ESTIMASI: Journal of Statistics and Its Application
Vol. 6, No. 1, Januari, 2025, Hal. 10-20
DOI: ….

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

Ainul Fajri[1*], Andi Kresna Jaya[2], La Podje Talangko[3]
[123]Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin, Makassar, 90245, Indonesia
*** Corresponding author, email: ainulfajri1999@gmail.com**

Abstract

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

Keywords*: Overdispersion, Neonatal Mortality, ZINB Regression, MLE.*

Abstrak

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

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

1. Pendahuluan

Analisis yang digunakan untuk memodelkan hubungan antara variabel respon dengan satu atau lebih variabel prediktor adalah analisis regresi. Hubungan antara variabel respon dan variabel prediktor dinyatakan dalam persamaan regresi yang dapat berbentuk hubungan linear atau non linear. Persamaan regresi linear digunakan untuk menganalisis variabel respon yang berupa data kontinu dan mengikuti distribusi normal, namun banyak ditemukan variabel respon yang tidak berdistribusi normal. Untuk mengatasi hal tersebut dikembangkan *Generalized Linear Model* (GLM). *Generalized Linear Model* digunakan sebagai perluasan model regresi umum dengan variabel responnya berdistribusi keluarga eksponensial [7].

Jika variabel respon yang digunakan merupakan data diskrit dan berdistribusi Poisson, maka dapat digunakan model regresi Poisson untuk pembentukan model regresi. Analisis data menggunakan regresi Poisson harus memenuhi asumsi seperti nilai varians dan nilai rata-rata memiliki nilai yang sama atau equidispersi [7]. Akan tetapi, pada penerapannya sering terjadi overdispersi, yaitu nilai varians lebih besar dari nilai rata-ratanya. Data yang mengandung overdispersi dan dianalisis dengan regresi Poisson menghasilkan galat baku yang lebih kecil dari nilai sesungguhnya (*underestimate*). Hal ini menyebabkan kesimpulan yang diperoleh menjadi tidak valid [6]. Overdispersi pada regresi Poisson dapat terjadi karena banyaknya pengamatan yang bernilai nol pada variabel respon (*excess zeros*), penanganan model yang dapat digunakan untuk mengatasi overdispersi akibat *excess zeros* pada variabel respon untuk data diskrit yaitu *Zero Inflated Negative Binomial* (ZINB).

Model regresi *Zero Inflated Negative Binomial* (ZINB) merupakan model yang dibentuk dari distribusi campuran Poisson gamma [3]. Penelitian ini menerapkan model regresi *Zero Inflated Negative Binomial* (ZINB) pada kasus kematian neonatal di Kota Makassar tahun 2018. Jumlah kematian neonatal adalah banyaknya kematian bayi yang terjadi saat bayi baru lahir hingga bayi berusia di bawah 28 hari, yang mana data tersebut merupakan data jumlahan sehingga tergolong ke dalam data diskrit. Kematian neonatal umumnya disebabkan oleh faktor yang dibawa anak sejak lahir yang diperoleh dari orang tua pada saat konsepsi atau selama kehamilan [1].

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

2. Material dan Metode

2.1 *Generalized Linear Model* (GLM)

Generalized Linear Model (GLM) merupakan perluasan dari model regresi ketika variabel respon mengikuti distribusi keluarga eksponensial yaitu normal, binomial, Poisson, geometrik, binomial negatif, eksponensial, dan gamma. McCullagh dan Neldar [6] menyatakan terdapat tiga komponen utama yang harus ada dalam suatu *Generalized Linear Model*, yaitu:

1. Komponen acak, yaitu suatu komponen yang mengidentifikasi distribusi dari variabel respon (Y) berasal dari keluarga eksponensial. Bentuk keluarga distribusi eksponensial dapat dinyatakan dalam fungsi peluang berikut:

$$f(y) = \exp[a(y)b(\theta) + c(\theta) + d(y)] \tag{1}$$

2. Komponen sistematik, meliputi variabel-variabel prediktor dari model untuk menjelaskan variabel-variabel yang berhubungan dalam sebuah model linear.
$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_1 x_{1i} + \cdots + \beta_1 x_{pi} \quad (2)$$
3. Fungsi penghubung (*link function*), yaitu fungsi yang menghubungkan rata-rata dari variabel respon (Y) dengan variabel-variabel prediktor melalui persamaan linear. Fungsi penghubung $g(.)$ menghubungkan $E(Y_i)$ dengan variabel prediktor, yaitu:
$$g(\lambda_i) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ji}\,,\ \ i = 1,2,\dots,n \quad (3)$$

