

PERBANDINGAN REGRESI ZERO INFLATED POISSON (ZIP) DAN REGRESI ZERO INFLATED NEGATIVE BINOMIAL (ZINB) PADA DATA OVERDISPERSI

Samson, Setyo Wira Rizki, Hendra Perdana

INTISARI

Analisis regresi Poisson adalah regresi nonlinear yang biasanya digunakan untuk data diskrit dan memiliki asumsi ekuidispersi. Pada praktiknya sering terjadi pelanggaran terhadap asumsi ekuidispersi, salah satu dari pelanggaran tersebut adalah overdispersi (nilai ragam lebih besar dari nilai rata-rata). Salah satu penyebab terjadinya overdispersi adalah jumlah nilai nol yang berlebih (*Excess Zero*) pada variabel respon. *Excess zeros* dapat dilihat pada proporsi variabel respon yang bernilai nol lebih besar dari data diskrit lainnya. Terdapat banyak metode untuk mengatasi overdispersi, dua diantaranya adalah regresi Zero Inflated Poisson (ZIP) dan regresi Zero Inflated Negative Binomial (ZINB). Tujuan pada penelitian ini adalah untuk mengetahui model regresi yang lebih baik digunakan pada data yang mengalami overdispersi. Data yang digunakan dalam menganalisis regresi ZIP dan ZINB adalah data hipertensi dalam kehamilan yang menyebabkan kematian ibu di Provinsi Kalimantan Barat tahun 2014 sampai 2019. Berdasarkan hasil penelitian, diketahui bahwa nilai Akaike Information Criterion (AIC) pada regresi ZIP 273.011 lebih kecil dari nilai AIC regresi ZINB 275.01. Sehingga regresi ZIP lebih baik digunakan serta faktor yang menyebabkan hipertensi dalam kehamilan adalah persentase ibu hamil melaksanakan program KI.

Kata kunci : Overdispersi, Regresi Poisson, Regresi ZIP, Regresi ZINB.

PENDAHULUAN

Suatu metode analisis statistik yang digunakan untuk melihat pengaruh antara dua atau lebih banyak variabel adalah analisis regresi. Hubungan variabel tersebut bersifat fungsional yang diwujudkan dalam suatu model matematis. Pada analisis regresi, variabel dibedakan menjadi dua bagian, yaitu variabel respon (*response variable*) dan variabel bebas (*independent variable*). Analisis regresi umumnya digunakan untuk menganalisis data variabel respon yang berupa data kontinu [1]. Namun dalam beberapa aplikasinya, variabel respon yang akan dianalisa dapat berupa data diskrit. Salah satu contoh dimana variabel responnya diskrit adalah banyaknya kejadian yang jarang terjadi (*rare event*). Misalkan meningkatnya angka kecelakaan lalu lintas setiap tahun, atau meningkatnya angka kematian balita atau ibu setiap tahunnya. Salah satu model regresi yang dapat digunakan untuk menjelaskan hubungan antara variabel respon Y yang berupa data diskrit dengan variabel bebas X berupa data diskrit, kontinu, kategorik atau campuran adalah model regresi Poisson.

Analisis regresi Poisson digunakan untuk menganalisis hubungan antara sebuah variabel dependen yang menyatakan data terhitung atau data *count* berdistribusi Poisson dengan satu atau lebih variabel independen. Data *count* yang dimaksud misalnya adalah banyaknya kejadian dalam interval waktu, ruang, atau volume tertentu. Dalam analisis regresi Poisson, variabel independen yang digunakan dapat berjenis kontinu. Salah satu asumsi yang harus dipenuhi dalam model regresi Poisson adalah nilai ekspektasi (*mean*) dan variansinya sama (*equidispersi*) yaitu $E(y) = Var(y) = \mu$ namun dalam analisis data diskrit dengan menggunakan model regresi Poisson terkadang terjadi pelanggaran asumsi tersebut, dimana nilai variansinya lebih besar dari nilai *mean* (overdispersi) atau varian lebih kecil dari nilai *mean* (*underdispersi*). Untuk menangani masalah

overdispersi, diperlukan model-model statistik diantaranya adalah model regresi *Zero-Inflated Poisson* (ZIP) dan model regresi *Zero-Inflated Negative Binomial* (ZINB).

