rata.
c. Terdapat *zero-inflated* (nilai 0 berlebih).
d. Terdapat outlier pada data.

Overdispersi dapat ditemukan dengan membagi *residual deviance* dengan derajat bebasnya. Terjadi overdispersi dalam data jika nilai dari hasil bagi tersebut lebih dari satu. Selain itu, overdispersi dapat ditentukan dengan membandingkan nilai varians dengan rata-ratanya, jika nilai varians melebihi rata-rata, dikatakan terjadi overdispersi dalam data.

2.6 *Zero inflated*

Menurut Winkelmann (2008), salah satu permasalahan pada regresi poisson adalah terlalu banyak nilai nol (*zero inflated*). *Zero inflated* merupakan salah satu penyebab terjadinya overdispersi. Pada variabel terikat, data diskrit mungkin banyak mengandung nilai nol. Dalam beberapa keadaan, nilai nol memiliki kepentingan pada penelitian. Data diskrit harus dimasukkan dalam analisis jika nilai nol berpengaruh pada data. Variabel terikat dengan nilai nol yang lebih tinggi dari data diskrit lainnya merupakan contoh dari *zero inflated*. Selanjutnya, jika data mengandung *zero inflated*, regresi poisson tidak lagi akurat dalam menggambarkan data yang sebenarnya (Hinde dan Demetrio, 1998).

2.7 Distribusi Negatif Binomial (NB)

Distribusi NB sebagai pendekatan suatu percobaan sampai terjadi suatu kejadian sukses, dengan setiap pengulangannya saling bebas di mana peluang gagal yaitu $1-p$ dan peluang sukses adalah p. Jika peubah acak Y menyatakan jumlah percobaan yang dibutuhkan sampai terjadi k sukses, maka sebaran peluang peubah acak Y disebut mengikuti sebaran NB dengan fungsi peluang sebagai berikut:

$$P(y;p) = \binom{y-1}{k-1} p^k (1-p)^{x-k} \quad ; \quad y = k, k+1, \ldots \tag{2.7}$$

dimana:

k: Jumlah kejadian sukses

y: Jumlah percobaan sampai mendapatkan sukses ke-k

p: peluang sukses

$1-p$: peluang gagal

2.8 Regresi Negatif Binomial (NB)

Regresi NB adalah metode analisis statistika yang digunakan untuk memodelkan hubungan antara variabel terikat yang berdistribusi NB dan variabel bebas. Regresi NB merupakan salah satu model terapan dari GLM. Sebagai penerapan dari GLM maka distribusi NB memiliki tiga komponen yaitu komponen random, komponen sistematik dan fungsi penghubung. Berikut ini adalah rata-rata dan ragam sebaran NB (Greene, 2008).

Rata-rata :

$$\mathrm{E}(y_i) = \mu_i \tag{2.8}$$

Ragam :

$$\mathrm{VAR}(y_i) = \mu_i + \boldsymbol{\kappa}\mu_i{}^2 \tag{2.9}$$

Fungsi peluang pada model regresi NB dapat dijabarkan pada persamaan (2.10) berikut ini:

$$P(y_i;\mu_i,\boldsymbol{\kappa}) = \frac{\Gamma\left(y_i+\frac{1}{\boldsymbol{\kappa}}\right)}{\Gamma\left(\frac{1}{\boldsymbol{\kappa}}\right)y_i!}\left(\frac{1}{1+\boldsymbol{\kappa}\mu_i}\right)^{\frac{1}{k}}\left(\frac{\boldsymbol{\kappa}\mu_i}{1+\boldsymbol{\kappa}\mu_i}\right)^{y_i}, \quad i=0,1,2,\ldots, \qquad (2.10)$$

dimana:

y_i: jumlah kejadian

μ_i: rata-rata jumlah kejadian

Pada saat $\boldsymbol{\kappa} \to 0$ maka sebaran NB memiliki ragam yang mendekati rata-rata. Sebaran NB akan mendekati suatu sebaran poisson yang mengasumsikan rata-rata dan ragam yang sama apabila n mendekati ∞. Dalam model regresi NB, y_i adalah variabel yang berupa data hitung. Menurut Hilbe (2011), model regresi NB pada umumnya menggunakan fungsi penghubung logaritma atau *log link* yaitu:

$$\ln \mu_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$$

Model regresi NB dapat menggunakan l*og link* karena $\ln \mu_i$ dan $\boldsymbol{x}_i^T \boldsymbol{\beta}$ akan terdefinisi di dalam interval (0,∞). Setelah diperoleh fungsi penghubung yang tepat, maka selanjutnya dapat dinyatakan model regresi NB untuk memodelkan data hitung yaitu:

$$\ln(\mathrm{E}(Y_i|\, X_i)) = \ln(\mu_i) = \boldsymbol{x}_i^T \boldsymbol{\beta} \qquad (2.11)$$

Sehingga didapatkan:

$$\mu_i = \exp \boldsymbol{x}_i^T \boldsymbol{\beta} \qquad (2.12)$$

2.9 Regresi *Zero Inflated Negative Binomial* (ZINB)

Model regresi ZINB adalah model distribusi campuran poisson gamma. Menurut Garay, Hashimoto, Ortega, dan Lachos (2011), model ini bisa digunakan untuk merepresentasikan data diskrit atau data cacah dengan variabel terikat mengalami *zero inflated* dan overdispersi. Nilai variabel terikat dapat terjadi dalam salah satu dari dua cara. Jika y_i adalah variabel acak dengan i=1,2,...,n, maka nilai variabel terikat terjadi dalam dua cara. Keadaan awal regresi ZINB yang dikenal sebagai binomial *state* memiliki distribusi NB, sedangkan keadaan kedua menghasilkan pengamatan bernilai nol, yang dikenal sebagai *zero state.* Berikut ini adalah fungsi kepadatan peluang model ZINB:

$$P(y_i;\mu_i,\boldsymbol{\kappa}) = \begin{cases} \pi_i + (1-\pi_i)\left(\frac{1}{1+\boldsymbol{\kappa}\mu_i}\right)^{\frac{1}{\boldsymbol{\kappa}}} \;\; ; \;\; untuk\; y_i = 0 \\ (1-\pi_i)\frac{\Gamma\left(y_i+\frac{1}{\boldsymbol{\kappa}}\right)}{\Gamma\left(\frac{1}{\boldsymbol{\kappa}}\right)y_i!}\left(\frac{1}{1+\boldsymbol{\kappa}\mu_i}\right)^{\frac{1}{\boldsymbol{\kappa}}}\left(\frac{\boldsymbol{\kappa}\mu_i}{1+\boldsymbol{\kappa}\mu_i}\right)^{y_i}; \\ untuk\; y_i > 0 \end{cases} \tag{2.13}$$

dimana $0 \le \pi_i \le 1$, $\mu_i \ge 0$, $\boldsymbol{\kappa}$ adalah parameter dispersi dengan $\frac{1}{\boldsymbol{\kappa}} > 0$, $\Gamma(.)$ adalah fungsi gamma. Ketika $\pi_i = 0$, variabel acak y_i memiliki distribusi binomial negatif dengan rata-rata μ_i dan parameter dispersi κ, sehingga $Y_i \sim NB\;(\mu_i\;,\kappa)$. Diasumsikan bahwa μ_i dan π_i bergantung pada vektor dari variabel x_i yang dapat didefinisikan sebagai berikut:

$$\mu_i = e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}$$

$$\frac{\pi_i}{1-\pi_i} = e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}$$

$$\pi_i = e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}} - \pi_i e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}$$

$$\pi_i\left(1+e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}\right) = e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}$$

$$\pi_i = \frac{e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}}{1+e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}} \tag{2.14}$$

Model regresi ZINB dapat dinyatakan seperti pada persamaan berikut:

Model untuk *NB state* $\hat{\mu}_i$

$$\ln \mu_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \; ; i = 1,2,\dots,n \; dan \; j = 1,2,\dots,p \tag{2.15}$$

