

EKSPLORASI MODEL TINGKAT KECELAKAAN LALU-LINTAS JALAN TOL DENGAN TEKNIK GLM (*GENERALIZED LINEAR MODELING*)

Bambang Haryadi

Jurusan Teknik Sipil, Fakultas Teknik, Universitas Negeri Semarang (UNNES)
Gedung E4, Kampus Sekaran Gunungpati Semarang 50229, Telp. (024) 8508102

Abstract: The objectives of the presents study was to explore mathematical models that could be used to estimate the number of accidents on inter-urban toll roads. Preditive models were developed by relating traffic exposure (average daily traffic and road section length) with the number of accidents per unit of time. Accident, roadway section and traffic volume data were obtained from Jagorawi, Jakarta-Cikampek, padaleunyi, and palikanci toll road. Accident rate models were developed from those data using *generalized linear modelling* (GLM) techniques. The conclusions from the study were: (1) *linier regression model* was not appropriate to be used to predict accidents number, because accident occurence did not follow normal distribution, (2) *Poisson regression* possessed accident occurence characteristics: discrete, rare and random, and (3) *Negative Binomial distribution* was more appropriate to represent accident occurence phenomenon with overdispersion.

Keywords: accident rate, *generalized linear modelling*, *Poisson regression*, *negative binomial regression*.

Abstrak: Penelitian ini bertujuan untuk mengeksplorasi model matematis yang dapat digunakan untuk meramalkan jumlah kecelakaan, pada suatu ruas jalan tol antar kota. Model prediksi dikembangkan dengan menghubungkan paparan lalu lintas, yang dinyatakan dengan volume lalu lintas dan panjang ruas jalan, dengan keselamatan, yang dinyatakan dalam jumlah kecelakaan per satuan waktu. Data jalan, kecelakaan, dan lalu lintas selama dua tahun, diambil dari jalan tol antar kota Jagorawi, Jakarta-Cikampek, Padaleunyi dan Palikanci. Model tingkat kecelakaan dikembangkan dari data tersebut dengan teknik *generalized linear modelling* (GLM), dan dikalibrasi dengan menggunakan teknik-teknik statistik. Beberapa kesimpulan yang dapat ditarik dari penelitian ini adalah: (1) Model regresi linier dengan distribusi normal tidak memadai untuk memprediksi kecelakaan; tingkat kecelakaan tidak terbukti mengikuti distribusi normal; (2) Penggunaan regresi Poisson lebih mencerminkan karakteristik kejadian kecelakaan: diskrit, acak, langka; dan (3) Distribusi Binomial Negatif paling baik merepresentasikan kejadian kecelakaan dengan adanya gejala *overdispersi* pada data.

Kata kunci: tingkat kecelakaan, *generalized linear modelling*, regresi poisson, regresi binomial negatif.

PENDAHULUAN

Tingkat Kecelakaan (TK) (*accident rate*) adalah persamaan matematis yang menyatakan hubungan antara faktor-faktor dominan yang mempengaruhi kecelakaan. Dengan menggunakan nilai TK memungkinkan kita melakukan perbandingan tingkat keselamatan antar berbagai lokasi yang berbeda. Hal ini dimungkinkan dengan cara menormalisasi jumlah kecelakaan yang terjadi di suatu ruas jalan dengan periode waktu, panjang ruas, dan

volume lalu lintas. Sebagai gambaran, apabila suatu ruas jalan mempunyai banyak kecelakaan dan sekaligus mempunyai volume lalu lintas yang sangat besar, ruas tersebut bisa mempunyai TK yang kecil dan oleh sebab itu dapat dikategorikan lebih aman dibandingkan ruas jalan lain dengan jumlah kecelakaan yang lebih sedikit tetapi dengan LHR (lalu lintas harian rata-rata) yang jauh lebih rendah. TK biasanya dinyatakan sebagai perbandingan antara jumlah

kecelakaan dibagi dengan jumlah panjang perjalanan atau *vehicle-km traveled* (VKT).

