

PENERAPAN REGRESI *ZERO-INFLATED NEGATIVE BINOMIAL* (ZINB) PADA DATA KECELAKAAN LALU LINTAS DI KOTA PONTIANAK

Anisa Nuraeni, Shantika Martha, Siti Aprizkiyandari

INTISARI

Model regresi poisson menganalisis sebuah hubungan antara peubah respon diskrit dengan satu atau lebih peubah prediktor yang diasumsikan dengan nilai mean dan varians yang sama yang disebut sebagai equidispersi. Ketika nilai varian lebih dari nilai mean atau yang disebut sebagai overdispersi, regresi poisson tidak layak untuk digunakan. Salah satu penyebab dari overdispersi adalah banyaknya nilai 0 pada sebuah data variabel Y. Untuk mengatasi masalah tersebut dapat menggunakan metode Zero Inflated Negative Binomial (ZINB). Tujuan penelitian ini yaitu menerapkan metode ZINB pada data kecelakaan lalu lintas di Kota Pontianak karena data mengandung banyak nilai 0 pada variabel Y, dan untuk melihat faktor-faktor apa saja yang berpengaruh terhadap jumlah kematian akibat kecelakaan lalu lintas di kota Pontianak. Penelitian ini menggunakan data sekunder yang tersedia di Dirjen Lalu Lintas Polisi Resort di Kota Pontianak yang terdiri dari jumlah kematian akibat kecelakaan lalu lintas (Y) sebagai variable respon dan variable bebas berupa faktor manusia (X_1), faktor kendaraan (X_2), faktor jalan (X_3), dan faktor lingkungan (X_4). Berdasarkan hasil analisis dari model ZINB diperoleh bahwa faktor yang paling berpengaruh untuk kasus kematian akibat kecelakaan lalu lintas adalah faktor manusia dan faktor kendaraan itu sendiri.

Kata kunci: Regresi Poisson, Overdispersi, Zero Inflated Negative Binomial (ZINB).

PENDAHULUAN

Kecelakaan lalu lintas menjadi masalah yang sangat mendunia tidak terkecuali dengan negara Indonesia [1]. Kecelakaan lalu lintas menurut UU RI No. 22 tahun 2009 adalah suatu peristiwa di jalan raya tidak diduga dan tidak disengaja melibatkan kendaraan dengan atau tanpa pengguna jalan lain yang mengakibatkan korban manusia dan/atau kerugian harta benda seseorang. Kecelakaan lalu lintas adalah kejadian yang paling sering dialami setiap harinya dan sering diberitakan di berbagai media dan termasuk tingkat kematian tertinggi. Berdasarkan data POLRI pada tahun 2009 diketahui bahwa jumlah kecelakaan lalu lintas terjadi sejumlah 62.960 kasus dengan jumlah korban sebanyak 106.384 jiwa di antaranya 19.979 jiwa korban meninggal dunia, 23.469 jiwa korban luka berat dan 62.936 jiwa korban luka ringan. Berbagai usaha telah dilakukan untuk mengurangi jumlah kecelakaan namun belum ada perubahan yang berarti.

Kota Pontianak merupakan Ibu Kota Kalimantan Barat yang memiliki jumlah kendaraan yang cukup ramai dengan pertumbuhan penduduk yang juga pesat. Setiap harinya terjadi peningkatan jumlah kendaraan di jalan yang menyebabkan meningkatnya jumlah kecelakaan lalu lintas. Penelitian dan penyuluhan akan kesadaran dalam berlalu lintas sering dilakukan guna menekan jumlah kecelakaan lalu lintas [2]. Banyak faktor yang menyebabkan tingkat kecelakaan lalu lintas seperti faktor manusia contohnya kondisi mengantuk, tidak fokus dalam membawa kendaraan, melamun dan sebagainya, faktor kendaraan seperti motor yang tiba tiba mati atau mogok, faktor lingkungan seperti cuaca yang ekstrim maupun faktor jalan yang ada. Ada beberapa faktor yang memiliki pengaruh yang signifikan terhadap penyebab kecelakaan lalu lintas. Ilmu statistika yang dapat digunakan untuk memodelkan data dari banyaknya kejadian dalam waktu atau wilayah tertentu disebut Regresi *Poisson*. Ciri-ciri Regresi *Poisson* adalah mengalami *equidispersi* yakni keadaan dimana nilai *mean* dan *varians* dari variabel respon memiliki nilai yang sama. Dalam penelitian ada kemungkinan mengalami overdispersi, yakni keadaan dimana nilai varians lebih besar dari pada nilai mean. Penyebab terjadinya overdispersi adalah terlalu banyak nilai nol pada data (*excess zeros*) pada variabel respon. Analisis regresi yang cocok dalam menangani overdispersi salah satunya adalah Regresi *Zero Inflated Negative Binomial* (ZINB). Tujuan Penelitian ini adalah untuk menangani data yang mengalami overdispersi pada regresi poisson dan untuk

mengetahui dan menentukan faktor-faktor mana saja yang berpengaruh secara signifikan terhadap jumlah kematian akibat kecelakaan lalu lintas dengan metode *Zero-Inflated Negative Binomial* (ZINB).

