

Penerapan Model Regresi *Zero-Inflated Poisson* pada Kasus Kematian Bayi di Kota Bandung Tahun 2020

Nanda Shofia Nur Ihsan*, Nusar Hajarisman

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Islam Bandung, Indonesia.

*shofianur353@gmail.com, nusarhajarisman@unisba.ac.id

Abstract. Poisson regression is used to analyze the relationship between predictor variables and response variables in the form of discrete data. In Poisson regression, one of the problems often faced is that the response has more than 50% zero data (zero inflation). So, the regression model to overcome the problem where the data contains too many zero values in discrete data and Poisson distribution is the Zero Inflated Poisson (ZIP) regression model. The purpose of this study is to obtain the application of the Zero-Inflated Poisson regression model to infant mortality cases in Bandung City in 2020. The data used in this study contains 80 observations based on UPT puskesmas in Bandung City with infant mortality data in Bandung city in 2020 and the factors that are thought to influence it. Parameter estimation of the ZIP regression model was carried out using the Maximum Likelihood Estimation (MLE) method and solved using the Expectation-Maximization (EM) algorithm, followed by testing the significance of the model using the likelihood ratio test and testing the significance of the parameters with the Wald test. From the results of hypothesis testing using a real level of 5%, the variable that affects infant mortality cases in the Poisson State model is unhealthy housing, while the variable that affects the chance of infant mortality cases being zero in the Zero State model is low birth weight.

Keywords: *Poisson, Zero Inflation, Zero-Inflated Poisson.*

Abstrak. Regresi Poisson digunakan untuk menganalisis hubungan antara variabel prediktor dengan variabel respon yang berupa data diskrit. Pada regresi poisson salah satu permasalahan yang sering dihadapi yaitu responnya mempunyai data nol yang lebih dari 50% (*zero inflation*). Sehingga, model regresi untuk mengatasi masalah di mana data mengandung terlalu banyak nilai nol pada data diskrit dan berdistribusi Poisson adalah model regresi *Zero Inflated Poisson* (ZIP). Tujuan dari penelitian ini yaitu memperoleh penerapan model regresi *Zero-Inflated Poisson* pada kasus kematian bayi di Kota Bandung tahun 2020. Data yang digunakan dalam penelitian ini berisi 80 pengamatan berdasarkan UPT puskesmas di Kota Bandung dengan data angka kematian bayi di kota Bandung pada tahun 2020 beserta faktor-faktor yang diduga mempengaruhinya. Penaksiran parameter model regresi ZIP dilakukan dengan metode *Maximum Likelihood Estimation* (MLE) dan diselesaikan menggunakan algoritma Ekspektasi-Maksimalisasi (EM), dilanjutkan dengan pengujian signifikansi model dengan menggunakan *likelihood ratio test* dan pengujian signifikansi parameter dengan uji Wald. Dari hasil pengujian hipotesis menggunakan taraf nyata sebesar 5% variabel yang mempengaruhi kasus kematian bayi pada model Poisson *State* yaitu rumah tidak sehat, sedangkan variabel yang mempengaruhi peluang kasus kematian bayi bernilai nol pada model *Zero State* yaitu berat badan bayi lahir rendah.

Kata Kunci: *Poisson, Zero Inflation, Zero-Inflated Poisson.*

A. Pendahuluan

Analisis regresi merupakan kajian terhadap hubungan antara variabel respon dengan variabel prediktor, dimana variabel prediktor merupakan variabel yang tidak dipengaruhi oleh variabel lain, sedangkan variabel respon merupakan variabel yang masih dipengaruhi oleh nilai variabel prediktor (5). Analisis regresi biasanya digunakan untuk menganalisis variabel respon pada data kontinu. Namun, dalam beberapa aplikasi, variabel respon yang akan dianalisis dapat berupa data diskrit dan berdistribusi Poisson. Model regresi yang digunakan untuk menjelaskan hubungan antara variabel prediktor dengan variabel respon yang berupa data diskrit dan distribusi Poisson adalah model regresi Poisson (13).

Dalam regresi Poisson, salah satu masalah yang sering terjadi yaitu responnya mempunyai data nol yang berlebih (*zero inflation*) (13). *Zero Inflation* adalah jumlah nilai nol yang merupakan lebih dari 50% dalam variabel respon (4). Data dengan nilai nol yang lebih dari 50% ini dapat menyebabkan kesimpulan yang tidak akurat. Regresi Poisson dapat digunakan untuk menganalisis data diskrit, tetapi masih belum dapat mengatasi masalah pada variabel respon yang memiliki nilai nol berlebih (*zero inflation*). Sehingga, model regresi untuk mengatasi masalah di mana data mengandung terlalu banyak nilai nol pada data diskrit adalah model regresi *Zero Inflated Poisson* (ZIP) (10).