Bagian penting dalam penilaian resiko kecelakaan adalah identifikasi ukuran paparan (*exposure*) yang tepat terhadap resiko yang sedang dikaji (Qin, et al, 2005). Jumlah perjalanan atau paparan menunjukkan peluang terjadinya masing-masing kecelakaan (Sacomano & Buyco, 1988). Arus lalulintas merupakan ukuran paparan yang paling umum. Misalnya, pengelola jalan tol PT Jasamarga mempunyai tolok ukur keselamatan, yaitu tingkat kecelakaan (TK), dengan definisi sebagai berikut:
$$TK = \frac{JK \times 100 \text{ juta kendaraan}}{LHR \times L \times \text{jumlah hari}}$$

Dimana JK adalah jumlah kecelakaan, LHR adalah lalu lintas harian rata-rata, dan L adalah panjang ruas jalan.

Kekurangan dari pendekatan ini adalah asumsi adanya hubungan linier antara jumlah kecelakaan dengan VKT. Pada kenyataannya terdapat bukti bahwa hubungan tersebut tidak linier, dan asumsi linieritas melemahkan akurasi prediksi kecelakaan (Qin, dkk., 2005). Sebagai alternatif, perlu eksplorasi bentuk hubungan antara jumlah kecelakaan dengan jumlah dan panjang perjalanan tanpa disaumsikan secara apriori adanya hubungan linier antara jumlah kecelakaan dengan (*exposure*) paparan perjalanan tersebut dengan menggunakan data aktual. Tulisan ini bertujuan untuk mengkaji hubungan antara jumlah kejadian kecelakaan dengan volume lalulintas harian rata-rata (LHR). Termasuk didalamnya adalah kajian mengenai asumsi distribusional dan bentuk fungsional.

METODOLOGI PENELITIAN

Pemodelan matematis adalah teknik untuk membuat metode kuantifiabel untuk

memprediksi terjadinya kejadian tertentu. Model prediksi kecelakaan adalah persamaan yang menyatakan frekuensi kecelakaan sebagai fungsi dari arus lalu lintas dan karakteristik jalan yang lain. Banyak isu yang muncul berkaitan dengan pemodelan maupun dengan sifat daari kecelakaan lalu lintas itu sendiri. Sejumlah isu penting dibahas dalam bagian berikut.

GLM (*Generalized Linear Modeling*) merupakan metode yang paling langsung yang digunakan untuk mengembangkan model matematis. GLM biasanya terdiri dari tiga komponen: komponen random, komponen sistematis, dan fungsi penghubung yang menghubungkan dua komponen untuk menghasilkan prediktor linier (Lord & Persaud, 2000). Asumsi penting dalam pemodelan linier tergeneralisasi adalah bahwa *random error* terjadi hanya pada variabel terikat dan bahwa variabel-variabel bebas diketahui tanpa *error* (Maher & Summersgill, 1996). Ini merupakan asumsi penting yang harus selalu diperhatikan karena tidak semua variabel yang diperlukan yang berkontribusi pada kecelakaan kendaraan diketahui tanpa *error*. Untuk variabel geometri dan kontrol seperti jumlah lajur dan keberadaan median, variabel ini diketahui tanpa *error*, tetapi tidak demikian dengan semua variabel karakteristik lalu lintas seperti volume dan persen kendaraan berat.

Pilihan distribusi probabilitas dalam GLM fleksibel untuk komponen random, yang membuat model ini efektif untuk keselamatan lalu lintas dimana jumlah kecelakaan dan variabel lain mengikuti distribusi Poisson atau binomial negatif dan variabel yang lain mengikuti distribusi normal.

Model Linier

Secara historis mayoritas penelitian tentang hubungan kecelakaan lalu lintas dengan faktor penyebabnya, seperti misalnya volume lalu lintas, menggunakan analisis regresi linier konvensional, yang mengasumsikan bahwa variabel terikat adalah kontinu dan berdistribusi normal dengan varian *error* konstan. Pada umumnya koefisien regresi didapatkan dengan metode kuadrat terkecil tradisional (*ordinary least squares*). Metode ini menghasilkan *point estimator*, β , yang mempunyai varian minimum. Pendekatan analisis varian (ANOVA) biasanya digunakan dan memisahkan *sum of squares* (jumlah kuadrat) dan derajat kebebasan yang berkaitan dengan variabel terikat. *Mean squared error*, MSE, dapat diperoleh dari tabel ANOVA dan merupakan taksiran varian (σ^2) tak-bias. Varian *error terms* (ϵ_i) juga merupakan indikasi varian dari distribusi probabilitas variabel terikat.

