

## PEMODELAN REGRESI BINOMIAL NEGATIF UNTUK MENGATASI *OVERDISPERSION* PADA REGRESI POISSON

**Rena Muntafiah<sup>1</sup>, Rochdi Wasono<sup>2</sup>, Moh. Yamin Darsyah<sup>3</sup>**

<sup>1,2,3</sup>Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Muhammadiyah Semarang  
Alamat e-mail : Rena\_fia@yahoo.co.id

### ABSTRAK

Data dari Kepolisian Republik Indonesia pada tahun 2009 menunjukkan bahwa setiap 9,1 menit sekali terjadi satu kecelakaan di jalan raya. Menurut WHO Saat ini kecelakaan lalu lintas menjadi penyebab pembunuh manusia terbesar di Dunia setelah penyakit jantung dan TBC. Banyaknya korban kecelakaan lalu lintas tidak hanya mengakibatkan kerugian materiil tetapi juga kerugian fisik dan psikis. Pada kenyataannya faktor kelalaian manusia merupakan penyebab utama terjadinya kecelakaan lalu lintas. Selain faktor manusia faktor jalan seperti jalan lurus (X1), tikungan (X2), persimpangan perempatan atau pertigaan (X3), jenis aspal (X4), penerangan gelap (X5), bundaran (X6) dan jalan berlubang (X7) juga diduga menjadi penyebab terjadinya kecelakaan lalu lintas. Kecelakaan lalu lintas dapat digambarkan sebagai peristiwa yang jarang terjadi, bersifat acak (random) dan diskrit, maka dari itu untuk meneliti kasus kecelakaan lalu lintas dapat digunakan metode *Generalized Linier Model* (GLM) dengan sebaran Poisson untuk menghasilkan pemodelan yang lebih baik. Pada penelitian ini digunakan metode Regresi Poisson untuk menganalisa data cacahan dengan variabel respon berdistribusi Poisson atau menyatakan kejadian yang relatif jarang terjadi dan bersifat diskrit. Tetapi pada kenyataannya, data diskrit seringkali mengalami *overdispersion* (dimana varians data lebih besar daripada mean). Adanya *overdispersion* dalam model Poisson menyebabkan nilai *deviance* model menjadi sangat besar. Metode Regresi Binomial Negatif dapat mengatasi *overdispersi* pada Regresi Poisson karena memiliki parameter dispersi ( $\kappa$ ). Kemungkinan fungsi hubungan dalam penelitian ini menggunakan pendekatan GLM yaitu;  $\mu = \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)$ . Hasil dari pemodelan diatas didapatkan pemodelan interaksi antara faktor persimpangan (4/3) dengan bundaran adalah model yang memiliki nilai devians terkecil. Sehingga model terbaik pada penelitian ini adalah  $\mu = \exp(\beta_0 + \beta_3 X_3 + \beta_6 X_6) = \exp(2.9239 + 0.0461 X_3 + 0.0587 X_6)$ .

**Kata Kunci :** Regresi Poisson, *Overdispersion*, Regresi Binomial Negatif, kecelakaan lalu lintas

### PENDAHULUAN

Kecelakaan lalu lintas merupakan suatu peristiwa di jalan yang tidak diduga dan tidak disengaja melibatkan kendaraan dengan atau tanpa pengguna jalan lain yang mengakibatkan korban manusia dan/kerugian harta benda (Pasal 1 Angka 24 UU Nomor 22 Tahun 2009). Menurut WHO Saat ini kecelakaan lalu lintas menjadi penyebab pembunuh manusia terbesar di Dunia setelah penyakit jantung

dan TBC. Hal ini di buktikan pada tahun 2013 jumlah kejadian kecelakaan di Kabupaten Blora sebesar 394 dengan korban meninggal sebesar 51, luka berat 206 dan luka ringan sebesar 444 dengan kerugian materiil sebesar 325.250.000 rupiah<sup>[1]</sup>. Sebaran Poisson merupakan pilihan yang sesuai dengan sifat kecelakaan lalu lintas yang dapat digambarkan sebagai peristiwa yang sangat jarang terjadi, bersifat acak (*random*), diskrit dan non-negatif<sup>[2]</sup>.