2.2 Regresi Binomial Negatif

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

1. Komponen Acak

   Pada model binomial negatif variabel respon diasumsikan berdistribusi binomial negatif yang dihasilkan dari distribusi campuran Poisson gamma [8]. Distribusi peluang campuran Poisson-gamma dapat diperoleh dengan cara:
$$P(Y = y_i) = \int_0^\infty \mathrm{Poisson}(Y|\lambda_i)\mathrm{Gamma}(\lambda_i|\alpha,\beta)\,d\lambda_i$$
$$= \frac{\beta^{-\alpha}}{\Gamma(\alpha)y_i!}\int_0^\infty e^{-\lambda_i\left(1+\frac{1}{\beta}\right)}\lambda_i^{y_i+\alpha-1}\,d\lambda_i \quad (4)$$

2. Komponen Sistematik

   Kontribusi variabel bebas dalam model regresi binomial negatif dinyatakan dalam bentuk kombinasi linier antara parameter $\boldsymbol{\eta}$ dengan parameter regresi yang diestimasi yaitu:
$$\boldsymbol{\eta} = \boldsymbol{X\beta} \quad (5)$$
   dengan $\boldsymbol{\eta}$ adalah vektor berukuran $n \times 1$ dari observasi, $\boldsymbol{X}$ adalah matriks berukuran $n \times c$ yang elemennya terdiri dari variabel prediktor, $\boldsymbol{\beta}$ adalah vektor berukuran $c \times 1$ yang elemennya terdiri dari parameter regresi, dengan $c = p + 1$.

3. Fungsi Penghubung

   Nilai rata-rata dari variabel respon adalah diskrit dan bernilai positif. Maka untuk mentransformasikan nilai η_i (bilangan riil) ke rentang yang sesuai dengan rentang pada variabel respon diperlukan suatu fungsi penghubung $g(.)$ yaitu:
$$g(\lambda_i) = \ln(\lambda_i) = \boldsymbol{x}_i^T\boldsymbol{\beta}$$
$$\lambda_i = \exp(\boldsymbol{x}_i^T\boldsymbol{\beta}) \quad (6)$$

2.3 Regresi Zero *Inflated Negative Binomial*

Model regresi *Zero Inflated Negative Binomial* (ZINB) merupakan model yang dibentuk dari distribusi campuran Poisson gamma. Jika y_i adalah variabel acak dengan $i = 1,2,\dots,n$ nilai nol pada observasi diduga muncul dalam dua cara yang sesuai untuk keadaan (*state*) yang terpisah. Menurut Garay dkk [2], regresi *Zero Inflated Negative*

Binomial (ZINB) dengan keadaan pertama disebut *zero state* terjadi dengan probabilitas π_i dan menghasilkan hanya observasi bernilai nol, sementara keadaan kedua disebut *negative binomial state* terjadi dengan probabilitas $(1-\pi_i)$ dan berdistribusi binomial negatif. Fungsi peluang model regresi *Zero Inflated Negative Binomial* (ZINB) dapat dinyatakan seperti pada persamaan (7).

$$P(Y_i = y_i) = \begin{cases} \pi_i + (1-\pi_i)\left(\frac{1}{1+\kappa\lambda_i}\right)^{\frac{1}{\kappa}}, & \text{untuk } y_i = 0 \\ (1-\pi_i)\frac{\Gamma\left(y_i+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y_i!}\left(\frac{1}{1+\kappa\lambda_i}\right)^{\frac{1}{\kappa}}\left(\frac{\kappa\lambda_i}{1+\kappa\lambda_i}\right)^{y_i}, & \text{untuk } y_i > 0 \end{cases} \quad (7)$$

dengan $0 \leq \pi_i \leq 1, \lambda_i \geq 0$, κ adalah parameter dispersi dengan $\frac{1}{\kappa} > 0$ dan $\Gamma(.)$ adalah fungsi gamma. Ketika $\pi_i = 0$, variabel acak y_i memiliki distribusi binomial negatif dengan rata-rata λ_i dan parameter dispersi κ, sehingga $Y_i \sim NB\ (\lambda_i, \kappa)$. Fungsi penghubung logit digunakan Ketika parameter model regresi bernilai 0 dan 1 dan fungsi penghubung ln digunakan jika parameter diharapkan bernilai positif.