Dalam penelitian ini membahas model *Zero Inflated Poisson* untuk masalah kesehatan, khususnya yang berkaitan dengan angka kematian bayi di Kota Bandung. Kematian atau mortalitas merupakan salah satu dari tiga faktor demografi yang dapat mempengaruhi jumlah dan komposisi umur penduduk. Salah satu angka kematian yang dikenal adalah kematian bayi (*infant mortality*) yang dapat didefinisikan sebagai kematian bayi dibawah usia satu tahun (3). Kematian bayi merupakan salah satu indikator utama untuk menilai tingkat kesehatan dan kesejahteraan masyarakat, karena jika angka kematian bayi di suatu daerah tinggi, artinya status kesehatan di daerah tersebut rendah (7). Kematian bayi di Kota Bandung telah mencapai 894 jiwa sejak tahun 2014 hingga 2020, dimana dari tahun ke tahun kasus kematian bayi di Kota Bandung terjadi penurunan (2). Meski kasus kematian bayi mengalami penurunan pada tahun 2020, namun angka kematian bayi di Kota Bandung masih cukup tinggi. Sehingga, diperlukan upaya penanggulangan secara komprehensif. Dengan mengetahui penyebab kematian bayi, maka dapat mengontrol faktor - faktor yang mempengaruhi kematian bayi dan pada akhirnya dapat digunakan untuk menentukan upaya penurunan angka kematian bayi di Indonesia khususnya di Kota Bandung.

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut: “Bagaimana penerapan model regresi *Zero-Inflated Poisson* (ZIP) pada kasus kematian bayi di Kota Bandung tahun 2020?”. Selanjutnya, tujuan dalam penelitian ini adalah untuk memperoleh penerapan model regresi *Zero-Inflated Poisson* pada kasus kematian bayi di Kota Bandung tahun 2020.

B. Metodologi Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang dipublikasi oleh Dinas Kesehatan Kota Bandung pada *website* resmi Open Data Jabar (<https://opendata.jabarpov.go.id/id>). Data yang diambil berisi 80 pengamatan berdasarkan UPT puskesmas di Kota Bandung. Variabel yang diamati dalam penelitian ini yaitu satu variabel respon (Angka Kematian Bayi) dan tiga variabel prediktor (Bayi diberi ASI Eksklusif, Berat Badan Bayi Lahir Rendah, Rumah Tidak Sehat)

Metode Analisis Data

Zero Inflation

Salah satu permasalahan pada regresi Poisson adalah kelebihan nilai nol pada variabel respon atau yang biasa dikenal dengan *Zero Inflation* (13). *Zero Inflation* terjadi ketika variabel respon dari suatu data memiliki nilai nol lebih dari 50% dibandingkan nilai data diskrit lainnya (4). Menurut Famoye dan Singh (4), besarnya persentase data yang bernilai nol dapat mempengaruhi pada ketepatan dari inferensia. Selain itu, regresi Poisson juga menjadi tidak tepat untuk memodelkan data yang sebenarnya.

Model Regresi Zero Inflated Poisson

Menurut Jansakul dan Hinde (9) model regresi ZIP adalah salah satu metode yang digunakan ketika ada lebih banyak pengamatan yang bernilai nol daripada data diskrit lainnya. Lambert (10) menjelaskan jika y_i adalah variabel random dengan distribusi ZIP, maka nilai nol pada pengamatan diduga muncul dengan dua cara yang tepat untuk memisahkan keadaan. Keadaan pertama disebut *zero state* (0) yang terjadi dengan peluang π dan hanya menghasilkan pengamatan bernilai nol, sedangkan keadaan kedua disebut *Poisson state* yang terjadi dengan peluang $1 - \pi$ dan memiliki distribusi Poisson dengan rata-rata μ_i (12).

Lambert (10) mendefinisikan fungsi masa peluang untuk model regresi ZIP sebagai:

$$P(Y_i = y_i) = \begin{cases} \pi_i + (1 - \pi_i)e^{-\mu_i} & ; \text{untuk } y_i = 0 \\ \frac{(1 - \pi_i)e^{-\mu_i}\mu_i^{y_i}}{y_i!} & ; \text{untuk } y_i > 0 \end{cases} \quad (1)$$

Lambert (10) menunjukkan model menggunakan link function untuk μ (*Poisson state*) dan π (*Zero state*), pada model regresi ZIP menggunakan fungsi hubung log dan fungsi hubung logit:

1. Untuk model *Poisson state*

Fungsi hubung *log* dimana $\eta_i = \log(\mu_i)$, sehingga (6):

$$\log(\mu_i) = \mathbf{X}\boldsymbol{\beta} \quad (2)$$

2. Untuk model *Zero state*

Fungsi hubung logit dimana $\eta_i = \text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right)$, sehingga (6):

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{X}\boldsymbol{\gamma} \quad (3)$$

di mana:

$\mathbf{X}$ = matriks kovariat yang berisi nilai-nilai variabel prediktor yang masing-masing mempengaruhi mean Poisson pada *Poisson state* dan peluang pada *zero state*

$\boldsymbol{\beta}$ = vektor dari parameter regresi yang diperkirakan sebagai variabel yang mempengaruhi keadaan Poisson

$\boldsymbol{\gamma}$ = vektor parameter regresi yang diperkirakan sebagai variabel yang mempengaruhi keadaan nol (*zero state*)

π = peluang pengamatan bernilai nol.