Model untuk *zero state* $\hat{\pi}_i$

$$\text{logit}\, \pi_i = \gamma_0 + \sum_{j=1}^{p} \gamma_j x_{ij} \; ; \; i = 1,2,\dots,n \; dan \; j = 1,2,\dots,p \tag{2.16}$$

dimana:

p: jumlah variabel bebas

n: jumlah pengamatan

μ_i: rata-rata NB

π_i: peluang data bernilai 0

β_j: parameter model regresi ZINB yang diduga

γ_j: parameter model regresi ZINB yang diduga

2.10 Metode *Maximum Likelihood Estimation* (MLE)

Berdasarkan fungsi peluang untuk y_i yang ada pada persamaan (2.13), akan dilakukan substitusi persamaan (2.14) dalam persamaan (2.13) dan menghasilkan :

$$P(Y_i = y_i) = \begin{cases} \dfrac{e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}}{1+e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}} + \dfrac{1}{1+e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}}\left(\dfrac{\mathbf{1}}{1+\boldsymbol{\kappa}\, e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\boldsymbol{\kappa}}\right)^{\frac{\mathbf{1}}{\boldsymbol{\kappa}}} ;\ untuk\ y_i = 0 \\ \left(\dfrac{1}{1+e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}}\right)\dfrac{\Gamma\left(y_i+\frac{\mathbf{1}}{\boldsymbol{\kappa}}\right)}{\Gamma\left(\frac{\mathbf{1}}{\boldsymbol{\kappa}}\right)\Gamma(y_i+1)}\left(\dfrac{\mathbf{1}}{1+\boldsymbol{\kappa}\, e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{\mathbf{1}}{\boldsymbol{\kappa}}}\left(\dfrac{\boldsymbol{\kappa}\, e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}}{1+\boldsymbol{\kappa}\, e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}}\right)^{y_i} ; \\ untuk\ y_i > 0 \end{cases} \tag{2.17}$$

Bentuk fungsi likelihood model ZINB dapat dituliskan sebagai berikut:

$$L(\boldsymbol{\kappa},\boldsymbol{\beta},\boldsymbol{\gamma}) = \begin{cases} \prod \dfrac{e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}} + \left(\frac{\mathbf{1}}{1+\boldsymbol{\kappa}\, e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{\mathbf{1}}{\boldsymbol{\kappa}}}}{1+e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}} ;\ untuk\ y_i = 0 \\ \prod \dfrac{\Gamma(y_i+\frac{\mathbf{1}}{\boldsymbol{\kappa}}) + \left(\frac{\mathbf{1}}{1+\boldsymbol{\kappa}\, e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{\mathbf{1}}{\boldsymbol{\kappa}}}\left(\frac{\boldsymbol{\kappa}\, e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}}{1+\boldsymbol{\kappa}\, e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}}\right)^{y_i}}{1+e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}\Gamma\left(\frac{\mathbf{1}}{\boldsymbol{\kappa}}\right)\Gamma(y_i+1)} ; \\ untuk\ y_i > 0 \end{cases} \tag{2.18}$$

Selanjutnya dari persamaan (2.18), dibuat persamaan ln *likelihood* dan dihasilkan:

$$\ln L(\boldsymbol{\kappa},\boldsymbol{\beta},\boldsymbol{\gamma}) = \begin{cases} \sum_{i=1}^{n} \ln\left(e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}} + \left(\frac{\mathbf{1}}{1+\boldsymbol{\kappa}\, e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{\mathbf{1}}{\boldsymbol{\kappa}}}\right) - \sum_{i=1}^{n} \ln\left(1+e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}\right) ;\ untuk\ y_i = 0 \\ -\sum_{i=1}^{n} \ln\left(1+e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}\right) + \sum_{i=1}^{n} \ln\left[\Gamma\left(y_i + \frac{\mathbf{1}}{\boldsymbol{\kappa}}\right)\right] - \sum_{i=1}^{n} \ln[\Gamma(y_i + 1)] - \\ \sum_{i=1}^{n} \ln\left[\Gamma\left(\frac{\mathbf{1}}{\boldsymbol{\kappa}}\right)\right] + y_i \sum_{i=1}^{n} \ln\left(\frac{e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}}{1+e^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}}\right)^{y_i} + \frac{\mathbf{1}}{\boldsymbol{\kappa}}\sum_{i=1}^{n} \ln\left(\frac{\mathbf{1}}{1+\boldsymbol{\kappa}\, e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{\mathbf{1}}{\boldsymbol{\kappa}}} ; \\ untuk\ y_i > 0 \end{cases} \tag{2.19}$$

