

ANALISIS KASUS KEMATIAN YANG DISEBABKAN OLEH KECELAKAAN LALU LINTAS DENGAN METODE REGRESI ZERO INFLATED POISSON

Magdalena Pelangi, Naomi Nesyana Debararaja, Setyo Wira Rizki

INTISARI

Regresi Poisson adalah suatu metode statistika yang digunakan untuk melakukan analisis terhadap data diskrit, data tersebut menyatakan banyaknya suatu kejadian yang jarang terjadi dalam selang waktu tertentu. Salah satu ciri dari regresi Poisson adalah equidispersi, yakni keadaan dimana nilai mean dan varians dari variabel terikat bernilai sama, namun dalam analisis data sering ditemukan suatu keadaan yang disebut overdispersi. Salah satu penyebab terjadinya overdispersi adalah adanya terlalu banyak nilai nol pada variabel terikat. Salah satu metode yang digunakan dalam menganalisis masalah overdispersi adalah metode regresi Zero Inflated Poisson (ZIP). Data yang dianalisis ialah kecelakaan lalu lintas di Kota Pontianak tahun 2018. Penelitian ini bertujuan untuk menentukan faktor yang paling berpengaruh terhadap kecelakaan lalu lintas di kota Pontianak. Berdasarkan nilai signifikan, regresi ZIP lebih baik digunakan untuk memodelkan data kecelakaan lalu lintas di kota Pontianak tahun 2018. Model regresi ZIP untuk data kecelakaan lalu lintas adalah: $\log(\mu_i) = -0,229 + 0,561 X_{2i}$ dan $\text{logit}(p_i) = 1,100 - 2,866 X_{2i}$. Hasil dari pengujian menunjukkan bahwa model regresi ZIP pada kematian akibat kecelakaan lalu lintas, faktor manusia (X_2) yang memberikan pengaruh paling signifikan diantara faktor-faktor lainnya.

Kata Kunci: Regresi Poisson, Overdispersi, Zero Inflated Poisson (ZIP)

PENDAHULUAN

Kecelakaan merupakan masalah kesehatan yang sangat serius di dunia, masalah yang sama juga dihadapi di berbagai negara termasuk Indonesia [1]. Kecelakaan lalu lintas menurut Undang-Undang Republik Indonesia Nomor 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 (UU RI, 2009). Kota Pontianak merupakan salah satu kota yang cukup ramai di Indonesia, hal ini dikarenakan jumlah penduduk yang banyak dan bertambah setiap tahunannya [2].

Ada beberapa faktor yang menjadi penyebab terjadinya kecelakaan lalu lintas, yaitu faktor manusia, kendaraan, lingkungan, dan kondisi jalan. Kematian akibat kecelakaan lalu lintas di kota pontianak relatif jarang terjadi, sehingga menghasilkan banyak nilai nol dalam beberapa pengamatan yang menyebabkan overdispersi. Model regresi *Zero-Inflated Poisson* (ZIP) merupakan salah satu metode yang dapat digunakan untuk menganalisis masalah overdispersi [3].

Tujuan penelitian ini adalah memperoleh model regresi Poisson dan model regresi ZIP, selanjutnya membandingkan regresi Poisson dengan regresi ZIP untuk memperoleh nilai terbaik berdasarkan nilai *log-likelihood* yang diperoleh. Setelah memperoleh nilai terbaik dari kedua model tersebut, maka diketahuilah faktor yang sangat mempengaruhi tingkat kematian akibat kecelakaan lalu lintas di Kota Pontianak melalui terbaik berdasarkan nilai *log-likelihood* yang diperoleh.

Model Regresi Poisson

Model Poisson merupakan model standar untuk data diskrit dan termasuk dalam model regresi nonlinear [4]. Model regresi Poisson ditulis sebagai berikut [5];

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

Pengujian kesesuaian model dengan menggunakan *goodness of fit test* pada analisis devians [6]. Analisis devians merupakan salah satu analisis yang digunakan dalam analisis regresi pada

pembentukan suatu model devians yang dapat diartikan sebagai logaritma dari uji *likelihood*. Perumusan hipotesis pengujian kesesuaian model regresi Poisson adalah sebagai berikut:

$$H_0 : \beta_i = 0, \quad 0 < i < k \text{ (pengaruh variabel ke-}i\text{ tidak signifikan)}$$

$$H_1 : \beta_i \neq 0, \quad 0 < i < k \text{ (pengaruh variabel ke-}i\text{ signifikan)}$$