Model regresi *Zero Inflated Negative Binomial* (ZINB) dapat dinyatakan seperti pada persamaan (8) dan (9) yang dituliskan pada persamaan berikut:
Model untuk *negative binomial state* $\hat{\lambda}_i$

$$\ln \hat{\lambda}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij},\ i = 1,2,\dots,n \text{ dan } j = 1,2,\dots,p \quad (8)$$

Model untuk *zero state* π_i

$$\text{logit } \hat{\pi}_i = \ln\left(\frac{\pi}{1-\pi}\right) = \hat{\gamma}_0 + \sum_{j=i}^{p} \hat{\gamma}_j x_{ij},\ \ i = 1,2,\dots,n \text{ dan } j = 1,2,\dots,p \quad (9)$$

dengan:
p : jumlah variabel prediktor
n : jumlah pengamatan
$\hat{\beta}_0, \hat{\gamma}_0$: intersep
$\hat{\beta}_j, \hat{\gamma}_j$: parameter ke-j dari regresi ZINB

2.4 Uji Serentak Parameter Model

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

Hipotesis yang digunakan adalah sebagai berikut:
H_0: $\beta_1 = \beta_2 = \cdots = \beta_p = \gamma_1 = \gamma_2 = \cdots = \gamma_p = 0$
H_1: minimal ada satu $\beta_j \neq 0$ atau $\gamma_j \neq 0, j = 1,2,\dots,p$
Statistik uji:

$$\begin{aligned} G &= -2 \ln \left[\frac{L(\widehat{\omega})}{L(\widehat{\Omega})} \right] \\ &= -2\left[\ln L(\widehat{\omega}) - \ln L(\widehat{\Omega})\right] \end{aligned} \quad (10)$$

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

Jika $G_{hitung} > \chi^2_{\alpha;p}$ dengan p jumlah parameter, maka H_0 ditolak yang artinya variabel prediktor secara bersama-sama memiliki pengaruh terhadap variabel respon [5].

2.5 Uji Parsial Parameter Model

Uji parsial digunakan untuk mengetahui variabel-variabel prediktor yang berpengaruh terhadap variabel respon. Statistik uji yang digunakan untuk uji parsial, yaitu uji Wald [5].

1. Uji Parsial Parameter $\boldsymbol{\beta}$

Parameter yang diuji pada pengujian ini mencakup seluruh parameter $\boldsymbol{\beta}$ secara parsial untuk mengetahui parameter mana saja yang memberikan pengaruh signifikan terhadap model. Hipotesis yang digunakan adalah sebagai berikut:
H_0 : $\beta_j = 0,\ j = 1,2, \dots ,p$
H_1 : minimal ada satu $\beta_j \neq 0,\ j = 1,2, \dots ,p$
Statistik uji:

$$W(\hat{\beta}_j) = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2 \quad (11)$$

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

Jika $W(\hat{\beta}_j) > \chi^2_{\alpha;p}$, maka H_0 ditolak artinya variabel prediktor memiliki pengaruh yang signifikan terhadap variabel respon Y [5].

2. Uji Parsial Parameter $\boldsymbol{\gamma}$

Parameter yang diuji pada pengujian ini mencakup seluruh parameter $\boldsymbol{\gamma}$ secara parsial untuk mengetahui parameter mana saja yang memberikan pengaruh signifikan terhadap model. Hipotesis yang digunakan adalah sebagai berikut:
H_0 : $\gamma_j = 0,\ j = 1,2, \dots ,p$
H_1 : minimal ada satu $\gamma_j \neq 0,\ j = 1,2, \dots ,p$
Statistik uji:

$$W(\hat{\gamma}_j) = \left(\frac{\hat{\gamma}_j}{SE(\hat{\gamma}_j)}\right)^2 \quad (12)$$

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

Jika $W(\hat{\gamma}_j) > \chi^2_{\alpha;p}$, maka H_0 ditolak artinya variabel prediktor memiliki pengaruh yang signifikan terhadap variabel respon Y [5].

2.6 Pemilihan Model Terbaik

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

$$AIC = 2c - 2\ln L(\hat{\boldsymbol{\theta}}) \tag{13}$$

dengan $L(\hat{\boldsymbol{\theta}})$ adalah nilai fungsi *likelihood* dan $c = p + 1$ adalah banyaknya parameter [4].