Menurut Ariawan dkk. (2012), fungsi likelihood pada persamaan (2.19) tidak dapat diselesaikan dengan teknik numerik biasa karena penjumlahan fungsi ln *likelihood* tidak linier. Untuk menentukan estimasi suatu parameter menggunakan kerangka metode MLE dari suatu fungsi distribusi dengan informasi data yang tidak mencukupi atau data yang hilang, maka digunakanlah algoritma EM (*Expectation Maximization*). Misalkan peubah Y berkaitan dengan peubah indikator w yaitu:

$$w_i = \begin{cases} 1, \text{jika } \mathrm{y_i} \text{ berasal dari } zero\ state \\ 0, \text{jika } \mathrm{y_i} \text{ berasal dari } NB\ state \end{cases} \tag{2.20}$$

Saat memasukkan variabel tambahan seperti variabel w_i pada persamaan (2.20), fungsi *likelihood* pada data yang menyertakan variabel laten dapat dimaksimalkan menggunakan teknik iteratif yang disebut algoritma EM. Tahap ekspektasi dan tahap maksimalisasi adalah dua tahap yang membentuk algoritma EM. Tahap ekspektasi merupakan tahap perhitungan untuk menentukan ekspektasi dari fungsi ln *likelihood*, selanjutnya tahap maksimalisasi yaitu langkah perhitungan untuk menemukan estimasi parameter yang memaksimalkan fungsi ln *likelihood* dari hasil tahap ekspektasi sebelumnya.

$$\ln L(\beta,\gamma|y_i,w_i) = \sum_{i=1}^{n}\{w_i\boldsymbol{z}_i^T\boldsymbol{\gamma} - \ln(1+\exp\boldsymbol{z}_i^T\boldsymbol{\gamma}) + (1-w_i)\,\mathbf{ln}\left(g\left(y_i;\beta,\frac{\mathbf{1}}{\boldsymbol{\kappa}}\right)\right)\} \tag{2.21}$$

dimana

$$g\left(y_i;\beta,\frac{\mathbf{1}}{\boldsymbol{\kappa}}\right) = \frac{\Gamma\left(\frac{\mathbf{1}}{\boldsymbol{\kappa}}+y_i\right)}{\Gamma(y_i+1)\Gamma\left(\frac{\mathbf{1}}{\boldsymbol{\kappa}}\right)}\left(\frac{\boldsymbol{\kappa}\mu_i}{1+\boldsymbol{\kappa}\mu_i}\right)^{y_i}\left(\frac{\mathbf{1}}{1+\boldsymbol{\kappa}\mu_i}\right)^{\frac{\mathbf{1}}{\boldsymbol{\kappa}}} \tag{2.22}$$

Persamaan (2.21) akan dimaksimumkan menggunakan algoritma EM, dengan parameter β dan γ dapat diestimasi secara terpisah.

Langkah pertama dalam Algoritma EM adalah E-Step (tahap ekspektasi). Dalam tahapan ini, (s) digunakan untuk menyimbolkan iterasi (Garay, 2011).

$$\widehat{w}_i^{(s)} = \begin{cases} \left[1+e^{-\boldsymbol{z}_i^T\widehat{\boldsymbol{\gamma}}^{(\boldsymbol{s})}}\left(\frac{1}{1+\widehat{\boldsymbol{\kappa}}^{(s)}+e^{\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}^{(\boldsymbol{s})}}}\right)^{\frac{1}{\widehat{\kappa}}^{(s)}}\right]^{-1} &, jika\ y_i = 0 \\ 0,\ jika\ y_i > 0 \end{cases} \tag{2.23}$$