Model regresi Poisson berasal dari distribusi Poisson dengan parameter intensitas  $\mu$  yang bergantung pada variabel prediktor. Dalam model Regresi Poisson terdapat beberapa asumsi yang harus dipenuhi yakni equidispersi yang berarti nilai variansi dari variabel respon Y yang diberikan oleh  $X = x$  harus sama dengan nilai *meannya* yaitu  $\text{var}(Y | x) = E(Y | x) = \mu$ . Tetapi pada kenyataannya, data diskrit seringkali mengalami *overdispersion* (dimana variansi data lebih besar daripada *mean*). Adanya *overdispersion* dalam model Poisson menyebabkan nilai *deviance* model menjadi sangat besar.

Pengujian taksiran dispersi ( $\alpha$ ) secara statistik dapat dilakukan dengan uji nilai Devians, hipotesis yang digunakan yaitu:  $H_0 : \alpha = 1$   $H_1 : \alpha > 1$ . Statistik uji nilai Deviansnya:  $G = 2 \sum_{i=1}^n [y_i \ln (\frac{y_i}{\hat{y}_i}) - (y_i - \hat{y}_i)]$ . Penolakan terhadap  $H_0$  jika  $G > \chi^2_{n-k-1}$ . Hal ini berarti terjadi *overdispersi* pada Regresi Poisson. Berdasarkan teori tersebut penulis akan melakukan penelitian menggunakan metode Regresi Binomial Negatif untuk memodelkan faktor jalan yang berpengaruh pada jumlah kasus kecelakaan di Kabupaten Blora.

Pemilihan model terbaik pada penelitian ini menggunakan nilai *Pearson Chi-Square* dan Perumusan hipotesis pada pengujian dengan *Likelihood Ratio* sebagai berikut:  $H_0$ : model Regresi Poisson = model Regresi Binomial Negatif,  $H_1$ : model Regresi Binomial Negatif lebih baik dari model Regresi Poisson, uji statistik *Likelihood Ratio*  $T = 2(l_1 - l_0)$ . Dengan  $l_1$  dan  $l_0$  adalah *log-likelihood* masing-masing model pada hipotesis. Penolakan  $H_0$  pada taraf signifikansi  $\alpha$  jika  $T > \chi^2_{(1-2\alpha, 1)}$  [3].

**Tabel 1** Perbandingan Regresi Poisson dengan Regresi Binomial Negatif

Kriteria	Regresi Poisson	Regresi Binomial Negatif
Peubah respon	$Y_{iid} \sim Poi(\mu_i)$	$Y_{iid} \sim BN(\mu_i, \kappa)$
Fungsi distribusi peluang	$p_{y;\mu} = \frac{e^{-\mu} \mu^y}{y!}$ (Myers et al. [3])	$f_{y,\kappa} = \frac{\Gamma(\kappa+1) \mu^y}{y! \Gamma(\kappa+1) (\kappa\mu+1) \kappa}$
Rata-rata dan Ragam	$EY_i = VarY_i = \mu_i$	$EY_i = \mu_i$ $VarY_i = \mu_i + \kappa \mu_i^2$
Penaksir parameter Fungsi Log Likelihood	<i>Maximum Likelihood Estimator</i> $\hat{y}_i \ln \mu_i; \beta - n_i = 1$ $\hat{t}_i \mu_i; \beta - \ln y_i! n_i = 1$	<i>Maximum Likelihood Estimator</i> $\log \Gamma(\kappa+1) + y_i$ $-\log \Gamma(\kappa+1 + y_i) \log$ $\kappa \mu_i + \kappa \mu_i + \kappa - 1 \log$ $1 + \kappa \mu_i$
Model Regresi	$\mu_i = x_i' \beta + \epsilon_i$	$\mu_i = x_i' \beta + \epsilon_i$

**METODE PENELITIAN**

**Sumber Data dan Variabel Penelitian**

Data yang digunakan dalam penelitian ini adalah data sekunder dari SATLANTAS POLRES Kabupaten Blora tahun 2013. Variabel yang digunakan di dalam penelitian ini yaitu;