Pada penelitian ini, kasus yang digunakan adalah hipertensi dalam kehamilan yang menyebabkan terjadinya kematian ibu. Kematian ibu terbesar di Indonesia masih didominasi oleh lima penyebab utama kematian yaitu perdarahan, hipertensi dalam kehamilan, infeksi, partus macet dan abortus. Proporsi perdarahan dan infeksi cenderung mengalami penurunan, sedangkan hipertensi dalam kehamilan proporsinya semakin meningkat. Berdasarkan data Dinas Kesehatan (Dinkes) Provinsi Kalimantan Barat, dari tahun 2014 sampai 2019, Angka Kematian Ibu (AKI) di Kalimantan Barat masih terus mengalami kenaikan. Sejak tahun 2013 lebih dari 25% kematian ibu di Indonesia disebabkan oleh hipertensi dalam kehamilan. Hipertensi dalam kehamilan merupakan 5-10% penyulit dalam kehamilan dan merupakan salah satu dari tiga penyebab tertinggi mortalitas dan morbiditas ibu bersalin.

Berdasarkan latar belakang yang telah diuraikan, maka judul yang dikaji dalam penelitian ini adalah “Perbandingan Regresi *Zero Inflated Poisson* (ZIP) Dan Regresi *Zero Inflated Negative Binomial* (ZINB) Pada Data Overdispersi”. Tujuan pada penelitian ini adalah untuk menentukan model regresi ZIP dan model regresi ZINB serta untuk mengetahui model regresi yang lebih baik dalam kasus hipertensi kehamilan di Provinsi Kalimantan Barat.

OVERDISPERSI

Overdispersi pada regresi Poisson dapat dilihat dari nilai taksiran dispersi yaitu nilai devians dan pearson *chi-square* dibagi dengan derajat bebasnya bernilai lebih besar dari satu. Fenomena overdispersi dinyatakan dengan[2]:

$$\text{Var}(Y) > E(Y)$$

Overdispersi dalam regresi Poisson dapat diketahui dengan membandingkan nilai mean dan varians yang terbentuk dari variabel respon, jika nilai varians lebih lebih besar dari pada nilai meannya maka model regresi Poisson tersebut memiliki masalah overdispersi.

Hubungan parameter dispersi (ϕ) dengan varians dan mean dalam regresi Poisson adalah:

$$\text{Var}(Y) = \phi \mu$$

dimana

$$\phi = \frac{\text{Nilai Deviance}}{\text{Derajat Bebasnya}}$$

Excess Zeros

Masalah lain yang sering terjadi pada data berdistribusi poisson adalah data yang banyak mengandung nilai nol (hal ini dikenal dengan *excess zero*). *Excess zeros* adalah salah satu penyebab terjadinya overdispersi. *Excess zeros* dapat dilihat pada proporsi variabel respon yang bernilai nol lebih besar dari data diskrit lainnya. Selain itu regresi *Poisson* juga menjadi tidak tepat lagi menggambarkan data yang sebenarnya. Model regresi yang dapat digunakan untuk mengatasi overdispersi yang disebabkan oleh *excess zeros* diantaranya adalah model regresi ZIP dan ZINB. *Excess zeros* dapat dilihat pada proporsi variabel respon yang bernilai nol lebih besar dari data diskrit lainnya.

Regresi *Zero-Inflated Poisson*

Salah satu metode yang diusulkan untuk data *excess zero* dan mengalami overdispersi yaitu model regresi *Zero Inflated Poisson* (ZIP). Model regresi ZIP adalah suatu metode yang digunakan untuk menganalisis lebih banyak observasi bernilai nol daripada yang ditaksir [3]. Jika y_i adalah

variabel random independen yang mempunyai distribusi ZIP, nilai nol pada observasi diduga muncul dalam dua cara yang sesuai untuk keadaan (*state*) yang terpisah [4]. Keadaan pertama disebut *zero state* terjadi dengan probabilitas ω_i dan menghasilkan hanya observasi bernilai nol. Keadaan kedua disebut *Poisson State* terjadi dengan probabilitas $(1 - \omega_i)$ dan berdistribusi Poisson dengan mean μ_i . Proses dua keadaan ini memberikan distribusi campuran dua komponen dengan fungsi probabilitas sebagai berikut [3]:

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

yang dinotasikan dengan $Y_i \sim ZIP(\mu, \omega)$

Hubungan model untuk μ dan ω adalah sebagai berikut:

$$\log(\mu) = X\beta \text{ dan } \text{logit}(\omega) = \log\left(\frac{\omega}{1-\omega}\right) = X\gamma \quad (2)$$

X adalah matriks variabel yang memuat himpunan-himpunan yang berbeda dari faktor eksperimen yang berhubungan dengan peluang pada *zero state* dan *mean Poisson* pada *Poisson state*, sedangkan β dan γ adalah parameter regresi yang akan ditaksir.