Varian digunakan untuk menghitung koefisien determinasi, R^2 , yang merepresentasikan proporsi variabilitas yang dijelaskan oleh fungsi regresi. Koefisien determinasi merupakan metode yang paling umum untuk menentukan kualitas model yang dicari dan mempunyai rentang nilai antara nol dan satu. Nilai R^2 mendekati nol mengindikasikan tidak adanya hubungan linier yang kuat antara variabel bebas dan terikat. Nilai R^2 mendekati satu mengindikasikan kecocokan linier yang kuat dimana model menjelaskan variabilitas pada data. R^2 harus digunakan dengan hati-hati untuk menjamin interpretasi yang benar dan digabungkan dengan pemeriksaan dengan diagram pencar atau *scatter plot* (Garber & Ehrhart, 2000). R^2 merupakan parameter yang bermanfaat hanya dalam model regresi linier; ia tidak aplikabel

pada kasus lain selain distribusi normal. Nilai R^2 yang rendah tidak hanya berarti model tidak cocok untuk data yang bersangkutan, tetapi juga tidak ada hubungan linier antara variabel-variabel yang dikaji, dan harus digunakan bentuk-bentuk fungsi atau distribusi yang lain.

Regresi Poisson

Beberapa peneliti percaya bahwa koefisien model prediksi kecelakaan tidak dapat dengan benar ditaksir dengan metode regresi kuadrat-terkecil biasa atau kuadrat-terkecil terbobot. Hal ini dikarenakan jumlah kecelakaan bersifat diskrit, non-negatif dan kenyataan bahwa varian jumlah kecelakaan makin membesar seiring dengan semakin besarnya arus lalu lintas tetapi tidak secara linier (Lord & Persaud, 2000). Model regresi normal atau lognormal konvensional tidak memiliki sifat-sifat statistis yang diperlukan untuk mendeskripsikan kecelakaan kendaraan. Masalah utama dengan pemodelan linier/multilinier adalah bahwa model ini mempunyai kemungkinan untuk memprediksi kecelakaan negatif, dimana tidak mungkin terjadi dalam kenyataan (Miaou *et al.*, 1992). Suatu lokasi tanpa kejadian kecelakaan dapat terjadi, tetapi tidak ada lokasi dengan jumlah kecelakaan negatif. Hubungan antara kecelakaan dengan faktor-faktor yang berkaitan tidak selalu refleksi perilaku linier yang berakibat regresi multi-linier tidak tepat untuk analisis sebab-sebab kecelakaan (Saccomanno & Buyco, 1988). Di lain pihak, dengan semakin mudahnya program pemodelan diakses, canggih dan *user friendly*, para peneliti mulai menaksir koefisien model dengan menggunakan metode *maximum-likelihood* (kecocokan maksimum) untuk kalibrasi model-model linier tergeneralisasi (*generalized linear models*).

Penggunaan tipe distribusi lain juga menjadi semakin populer. Pilihan model yang disukai adalah distribusi Poisson. Pilihan fungsi sesuai dengan sifat kecelakaan adalah fungsi eksponensial, dimana sudah secara luas digunakan oleh statistikawan dan ekonometrikawan (Miaou *et al.*, 1992).

Regresi Poisson adalah suatu pendekatan pemodelan nonlinier dimana variabel terikatnya adalah hitungan (*count*), atau kejadian diskrit yang langka (Nater *et al.*, 1999). Data hitungan telah dianalisis dengan regresi linier biasa, tetapi ada keuntungan dengan menggunakan regresi Poisson yaitu dari kenyataan bahwa distribusi ini digunakan atau dirancang untuk distribusi variabel terikat diskrit dan yang seringkali sangat menceng (*skewed*).