Dengan Mean dan variansi dari distribusi Zero Inflated Poisson, yaitu: $E(Y_i) = (1 - \pi_i)\mu_i$ dan $\text{Var}(Y_i) = (1 - \pi_i)\mu_i(1 + \mu_i\pi_i)$

Penaksiran Parameter Zero Inflated Poisson (ZIP)

Penaksiran parameter regresi ZIP dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Metode MLE merupakan salah satu metode penaksiran parameter yang dapat digunakan untuk mengestimasi parameter model dengan distribusi yang telah diketahui.

$$\mu_i = \exp(x_i^T \boldsymbol{\beta}), \pi_i = \frac{e^{x_i^T \boldsymbol{\gamma}}}{1 + e^{x_i^T \boldsymbol{\gamma}}}, \text{ dan } (1 - \pi_i) = \frac{1}{1 + e^{x_i^T \boldsymbol{\gamma}}} \quad (4)$$

Selanjutnya persamaan (4) disubstitusikan ke dalam persamaan (1) yaitu (10):

$$P(Y_i = y_i) = \begin{cases} \frac{e^{x_i^T \boldsymbol{\gamma}}}{1 + e^{x_i^T \boldsymbol{\gamma}}} + \frac{1}{1 + e^{x_i^T \boldsymbol{\gamma}}} \left(\exp(-e^{x_i^T \boldsymbol{\beta}}) \right), & \text{untuk } y_i = 0 \\ \frac{\frac{1}{1 + e^{x_i^T \boldsymbol{\gamma}}} \left(\exp(-e^{x_i^T \boldsymbol{\beta}}) \right) (e^{x_i^T \boldsymbol{\beta}})^{y_i}}{y_i!}, & \text{untuk } y_i \geq 0 \end{cases} \quad (5)$$

Dari persamaan (5), dengan mengalikan semua fungsi probabilitas dari Y_i sehingga fungsi *likelihood* ZIP diperoleh sebagai berikut:

$$L(\beta, \gamma | y_i) = \prod_{i=1}^n P(y_i; \beta, \gamma)$$

$$L(\beta, \gamma | y_i) = \begin{cases} \prod_{i=1}^n \left[\frac{e^{x_i^T \gamma} + \exp(-e^{x_i^T \beta})}{1 + e^{x_i^T \gamma}} \right], & \text{untuk } y_i = 0 \\ \prod_{i=1}^n \left[\frac{\exp(-e^{x_i^T \beta}) (e^{x_i^T \beta})^{y_i}}{(1 + e^{x_i^T \gamma}) y_i!} \right], & \text{untuk } y_i \geq 0 \end{cases} \quad (6)$$

Selanjutnya dari persamaan (6) dibuat persamaan *ln likelihood* maka diperoleh:
Untuk $y_i = 0$, fungsi $\ln L(\beta, \gamma | y_i)$ sebagai berikut:

$$\ln L(\beta, \gamma | y_i) = \sum_{i=1}^n \ln(e^{x_i^T \gamma} + \exp(-e^{x_i^T \beta})) - \sum_{i=1}^n \ln(1 + e^{x_i^T \gamma}) \quad (7)$$

Untuk $y_i > 0$, fungsi $\ln L(\beta, \gamma | y_i)$ sebagai berikut:

$$\ln L(\beta, \gamma | y_i) = \sum_{i=1}^n \left((e^{x_i^T \beta}) y_i - e^{x_i^T \beta} \right) - \sum_{i=1}^n \ln(1 + e^{x_i^T \gamma}) - \sum_{i=1}^n \ln y_i! \quad (8)$$

Penjumlahan fungsi *ln likelihood* dalam persamaan (7) dan (8) akan merumitkan perhitungan karena tidak diketahui nilai nol mana yang berasal dari keadaan nol (*zero state*) dan mana yang berasal dari keadaan Poisson, maka fungsi *ln likelihood* ini tidak dapat diselesaikan menggunakan metode numerik biasa.

Untuk memaksimalkan fungsi *ln likelihood* pada persamaan (7) dan (8) maka menggunakan algoritma EM (*Expectation Maximization*).

Misal variabel Y berkaitan dengan variabel indikator Z yaitu:

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

Di mana:

Z = variabel random diskrit dari suatu fungsi distribusi dengan informasi data yang tidak lengkap atau data yang hilang (*missing*).