dan juga

$$Q\left(\beta,\gamma,\hat{\beta}^{(s)},\hat{\gamma}^{(s)}\right) = \ln L(\boldsymbol{\theta}|\, y_i, w_i)|y_i, \widehat{\boldsymbol{\theta}}^{(s)})$$

$$= \textstyle\sum_{i=1}^{n} Q_{1i}(\boldsymbol{\gamma}|\, \widehat{\boldsymbol{\theta}}^{(s)}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\beta},\boldsymbol{\kappa}|\, \widehat{\boldsymbol{\theta}}^{(s)}) \quad (2.24)$$

dimana

$$Q_{1i}(\boldsymbol{\gamma}|\widehat{\boldsymbol{\theta}}^{(s)}) = \widehat{w}_i^{(s)} \boldsymbol{z}_i^T \boldsymbol{\gamma} - \ln\left(1 + e^{\boldsymbol{z}_i^T \boldsymbol{\gamma}}\right) \quad (2.25)$$

dan

$$Q_{2i}(\boldsymbol{\beta},\boldsymbol{\kappa}|\hat{\beta}^{(s)},\hat{\gamma}^{(s)}) = (1-\widehat{w}_i^{(s)}) \ln\left[\frac{\Gamma(\boldsymbol{\kappa}+y_i)}{\Gamma(\boldsymbol{\kappa})\Gamma(y_i+1)}\left(\frac{e^{x_i^T\boldsymbol{\beta}}}{e^{x_i^T\boldsymbol{\beta}}+\kappa}\right)^{y_i}\left(\frac{\kappa}{e^{x_i^T\boldsymbol{\beta}}+\kappa}\right)^{\kappa}\right] \quad (2.26)$$

Tahapan selanjutnya masuk pada tahap M-Step (maksimalisasi). Dengan $Q\left(\beta,\gamma,\hat{\beta}^{(s)},\hat{\gamma}^{(s)}\right)$ fungsi yang memaksimumkan parameter yang akan diduga. Dengan menggunakan metode iterasi *Newton Raphson* untuk memaksimalkan $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ dari hasil tahap E-Step (ekspektasi) dengan menghitung $\beta^{(m+1)}$ dan $\gamma^{(m+1)}$. Iterasi dilakukan hingga konvergen dengan syarat $\left|\hat{\beta}^{(m+1)} - \hat{\beta}^{(m)}\right|$ dan $\left|\hat{\gamma}^{(m+1)} - \hat{\gamma}^{(m)}\right|$ kurang dari ε dengan ε yang telah ditetapkan sebelumnya.

2.11 Pengujian Parameter Model

Pengujian signifikansi parameter model regresi bertujuan untuk mengetahui apakah terdapat hubungan linier antara variabel terikat dan variabel bebas. Pengujian signifikansi dapat dilakukan secara serentak (simultan) maupun individual (parsial).

2.11.1 Pengujian secara Simultan

Pengujian secara simultan digunakan untuk menguji peranan variabel bebas di dalam model secara bersama-sama yang mencakup seluruh parameter β dan γ.

Untuk melakukan uji simultan parameter model dapat dilakukan dengan statistik uji *ratio likelihood* atau uji G dengan ketentuan sebagai berikut:

Hipotesis yang digunakan adalah sebagai berikut:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_p = \gamma_1 = \gamma_1 = \cdots = \gamma_p = 0$ (tidak ada variabel bebas yang berpengaruh terhadap variabel terikat)

H_1: Paling sedikit ada satu $\beta_j \neq 0$ atau $\gamma_j \neq 0$, $j = 1,2, \dots, p$ (paling sedikit ada satu variabel bebas yang berpengaruh terhadap variabel terikat)

Menurut Hosmer-Lemeshow (2000),pengujian parameter model secara simultan dilakukan dengan menggunakan statistik uji rasio likelihood. Uji G adalah fungsi dari L_0 dan L_1 yang berdistribusi *chi-square* (χ^2) dengan derajat bebas p yang didefinisikan sebagai berikut:

$$G = -2[L_0 - L_1] \tag{2.27}$$

dimana:

L_0: ln *likelihood* model tanpa variabel bebas (model intersept)

L_1: ln *likelihood* model dengan variabel bebas (model penuh)

Kriteria pengujian:

H_0 ditolak apabila statistik uji G lebih besar dari $\chi 2_{\mathrm{p}(\alpha)}$.