Distribusi Poisson memodelkan probabilitas kejadian diskrit dengan

$$P(Y) = \frac{e^{-\mu} \mu^Y}{Y!}$$

dan dapat diturunkan sebagai limit distribusi binomial dimana jumlah ulangan (*trial*), n , mendekati tak-hingga dan probabilitas sukses tiap ulangan, p , mendekati nol sedemikian rupa sehingga $np = \lambda$ (Rice, 1988). Dimana Y adalah jumlah kejadian pada periode yang dipilih dan μ adalah jumlah rata-rata kejadian pada periode yang dipilih. Model regresi Poisson mengasumsikan bahwa jumlah kejadian rata-rata merupakan fungsi variabel-variabel regresor. Untuk menaksir frekuensi kecelakaan, ia diasumsikan berdistribusi Poisson dengan

$$P(Y = Y_i) = \frac{e^{-\mu(X_i, \beta)} [\mu_i(X_i, \beta)]^{Y_i}}{Y_i!}$$

. Y_i sama dengan jumlah kecelakaan pada ruas jalan 'i' untuk periode waktu yang dipilih. β adalah vektor parameter yang ditaksir. $\mu_i = \mu(X_i, \beta)$ adalah jumlah kecelakaan rata-rata pada ruas 'i'

yang merupakan fungsi dari satu set variabel bebas X . X_i adalah vektor variabel bebas untuk ruas i . Fungsi $\mu_i = \mu(X_i, \beta)$, yang menghubungkan rerata distribusi pada variabel bebas, adalah fungsi penghubung $\mu_i = \mu(X_i, \beta) = e^{X_i \beta}$. Variabel bebas adalah item-item seperti arus lalu lintas dan karakteristik geometrik. Vektor X , yang terdiri dari variabel bebas mempunyai nilai 1 sebagai faktor (*term*) pertamanya sehingga faktor pertama dalam vektor β adalah interseptor atau konstan. Apabila lokasi yang dikaji berupa suatu panjang jalan, bukan berupa persimpangan, biasanya diasumsikan bahwa μ_i juga proporsional terhadap panjang L_i maupun terhadap periode waktu, sehingga λ_i adalah dalam bentuk kecelakaan per kilometer per tahun.

Satu masalah utama adalah fenomena overdispersi dimana asumsi struktur *error* Poisson murni dapat dilihat tidak memadai. Model binomial negatif sering dipilih untuk mengatasi masalah ini sebagai ekstensi dari model Poisson. Tetapi seringkali varian lebih besar dari rerata terjadi disebabkan sebagian karena tidak dimasukkannya semua variabel yang relevan ke dalam model (Knuiman *et al.*, 1993). Bila yang terjadi varian lebih besar dibandingkan dengan rerata hal ini disebut dengan overdispersi.

Regresi Binomial Negatif

Distribusi binomial negatif merupakan ekstensi alamiah dari distribusi Poisson, yang memperhitungkan variabilitas yang berlebihan yang kadang ditemukan dalam model prediksi kecelakaan. Distribusi ini mulai diminati untuk digunakan dalam penelitian transportasi, digunakan untuk membantu mengatasi masalah yang terjadi dengan pemodelan Poisson,

khususnya varian dimungkinkan untuk berbeda dari rerata dalam regresi binomial negatif (Hadi *et al.*, 1995). Kedua distribusi berkaitan dengan sekuen Bernoulli. Model binomial negatif dapat dianggap sebagai distribusi yang lebih umum untuk data hitungan dibandingkan dengan model Poisson disebabkan karena faktor pengganggu (*disturbance term*) yang membantu untuk mengatasi masalah overdispersi yang terjadi pada pemodelan Poisson (Allison, 1999). Koefisien beta pada model ditaksir dengan metode *quasi-likelihood* (Knuiman *et al.*, 1993). Estimasi *maximum likelihood* juga merupakan cara yang efisien untuk menaksir parameter dalam regresi binomial negatif. Model regresi binomial mempunyai bentuk $\text{Log } \lambda_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \sigma \varepsilon_i$, dimana variabel terikat Y diasumsikan mengikuti distribusi Poisson dengan nilai harapan λ_i kondisional terhadap ε_i (Allison, 1999). Nilai harapan ε_i diasumsikan mengikuti distribusi gamma standard. Selanjutnya distribusi *unconditional* Y_i mengikuti distribusi negatif binomial (Allison, 1993).