Masalahnya adalah jika nilai variabel respon $y_i = 1, 2, 3, \dots$, maka nilai $z_i = 0$. Sebaliknya, jika nilai variabel respon adalah $y_i = 0$, maka z_i bisa jadi 0 atau mungkin 1. Oleh karena itu, nilai z_i dianggap hilang. Untuk mengatasi masalah ini, penaksiran parameter dilakukan dengan menggunakan algoritma EM. Langkah-langkahnya adalah sebagai berikut (11):

1. Menentukan distribusi dari variabel Z
 $P(z_i = 1) = \pi_i$ dan
 $P(z_i = 0) = P(y_i \sim \text{Poisson}(\mu_i)) = 1 - \pi_i$
 Sehingga $z_i \sim \text{Bernoulli}(1, \pi_i)$, $E(z_i) = \pi_i$ dan $\text{var}(z_i) = \pi_i(1 - \pi_i)$
2. Membentuk distribusi gabungan antara fungsi distribusi y_i (*Poisson*) dan fungsi distribusi z_i (*Bernoulli*) dengan diasumsikan y_i dan z_i saling bebas, maka:

$$f(y_i, z_i | \mu_i, \pi_i) = f(z_i) f(y_i | z_i) = f(z_i | 1, \pi_i) f(y_i | z_i, \mu_i)$$

$$= (1 - \pi_i)^{(1-z_i)} (\pi_i)^{z_i} \left(\frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!} \right)^{(1-z_i)} \quad (9)$$

Kemudian substitusikan persamaan (4) ke persamaan (9) hingga didapat persamaan *ln likelihood*:

$$\begin{aligned} \ln L(\beta, \gamma | y, z) = & \sum_{i=1}^n \left[z_i x_i^T \gamma - \ln(1 + \exp(x_i^T \gamma)) \right] \\ & + \sum_{i=1}^n (1 - z_i) (y_i x_i^T \beta - \exp(x_i^T \beta)) - \sum_{i=1}^n (1 - z_i) \ln y_i! \end{aligned} \quad (10)$$

Persamaan (10) akan dimaksimumkan dengan algoritma EM, dimana β dan γ dapat diduga secara terpisah, dengan menuliskan persamaan (10) menjadi:

$$\ln L(\beta, \gamma | y, z) = \ln L(\beta, y, z) + \ln L(\gamma, y, z) - \sum_{i=1}^n (1 - z_i) \ln y_i! \quad (11)$$

dengan

$$\ln L(\gamma, y, z) = \sum_{i=1}^n \left[z_i x_i^T \gamma - \ln(1 + \exp(x_i^T \gamma)) \right] \quad (12)$$

dan

$$\ln L(\beta, y, z) = \sum_{i=1}^n (1 - z_i) (y_i x_i^T \beta - \exp(x_i^T \beta)) \quad (13)$$

3. Tahap ekspektasi

Mengganti variabel z_i dengan $z_i^{(t)}$ yang merupakan ekspektasi dari z_i

$$z_i^{(t)} = E(z_i | y_i, \gamma^{(t)}, \beta^{(t)}) = P(z_i = 1 | y_i, \gamma^{(t)}, \beta^{(t)})$$

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

$$\text{Untuk } y_i > 0, z_i^{(t)} = 0$$

Sehingga persamaan (12) dan (13) menjadi:

$$\ln L(\gamma^{(t)}, y, z^{(t)}) = \sum_{i=1}^n \left[z_i^{(t)} x_i^T \gamma^{(t)} - \ln(1 + \exp(x_i^T \gamma^{(t)})) \right] \quad (14)$$

$$\ln L(\beta^{(t)}, y, z^{(t)}) = \sum_{i=1}^n \left(1 - z_i^{(t)} (y_i x_i^T \beta^{(t)} - \exp(x_i^T \beta^{(t)})) \right) \quad (15)$$

4. Tahap maksimalisasi

Memaksimumkan β dan γ pada persamaan (14) dan (15) dengan menghitung $\beta^{(t+1)}$ dan $\gamma^{(t+1)}$ dengan metode *Newton raphson*. Misalkan $\beta^{(t)}$ dan $\gamma^{(t)}$ adalah aproksimasi metode maksimum *likelihood* untuk mengestimasi $\hat{\beta}$ dan $\hat{\gamma}$. Dengan menggunakan metode *Newton raphson* maka:

$$\beta^{(t+1)} = \beta^{(t)} - (H^{(t)})^{-1} U^{(t)} \quad (16)$$

dan

$$\gamma^{(t+1)} = \gamma^{(t)} - (H^{(t)})^{-1} U^{(t)} \quad (17)$$

di mana H adalah turunan kedua dari $\ln L(\beta^{(t)}, y, z^{(t)})$ dan $\ln L(\gamma^{(t)}, y, z^{(t)})$, U adalah turunan pertama dari $\ln L(\beta^{(t)}, y, z^{(t)})$ dan $\ln L(\gamma^{(t)}, y, z^{(t)})$.