2.11.2 Pengujian secara Parsial

Pengujian secara parsial digunakan untuk mengetahui pengaruh variabel-variabel bebas terhadap variabel terikat. Statistik uji yang digunakan untuk uji parsial, yaitu uji Wald.

a. Uji Parsial Parameter β

Pengujian ini menguji seluruh parameter β secara parsial untuk mengetahui parameter apa saja yang memberikan pengaruh signifikan terhadap model.

Hipotesis yang digunakan adalah sebagai berikut:

H_0: $\beta_j = 0,\ j\ =\ 1,2, \dots\ , p$ (variabel bebas tidak berpengaruh terhadap variabel terikat)

H_1: Paling sedikit ada satu $\beta_j \neq 0,\ j\ =\ 1,2, \dots\ , p$ (variabel bebas berpengaruh terhadap variabel terikat)

Menurut Hosmer-Lemeshow (2000), statistik uji yang digunakan:

$$W(\hat{\beta}_j)\ =\ \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2 \tag{2.28}$$

dimana:

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

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

Kriteria pengujian:

Jika $W(\hat{\beta}_j) > \chi^2_{\alpha;1}$, maka H_0 ditolak yang artinya variabel bebas memiliki pengaruh yang signifikan terhadap variabel terikat Y.

b. Uji Parsial Parameter γ

Pengujian ini menguji seluruh parameter γ secara parsial untuk mengetahui parameter apa saja yang memberikan pengaruh signifikan terhadap model.

Hipotesis yang digunakan adalah sebagai berikut:

H_0: $\gamma_j = 0,\ j\ =\ 1,2, \dots\ , p$ (variabel bebas tidak berpengaruh terhadap variabel terikat)

H_1: Paling sedikit ada satu $\gamma_j \neq 0,\ j\ =\ 1,2, \dots\ , p$ (variabel bebas berpengaruh terhadap variabel terikat).

Menurut Hosmer-Lemeshow (2000), statistik uji yang digunakan:

$$W(\hat{\gamma}_j)\ =\ \left(\frac{\hat{\gamma}_j}{SE(\hat{\gamma}_j)}\right)^2 \tag{2.29}$$

dimana:

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

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

Kriteria pengujian:

Jika $W(\hat{\gamma}_j) > \chi^2_{\alpha;1}$, maka H_0 ditolak yang artinya variabel bebas memiliki pengaruh yang signifikan terhadap variabel terikat Y.

2.12 *Akaike Information Criterion* (AIC)

Menurut Akaike (1978), nilai *Akaike Information Criterion* (AIC) terkecil merupakan salah satu kriteria terbaik dalam pemilihan model. Nilai AIC ditentukan berdasarkan nilai *maximum likelihood* dan banyaknya parameter yang digunakan untuk membuat model regresi. Ketika digunakan dengan data yang dimiliki secara mutlak, kriteria AIC dapat menunjukkan seberapa akurat model tersebut. Berikut adalah definisi dari kriteria AIC:

$$AIC = -2 \ln L(\theta) + 2c \tag{2.30}$$

dimana:

$L(\theta)$: nilai *maximum likelihood*

c: jumlah parameter

2.13 Jumlah Kematian Bayi dan Faktornya

Angka Kematian Bayi (AKB) adalah banyaknya bayi yang meninggal sebelum mencapai usia yang ditentukan yaitu satu tahun untuk setiap 1.000 kelahiran hidup (Dinkes, 2022). Untuk tujuan *Sustainable Development Goals* (SDGs) yang harus dicapai, AKB digunakan untuk merepresentasikan tingkat pembangunan kesehatan suatu bangsa dan kualitas hidup masyarakat. Kematian bayi dapat terjadi karena berbagai sebab antara lain asfiksia, BBLR, tetanus, dan kesulitan menyusui, dll.