2.7 Metode Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari buku Profil Dinas Kesehatan Kota Makassar tahun 2018 yang dipublikasikan oleh Dinas Kesehatan Kota Makassar. Jumlah pengamatannya terdiri dari 46 puskesmas yang tersebar di 14 kecamatan di Kota Makassar. Variabel yang digunakan dalam penelitian ini terdiri dari variabel respon (y) dan variabel prediktor (x). Variabel respon dalam penelitian ini adalah jumlah kematian neonatal, yaitu jumlah kematian bayi di bawah usia 28 hari. Sedangkan variabel prediktornya, yaitu: jumlah bayi dengan berat badan lahir rendah (x_1), jumlah bayi yang diberi ASI eksklusif (x_2), jumlah kunjungan neonatal (x_3).

Langkah-langkah yang dilakukan untuk menganalisis data adalah sebagai berikut:

1. Memeriksa distribusi variabel respon dengan uji *Kolmogorov-Smirnov*
2. Memeriksa proporsi nilai nol pada variabel respon.
3. Memeriksa overdispersi menggunakan nilai *deviance* sesuai dengan persamaan
4. Mengetahui fungsi peluang model regresi ZINB.
5. Menyusun algoritma untuk proses estimasi parameter model regresi ZINB berdasarkan fungsi *likelihood* yang sudah diketahui. Estimasi parameter model regresi ZINB dilakukan menggunakan metode MLE dan diselesaikan menggunakan algoritma EM.
6. Mendapatkan model regresi ZINB pada data jumlah kematian neonatal di Kota Makassar tahun 2018.
7. Melakukan pengujian signifikansi parameter model regresi ZINB secara simultan dengan uji rasio *likelihood* dan secara parsial dengan uji Wald.

8. Menentukan model terbaik, yaitu antara model Binomial Negatif dan model ZINB dengan menggunakan kriteria AIC.
9. Menginterpretasi model yang diperoleh dan menarik kesimpulan.

3. Hasil dan Diskusi

Sebelum melakukan analisis data, terdapat asumsi yang harus dipenuhi, yaitu variabel respon berdistribusi Poisson dan overdispersi. Oleh karena itu, perlu dilakukan uji kecocokan distribusi Poisson dan uji overdispersi.

3.1 Uji Kecocokan Distribusi Poisson

Uji kecocokan distribusi Poisson bertujuan untuk mengetahui data berdistribusi Poisson atau tidak. Salah satu uji kecocokan distribusi Poisson yang dapat digunakan adalah uji *Kolmogorov-Smirnov*. Hasil uji kecocokan distribusi Poisson berdasarkan output SPSS ditunjukkan pada Tabel 1.

Tabel 1. Uji Kecocokan Distribusi Poisson

Variabel Respon	Statistik Uji	*P-value*
Kematian Neonatal (Y)	0.180	0.100

Sumber: *Data diolah, 2021*

Tabel 1 menunjukkan nilai statistik uji $D_{hitung} = 0.180 < D_{tabel} = 0.240$ untuk variabel respon Y dan dapat juga dilihat nilai p-value sebesar 0.100 yang lebih besar dari nilai α = 0.05 yang berarti H_0 diterima. Artinya, variabel jumlah kematian neonatal di Kota Makassar berdistribusi Poisson sehingga dapat dianalisis menggunakan regresi Poisson.

3.2 Pemeriksaan *Excess Zeros* pada Variabel Respon

Pemeriksaan *excess zeros* dilakukan dengan menghitung persentase amatan yang bernilai nol pada variabel respon. Hasil pemeriksaan *excess zeros* pada variabel respon ditunjukkan pada Tabel 2.

Tabel 2. Pemeriksaan *excess zeros* pada variabel respon

Nilai	Jumlah	Persentase (%)
0	28	60.87
1	8	17.39
2	3	6.52
3	5	10.87
4	1	2.17
6	1	2.17
Total	**46**	**100.00**

Sumber: *Data diolah, 2021*

Berdasarkan Tabel 2 terlihat bahwa nilai nol mempunyai proporsi sebesar 60.87% yang melebihi proporsi nilai pada variabel respon lainnya, sehingga variabel respon pada data jumlah kematian neonatal di Kota Makassar tahun 2018 mengalami *excess zeros*.