Distribusi negatif binomial didasarkan pada variabel random binomial negatif dimana jumlah sukses tetap dan jumlah ulangan random (Devore, 1999). Variabel random yang menjadi perhatian adalah X =jumlah kegagalan yang mendahului sukses ke- r . X mempunyai kemungkinan nilai 0, 1, 2, ... Fungsi massa probabilitas untuk distribusi binomial negatif dapat dituliskan sebagai

$$P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r} \quad \text{dimana } k=r, r+1, \dots$$

Misalnya urutan ulangan bebas dilakukan sampai seluruhnya ada r sukses; dengan X menyatakan jumlah total ulangan. Untuk mendapatkan $P(X=k)$, kita dapat melakukan

dengan cara berikut: Dari asumsi independen setiap sekuen mempunyai probabilitas $p^r (1-p)^{k-r}$. Ulangan terakhir sukses, dan $r-1$ sukses sisanya dapat diberikan pada ulangan $k-1$ sisanya dengan cara $\binom{k-1}{r-1}$ (Rice, 1988). Jika kejadian ke- r terjadi pada ulangan ke- k , akan ada tepat $r-1$ kejadian sebelum ulangan ke- $n-1$ dan pada ulangan ke- k kejadiannya juga terjadi. 'X' biasanya didefinisikan sebagai jumlah total ulangan dalam distribusi, tetapi kadang-kadang didefinisikan sebagai jumlah total kegagalan dalam distribusi (Rice, 1988). Cara penulisan fungsi massa probabilitas memungkinkan hubungan antara distribusi binomial dan distribusi binomial negatif. Kedua distribusi terdiri dari sekuen ulangan independen.

Dalam distribusi binomial negatif, rerata tidak harus sama dengan varian. Rerata untuk variabel random binomial negatif sama dengan:

$$\mu = E(x) = \frac{r}{p}$$

$$\sigma^2 = V(x) = \frac{r(1-p)}{p^2} \quad (\text{Montgomery \& Runger, 2003}).$$

DATA

Data penelitian diambil dari tiga lokasi jalan tol, yaitu Jalan Tol Jagorawi (Jakarta-Bogor-Ciawi), Jalan Tol Jakarta-Cikampek, dan Jalan Tol Padaleunyi (Padalarang-Cileunyi). Tipe data yang digunakan dalam penelitian adalah data ruas jalan, data lalu lintas, dan data kecelakaan. Data dikumpulkan dari pengelola masing-masing jalan tol, yaitu PT Jasamarga cabang yang bersangkutan. Secara keseluruhan, dari ketiga jalan tersebut ada 57 ruas jalan yang masing-masing relatif homogen sepanjang ruas baik dalam aspek lalu-lintas

maupun karakteristik geometri jalan. Untuk keperluan analisis digunakan periode data dua tahun agar diperoleh data jumlah kecelakaan yang memadai untuk analisis.

HASIL

Model Regresi Linier

Ada kemungkinan hubungan antara volume versus jumlah kecelakaan total maupun

panjang ruas versus jumlah kecelakaan total mengikuti distribusi normal tetapi tidak menjelaskan sebagian besar variasi dalam data, maka diperlukan model dengan mengkombinasikan kedua variabel bebas dalam satu model. Berikut ini adalah hasil keluaran regresi linier dan anova hubungan kecelakaan total dengan panjang ruas dan LHR dari fungsi *lm (linear model)* dari perangkat lunak R .

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.459e+01	1.695e+01	1.451	0.1526
Panjang	5.217e+00	2.008e+00	2.598	0.0120 *
lhrt	1.298e-04	2.043e-04	0.635	0.5280

Residual standard error: 44.75 on 54 degrees of freedom
 Multiple R-Squared : 0.1131, Adjusted R-squared: 0.08027
 F-statistic : 3.444 on 2 and 54 DF, p-value: 0.03911

Bentuk model yang dihasilkan dari output fungsi *lm* perangkat lunak R adalah:
 $Total = 24.59 + 5.217 * Panjang + 0.0001298 * lhrt$.