5. Tahap ke-3 dan ke-4 ini dilakukan berulang hingga diperoleh penaksir parameter yang konvergen yaitu pada saat $|\beta^{(t)} - \beta^{(t+1)}| \leq \varepsilon$ dan $|\gamma^{(t)} - \gamma^{(t+1)}| \leq \varepsilon$, di mana $\varepsilon = 10^{-6}$.

Pengujian Signifikansi Model

Untuk menguji signifikansi model menggunakan statistik uji G^2 yang merupakan

statistik uji rasio kemungkinan maksimum (*likelihood ratio test*) (1):

Hipotesis yang digunakan :

$H_0: \beta_1 = \dots = \beta_k = \gamma_1 = \dots = \gamma_k = 0$ (tidak ada variabel prediktor ke- j yang berpengaruh secara signifikan terhadap variabel respon)

H_1 : paling sedikit terdapat satu $\beta_j \neq 0$ dan $\gamma_j \neq 0$, dengan $j = 1, \dots, k$ (paling sedikit ada satu variabel prediktor yang berpengaruh secara signifikan terhadap variabel respon)

Statistik uji yang digunakan:

$$G^2 = -2 \ln \left(\frac{L_0}{L_1} \right) \sim \chi^2_{(k)} \quad (18)$$

di mana:

L_0 = *likelihood* model tanpa variabel prediktor (model intersep)

L_1 = *likelihood* model dengan variabel prediktor (model penuh)

Kriteria ujinya yaitu H_0 ditolak jika statistik uji $G^2 > \chi^2_{(\alpha; k)}$ dengan k adalah jumlah variabel prediktor.

Pengujian Signifikansi Parameter

Pengujian ini bertujuan untuk melihat signifikansi parameter koefisien regresi dari masing-masing parameter regresi *Zero Inflated Poisson* sebagai berikut:

1. Untuk Poisson State

$H_0: \beta_j = 0$ (variabel prediktor ke- j tidak berpengaruh secara signifikan terhadap variabel respon)

$H_1: \beta_j \neq 0$ (variabel prediktor ke- j berpengaruh secara signifikan terhadap variabel respon)

2. Untuk Zero State

$H_0: \gamma_j = 0$ (variabel prediktor ke- j tidak berpengaruh secara signifikan terhadap peluang variabel respon bernilai nol)

$H_1: \gamma_j \neq 0$ (variabel prediktor ke- j berpengaruh secara signifikan terhadap peluang variabel respon bernilai nol)

Statistik uji yang digunakan adalah statistik uji Wald (8):

$$W_j = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right) \text{ atau } W_j = \left(\frac{\hat{\gamma}_j}{SE(\hat{\gamma}_j)} \right) \quad (19)$$

dimana $\hat{\beta}_j$ dan $\hat{\gamma}_j$ adalah nilai taksiran parameter β_j dan γ_j , sedangkan $SE(\hat{\beta}_j)$ dan $SE(\hat{\gamma}_j)$ adalah taksiran *standard error* dari taksiran parameter β_j dan γ_j yang diperoleh dari matriks taksiran variansi-kovariansi dari $\hat{\beta}_j$ dan $\hat{\gamma}_j$ yang dinyatakan dengan $V(\hat{\beta})$ dan $V(\hat{\gamma})$ diperoleh dari minus invers dari matriks Hessian pada saat proses iterasinya sudah konvergen yaitu sebagai berikut:

$$V(\hat{\beta}) \approx -[H(\hat{\beta})]^{-1}$$

dan

$$V(\hat{\gamma}) \approx -[H(\hat{\gamma})]^{-1}$$

Elemen diagonal utama ke- j masing-masing matriks $V(\hat{\beta})$ dan $V(\hat{\gamma})$ menunjukkan taksiran variansi dari $\hat{\beta}_j$ dan $\hat{\gamma}_j$ yang dinyatakan dengan $Var(\hat{\beta}_j)$ dan $Var(\hat{\gamma}_j)$. Dari matriks $V(\hat{\beta})$ dan $V(\hat{\gamma})$ diperoleh taksiran standar error dari $\hat{\beta}_j$ dan $\hat{\gamma}_j$ yaitu:

$$SE(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)} \text{ atau } SE(\hat{\gamma}_j) = \sqrt{Var(\hat{\gamma}_j)} \quad (20)$$

Sehingga kriteria uji untuk statistik Wald yang mengikuti distribusi normal baku adalah tolak H_0 jika $|W_j| > z_{1-\frac{\alpha}{2}}$