REGRESI LINIER

Analisis regresi linier adalah sebuah teknik untuk membangun persamaan dan menggunakan persamaan tersebut untuk membuat perkiraan. Analisis regresi sering disebut sebagai analisis prediksi atau perkiraan [3]. Regresi linier sederhana merupakan bagian regresi yang mencakup hubungan linier antara satu variabel terikat dengan satu variabel bebas. Bentuk umum persamaan regresi linier sederhana yang menunjukkan hubungan antara dua variabel, yaitu X sebagai variabel bebas dan variabel Y sebagai variabel terikat dari suatu populasi adalah sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad (1)$$

Keterangan:

Y_i = variabel terikat

X_i = variabel bebas

β_0 = jarak titik pangkal dengan titik potong garis regresi dengan sumbu Y

β_1 = kemiringan (slope) garis regresi

ϵ_i = nilai kesalahan.

MODEL REGRESI POISSON

Regresi Poisson merupakan suatu bentuk analisis regresi yang digunakan untuk memodelkan data yang berbentuk jumlah, misalkan data tersebut dilambangkan dengan Y, yaitu banyaknya kejadian yang terjadi dalam suatu periode waktu dan atau wilayah tertentu. Model regresi Poisson merupakan model standar untuk data diskrit dan termasuk dalam model regresi nonlinear [4]. Poisson adalah suatu bentuk model linier umum dimana variabel prediktor dimodelkan sebagai distribusi Poisson. Model regresi Poisson ditulis sebagai berikut [5];

$$y_i = \mu_i + \epsilon_i, \quad i = 1, 2, \dots, n \quad (2)$$

y_i adalah jumlah kejadian dan μ_i adalah rata-rata jumlah kejadian yang berdistribusi Poisson. Uji yang digunakan untuk melihat adakah model tersebut dianggap kurang sesuai dan ada model lain yang lebih sesuai dan cocok dalam menggambarkan hubungan antara variabel respon dan prediktor disebut uji kesesuaian model. Pengujian kesesuaian model dengan menggunakan *goodness of fit* disebut devians [6]. Analisis *deviance* merupakan salah satu analisis yang digunakan dalam analisis regresi pada pembentukan suatu model, *deviance* dapat diartikan sebagai logaritma dari uji likelihood nya.

Hipotesis:

$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$

H_1 : Paling sedikit ada satu $\beta_j \neq 0$, dengan $j=1,2,\dots,p$

Statistik Uji:

$$D = -2 \log \left[\frac{L(y; \mu)}{L(y; y)} \right]$$

Overdispersi

Model regresi Poisson mensyaratkan equidispersi, yaitu kondisi di mana nilai mean dan variansnya dari variabel terikat bernilai sama. Namun, adakalanya terjadi overdispersi berarti varians lebih besar daripada mean [6]. Maka model regresi Poisson tidak cocok untuk data tersebut. Fenomena *overdispersion* dinyatakan dengan [7]:

$$Var(Y) > E(Y)$$

Jika rasio menghasilkan nilai yang lebih besar dari satu, maka cara lain yang dapat digunakan untuk mendeteksi *overdispersion* yaitu:

Deviance

$$\theta_1 = \frac{D}{db} > 1; D = 2 \sum_{i=1}^n \left\{ y_i \log \frac{y_i}{\mu_i} \right\} \tag{3}$$

Pearson Chi-Square

$$\theta_2 = \frac{x^2}{db} > 1; x^2 = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\sigma_i} \tag{4}$$