Taksiran parameter untuk panjang ruas signifikan bertanda positif yang berarti semakin panjang ruas jalan semakin banyak terjadi kecelakaan. Koefisien untuk LHR tidak signifikan, juga positif yang berarti semakin besar volume lalu lintas semakin banyak kecelakaan yang terjadi. Dengan mengkombinasikan kedua variabel bebas dalam satu model lebih banyak variabilitas yang dalam model yang dijelaskan, yaitu sebesar 11.31 persen, lebih besar dibandingkan bila hanya menggunakan variabel bebas volume (0.22 persen) atau hanya menggunakan variabel bebas panjang ruas (10.65 persen).

Kedua variabel bebas, *Panjang* dan *lhrt* bertanda positif seperti yang diharapkan, tetapi adanya intersep merupakan suatu masalah. Positif intersep menunjukkan bahwa walaupun tidak terdapat lalu lintas dan panjang ruas sama dengan nol model meramalkan terjadinya lebih dari 24 kejadian kecelakaan per dua tahun. Hal ini tentu saja tidak mungkin dalam

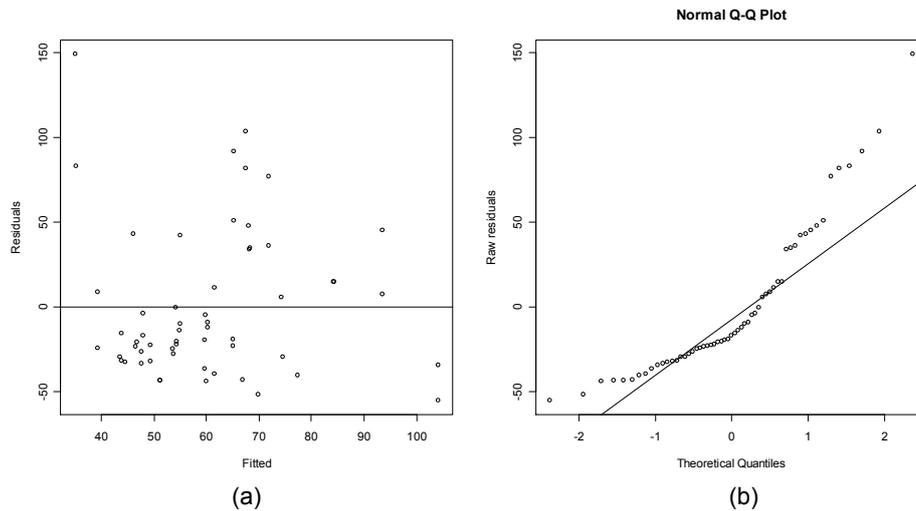
kenyataan, oleh karena itu model ini tidak dapat digunakan untuk menunjukkan hubungan antara jumlah kecelakaan total dengan panjang ruas dan volume lalu lintas. Dari analisis statistik diketahui bahwa hanya parameter panjang ruas yang signifikan diatas lima persen, sedangkan intersep dan parameter untuk LHR tidak signifikan dan tidak banyak membantu dalam menjelaskan variasi dalam data.

Untuk menguji apakah asumsi model terpenuhi, dikaji hubungan antara nilai prediksi dan nilai residu seperti disajikan pada Gambar 1(a) plot residu memperlihatkan ketidakseimbangan antara titik-titik di atas dan di bawah garis nol, tetapi tidak menunjukkan adanya penyimpangan linieritas dan kesamaan varian yang substansial pada data. Tidak terdapat pola yang nyata pada titik-titik data dan dengan sedikit pengecualian titik tersebut kurang lebih terdistribusi seimbang antara nilai positif dan nilai negatif, meskipun dengan jelas terlihat di bawah garis nol titik tersebar pada rentang 0 hingga -50 sedangkan di atas garis nol posisi nilai residual lebih menyebar pada rentang nilai antara 0 hingga 100 bahkan ada

satu titik menyebar jauh dari yang lain pada nilai 150.

Normal quantile plot untuk kecelakaan total, panjang ruas dan LHR ditunjukkan pada Gambar 1(b). Garis lurus pada Gambar 1(b) menunjukkan normalitas sedang garis berupa titik-titik merepresentasikan data aktual. Dari

gambar terlihat adanya penyimpangan dari data terhadap normalitas. Pada bagian awal penyimpangan ada di bagian atas garis normalitas kemudian berangsur turun menyimpang di bagian bawah garis normalitas di bagian tengah dan kemudian naik lagi menyimpang di bagian atas garis normalitas.