3.3 Uji Overdispersi

Regresi Poisson memiliki asumsi khusus yang harus dipenuhi, yaitu asumsi *equidispersi* dengan nilai mean dan variansi dari data bernilai sama. Uji overdispersi dilakukan dengan menggunakan uji *Pearson Chi Square.* Hasil uji overdispersi berdasarkan sintaks RStudio ditunjukkan pada Tabel 3.

Tabel 3. Uji Overdispersi

Deviance	**Derajat Bebas**	$\boldsymbol{\phi}$
82.61	42	1.966

Sumber: *Data diolah, 2021*

Tabel 3 menunjukkan bahwa nilai statistik uji $\phi = 1.966 > 1$ yang berarti variabel jumlah kematian neonatal di Kota Makassar mengalami overdispersi. Karena variabel respon mengalami overdispersi dan berdistribusi Poisson, maka dapat dilakukan analisis regresi *Zero Inflated Negative Binomial* (ZINB).

3.4 Pemodelan Regresi *Zero Inflated Negative Binomial* (ZINB) pada Data Jumlah Kematian Neonatal di Kota Makassar Tahun 2018

Pemodelan data jumlah kematian neonatal di Kota Makassar tahun 2018 dilakukan menggunakan model regresi *Zero Inflated Negative Binomial* (ZINB). Berdasarkan sintaks RStudio, diperoleh hasil estimasi parameter seperti pada Tabel 4.

Tabel 4. Estimasi Parameter κ, $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$

Parameter	**Estimasi**
$\hat{\kappa}$	0.25186
$\hat{\beta}_0$	-1.17093
$\hat{\beta}_1$	0.04583
$\hat{\beta}_2$	-0.00086
$\hat{\beta}_3$	0.00087
$\hat{\gamma}_0$	-30.4139
$\hat{\gamma}_1$	0.34375
$\hat{\gamma}_2$	-0.00987
$\hat{\gamma}_3$	0.01534

Sumber: *Data diolah, 2021*

Model regresi ZINB dibentuk dari hasil estimasi parameter pada Tabel 4 sehingga diperoleh model berikut:

$$\ln(\hat{\lambda}_i) = -1.17093 + 0.04583\boldsymbol{x}_{i1} - 0.00086\boldsymbol{x}_{i2} + 0.00087\boldsymbol{x}_{i3}$$
$$\text{logit}(\hat{\pi}_i) = -30.4139 + 0.34375\boldsymbol{x}_{i1} - 0.00987\boldsymbol{x}_{i2} + 0.01534\boldsymbol{x}_{i3}$$

dari model di atas, dilakukan uji serentak untuk mengetahui apa seluruh variabel prediktor berpengaruh terhadap variabel respon.

Hasil uji serentak untuk model regresi ZINB didapatkan nilai statistik uji G sebesar 43.250. Dengan tingkat signifikan $\alpha = 0.05$ diperoleh $\chi^2_{(0,05;3)} = 7.814 < 43.250$ maka H_0 ditolak yang artinya minimal ada satu variabel prediktor yang berpengaruh signifikan terhadap variabel respon jumlah kematian neonatal. Oleh karena itu, diperlukan pengujian parsial untuk mengetahui variabel apa yang berpengaruh signifikan terhadap variabel respon.

Hasil uji parsial parameter $\boldsymbol{\beta}$ menunjukkan bahwa dengan tingkat signifikan sebesar $\alpha = 0.05$ dan nilai $\chi^2_{(0,05;1)} = 3.841$ maka variabel yang berpengaruh signifikan terhadap jumlah kematian neonatal adalah jumlah bayi yang lahir dengan Berat Badan Lahir Rendah (BBRL). Sedangkan hasil dari uji parsial parameter $\boldsymbol{\gamma}$ menunjukkan bahwa dengan tingkat signifikan sebesar $\alpha = 0.05$ dan nilai $\chi^2_{(0.05;1)} = 3.841$ maka variabel yang berpengaruh signifikan terhadap jumlah kematian neonatal adalah jumlah bayi yang lahir dengan Berat Badan Lahir Rendah (BBRL).