Estimasi parameter regresi ZIP menggunakan metode *Maksimum Likelihood Estimation* (MLE) [5]. Metode ini biasanya digunakan untuk menaksir parameter suatu model yang diketahui fungsi densitasnya. Fungsi *log-likelihood* untuk model regresi ZIP:

$$\ln L(\beta, \gamma | y_i) = \begin{cases} \sum_{i=1}^n \ln(e^{x_i^T \gamma} + \exp(-e^{x_i^T \beta})) - \sum_{i=1}^n \ln(1 + e^{x_i^T \gamma}), & \text{untuk } y_i = 0 \\ \sum_{i=1}^n (x_i^T \beta)y_i - e^{x_i^T \beta} - \sum_{i=1}^n \ln(1 + e^{x_i^T \gamma}) - \sum_{i=1}^n \ln y_i!, & \text{untuk } y_i > 0 \end{cases} \quad (3)$$

Penjumlahan fungsi *log-likelihood* pada Persamaan (3) akan membuat sulit perhitungan karena tidak diketahui nilai nol mana yang berasal dari *zero state* dan mana yang berasal dari *Poisson state*, sehingga fungsi *likelihood* ini tidak dapat diselesaikan dengan metode numerik biasa. Dalam memaksimalkan fungsi *log-likelihood* digunakan algoritma EM (*Expectation Maximization*). Algoritma EM merupakan salah satu metode yang digunakan untuk menemukan estimasi suatu parameter melalui kerangka metode MLE dari suatu fungsi distribusi yang dengan informasi data yang tidak lengkap atau data hilang (*missing*).

Misalkan variabel Y berkaitan dengan variabel indikator Z yaitu :

$$Z = \begin{cases} 1, & \text{jika } y_i \text{ berasal dari } \textit{zero state} \\ 0, & \text{jika } y_i \text{ berasal dari } \textit{Poisson state} \end{cases} \quad (4)$$

Permasalahannya adalah jika nilai variabel respon $y_i = 1, 2, 3, \dots$, maka nilai $z_i = 0$ sedangkan jika nilai variabel respon $y_i = 0$, maka nilai z_i mungkin 0 mungkin 1. Oleh karena itu, nilai z_i dianggap hilang. Untuk mengatasi hal ini dilakukan estimasi parameter dengan algoritma EM.

Regresi *Zero-Inflated Negative Binomial*

Selain model regresi ZIP, model regresi yang digunakan untuk mengatasi overdispersi yang disebabkan oleh *excess zeros* adalah regresi ZINB. Regresi ZINB merupakan model yang dibentuk dari distribusi campuran *Poisson gamma* [6]. Distribusi campuran *Poisson gamma* terbentuk jika suatu distribusi Poisson (μ) dengan μ merupakan nilai variabel random yang berdistribusi gamma maka akan dihasilkan distribusi campuran *Poisson gamma* yang dinamakan distribusi binomial negatif. Fungsi kepadatan peluangnya adalah [2]:

$$f(y | \alpha, \beta) = \frac{\Gamma(y+\alpha)}{y!\Gamma(\alpha)} \left(\frac{1}{1+\beta}\right)^\alpha \left(1 - \frac{1}{1+\beta}\right)^y \quad y = 0, 1, 2, \dots \quad (5)$$

dengan rata-rata dan variansi distribusi binomial negatif yaitu:

$$E[Y] = \alpha\beta \text{ dan } V[Y] = \alpha\beta + \alpha\beta^2$$

Untuk membentuk suatu model regresi pada distribusi binomial negatif, maka nilai parameter dari distribusi campuran *Poisson gamma* dinyatakan dalam bentuk $\mu = \alpha\beta$ dan $k = 1/\alpha$ sehingga diperoleh rata-rata dan variansi dalam bentuk [7]:

$$E(Y) = \mu \text{ dan } V(Y) = \mu + k\mu^2$$

Regresi ZINB dengan keadaan pertama disebut *zero state* terjadi dengan probabilitas p_i dan menghasilkan hanya observasi bernilai nol. Keadaan kedua disebut *Negative Binomial State* terjadi dengan probabilitas $(1 - p_i)$ dan berdistribusi Binomial Negatif dengan mean μ dengan $0 \leq p_i \leq 1$ [7]. Proses dua keadaan ini dengan variabel Y_i memberikan distribusi campuran dua komponen dan didapat fungsi probabilitas sebagai berikut:

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i) \left(\frac{1}{1+k\mu_i}\right)^{1/k}, & \text{untuk } y_i = 0 \\ (1 - p_i) \frac{\Gamma(y_i+1/k)}{\Gamma(1/k)\Gamma(y_i+1)} \left(\frac{1}{1+k\mu_i}\right)^{1/k} \left(\frac{k\mu_i}{1+k\mu_i}\right)^{y_i}, & \text{untuk } y_i = 1, 2, \dots \end{cases} \quad (6)$$

Estimasi parameter model regresi ZINB menggunakan metode MLE dengan prosedur Algoritma EM dan Newton Rhapsion [8]. Metode ini biasanya digunakan untuk menduga parameter suatu model yang diketahui fungsi densitasnya. Fungsi *log-likelihood* untuk model regresi ZINB:

$$\ln L(\theta | y_i) = \begin{cases} \sum_{i=1}^n \ln \left\{ e^{x_i^T \gamma} + \left(\frac{1}{1+ke^{x_i^T \gamma}}\right)^{\frac{1}{k}} \right\} - \sum_{i=1}^n \ln [1 + e^{z_i^T Y}], & \text{untuk } y_i = 0 \\ - \sum_{i=1}^n \ln [1 + e^{z_i^T \gamma}] + \sum_{i=1}^n \ln \left[\Gamma\left(\frac{1}{k} + y_i\right) \right] - \sum_{i=1}^n \ln [\Gamma(y_i + 1)] - \\ \sum_{i=1}^n \ln \left[\Gamma\left(\frac{1}{k}\right) \right] + y_i \sum_{i=1}^n \ln \left(\frac{e^{x_i^T \gamma}}{1+e^{x_i^T \gamma}} \right)^{y_i} + \frac{1}{k \sum_{i=1}^n \ln \left(\frac{1}{1+ke^{x_i^T \gamma}} \right)^{\frac{1}{k}}} \end{cases} \quad (7)$$

, untuk $y_i = 1, 2, \dots$

Penjumlahan fungsi *log-likelihood* pada Persamaan (7) tidak linear, sehingga fungsi *log-likelihood* ini tidak dapat diselesaikan dengan metode numerik biasa.

Misalkan variabel $y_i (i = 1, 2, \dots, n)$ berkaitan dengan vektor variabel indikator $W = (w_1, \dots, w_n)^T$ yaitu:

$$w_i = \begin{cases} 1, & \text{jika } y_i \text{ berasal dari } zero \text{ state} \\ 0, & \text{jika } y_i \text{ berasal dari } Negative \text{ Binomial state} \end{cases} \quad (8)$$

dengan $i = 1, 2, \dots, n$ jika nilai variabel respon $y_i = 1, 2, \dots$ maka nilai $w_i = 0$. Sedangkan jika nilai variabel respon $y_i = 0$ maka nilai w_i mungkin 0 mungkin 1. Oleh karena itu, nilai w_i dianggap hilang. Peluang dari w_i dapat dinyatakan [8]:

$$\begin{aligned} P(w_i = 1) &= p_i \\ P(w_i = 0) &= 1 - p_i \end{aligned}$$

dengan $i = 1, 2, \dots, n$. Sehingga distribusi dari variabel W adalah $w_i \sim \text{Binomial}(n, p_i)$ mempunyai rata-rata dan variansi $E(w_i) = np_i$ dan $V(w_i) = np_i(1 - p_i)$.

PEMBAHASAN

Dalam penelitian ini, sumber data yang digunakan adalah data sekunder dari Dinas Kesehatan Provinsi Kalimantan Barat. Data tersebut adalah data hipertensi dalam kehamilan yang menyebabkan kematian ibu tahun 2014 sampai 2019 dari tiap 14 Kabupaten/Kota di Provinsi Kalimantan Barat. Variabel-variabel yang digunakan dalam penelitian ini adalah:

1. Hipertensi dalam kehamilan (Y),
2. Persentase ibu hamil melaksanakan program K1 (X_1),
3. Persentase ibu hamil melaksanakan program K4 (X_2),
4. Persentase ibu nifas mendapatkan Vitamin A (X_3),
5. Persentase persalinan ditolong tenaga Kesehatan (X_4), dan
6. Persentase ibu hamil mendapatkan Fe (X_5).