Bayi baru lahir dengan berat badan kurang dari 2.500 gram tergolong memiliki BBLR. Hampir semua kasus BBLR memerlukan perawatan di unit perawatan kritis bayi baru lahir (NICU). Bayi dengan BBLR lebih mungkin meninggal dalam 28 hari pertama kehidupan. BBLR lebih rentan mengalami stunting, IQ rendah, yang dapat membahayakan kualitas sumber daya manusia di masa depan, bahkan membahayakan nyawanya. Sedangkan bayi baru lahir BBLR dapat mengakibatkan obesitas, diabetes, penyakit jantung, dan penyakit tidak menular lainnya (WHO, 2019).

Pelayanan Kesehatan Bayi (PKB) pada bayi baru lahir minimal diberikan sebanyak empat kali, yaitu antara usia 29 hari-2 bulan, 1 kali pada umur 3-5 bulan, 1 kali pada umur 6-8 bulan, dan 1 kali pada umur 9-11 bulan Pelayanan kesehatan bayi tersebut meliputi, pemantauan pertumbuhan, pemberian vitamin A pada bayi umur 6-11 bulan, pemberian imunisasi dasar (BCG, DPT/HB/HiB1-3, Polio 1-4, dan Campak), Stimulasi Deteksi Intervensi Dini Tumbuh Kembang (SDIDTK), pemberian ASI eksklusif dan Makanan Pendamping ASI (MPASI). Unsur lainnya adalah ASI Eksklusif, yaitu pemberian ASI eksklusif pada bayi usia 0 sampai 6 bulan tanpa makanan tambahan apapun (Depkes RI, 2004).

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat

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

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari profil dinas kesehatan kabupaten Lampung Selatan tahun 2021 Jumlah pengamatan pada penelitian ini adalah 27 pengamatan berdasarkan jumlah puskesmas di Lampung Selatan. Variabel terikat yang digunakan adalah jumlah kematian bayi (Y), sedangkan variabel bebas yang digunakan adalah sebanyak 4 variabel yaitu BBLR (X_1), persalinan ditolong tenaga kesehatan (X_2), pelayanan kesehatan bayi (X_3), dan pemberian ASI eksklusif (X_4).

3.3 Metode Penelitian

Dalam penelitian ini dilakukan dengan menggunakan bantuan *software R* Langkah-langkah dalam penelitian ini adalah sebagai berikut:

1. Melakukan analisis deskriptif variabel.
2. Melakukan uji asumsi multikolinearitas pada variabel-variabel bebas menggunakan kriteria uji VIF, apabila nilai VIF<10 maka tidak terjadi multikolinearitas.
3. Melakukan uji overdispersi dengan melihat nilai *Residual Devians* yang dibagi dengan derajat bebasnya atau nilai ragam yang lebih besar daripada rata-ratanya.
4. Pemeriksaan *zero-inflated* pada variabel terikat, jika nilai nol lebih besar dari 50% maka dikatakan terjadi *zero-inflated*.
5. Memeriksa sebaran variabel terikat apakah mengikuti sebaran poisson.
6. Melakukan pendugaan parameter regresi NB dan melakukan pengujian signifikansi parameter model regresi dengan uji signifikansi parameter secara simultan menggunakan uji G dan secara parsial menggunakan uji Wald.
7. Melakukan pendugaan parameter regresi ZINB dan melakukan pengujian signifikansi parameter model regresi dengan uji signifikansi parameter secara simultan menggunakan uji G dan secara parsial menggunakan uji Wald.
8. Membandingkan nilai AIC dari kedua model yang terbentuk dan menentukan model terbaik menggunakan kriteria nilai AIC.