Rumus yang digunakan sebagai berikut:

$$\begin{aligned} D &= -2 \log \left[\frac{L(y_i; \hat{\beta})}{L(y_i; \hat{\lambda})} \right] \\ &= -2 \log \left[L(y_i; \hat{\beta}) - L(y_i; \hat{\lambda}) \right] \\ &= -2 \log \left[-L(y_i; \hat{\lambda}) - L(y_i; \hat{\beta}) \right] \\ &= -2 \sum_{i=1}^n \left[(y_i \log y_i - y_i - \log(y_i)!) - (y_i \log \hat{y}_i - \hat{y}_i - \log(y_i)!) \right] \\ &= -2 \sum_{i=1}^n \left[y_i \log \left(\frac{y_i}{\hat{y}_i} \right) - (y_i - \hat{y}_i) \right] \end{aligned}$$

Overdispersi

Salah satu asumsi yang harus dipenuhi dalam model regresi Poisson adalah variansi sama dengan meannya, yaitu: $E(y) = Var(y) = \mu$. Data dikatakan overdispersion jika taksiran dispersi lebih besar dari 1 dan underdispersion jika taksiran dispersi kurang dari 1. Namun dalam analisis data diskrit dengan model regresi Poisson kadang dapat terjadi pelanggaran asumsi, dimana nilai variansi lebih besar dari nilai mean yang disebut overdispersi [5]. Fenomena overdispersi dinyatakan dengan [7]:

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

Pada keadaan ini overdispersi dapat ditangani dengan menggunakan beberapa model salah satunya ialah model Regresi *Zero Inflated Poisson*.

Regresi *Zero Inflated Poisson*

Model regresi ZIP adalah salah satu alternatif metode untuk menganalisis data dengan banyak nilai nol yang terkandung pada variabel respon. Banyaknya nilai nol pada data dapat mengakibatkan pelanggaran pada asumsi kesamaan mean dan variansi pada distribusi Poisson.

Untuk setiap pengamatan y_i yang saling bebas, $i = 1, \dots, n$, dan

$$Y_i \sim \begin{cases} 0, & \text{dengan peluang } \pi_i \\ \text{poisson}(\lambda_i), & \text{dengan peluang } (1 - \pi_i) \end{cases} \quad (2)$$

maka:

$$P(Y = y_i) = \begin{cases} \pi_i + (1 - \pi_i)e^{-\lambda_i}, & \text{untuk } y_i = 0 \\ \frac{(1 - \pi_i)e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, & \text{untuk } y_i > 0 \end{cases} \quad (3)$$

Dengan parameter $\lambda = (\lambda_1, \dots, \lambda_n)^T$ dan $\pi = (\pi_1, \dots, \pi_n)^T$ yang memenuhi:

$$\log(\lambda) = X\beta \quad (4)$$

dan

$$\text{logit}(\pi) = X\gamma \quad (5)$$

Pengujian Kesesuaian Analisis Regresi Zero-Inflated Poisson

Pengujian kesesuaian model regresi ZIP dilakukan menggunakan pengujian *likelihood ratio*. Hipotesis untuk pengujian kesesuaian model adalah sebagai berikut:

$$\begin{aligned} H_0: \beta_1 = \beta_2 = \dots = \beta_i = 0 \text{ atau } \gamma_1 = \gamma_2 = \dots = \gamma_i = 0 \quad i = 0, 1, 2, \dots, k \\ H_0: \text{ paling sedikit ada satu } \beta_i \neq 0 \text{ atau } \gamma_i \neq 0, \quad i = 0, 1, 2, \dots, k \end{aligned}$$

Dimana $k + 1$ adalah jumlah parameter, β_i adalah parameter model log ke- i , dan γ_i adalah parameter model logit ke- i . Perhitungan statistik uji untuk pengujian kesesuaian model sebagai berikut [8]:

$$\begin{aligned} D &= -2 \log \left[\frac{L(y_i; \hat{\omega})}{L(y_i; \hat{\Omega})} \right] \\ &= \left(2 \sum_{i=1}^n z_i x_i^T \hat{\gamma} - \log(1 + \exp(x_i^T \hat{\gamma})) + 2 \sum_{i=1}^n (1 - z_i) (y_i x_i^T \hat{\beta}) - \exp(x_i^T \hat{\beta}) \right) - \\ &\quad \left(2 \sum_{i=1}^n z_i \hat{y}_0 - \log(1 + x_i^T \hat{y}_0) + 2 \sum_{i=1}^n (1 - z_i) (y_i x_i^T \hat{\beta}_0 - \exp(x_i^T \hat{\beta}_0)) \right) \end{aligned}$$