Berikut adalah statistik deskriptif data hipertensi dalam kehamilan yang menyebabkan kematian ibu:

Tabel 1 Statistik Deskriptif

| Variabel | N | Minimal | Maksimal | Rata-rata | Varians | Std. Deviasi |
|----------|----|---------|----------|-----------|---------|--------------|
| Y | 84 | 0 | 6 | 1.524 | 1.915 | 1.384 |
| X_1 | 84 | 79 | 105 | 92.99 | 34.349 | 5.861 |
| X_2 | 84 | 56 | 98 | 82.9 | 104.497 | 10.222 |
| X_3 | 84 | 66 | 192 | 83.82 | 220.703 | 14.856 |
| X_4 | 84 | 60 | 99 | 81.13 | 94.067 | 9.699 |
| X_5 | 84 | 48 | 98 | 81.81 | 116.469 | 10.792 |

Setelah diperoleh statistik deskriptif dari variabel responden (Y) dan variabel bebas (X_1, X_2, X_3, X_4, X_5) tersebut, maka variabel Y akan ditentukan nilai varians dan rataannya untuk mengetahui data tersebut mengalami overdispersi yaitu nilai nilai varians lebih besar dari nilai rataannya. Pada tabel 1, dapat dilihat dari variabel respon bahwa nilai varians 1.915 dan nilai rataannya 1.524, ini menunjukkan data mengalami overdispersi.

Model regresi ZIP bertujuan untuk memperbaiki model regresi poisson yang mengalami overdispersi dan memiliki banyak nilai nol pada variabel respon. Adapun pendugaan parameter dapat dilakukan dengan bantuan paket *Software RStudio* sehingga didapatkan hasil sebagai berikut:

Tabel 2 Estimasi Parameter Model Regresi ZIP

| Parameter | Estimasi | SE | <i>P-value</i> | Signifikansi(5%) |
|------------|----------|----------|----------------|------------------|
| β_0 | -2.55107 | 2.11570 | 0.228 | Tidak Signifikan |
| β_1 | 0.07592 | 0.03782 | 0.045 | Signifikan |
| β_2 | -0.04133 | 0.02559 | 0.106 | Tidak Signifikan |
| β_3 | 0.01203 | 0.01834 | 0.512 | Tidak Signifikan |
| β_4 | -0.03733 | 0.02205 | 0.091 | Tidak Signifikan |
| β_5 | 0.01843 | 0.03096 | 0.552 | Tidak Signifikan |
| γ_0 | 24.76861 | 15.39693 | 0.108 | Tidak Signifikan |
| γ_1 | -0.06194 | 0.35005 | 0.860 | Tidak Signifikan |
| γ_2 | -0.17765 | 0.20595 | 0.388 | Tidak Signifikan |
| γ_3 | 0.12278 | 0.17109 | 0.473 | Tidak Signifikan |
| γ_4 | -0.39494 | 0.35224 | 0.262 | Tidak Signifikan |
| γ_5 | 0.15852 | 0.12824 | 0.216 | Tidak Signifikan |

Berdasarkan hasil pengolahan data pada Tabel 2 didapatkan keseluruhan model ZIP untuk model *log* dan *logit* sebagai berikut:

$$\text{Log}(\mu_i) = -2.55107 + 0.07592X_1 - 0.04133X_2 + 0.01203X_3 - 0.03733X_4 + 0.01843X_5$$

dan

$$\text{Logit}(p_i) = 24.76861 - 0.06194X_1 - 0.17765X_2 + 0.12278X_3 - 0.39494X_4 + 0.15852X_5.$$

Pada tabel 2 menunjukkan bahwa faktor yang memberikan pengaruh terhadap hipertensi kehamilan yang menyebabkan kematian ibu di provinsi Kalimantan Barat adalah persentase ibu hamil melaksanakan program K1. Interpretasi yang diperoleh dari nilai odd rasio pada model *log* adalah untuk setiap peningkatan 1% ibu yang melaksanakan kunjungan kehamilan program K1 akan menambah rata-rata hipertensi kehamilan sebesar 7,89%. Interpretasi yang diperoleh dari nilai odd rasio pada model *logit* adalah untuk setiap peningkatan 1% ibu yang melaksanakan kunjungan kehamilan akan mengurangi rata-rata jumlah kematian ibu sebesar -6,01%.