Gambar 1. (a) Nilai prediksi vs. residu untuk kecelakaan total, panjang ruas dan volume lalulintas, (b) Normal quantile plot untuk kecelakaan total, panjang ruas dan volume lalulintas dengan regresi linier.

Model Poisson

Model regresi linier berdasarkan pada asumsi distribusi normal yang merupakan distribusi kontinu. Hal ini menjadi alasan untuk melakukan eksplorasi terhadap distribusi lain karena data kecelakaan lalulintas pada suatu ruas jalan pada periode waktu tertentu merupakan data hitungan (*count*) diskrit non-negatif yang tidak mengikuti distribusi normal.

Oleh karena itu distribusi yang berdasarkan *count data* dikaji kemungkinan lebih tepat untuk memprediksi jumlah kecelakaan.

Fungsi *glm (generalized linear model)* dari perangkat lunak R digunakan untuk *fitting* model Poisson hubungan antara jumlah kecelakaan total dengan panjang ruas dan lalulintas harian rata-rata (LHR). Ringkasan disajikan berikut ini.

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.503e+00	5.224e-02	67.063	< 2e-16 ***
Panjang	7.792e-02	5.204e-03	14.973	< 2e-16 ***
lhrt	2.622e-06	5.934e-07	4.419	9.93e-06 ***

Signif. Codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 Null deviance : 1891.6 on 56 degrees of freedom
 Residual deviance : 1677.6 on 54 degrees of freedom
 AIC : 2004.7

Model yang dihasilkan adalah $Ln_{(Total)} = 3.503 + 0.07792(Panjang) + 0.000002622(LHR)$, atau

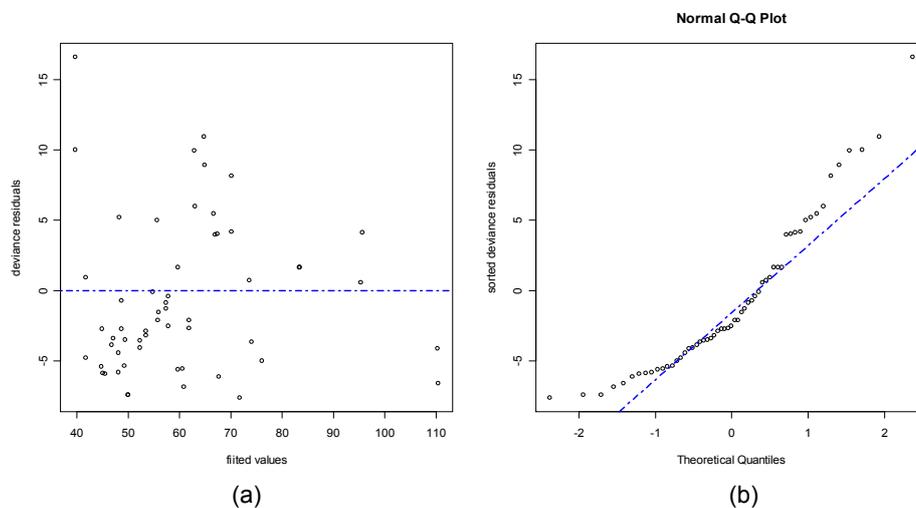
$$Total = e^{3.503} * e^{0.07792(Panjang)} * e^{0.000002622(LHR)}$$

Asumsi dasar dari distribusi Poisson adalah adanya kesamaan nilai rerata dan varian. Model yang dikembangkan dengan menggunakan distribusi Poisson menunjukkan nilai overdispersi yang besar, dimana

merupakan indikasi bahwa nilai rerata jauh berbeda dari nilai varian. Hal ini melanggar asumsi yang paling dasar. Dari keluaran perangkat lunak R diatas memperlihatkan gejala overdispersi yang besar. Nilai *residual deviance* adalah 1677.6 dengan 54 derajat kebebasan memberikan rasio *deviance* dibagi derajat kebebasan sebesar 31.07, sebuah angka perbandingan yang besar. Nilai satu mengindikasikan tidak ada masalah dengan overdispersi, semakin besar nilainya, semakin besar pula perbedaan nilai antara rerata dan

varian. Hal ini juga mengindikasikan bahwa data tidak cocok dengan jenis fungsi dari model.