Sedangkan pengujian parameter secara individu ada dua, yaitu pengujian parameter model log dan pengujian parameter logit. Berikut ini adalah perumusan hipotesis untuk pengujian parameter model log:

$$\begin{aligned} H_0: \beta_i = 0, \quad i = 0, 1, 2, \dots, k \\ H_1: \beta_i \neq 0 \end{aligned}$$

Dimana $k + 1$ adalah banyak parameter.

Statistik uji untuk pengujian parameter model log secara individu adalah sebagai berikut [8]:

$$\begin{aligned} D &= -2 \log \left[\frac{L(y_i; \hat{\omega})}{L(y_i; \hat{\Omega})} \right] \\ &= \left(2 \sum_{i=1}^n z_i x_i^T \hat{\gamma} - \log(1 + \exp(x_i^T \hat{\gamma})) \right) + 2 \sum_{i=1}^n (1 - z_i) (y_i x_i^T \hat{\beta} - \exp(x_i^T \hat{\beta})) + 2 \sum_{i=1}^n (1 - z_i) (y_i x_i^T \hat{\beta}_i) \end{aligned}$$

Perumusan hipotesis untuk pengujian parameter model logit secara individu adalah sebagai berikut:

$$\begin{aligned} H_0: \gamma_i = 0, \quad i = 0, 1, 2, \dots, k \\ H_1: \gamma_i \neq 0 \end{aligned}$$

Dimana $k + 1$ adalah banyak parameter.

Statistik uji untuk pengujian parameter model log secara individu adalah sebagai berikut [8]:

$$\begin{aligned} D &= -2 \log \left[\frac{L(y_i; \hat{\omega})}{L(y_i; \hat{\Omega})} \right] \\ &= \left(2 \sum_{i=1}^n z_i x_i^T \hat{\gamma} - \log(1 + \exp(x_i^T \hat{\gamma})) \right) + 2 \sum_{i=1}^n (1 - z_i) (y_i x_i^T \hat{\beta} - \exp(x_i^T \hat{\beta})) + \\ &\quad 2 \sum_{i=1}^n (1 - z_i) \log(y_i)! - 2 \sum_{i=1}^n z_i \hat{y}_0 - \log(1 + x_i^T \hat{y}_0) + \exp \hat{\gamma}_0 \end{aligned}$$

Kriteria pengujian untuk ketiga pengujian diatas adalah H_0 ditolak pada taraf signifikan α jika $D_{hitung} > \chi_{(v, \alpha)}^2$, dimana v adalah derajat bebas.

STUDI KASUS

Penelitian ini menggunakan data sekunder yang tersedia di Ditjen Lalu Lintas Polisi *Resort* di Kota Pontianak, datanya ialah jumlah kematian akibat kecelakaan lalu lintas (Y) sebagai variabel terikat 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 kematian akibat kecelakaan lalu lintas tahun 2018:

Tabel 1 Nilai Statistik Deskriptif dari Data Kematian akibat Kecelakaan Lalu Lintas Tahun 2018

Variabel	N	Minimal	Maksimal	Rata-rata	Varians	Std. Deviasi
Y	214	0	4	0,57	0,669	0,818
X_1	214	0	6	1,74	1,037	1,018
X_2	214	0	3	0,06	0,085	0,292
X_3	214	0	2	0,07	0,080	0,283
X_4	214	0	1	0,06	0,057	0,239

Setelah diperoleh Statistik Deskriptif dari variabel terikat (Y) dan variabel bebas (X_1, X_2, X_3, X_4) tersebut, maka variabel Y akan ditentukan nilai varians dan rata-ratanya untuk mengetahui data tersebut mengalami overdispersi atau tidak. Overdispersi yaitu nilai varians lebih besar dari nilai rata-rata, salah satu penyebab terjadinya overdispersi adalah terlalu banyak nilai nol pada variabel terikat.