Analisis regresi ZINB digunakan sebagai pembandingan dari regresi ZIP dengan tujuan untuk mengetahui model regresi mana yang baik digunakan dalam menganalisa data hipertensi kehamilan yang menyebabkan kematian ibu. Model Regresi ZINB adalah model regresi yang mengakomodasi *excess zero* dalam data.

Berdasarkan hasil pengolahan data pada Tabel 3 didapatkan keseluruhan model ZINB untuk model *log* dan *logit* sebagai berikut:

$$\text{Log}(\mu_i) = -2.55071 + 0.07591X_1 - 0.04133X_2 + 0.01203X_3 - 0.03733X_4 + 0.01844X_5$$

dan

$$\text{Logit}(p_i) = 24.77251 - 0.06207X_1 - 0.17758X_2 + 0.12276X_3 - 0.39486X_4 + 0.15850X_5$$

Pada tabel 3 menunjukkan bahwa faktor yang memberikan pengaruh terhadap hipertensi kehamilan yang menyebabkan kematian ibu di provinsi Kalimantan Barat adalah persentase ibu hamil melaksanakan program K1. Interpretasi yang diperoleh dari nilai odd rasio pada model *log* adalah untuk setiap peningkatan 1% ibu yang melaksanakan kunjungan kehamilan program K1 akan menambah rata-rata hipertensi kehamilan sebesar 7,89%. Interpretasi yang diperoleh dari nilai odd rasio pada model *logit* adalah untuk setiap peningkatan 1% ibu yang melaksanakan kunjungan kehamilan akan mengurangi rata-rata jumlah kematian ibu sebesar -6,02%.

Tabel 3 Estimasi Parameter Model Regresi ZINB

| Parameter | Estimasi | SE | <i>P-value</i> | Signifikansi(5%) |
|------------|----------|----------|----------------|------------------|
| β_0 | -2.55071 | 2.11592 | 0.228 | Tidak Signifikan |
| β_1 | 0.07591 | 0.03782 | 0.045 | Signifikan |
| β_2 | -0.04133 | 0.02559 | 0.106 | Tidak Signifikan |
| β_3 | 0.01203 | 0.01834 | 0.512 | Tidak Signifikan |
| β_4 | -0.03733 | 0.02205 | 0.091 | Tidak Signifikan |
| β_5 | 0.01844 | 0.03096 | 0.552 | Tidak Signifikan |
| γ_0 | 24.77251 | 15.39667 | 0.108 | Tidak Signifikan |
| γ_1 | -0.06207 | 0.35009 | 0.860 | Tidak Signifikan |
| γ_2 | -0.17758 | 0.20597 | 0.388 | Tidak Signifikan |

Lanjutan Tabel 3

| Parameter | Estimasi | SE | <i>P-value</i> | Signifikansi(5%) |
|------------|----------|---------|----------------|------------------|
| γ_3 | 0.12276 | 0.17107 | 0.473 | Tidak Signifikan |
| γ_4 | -0.39486 | 0.35227 | 0.262 | Tidak Signifikan |
| γ_5 | 0.15850 | 0.12824 | 0.216 | Tidak Signifikan |

PEMILIHAN MODEL TERBAIK

Evaluasi dalam menentukan model terbaik ditentukan berdasarkan nilai *Akaike's Information Criterion* (AIC) yang terkecil dari masing-masing model. Besarnya nilai AIC sejalan dengan nilai devians dari model. Semakin kecil nilai devians maka akan semakin kecil pula tingkat kesalahan yang dihasilkan model sehingga model yang diperoleh menjadi semakin tepat. Oleh karena itu, model terbaik adalah model dengan AIC terkecil. Berikut adalah tabel nilai AIC:

Tabel 4 Nilai AIC

| Kriteria Uji | Regresi ZIP | Regresi ZINB |
|--------------|-------------|--------------|
| AIC | 273.011 | 275.011 |

Berdasarkan Tabel 4, dapat disimpulkan bahwa nilai AIC dari regresi ZIP lebih kecil dibandingkan dengan regresi ZINB. Hal tersebut menunjukkan bahwa model regresi ZIP lebih baik digunakan pada data hipertensi kehamilan yang menyebabkan kematian ibu di Provinsi Kalimantan Barat.