- Y : Jumlah korban kecelakaan yang ada di Kabupaten Blora pada tahun 2013.
- X1 : Jumlah korban kecelakaan pada jalan yang lurus di Kabupaten Blora pada tahun 2013.
- X2 : Jumlah korban kecelakaan pada jalan tikungan di Kabupaten Blora pada tahun 2013.
- X3 : Jumlah korban kecelakaan pada jalan persimpangan (4/3) di Kabupaten Blora pada tahun 2013.
- X4 : Jumlah korban kecelakaan pada jenis aspal di Kabupaten Blora pada tahun 2013.

- X5 : Jumlah korban kecelakaan pada jalan dengan penerangan gelap di Kabupaten Blora pada tahun 2013.
- X6 : Jumlah korban kecelakaan pada jalan bundaran di Kabupaten Blora pada tahun 2013.
- X7 : Jumlah korban kecelakaan pada jalan yang berlubang di Kabupaten Blora pada tahun 2013.

### Metode Analisis

Langkah menganalisis data dalam penelitian ini yaitu dengan menggunakan metode Regresi Poisson. Kemudian memeriksa terjadinya overdispersi pada Regresi Poisson. Jika terjadi overdispersi pada Regresi Poisson, akan dilanjutkan menganalisis data dengan menggunakan metode Regresi Binomial Negatif.

## HASIL PENELITIAN

### Pembentukan Model Regresi Poisson

Dari 29 kemungkinan yang terbentuk dari pendekatan GLM menunjukkan bahwa seluruh kemungkinan model memiliki nilai devians lebih dari 1. Hal ini menunjukkan adanya indikasi *overdispersion* pada Regresi Poisson. Keadaan *overdispersion* pada model mengakibatkan taksiran parameter menjadi bias. Terdapat 8 nilai devians terkecil dari 29 kemungkinan model di atas salah satunya yaitu model  $\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6 + \beta_7 X_7)$  dari model tersebut dapat dijelaskan bahwa konstanta sebesar 2.7338 artinya jika kondisi jalan lurus, tikungan, persimpangan (4/3), bundaran dan jalan berlubang nilainya 0, maka korban kecelakaan lalu lintas pada faktor penyebab jalan nilainya sebesar 2.7338.

### Pengujian Kesesuaian Model Regresi Poisson

Dari tabel 2 di dibawah dapat diketahui bahwa seluruh nilai devians lebih besar dari nilai *chi-square*. Hal ini berarti kedelapan model Regresi Poisson tidak layak digunakan.

### Pengujian *Overdispersion*

Dari tabel 3 didapat kesimpulan bahwa delapan  $t_{wald}$  diatas, memiliki nilai yang lebih besar dari  $Z_{(1-\delta)} = 1.645$  sehingga  $H_0$  ditolak, yang berarti bahwa tidak terpenuhinya asumsi *equidispersion* pada Regresi Poisson.

### Pengujian Kesesuaian Model Regresi Binomial Negatif

Dari tabel 4 dapat diketahui bahwa nilai devians dari model Regresi Binomial Negatif lebih kecil dibandingkan dengan nilai devians dengan menggunakan model Regresi Poisson. Hal ini berarti model Regresi Binomial Negatif lebih layak digunakan dibandingkan dengan model Regresi Poisson. Dari model Regresi Binomial Negatif pada tabel 4 juga dapat dilihat bahwa seluruh nilai devians lebih besar dari nilai *chi-square*, hal ini menunjukkan bahwa seluruh model sesuai dengan model Regresi Binomial Negatif.