Dimana:

y_i : nilai peubah dari pengamatan ke- i

μ_i : penduga bagi respon rata-rata ke- i

σ_i : penduga bagi respon ke- i

db : $n - k - 1$

k : banyaknya parameter termasuk konstanta

n : banyaknya pengamatan

Jika θ_1 dan θ_2 bernilai lebih dari 1 terjadi overdispersi pada data

REGRESI ZERO-INFLATED NEGATIVE BINOMIAL

Model regresi ZINB adalah model yang terbentuk dari distribusi campuran *poisson gamma* [8]. Jika Y_i adalah variabel random independen diskrit dengan $i = 1, 2, 3, \dots, n$, nilai nol pada sebuah obeservasi diduga muncul dalam dua cara sesuai kedaannya (*state*) yang terpisah. Kedaan 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 negative dengan mean μ , dengan $0 \leq p_i \leq 1$. Proses dua keadaan ini dengan variabel Y_i memberikan distribusi campuran dua komponen. Fungsi kepadatan peluangnya adalah

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

dengan rata-rata dan variansi distribusi binomial negatif yaitu:

$$E[Y] = \alpha\beta \text{ dan } V[Y] = \alpha\beta + \alpha\beta^2 \tag{5}$$

Regresi ZINB dengan keadaan pertama disebut *zero state* terjadi dengan probabilitas p_i dan menghasilkan hanya observasi bernilai nol, sementara keadaan kedua disebut *Negative Binomial State* terjadi dengan probabilitas $(1 - p_i)$ dan berdistribusi Binomial Negatif dengan mean μ dengan $0 \leq p_i \leq 1$ [9]. Proses dua keadaan ini dengan variabel Y_i memberikan distribusi campuran dua komponen dan didapat fungsi probabilitas sebagai berikut:

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

dengan $i = 1, 2, 3 \dots, n$; $0 \leq p_i \leq 1$, $\mu_i \geq 0$, k adalah parameter dispersi dengan $1/k > 0$ dan $\Gamma(\cdot)$ adalah fungsi gamma. Diasumsikan bahwa parameter μ_i dan p_i masing – masing bergantung pada variabel x_i dan z_i , model dari regresi ZINB dibagi menjadi dua komponen model [7] yaitu:

1. Model data diskrit untuk μ_i adalah

$$\ln(\mu_i) = x_i^T \beta, \mu_i \geq 0, i = 1, \dots, n \quad (6)$$

2. Model zero-Inflation untuk p_i adalah

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = z_i^T \gamma, 0 \leq p_i \leq 1, i = 1, \dots, n \quad (7)$$

Uji Kesesuaian Model

Pengujian kesesuaian model regresi ZINB menggunakan Likelihood Ratio (LR) Test dengan prosedur pengujian

Hipotesis:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = \gamma_1 = \gamma_2 = \dots = \gamma_p = 0$$

$$H_1: \text{Paling sedikit ada satu } \beta_j \neq 0 \text{ atau } \gamma_j \neq 0, \text{ dengan } j = 1, 2, \dots, p$$

dimana β_j adalah parameter ke- j dari model $\ln(\mu_i) = x_i^T \beta$ dengan $i = 1, \dots, n$, γ_j adalah parameter ke- j dari model $\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = x_i^T \gamma$ dengan $i = 1, \dots, n$.

Statistika uji:

$$D = -2 \ln \left[\frac{L_0}{L_1} \right] = -2(\ln L_0 - \ln L_1)$$

L_0 : likelihood tanpa variabel bebas L_1 : likelihood dengan variabel bebas

Kriteria uji: Tolak H_0 pada taraf signifikansi α jika $D_{hitung} > \chi_{\alpha; 2p}^2$.

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data sekunder yang tersedia di Dirjen Lalu Lintas Polisi Resort di Kota Pontianak yang terdiri dari jumlah kematian akibat kecelakaan lalu lintas (Y) sebagai variabel respon dan variabel bebas berupa data faktor manusia (X_1), faktor kendaraan (X_2), faktor jalan (X_3) dan faktor lingkungan (X_4). Berikut adalah statistik deskriptif data kecelakaan lalu lintas tahun 2018.

Tabel 1 Statistik Deskriptif Data Kecelakaan Lalu Lintas di Kota Pontianak

Variabel	N	Minimum	Maximum	Sum	Mean	Std. Deviation	Variance
Y	214	0	4	163	0.53	0.794	0.630
X_1	214	0	13	1355	4.43	2.212	4.895
X_2	214	0	5	36	0.12	0.589	0.347
X_3	214	0	5	46	0.15	0.640	0.410
X_4	214	0	4	21	0.07	0.386	0.149
Valid N	214						

Sumber: Hasil Pengolahan Data 2021

Pada Tabel 1 dapat dilihat bahwa jumlah data pada semua variabel sama yakni 214 data dengan nilai minimum adalah 0 data. Nilai maksimum data adalah 13 pada variabel X_1 . Pada data tersebut juga dapat dilihat bahwa nilai varians pada data lebih besar dari nilai mean yang berarti terjadinya overdispersi pada data.