V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan yaitu memodelkan data jumlah kematian bayi di Lampung Selatan tahun 2021, maka dapat disimpulkan bahwa:

1. Model regresi ZINB pada data jumlah kematian bayi di Lampung Selatan pada tahun 2021 adalah sebagai berikut:

   Model untuk NB *state* $\widehat{\mu_i}$

   $$\widehat{\mu_i} = 0.240 + 0.127X_1 - 0.013X_2 + 0.013X_3 - 0.001X_4$$

   Model untuk *zero-inflated state* $\widehat{\pi_i}$

   $$\widehat{\pi_i} = \frac{\exp 6.524 + 0.934X_1 - 1.363X_2 + 1.619X_3 - 0.154X_4}{1 + \exp 6.524 + 0.934X_1 - 1.363X_2 + 1.619X_3 - 0.154X_4}$$

2. Berdasarkan model regresi ZINB yang telah terbentuk, variabel bebas yang signifikan adalah BBLR (X_1), persalinan ditolong tenaga kesehatan (X_2),dan pemberian ASI eksklusif (X_4).
3. Berdasarkan kriteria nilai AIC, model regresi ZINB merupakan model terbaik untuk memodelkan data jumlah kematian bayi di Lampung Selatan pada tahun 2021.

DAFTAR PUSTAKA

Akaike, H. 1978. A Bayesian Analysis of The Minimum AIC Procedure. *Annals of The Institute Statistical Mathematics*. **1**(1): 9-14.

Amaliana L., Sa'adah, U., dan Wayan, N.S.W. 2019. Performa Proporsi Zero-Inflated pada Regresi Zero Inflated Negative Binomial. *E-Jurnal Matematika.* **8**(2):78-87.

Astuti, C.C. 2015. Pemodelan Regresi Zero Inflated Negative Binomial (ZINB) Pada Kasus Tetanus Neonatorum di Provinsi Jawa Timur. (Thesis). Jurusan Statistika. Institut Teknologi Sepuluh Nopember Surabaya. Surabaya.

Ariawan, B., Suparti, dan Sudarno. 2012. Pemodelan Regresi Zero-Inflated Negative Binomial (ZINB) Untuk Data Respon Diskrit dengan Excess Zeros. *Jurnal GAUSSIAN*. **1**(1) : 55-64.

Cameron, C.A. 1998. *Regression Analysis of Count Data*. Cambridge University Press, New York.

Dewanti, N.P., Susilawati, M., dan Srinadi, I. G. 2019. Perbandingan Regresi Zero Inflated Poisson (ZIP) dan Regresi Zero Inflated Negative Binomial (ZINB) pada Data Overdispersion. *E-Jurnal Matematika.* **5**(4):133-138.

Dinas Kesehatan Kabupaten Lampung Selatan. 2022. *Profil Kesehatan Kabupaten Lampung Selatan Tahun 2021*. Dinas Kesehatan, Lampung Selatan.

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

Hardin, J.W. dan Hilbe, J.M. 2007. *Generalized Linear Models and Extensions*. A Stata Press Publication, Texas.

Hilbe. J.M. 2011. *Negatif Binomial Regression*. 2^{th} Edition. Cambridge University Press, New York.

Hinde, J. dan Demetrio, C.G. 1998. Overdispersion: Models and Estimation. *Computational Statistic and Data Analysis*. **27**(2): 151-170.

Hocking, R. 1996. *Methods and Application of Linear Models*. John Wiley dan Sons, New York.

Hosmer, D.W. dan Lemeshow, S. 2000. *Applied Logistic Regression*. 2^{nd} Edition. John Wiley dan Sons, New York.

Lambert, D. 1992. Zero Inflated Poisson Regression, With an Application to Defect in Manufacturing. *Technometric*, **34:** 1.

Mc Cullagh, P. dan Nelder, J.A. 1989. *Generalized Linear Models*. 2^{th} Edition. Chapman dan Hall, London.

Nelder, J.A. dan Wedderburn, R.W. 1972. Generalized Linear Models. *Journal of the Royal Statistical Society*. **135**(3): 370-384.

Sudjana. 2005. *Metoda Statistika*. Tarsito, Bandung.

Walpole, R.E. 1995. *Pengantar Statistika*. Edisi ke-3. PT. Gramedia Pustaka, Jakarta.

Winkelman, R. 2008. *Econometric Analysis of Count Data*. 5^{th} Edition. Springer, Berlin.

Zahro, J., Caraka, R.E., dan Herliansyah, R. 2018. *Aplikasi generalized linear model pada R*. Innosain, Yogyakarta.