C. Hasil Penelitian dan Pembahasan

Pemeriksaan Zero Inflation

Pemeriksaan *Zero Inflation* pada peubah respon dapat dilakukan dengan melihat jumlah amatan yang bernilai nol. Peubah respon dapat dikatakan *Zero Inflation* apabila persentase jumlah

amatan bernilai nol lebih dari 50%. Berdasarkan data angka kematian bayi di Kota Bandung tahun 2020 terdapat amatan keseluruhan berjumlah 80 dengan amatan bernilai nol berjumlah 59. Maka persentase angka nol adalah 73.75%, nilai tersebut lebih besar dari syarat terjadinya *zero inflation* yaitu 50%. Sehingga dapat disimpulkan bahwa terjadi *Zero Inflation* pada variabel respon yaitu pada data angka kematian bayi di Kota Bandung tahun 2020, maka dilakukan pemodelan dengan menggunakan model regresi *Zero Inflated Poisson*.

Estimasi Parameter Model Regresi ZIP

Nilai taksiran parameter dari regresi *Zero Inflated Poisson* yang hasilnya disajikan pada Tabel 1.

Tabel 1. Nilai Estimasi Parameter untuk Model Regresi ZIP

Parameter	Estimasi	Standard Error
β_0	-0.7481	0.6077
β_1	-0.00030	0.000881
β_2	0.007634	0.01290
β_3	0.000453	0.000186
γ_0	3.2010	1.3245
γ_1	0.004078	0.003959
γ_2	-0.1913	0.09562
γ_3	-0.00014	0.000698

Berdasarkan hasil perhitungan yang tersaji pada Tabel 1 dapat diperoleh nilai-nilai estimasi parameter untuk Poisson *State* dan *Zero State*. Dengan demikian model untuk Poisson *State* dan *Zero State* masing-masing adalah:

$$\log(\mu_i) = -0,7481 - 0,00030X_1 + 0,007634X_2 + 0,000453X_3$$

$$\text{logit}(\pi_i) = 3,2010 + 0,004078X_1 - 0,1913X_2 - 0,00014X_3$$

Pengujian Signifikansi Model (Uji Simultan)

Pengujian parameter secara simultan dilakukan dengan statistik uji G^2 . Hipotesis yang digunakan adalah:

$H_0: \beta_1 = \beta_2 = \beta_3 = \gamma_1 = \gamma_2 = \gamma_3 = 0$ (Tidak ada variabel prediktor ke- j pada model Poisson *State* dan *Zero State* yang berpengaruh secara signifikan terhadap variabel angka kematian bayi (Y))

H_1 : paling sedikit terdapat satu $\beta_j \neq 0$ atau $\gamma_j \neq 0$, dengan $j = 1, 2, 3$ (Paling sedikit ada satu variabel prediktor ke- j pada model Poisson *State* atau *Zero State* yang berpengaruh secara signifikan terhadap variabel angka kematian bayi (Y))

Berdasarkan *output software SAS* diperoleh nilai statistik uji G^2 sebesar 84,8 dan diperoleh nilai $\chi^2_{(0,05;3)}$ dari tabel Chi-Square sebesar 7,815. Dimana nilai $G^2 = 84,8 > \chi^2_{(0,05;3)} = 7,815$ maka dengan menggunakan $\alpha = 0,05$ H_0 ditolak. Artinya bahwa paling sedikit ada satu variabel prediktor ke- j pada model Poisson *State* atau *Zero State* yang berpengaruh secara signifikan terhadap variabel angka kematian bayi (Y).

Pengujian Signifikansi Parameter (Uji Parsial)

Pengujian signifikansi parameter dilakukan dengan statistik uji z (Wald) sebagai berikut:

1. Untuk model Poisson *State*:

$H_0 : \beta_j = 0$ dengan $j = 1, 2, 3$; (variabel prediktor ke- j tidak berpengaruh secara signifikan terhadap angka kematian bayi (Y)).

$H_1 : \beta_j \neq 0$ dengan $j = 1, 2, 3$; (variabel prediktor ke- j berpengaruh secara signifikan terhadap angka kematian bayi (Y)).

Dengan menggunakan $\alpha = 5\%$ diperoleh $Z_{1-\frac{\alpha}{2}} = Z_{0,975} = 1,96$, untuk statistik uji z (Wald) masing-masing variabel prediktor pada model Poisson *State* dan hasil ujinya

disajikan dalam tabel 2.

Tabel 2. Hasil Uji Wald Pada Model Poisson State

Parameter	$ Z_{hitung} $	$Z_{1-\frac{\alpha}{2}}$	Kriteria Uji	Kesimpulan
β_1	0.34	1,96	Terima H0	X1 tidak signifikan terhadap Y
β_2	0.59		Terima H0	X2 tidak signifikan terhadap Y
β_3	2.44		Tolak H0	X3 signifikan terhadap Y

Didapat kesimpulan untuk model Poisson State bahwa variabel rumah tidak sehat (X_3) berpengaruh secara signifikan terhadap variabel angka kematian bayi (Y). Sedangkan untuk variabel jumlah bayi diberi ASI eksklusif (X_1) dan jumlah berat badan bayi lahir rendah (X_2) tidak berpengaruh secara signifikan terhadap variabel angka kematian bayi (Y).