Tabel 2 Tabel Proporsi Nilai 0 pada Data

Variabel Y	Jumlah Data	Persentase
0	129	60%
1	56	26%
2	23	11%
3	4	2%
4	2	1%

Sumber : Hasil pengolahan data, 2021

Dari Tabel 2 dapat dilihat bahwa pada variabel Y data yang bernilai 0 adalah sebanyak 129 data dengan presentase 60% yang menunjukkan bahwa data mengalami excess zeros.

Pengujian Distribusi Variabel Respon

Uji distribusi variabel respon dilakukan untuk mengetahui apakah variabel respon mengikuti sebaran poisson atau tidak. Pemeriksaan sebaran variabel respon dilakukan menggunakan uji *Kolmogorof Smirnov*. Pemeriksaan distribusi variabel respon dilakukan pada 214 pengamatan. Hipotesis yang digunakan adalah:

H_0 : variabel respon berdistribusi poisson

H_1 : variabel respon tidak berdistribusi poisson

Tabel 3 Uji Distribusi *Poisson* Variabel Respon

Kolmogorov-Smirnov	
Kolmogorov-Smirnov Z	0,421
Asymp. Sig. (2-tailed)	0,994

Sumber : Hasil pengolahan data, 2021

Berdasarkan tabel 3 dapat diketahui bahwa nilai asymp sig nya adalah 0,994 lebih besar dari alpha yang berarti bahwa H_0 diterima artinya variabel respon berdistribusi poisson.

Pengujian Multikolinieritas Antar Variabel Bebas

Pengujian Multikolinieritas dilakukan menggunakan kriteria nilai VIF (Variance Inflation Factor). Jika nilai VIF lebih besar dari 10 maka menunjukkan adanya multikolinieritas antar variabel prediktor.

Tabel 4 Pengujian Multikolinieritas

Variabel	VIF
X_1	1,017
X_2	1,009
X_3	1,002
X_4	1,013

Sumber : Hasil pengolahan data, 2021

Tabel 4 menunjukkan bahwa nilai VIF pada masing-masing variabel lebih kecil dari 10, maka tidak terjadi multikolinearitas terhadap data yang di uji sehingga dapat dilanjutkan ke model regresi Poisson.

Model Regresi Poisson

Model regresi Poisson merupakan salah satu regresi yang menganalisis antara peubah respon berupa data diskrit dengan satu atau lebih peubah prediktor. Berikut ini disajikan Tabel 5 yang menunjukkan estimasi parameter dari model regresi poisson.

Tabel 5 Nilai Estimasi Parameter Model Regresi Poisson

Parameter	Estimasi	Std. Error	P-value	Sig	Log-likelihood
(Intercept)	-0.95415	0.19033	5.35 e-07	***	
X_1	0.15162	0.08177	0.0637	**	
X_2	0.78603	0.14681	8.60e-08	***	-207.5978
X_3	0.23544	0.40697	0.5629	Ns	
X_4	0.43295	0.33265	0.1931	Ns	

Sumber : Hasil pengolahan data, 2021

Berdasarkan Tabel 5 dapat diketahui variabel yang signifikan dalam variabel tersebut adalah X_2 dan X_1 dengan alpha 10% sedangkan faktor yang lainnya tidak berpengaruh karena faktor tidak signifikan. Dengan kata lain dapat dinyatakan bahwa faktor-faktor yang mempengaruhi kecelakaan lalu lintas di kota Pontianak adalah faktor manusia (X_1) dan kendaraan (X_2). Model regresi yang terbentuk adalah:

$$\mu_i = \exp(-0,95415 + 0,151x_{1i} + 0,786x_{2i})$$

MODEL REGRESI ZERO-INFLATED NEGATIVE BINOMIAL (ZINB)

Model regresi ZINB digunakan dalam mengatasi data atau model yang overdispersi. Seperti halnya regresi *poisson*, pada ZINB juga akan dilakukan uji estimasi parameter sebagai yang akan ditunjukkan pada Tabel 6.