### Pengujian parameter Regresi Binomial Negatif

Statistik uji yang digunakan dalam uji hipotesis ini adalah Uji Wald *Chi-Square*. Uji Wald *Chi-square* akan menolak hipotesis awal ( $H_0$ ) jika  $W > \chi^2_{(1,\alpha)}$ . Nilai  $\alpha = 0.05$  pada Uji Wald *Chi-Square*, didapatkan  $\chi^2_{(1,\alpha)} = 2.71$ . Dari tabel 5 terdapat tiga model yang memenuhi asumsi pengujian parameter Regresi Poisson yaitu model  $\mu = \exp(\beta_0 + \beta_3 X_3 + \beta_6 X_6)$  dan  $\mu = \exp(\beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6)$ . Pemilihan model terbaik pada Regresi Binomial Negatif adalah dengan melihat nilai devians yang terkecil. Nilai devians yang terkecil dari ketiga model diatas adalah 19.6171 pada pemodelan

fungsi  $\mu = \exp(\beta_0 + \beta_3 X_3 + \beta_6 X_6) = \exp(2.9239 + 0.0461 X_3 + 0.0587 X_6)$ .

**Tabel 2** Pengujian Kesesuaian Model Regresi Poisson

Model	$D(\hat{\beta})$	db	$\chi_{(db,5\%)}$	$D(\hat{\beta})/db$
$\mu = \exp(\beta_0 + \beta_3 X_3 + \beta_6 X_6)$	109.1317	14	96.9528	7.7951
$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_6 X_6)$	108.9443	13	<b>96.1921</b>	<b>8.3803</b>
$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_6 X_6)$	106.9947	13	91.6501	8.2304
$\mu = \exp(\beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6)$	93.3812	13	78.4516	7.1832
$\mu = \exp(\beta_0 + \beta_3 X_3 + \beta_6 X_6 + \beta_7 X_7)$	107.9483	13	95.8137	8.3037
$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6)$	92.8487	12	79.2488	7.7374
$\mu = \exp(\beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6 + \beta_7 X_7)$	90.0980	12	76.1691	7.5082
$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6 + \beta_7 X_7)$	84.9772	11	74.5298	7.7252

**Tabel 3** Pengujian *Overdispersion*

Model	$\hat{\alpha}$	s.s.e( $\hat{\alpha}$ )	$t_{wald}$
$\mu = \exp(\beta_0 + \beta_3 X_3 + \beta_6 X_6)$	0.2690	0.1152	2.335
$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_6 X_6)$	0.2403	0.1053	2.282
$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_6 X_6)$	0.2499	0.1089	2.295
$\mu = \exp(\beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6)$	0.2097	0.0966	2.171
$\mu = \exp(\beta_0 + \beta_3 X_3 + \beta_6 X_6 + \beta_7 X_7)$	0.2566	0.1110	2.312
$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6)$	0.2092	0.0963	2.172
$\mu = \exp(\beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6 + \beta_7 X_7)$	0.1737	0.0822	2.113
$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6 + \beta_7 X_7)$	0.1621	0.0781	2.075

**Tabel 4** Pengujian Kesesuaian Model Regresi Binomial Negatif

Model	$D(\hat{\beta})$	db	$\chi_{(db,5\%)}$
$\mu = \exp(\beta_0 + \beta_3 X_3 + \beta_6 X_6)$	19.6171	14	13.0118
$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_6 X_6)$	19.6831	13	13.7491
$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_6 X_6)$	19.6852	13	13.0118
$\mu = \exp(\beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6)$	20.1070	13	13.6771
$\mu = \exp(\beta_0 + \beta_3 X_3 + \beta_6 X_6 + \beta_7 X_7)$	19.6435	13	14.4533
$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6)$	20.0904	12	13.4521
$\mu = \exp(\beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6 + \beta_7 X_7)$	20.0753	12	14.2722
$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6 + \beta_7 X_7)$	20.1235	11	14.8843

**Tabel 5** Pengujian Parameter Regresi Binomial Negatif

Model	Uji Wald					
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_6$	$\beta_7$
$\mu = \exp(\beta_0 + \beta_3 X_3 + \beta_6 X_6)$	274.26			5.90	10.22	
$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_6 X_6)$	158.29	1.99	1.60		1.23	
$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_6 X_6)$	207.23	1.04		1.03	1.36	
$\mu = \exp(\beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6)$	183.61		3.32	3.59	9.28	
$\mu = \exp(\beta_0 + \beta_3 X_3 + \beta_6 X_6 + \beta_7 X_7)$	199.84			6.50	10.94	0.65
$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6)$	178.17	0.05	2.20	1.62	2.75	
$\mu = \exp(\beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6 + \beta_7 X_7)$	113.74		5.63	4.20	10.63	2.66
$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6 + \beta_7 X_7)$	109.63	0.89	5.79	4.47	6.28	3.62