KESIMPULAN DAN SARAN

Berdasarkan hasil pengolahan data menggunakan *Software Rstudio* dan analisis regresi ZIP dan regresi ZINB diperoleh kesimpulan sebagai berikut:

1. Model regresi ZIP

$$\text{Log}(\mu_i) = -2.55107 + 0.07592X_1 - 0.04133X_2 + 0.01203X_3 - 0.03733X_4 + 0.01843X_5$$

$$\text{Logit}(p_i) = 24.76861 - 0.06194X_1 - 0.17765X_2 + 0.12278X_3 - 0.39494X_4 + 0.15852X_5$$

Dengan nilai AIC 273.011

2. Model regresi ZINB

$$\text{Log}(\mu_i) = -2.55071 + 0.07591X_1 - 0.04133X_2 + 0.01203X_3 - 0.03733X_4 + 0.01844X_5$$

$$\text{Logit}(p_i) = 24.77251 - 0.06207X_1 - 0.17758X_2 + 0.12276X_3 - 0.39486X_4 + 0.15850X_5$$

Dengan nilai AIC 275.011

Berdasarkan kriteria nilai AIC dengan membandingkan model regresi ZIP dan model regresi ZINB dapat disimpulkan bahwa model regresi ZIP lebih baik digunakan untuk memodelkan hipertensi kehamilan yang menyebabkan kematian ibu serta mengandung banyak nilai nol dan mengalami overdispersi.

Penelitian selanjutnya untuk menyelesaikan kasus overdispersi dikarenakan *excess zero* pada regresi poisson dapat menggunakan model *Zero-Inflated Generalized Poisson* (ZIGP) atau *Zero-altered Poisson* (ZAP).

DAFTAR PUSTAKA

- [1]. Ruliana, P. Hendikawati, & A. Agoestanto. Pemodelan *Generalized Poisson Regression* (GPR) Untuk Mengatasi Pelanggaran *Equidispersi* Pada Regresi *Poisson* Kasus Campak di Kota Semarang Tahun 2013. *UNNES Journal of Mathematics*, 5(1) : 39-46; 2016.
- [2]. Kusuma, W., Kumalasari, D. dan Hadijati, M. Model Regresi *Zero-Inflated Poisson* pada Data *Overdispersion*, *Jurnal Matematika*, 3:71-85; 2013.
- [3]. Winkelman, R. *Econometric Analysis of Count Data* (5th ed.). Berlin : Springer; 2008.
- [4]. Lambert, D. *Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing*. *Technometrics*, 34(1) : 1-14; 1992.
- [5]. Jansakul, N. & J.P. Hinde. *Score Tests for Zero-Inflated Poisson Models*. *Computational Statistics & Data Analysis*, 40 : 75-96; 2002.
- [6]. Taufan, M., Suparti, & A. Rusgiyono. Analisis Faktor-Faktor Yang Mempengaruhi Banyaknya Klaim Asuransi Kendaraan Bermotor Menggunakan Model Regresi *Zero-Inflated Poisson* (Studi Kasus di PT. Asuransi Sinar Mas Cabang Semarang Tahun 2010). *Media Statistika*, 5(1): 49-61; 2012
- [7]. Hilbe, J.M. 2011. *Negative Binomial Regression* (2th ed.). New York : Cambridge University Press; 2011.
- [8]. Garay, A.M., E.M. Hashimoto, E.M.M. Ortega, & V.C. Lachos. *On Estimation And Influence Diagnostics for Zero-Inflated Negative Binomial Regression Models*. *Computational Statistics and Data Analysis*, 55 : 1304- 1318; 2011.
- [9]. Ariawan, B., Suparti, & Sudarmo. Pemodelan *Regresi Zero-Inflated Negative Binomial* (ZINB) Untuk Data Respon Diskrit dengan *Excess Zeros*. *Jurnal GAUSSIAN*, 1(1) : 55-64; 2012.

SAMSON : Jurusan Matematika FMIPA Untan, Pontianak,
pratamasamson@student.untan.ac.id

SETYO WIRA RIZKI : Jurusan Matematika FMIPA Untan, Pontianak,
setyo.wirarizki@math.untan.ac.id

HENDRA PERDANA : Jurusan Matematika FMIPA Untan, Pontianak,
hendra.perdana@ math.untan.ac.id