Pengujian Multikolinieritas Antar Variabel Bebas

Pengujian Multikolinieritas dilakukan menggunakan kriteria nilai VIF. Jika nilai *variance inflation factor* (VIF) lebih besar dari 10 menunjukkan adanya multikolinieritas antar variabel prediktor.

Tabel 2 Pungujian Multikolinieritas

Toleransi	VIF
0,985	1,016
0,986	1,015
0,988	1,012
0,992	1,008

Nilai toleransi pada masing-masing variabel lebih besar dari 0,10 dan nilai VIF pada masing-masing variabel lebih kecil dari 10,00, maka tidak terjadi multikolinieritas terhadap data yang di uji.

Model Regresi Poisson

Model regresi Poisson dapat dibentuk dari variabel tak bebas dan beberapa variabel bebas yang mempengaruhi, data kecelakaan lalu lintas dianalisis untuk memperoleh model regresi Poisson. Berikut ini disajikan Tabel 3 yang menunjukkan estimasi parameter dari model regresi Poisson.

Tabel 3 Nilai Estimasi Parameter Model Regresi Poisson

Parameter	Estimasi	SE	<i>P-value</i>	Sig	Log-likelihood Poisson
Y	-0,954	0,190	0,000	***	
X_1	0,151	0,081	0,063	.	
X_2	0,786	0,146	0,000	***	-207,5978
X_3	-0,235	0,406	0,562	ns	
X_4	0,432	0,332	0,193	ns	

Berdasarkan Tabel 3 dapat diketahui variabel yang tidak signifikan terhadap model yaitu variabel X_3 dan X_4 , sedangkan variabel X_1 dan X_2 signifikan. Terlihat pada tingkat signifikan yang dihasilkan dari masing-masing faktor, apabila tingkat signifikan $\geq 10\%$ (ns) maka faktor tersebut tidak mempengaruhi variabel dependen, sedangkan jika tingkat signifikan kurang dari 10% (****, ***, **, ".") maka

faktor tersebut berpengaruh terhadap variabel terikat. Dengan kata lain dapat dinyatakan bahwa faktor-faktor yang paling mempengaruhi kecelakaan lalu lintas di kota pontianak adalah faktor kendaraan dan manusia. Model regresi Poisson untuk data kecelakaan lalu lintas adalah:

$$\mu_i = \exp(-0,954 + 0,151X_{1i} + 0,786X_{2i})$$

Overdispersi pada Regresi Poisson

Ciri dari regresi Poisson adalah equidispersi, apabila variabel respon mengalami overdispersi maka model regresi Poisson tidak sesuai. Taksiran dispersi dapat diukur dengan nilai Residual *deviance* yang dibagi derajat bebas. Apabila nilai taksiran lebih dari 1 maka ada indikasi overdispesi. Berikut akan disajikan Tabel 4 tentang nilai taksiran disperse untuk model-model regresi Poisson yang telah diperoleh.

Tabel 4 Taksiran Dispersi Model-model Regresi Poisson

No	Model Regresi Poisson dengan Variabel Bebas	Residual deviance	derajat bebas	Taksiran dispersi	Keterangan
1	Model Nol	246,48	213	1,157	Overdispersi
2	Model Lengkap	224,96	209	1,076	

Pada Tabel 4 dapat diketahui bahwa model regresi Poisson yang didapatkan mengalami overdispersi yang mengakibatkan model regresi poisson menjadi tidak sesuai, sehingga perlu dilakukan analisis dengan model regresi yang lain yaitu ZIP.

Model Regresi Zero inflated Poisson (ZIP)

Model regresi *zero inflated* Poisson bertujuan untuk memperbaiki model regresi Poisson karena pada model regresi Poisson dimungkinkan adanya overdispersi pada variabel terikat. Pemodelan *zero inflated* Poisson melibatkan variabel bebas yang sama dengan model regresi Poisson sebelumnya. Seperti dalam model regresi Poisson, akan dilakukan estimasi parameter model regresi ZIP. Berdasarkan hasil pengolahan data dapat dilihat pada Tabel 5 Estimasi Parameter Model Regresi ZIP berikut:

Tabel 5 Estimasi Parameter Model Regresi ZIP

Parameter	Estimasi	SE	P-value	Sig	Log-likelihood ZIP
β_0	0,123	0,300	0,681	ns	-77,23
β_1	0,027	0,016	0,082	.	
β_2	0,101	0,048	0,036	*	
β_3	0,039	0,058	0,498	ns	
β_4	0,078	0,115	0,495	ns	
γ_0	1,695	1,453	0,243	ns	
γ_1	-0,427	0,275	0,12	ns	
γ_2	-0,161	1,23	0,896	ns	
γ_3	-13,902	162,433	0,932	ns	
γ_4	1,008	1,159	0,384	ns	

Berdasarkan Tabel 5 diatas didapatkan keseluruhan model ZIP untuk model log dan logit sebagai berikut:

$$\text{Log}(\mu_i) = -0,229 + 0,561X_{2i}$$

dan

$$\text{Logit}(p_i) = 1,100 - 2,866X_{2i}$$

Model logit regresi ZIP menunjukkan bahwa kematian akibat kecelakaan lalu lintas di kota Pontianak yang memberikan pengaruh yang paling signifikan di antara faktor-faktor yang lain ialah faktor manusia, karena variabel ini yang memiliki tingkat signifikan yang paling tinggi.

KESIMPULAN

1. Model Regresi Poisson untuk data kecelakaan lalu lintas adalah:

$$\mu_i = -0,954 + 0,151X_{1i} + 0,786X_{2i}$$
2. Model Regresi *Zero Inflated* Poisson (ZIP) untuk data kecelakaan lalu lintas adalah:

$$\text{Log}(\mu_i) = -0,229 + 0,561X_{2i}$$

dan

$$\text{Logit}(p_i) = 1,100 - 2,866X_{2i}$$

3. Berdasarkan kriteria nilai *log-likelihood* regresi poisson dan regresi ZIP, dapat disimpulkan bahwa model regresi ZIP lebih baik digunakan untuk memodelkan data kecelakaan lalu lintas di kota Pontianak tahun 2018, dengan kasus kematian yang disebabkan oleh kecelakaan tersebut mengandung banyak nilai nol dan mengalami overdispersi.
4. Hasil dari pengujian menunjukan bahwa model regresi ZIP yang terbentuk pada kematian akibat kecelakaan lalu lintas, faktor manusia (X_2) memberikan pengaruh yang paling signifikan diantara faktor-faktor lainnya.

DAFTAR PUSTAKA

- [1]. Utama, S. U., Magetsari, R., Pibadi, V. Estimasi Prevelensi Kecelakaan Lalu Lintas dengan Metode *Capture-Recapture*. *Berita Kedokteran Masyarakat*. 2008. **24**(1):16-26.
- [2]. Arfan, I. dan Wulandari. Studi Epidemiologi Kejadian Kecelakaan Lalu Lintas di Kota Pontianak. *Jurnal Vokasi Kesehatan*. 2018. **2**(2):90-96.
- [3]. Kusuma, W., Kumalasari, D. dan Hadijati, M. Model Regresi *Zero-Inflated* Poisson pada Data *Overdispersion*. *Jurnal Matematika*. 2013. **3**(2):71-85.
- [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]. Cahyandari, R. Pengujian Overdispersi pada Model Regresi. *Statistika*. 2014. **14**(2):69-76.
- [6]. Kleinbaum, Kupper, Muller dan Nizam. *Applied Regression Analysis and Other Multivariable Methods 3rd Editon*. London: Brooks/Cole Publishing Comp. 1998.
- [7]. Kusuma, W., Kumalasari, D. dan Hadijati, M. Model Regresi *Zero-Inflated* Poisson pada Data *Overdispersion*. *Jurnal Matematika*. 2013. **3**(2):71-85.
- [8]. Lestari, A., Puhadi. dan Ratna, M. Pemodelan Regresi *Zero-Inflated* Poisson (Aplikasi pada Data Pekerja Seks Komersial di Klinik Reproduksi Putat Jaya Surabaya). *Pythagoras*. 2009. **5**(2):57-72.

MAGDALENA PELANGI : Jurusan Matematika FMIPA UNTAN, Pontianak
 magdalenapelangi96@gmail.com

NAOMI NESSYANA DEBATARAJA : Jurusan Matematika FMIPA UNTAN, Pontianak
 naominessyana@math.untan.ac.id

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