**KESIMPULAN**

Berdasarkan data kecelakaan lalu lintas di Kabupaten Blora pada tahun 2013 yang diperoleh dari data Laka Lantas SATLANTAS POLRES Blora, yang telah dianalisis menggunakan *software* SAS 9.1.3 didapat kesimpulan sebagai berikut;

- a. Kriteria jalan yang dapat mempengaruhi jumlah kecelakaan di Kabupaten Blora pada tahun 2013 yaitu interaksi antara kriteria jalan persimpangan (4/3) dan jalan bundaran
- b. Pemodelan faktor jalan yang mempengaruhi jumlah kecelakaan di Kabupaten Blora pada tahun 2013 dengan menggunakan metode Regresi Poisson adalah  $\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6 + \beta_7 X_7)$  dari model tersebut dapat dijelaskan bahwa konstanta sebesar 2.7338 artinya jika kondisi jalan lurus, tikungan, persimpangan (4/3), bundaran dan jalan berlubang nilainya 0, maka korban kecelakaan lalu lintas pada faktor penyebab jalan nilainya sebesar 2.7338.
- c. Pemodelan faktor jalan yang mempengaruhi jumlah kecelakaan di Kabupaten Blora pada tahun 2013 dengan menggunakan metode Regresi Binomial Negatif adalah

$$\mu = \exp(\beta_0 + \beta_3 X_3 + \beta_6 X_6) = \exp(2.9239 + 0.0461 X_3 + 0.0587 X_6).$$

Interaksi antara variabel persimpangan (4/3) dan variabel jalan bundaran merupakan faktor penyebab terjadinya kecelakaan di setiap Kecamatan di Kabupaten Blora. Koefisien regresi pada kondisi persimpangan (4/3) sebesar 0.0461 artinya jika korban kecelakaan pada persimpangan (4/3) mengalami kenaikan satu satuan, maka korban kecelakaan pada faktor jalan akan mengalami kenaikan sebesar 0.0461 satuan dengan asumsi variabel prediktor lainnya bernilai tetap. Koefisien regresi variabel jalan bundaran sebesar 0.0587 artinya jika korban kecelakaan pada jalan bundaran mengalami kenaikan satu satuan, maka korban kecelakaan pada faktor jalan akan mengalami peningkatan sebesar 0.0587 satuan dengan asumsi variabel prediktor lainnya bernilai tetap.

- d. Penggunaan model fungsi hubungan pada Regresi Poisson untuk kasus kecelakaan di Kabupaten Blora pada tahun 2013 kurang layak di gunakan hal ini di buktikan dengan nilai devians pada model Regresi Poisson yang sangat besar, penggunaan Regresi Binomial Negatif lebih layak digunakan untuk mengatasi kasus kecelakaan di Kabupaten Blora dibandingkan dengan

Regresi Poisson, hal ini dibuktikan dengan nilai devians dari pemodelan Regresi Binomial Negatif yang lebih mendekati 1.

#### DAFTAR PUSTAKA

- [1] Satlantas Kabupaten Blora, 2013, *Data Kecelakaan Lalu Lintas*, Satuan Lalu Lintas, Blora.
- [2] Nugroho, Y., 2008, *Pengembangan Model untuk Memperkirakan Kinerja Keselamatan Jalan Tol Menggunakan Model Kombinasi Regresi Linier dan stimasi Bayes*, Jurusan Teknik Sipil Fakultas Teknik Universitas Diponegoro.
- [3] Ismail, N., & Jemain, A.A., 2007, Handling Overdispersion with Binomial Negative and Generalized Poisson Regression Models, *Casualty Actuarial Society Forum Casualty Actuarial Society – Arlington, Virginia: Winter*, 103 - 158. <http://www.casact.org/pubs/forum/07wforum/07w109.pdf> (20 April 2014).