2. Untuk model Zero State adalah sebagai berikut:

$H_0 : \gamma_j = 0$ dengan $j = 1,2,3$; (variabel prediktor ke- j tidak berpengaruh secara signifikan terhadap peluang variabel angka kematian bayi (Y) yang bernilai nol).

$H_1 : \gamma_j \neq 0$ dengan $j = 1,2,3$; (variabel prediktor ke- j berpengaruh secara signifikan terhadap peluang variabel angka kematian bayi (Y) yang bernilai nol).

Dengan menggunakan $\alpha = 5\%$ diperoleh $Z_{1-\frac{\alpha}{2}} = Z_{0,975} = 1,96$, untuk statistik uji z (Wald) masing-masing variabel prediktor pada model Zero State dan hasil ujinya disajikan dalam tabel 3.

Tabel 3. Hasil Uji Wald Pada Model Zero State

Parameter	$ Z_{hitung} $	$Z_{1-\frac{\alpha}{2}}$	Kriteria Uji	Kesimpulan
γ_1	1.03	1,96	Terima H0	X1 tidak signifikan terhadap Y
γ_2	2.00		Tolak H0	X2 signifikan terhadap Y
γ_3	0.20		Terima H0	X3 tidak signifikan terhadap Y

Didapat kesimpulan untuk model Zero State bahwa variabel jumlah bayi diberi ASI eksklusif (X_1) dan jumlah rumah tidak sehat (X_3) tidak berpengaruh secara signifikan terhadap peluang variabel angka kematian bayi (Y) yang bernilai nol. Sedangkan untuk variabel berat badan bayi lahir rendah (X_2) berpengaruh secara signifikan terhadap peluang variabel angka kematian bayi (Y) yang bernilai nol atau yang tidak terdapat kasus kematian bayi.

Interpretasi Model

Diperoleh nilai-nilai estimasi parameter untuk model Poisson State dan model Zero State sebagai berikut:

$$\log(\mu_i) = -0,7481 - 0,00030X_1 + 0,007634X_2 + 0,000453X_3$$

$$\logit(\pi_i) = 3,2010 + 0,004078X_1 - 0,1913X_2 - 0,00014X_3$$

Berdasarkan pengujian signifikansi parameter model regresi Zero Inflated Poisson maka hanya parameter yang signifikan saja yang dapat diinterpretasi sebagai berikut:

Dimana X_2 menyatakan berat badan bayi lahir rendah dan X_3 menyatakan jumlah rumah tidak sehat dari setiap 80 puskesmas di Kota Bandung tahun 2020.

Model $\log(\mu_i)$ atau Poisson State untuk β_3 sebesar 0,000453 menyatakan bahwa setiap kenaikan jumlah rumah tidak sehat (X_3) sebanyak 1 rumah tinggal dengan asumsi nilai variabel lain dianggap konstan, maka rata-rata kematian bayi (Y) cenderung bertambah sebesar $\mu_i = \exp(0,000453) = 1,0004 \approx 1$ kasus kematian bayi per 1000 peristiwa kelahiran hidup di Kota Bandung pada tahun 2020.

Sedangkan pada model logit atau Zero State untuk γ_2 sebesar $-0,1913$ menyatakan bahwa apabila suatu puskesmas terdapat penambahan 1 bayi yang mengalami berat badan bayi lahir rendah (X_2) dengan berat di bawah 2500gram dan asumsi nilai variabel lain dianggap konstan, maka kemungkinan tidak terjadinya kasus kematian bayi per 1000 peristiwa kelahiran

hidup di Kota Bandung pada tahun 2020 berkurang sebesar $\exp(-0,1913) = 0,8259$ kali lipat dibandingkan dengan puskesmas lainnya. Kondisi pada hasil penelitian ini menyatakan bahwa jika terjadi peningkatan 1 bayi yang mengalami berat badan lahir rendah, maka kemungkinan bayi yang tidak mengalami kematian akan menurun atau kemungkinan bayi yang mengalami kematian akan meningkat. Sehingga hasil ini sudah sesuai dengan teori yang menyatakan bahwa jika terjadi peningkatan bayi dengan berat badan lahir rendah, maka dapat menyebabkan penambahan kasus kematian pada bayi per 1000 peristiwa kelahiran hidup.