Tabel 6 Estimasi Parameter Model Regresi ZINB

Parameter	Estimasi	SE	P-value	Sig	Log-likelihood
β_0	-0,23	0,28949	0,42698	Ns	
β_1	-0,019	0,10171	0,84582	Ns	-203,2
β_2	0,56113	0,16887	0,00099	***	
β_3	-0,6295	0,44263	0,15495	Ns	
β_4	0,24589	0,48868	0,61484	Ns	
γ_0	1,0994	0,7427	0,1388	Ns	
γ_1	-1,0826	0,4623	0,0192	***	
γ_2	-2,8693	2,2409	0,2004	Ns	
γ_3	-14,949	6386704	0,9813	Ns	

γ_4	-0,9688	1,9594	0,621	Ns
------------	---------	--------	-------	----

Sumber : Hasil pengolahan data, 2021

Berdasarkan Tabel 6 diatas didapatkan full model ZINB untuk model log dan logit sebagai berikut:

$$\text{Log}(\mu_i) = 0,56113 X_{2i}$$

dan

$$\text{Logit}(p_i) = -1,0826X_{1i}$$

Faktor-faktor yang berpengaruh adalah faktor manusia dan faktor kendaraan karena memiliki nilai *p value* dibawah *alpha* 10%.

PENUTUP

Analisis data jumlah kematian akibat kecelakaan lalu lintas di kota Pontianak dengan variabel bebas berupa faktor manusia, faktor kendaraan, faktor jalan, dan faktor lingkungan memberi kesimpulan bahwa regresi poisson tidak cocok digunakan dalam data tersebut karena mengalami overdispersi pada model poisson, sehingga *Zero Inflated Negative Binomial* lebih cocok digunakan. Model Regresi *Zero Inflated Negative Binomial* nya sebagai berikut:

$$\text{Log}(\mu_i) = 0,56113 X_{2i}$$

dan

$$\text{Logit}(p_i) = -1,0826X_{1i}$$

Berdasarkan hasil analisis pada model regresi *Zero Inflated Negative Binomial* (ZINB), faktor paling berpengaruh dalam kasus kematian akibat kecelakaan lalu lintas adalah faktor manusia dan faktor kendaraan. Sehingga penulis menyarankan untuk meningkatkan kewaspadaan pengendara dalam membawa kendaraan dan selalu mengecek kendaraan sebelum berpergian agar tingkat kecelakaan lalu lintas di Kota Potianak bisa menurun dan tertib berlalu lintas.

DAFTAR PUSTAKA

- [1]. Utama, S. U., Magetsari, R., Pibadi, V. Estimasi Prevelensi Kecelakaan Lalu Lintas dengan Metode *Capture – Recapture*. Berita Kedokteran Masyarakat. 2018. 24(1): 16-26
- [2]. Arfan, I dan Wulandari Studi Epidmologi Kejadian Kecelakaan Lalu Lintas di Kota Pontianak. jurnal vokasi kesehatan 2018. 2 (2):90-96
- [3]. Kusnandar, D., Debararaja, N. N., Mara, M. N. dan Satyahadewi, N, *Metode Statistika Serta Aplikasinya dengan Minitab , Excel, Dan R*, UNTAN Press, Pontianak, 2009.
- [4]. Wulandari,S. P., Ulama, S. S. B. Dan Rahmawati, I. Pemodelan Resiko Penyakit Kaki Gajah (Filariasis) di Provinsi Papua dengan regresi zero Inflated Poisson. Forum Statistika dan Komputasi. 2010. 15 (1): 8-16
- [5]. Myers, R.H., Douglas C. Montgomery, G. Geoffrey Vining, & Timothy J, Robinson, *Generalized Linear Models with Applications in Engineering and The Sciences*, Second edition, New Jersey: John Wiley and Sons, 2010.
- [6]. Cahyandari, R., Pengujian Overdispersi pada Model Regresi, statistika, 2009. 14:69-76.
- [7]. Kleinbaum, Kupper, Muller dan Nizam, *Applied Regression Analysis and Other Multivariable Methods 3rd Edition*, London: Brooks/Cole Publishing Comp, 1998.
- [8]. Mc Cullagh, P. & Nelder, J.A, *Generalized Linear Models*. 2nd ed. London: Chapman & Hall, 1989.
- [9]. Hilbe, J.M., *Negative Binomial Regression*, Second Edition, New York: Cambridge University Press, 2011.

- [10]. Garay, A.M. and Hashimoto, E.M., On Estimation and Influence Diagnostics for Zero Inflated Negative Binomial Regression Models, Computational Statistics and Data Analysis, 2011, 55:1304-1318.

ANISA NURAENI

: Jurusan Statistika FMIPA UNTAN, Pontianak
anisanuraenistatistik@student.untan.ac.id

SHANTIKA MARTHA

: Jurusan Matematika FMIPA UNTAN, Pontianak
shantika.martha@math.untan.ac.id

SITI APRIZKIYANDARI

: Jurusan Matematika FMIPA UNTAN, Pontianak
siti.aprizkiyandari@faperta.untan.ac.id