Dengan tidak mengikutsertakan variabel prediktor yang tidak signifikan ke dalam model, maka didapatkan model sebagai berikut:

1. Model data diskrit untuk $\hat{\lambda}_i$
   $$\ln(\hat{\lambda}_i) = -1.17093 + 0.04583\boldsymbol{x}_{i1}$$
   Interpretasi model tersebut sebagai berikut:
   Setiap penambahan jumlah bayi yang lahir dengan Berat Badan Lahir Rendah (BBLR), maka akan meningkatkan rata-rata jumlah kasus kematian neonatal sebesar $\exp(0.04583) = 1.04689$ dari rata-rata jumlah kematian neonatal semula apabila variabel lain tidak dilibatkan dalam model.
2. Model *zero inflation* untuk $\hat{\pi}_i$
   $$\text{logit}(\hat{\pi}_i) = -30.4139 + 0.34375\boldsymbol{x}_{i1}$$
   Interpretasi model tersebut sebagai berikut:
   Setiap penambahan jumlah bayi yang lahir dengan Berat Badan Lahir Rendah (BBLR), maka akan meningkatkan peluang jumlah kasus kematian neonatal sebesar $\exp(0.34375) = 1.41022$ kali dari jumlah kematian neonatal semula apabila variabel lain tidak dilibatkan dalam model.

3.5 Pemilihan Model Terbaik

Untuk memilih model terbaik, nilai AIC untuk model Binomial Negatif dibandingkan dengan nilai AIC untuk model regresi ZINB. Berdasarkan sintaks RStudio, diperoleh nilai AIC seperti pada Tabel 5.

Tabel 5. Nilai AIC model regresi binomial negatif dan ZINB

Model Regresi	**AIC**
Binomial Negatif	123,2228
Zero Inflated Negative Binomial	119,6187

Sumber: *Data diolah, 2021*

Tabel 5 menunjukkan bahwa model regresi ZINB mempunyai nilai AIC yang lebih kecil daripada model regresi binomial negatif sehingga model regresi ZINB lebih baik daripada model regresi binomial negatif. Artinya, model regresi ZINB lebih baik dalam mengatasi overdispersi pada data jumlah kematian neonatal di Kota Makassar tahun 2018 dibandingkan dengan model regresi binomial negatif.

4. Kesimpulan

Model regresi *Zero Inflated Negative Binomial* menggunakan metode *Maximum Likelihood Estimation* (MLE) dengan Algoritma *Expectation Maximization* (EM) pada data jumlah kematian neonatal di Kota Makassar tahun 2018 adalah sebagai berikut:

$$\ln(\hat{\lambda}_i) = -1.17093 + 0.04583\boldsymbol{x_{i1}} - 0.00086\boldsymbol{x_{i2}} + 0.00087\boldsymbol{x_{i3}}$$
$$\text{logit}(\hat{\pi}_i) = -30.4139 + 0.34375\boldsymbol{x_{i1}} - 0.00987\boldsymbol{x_{i2}} + 0.01534\boldsymbol{x_{i3}}$$

Penelitian ini dapat dekembangkan dengan melakukan pemodelan dengan efek spasial dari tiap-tiap Kecamatan di Kota Makasssar seperti *Geographically Weighted Zero Inflated Negative Binomial* (GWZINB).

Daftar Pustaka

[1] Dinkes. *Profil Kesehatan Kota Makassar Tahun 2018*. Dinas Kesehatan Provinsi Kota Makassar, Makassar, 2019.

[2] Garay, A. M., Hashimoto, E. M., Ortega, E. M. M., & Lachos, V. H. On Estimation and Influence Diagnostics for Zero Inflated Negative Binomial Regression Model. *Computational Statistics and Data Analysis*, 55, 1304-1318, 2011.

[3] Hilbe, J. M. *Negative Binomial Regression*. Cambridge University Press, New York, 2011.

[4] Ilmi, F. M. *Pemodelan Kasus Malaria dan Filariasis di Jawa Timur Menggunakan Regresi Poisson Bivariat*. Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember, 2015.

[5] Kurniawan, I. *Model Regresi Poisson Terbaik Menggunakan Zero-Inflated Poisson (ZIP) dan Zero-Inflated Negative Binomial (ZINB).* Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang, 2017.

[6] McCullagh, P., & Nelder, J. A. *Generalized Linear Models. Second Edition*. Chapman and Hall, London, 1989.

[7] Myers, R. H., Montgomery, D. C., Vining, G. G., & Robinson, T. J. *Generalized Linear Models with Applications in Engineering and The Sciences. Second edition*. John Wiley and Sons, New Jersey, 2010.

[8] Simarmata, R. T., & Ispriyanti, D. Penanganan Overdispersi pada Model Regresi Poisson Menggunakan Model Regresi Binomial Negatif. *Media Statistika*, 4(2), 95-104, 2011.