D. Kesimpulan

Berdasarkan penerapan analisis dengan model regresi *Zero Inflated Poisson* (ZIP) yang sudah dilakukan terhadap data kematian bayi di Kota Bandung pada tahun 2020, diperoleh model regresi ZIP sebagai berikut:

$$\log(\mu_i) = -0,7481 - 0,00030X_1 + 0,007634X_2 + 0,000453X_3$$

$$\text{logit}(\pi_i) = 3,2010 + 0,004078X_1 - 0,1913X_2 - 0,00014X_3$$

Dimana X_2 menyatakan berat badan bayi lahir rendah dan X_3 menyatakan jumlah rumah tidak sehat dari setiap 80 puskesmas di Kota Bandung tahun 2020. Pada model $\log(\mu_i)$ atau *Poisson State* faktor yang mempengaruhi angka kematian bayi di Kota Bandung tahun 2020 yaitu hanya faktor rumah tidak sehat, sehingga nilai estimasi sebesar 0,000453 menyatakan bahwa setiap kenaikan jumlah rumah tidak sehat (X_3) sebanyak 1 rumah tinggal dengan asumsi nilai variabel lain dianggap konstan, maka rata-rata kematian bayi (Y) cenderung bertambah sebesar $\exp(0,000453) = 1,0004 \approx 1$ kasus kematian bayi per 1000 peristiwa kelahiran hidup di Kota Bandung pada tahun 2020. Sedangkan pada model logit atau *Zero State* faktor yang mempengaruhi peluang angka kematian bayi bernilai nol di Kota Bandung tahun 2020 yaitu hanya faktor berat badan bayi lahir rendah, sehingga nilai estimasi sebesar $-0,1913$ menyatakan apabila suatu puskesmas terdapat penambahan 1 bayi yang mengalami berat badan bayi lahir rendah (X_2) dengan asumsi nilai variabel lain dianggap konstan, maka kemungkinan tidak terjadinya kasus kematian bayi per 1000 peristiwa kelahiran hidup di Kota Bandung pada tahun 2020 berkurang sebesar $\exp(-0,1913) = 0,8259$ kali lipat dibandingkan dengan puskesmas lainnya.

Acknowledge

With my gratitude to Allah SWT for grace and guidance. Alhamdulillah, the author can complete this research on time. The author dedicates this research to two great people in life, namely Mom and Dad, thank you for always praying endlessly. Thank you for myself for struggling, and do not forget to also thank the author for the supervisor, boyfriend and all friends involved who always provide help, prayers and support to the author. The author is very grateful to have all of you in this life.

Daftar Pustaka

- [1] Dinas Kesehatan Kota Bandung. 2020. *Profil Kesehatan Kota Bandung 2020*. Bandung: Dinas Kesehatan Kota Bandung.
- [2] Dinas Kesehatan Provinsi Jawa Barat. 2020. *Profil Kesehatan Jawa Barat Tahun 2020*. Dinas Kesehatan Provinsi Jawa Barat.
- [3] Famoye, F. dan Singh, K.P. 2006. Zero Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data. *Journal of Data Science*. 4: 117-130.
- [4] Gujarati, Damodar, 2003, *Ekonometri Dasar*. Terjemahan: Sumarno Zain, Jakarta: Erlangga
- [5] Hajarisman, N. (2009). *Analisis Data Kategorik*. Bandung: Program Studi Statistika, Universitas Islam Bandung.
- [6] Herwandi., dan Amrie, M. A. 2021. *Faktor-Faktor Yang Mempengaruhi Angka Kematian Bayi Di Kabupaten Bulungan*. Jurnal Ekonomi Pembangunan Wilayah
- [7] Hosmer, D. W., & Lemeshow, S. 2000. *Applied Logistic Regression*. New York: John Wiley and Sons.

- [8] Jansakul, N & J.P. Hinde. 2002. *Score Tests for Zero-Inflated Poisson Models*. Computational Statistics & Data Analysis
- [9] Lambert, D. 1992. Zero Inflated Poisson Regression, with an Application to Detect in Manufacturing. *Technometrics*. Vol.34, No.1, 1-14.
- [10] Taufan, M., Suparti dan Rusgiyono, A. 2012. Analisis Faktor-faktor yang Mempengaruhi Banyaknya Klaim Asuransi Kendaraan Bermotor Menggunakan Model Regresi Zero-Inflated Poisson. *Media Statistika*.
- [11] Tawiah, K., Iddi, S., & Lotsi, A. (2020). *On zero-inflated hierarchical Poisson models with application to maternal mortality data*. *International Journal of Mathematics and Mathematical Sciences*, 2020, 1-8.
- [12] Winkelman, R. 2008. *Econometric Analysis of Count Data*. 5th Edition. Springer, Berlin.
- [13] Muhammad Rizq Nafisyah Alam, and Aceng Komarudin Mutaqin. 2023. "Pemodelan Distribusi Poisson-Sujatha Pada Data Frekuensi Klaim Asuransi Kendaraan Bermotor Di Indonesia." *Jurnal Riset Statistika*: 71–78.
<https://journals.unisba.ac.id/index.php/JRS/article/view/1944>.