Dari keluaran perangkat lunak, walaupun semua variabel signifikan pada taraf lebih dari 95 persen, tetapi dapat dilihat adanya kemungkinan nilai koefisien panjang ruas dan LHR dapat sama dengan nol. Gagasan bahwa volume dan panjang ruas tidak mempunyai pengaruh pada jumlah kecelakaan yang terjadi, berdasarkan pengamatan, merupakan ide yang sulit diterima. Karena asumsi model tidak benar hubungan ini invalid.



Gambar 2. (a) Nilai prediksi vs. residu, dan (b) *Normal quantile plot* untuk kecelakaan total, panjang ruas dan volume lalu lintas dengan regresi Poisson.

Model Binomial Negatif

Fungsi *generalized linear model* untuk binomial negatif (*glm.nb*) dari perangkat lunak R untuk mencocokkan model binomial negatif hubungan antara kecelakaan total dengan panjang ruas dan volume lalu lintas. Ringkasan luarannya adalah sebagai berikut.

Penggunaan distribusi binomial negatif untuk memodelkan panjang ruas, volume lalu lintas dan kecelakaan total mengatasi masalah overdispersi. Model yang dihasilkan mirip dengan model yang dihasilkan dengan

distribusi Poisson, tetapi masalah overdispersi hampir sepenuhnya dapat diatasi. Nilai intersep dan koefisien untuk panjang ruas dan volume sangat mirip, tetapi eliminasi masalah overdispersi membuat data lebih cocok dengan distribusi ini. Nilai deviance dibagi dengan derajat kebebasan adalah 1.133278, merupakan nilai yang sangat kecil, menjadikannya model yang sangat bagus untuk variabelnya. Nilai sebesar 1.0 menunjukkan tidak ada masalah varian lebih besar dari pada yang diijinkan.

Coefficients:

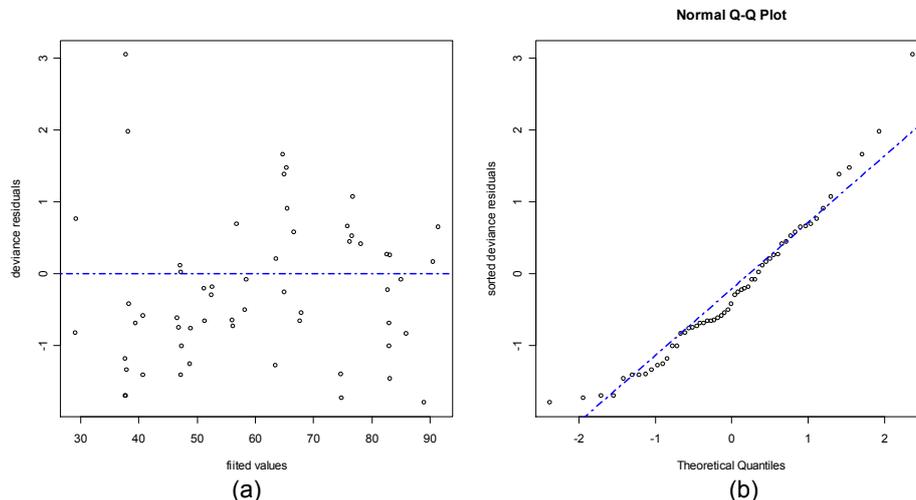
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.491e+00	2.732e-01	12.776	<2e-16 ***
Panjang	7.591e-02	3.221e-02	2.357	0.0184 *

lhrt	3.108e-06	3.285e-06	0.946	0.3442
Null deviance	: 67.734 on 56 degrees of freedom			
Residual deviance	: 61.197 on 54 degrees of freedom			
AIC	: 574.75			

Variabel panjang ruas signifikan pada level 95 persen, tetapi variabel LHR tidak signifikan. Sebagaimana dengan model yang dikembangkan dengan distribusi Poisson, baik koefisien panjang ruas maupun LHR mempunyai peluang untuk bernilai nol. Model, baik dengan menggunakan distribusi Poisson

maupun distribusi negatif binomial tidak memberikan metode yang baik untuk membangun tingkat kecelakaan. Untuk itu dilakukan transformasi logaritma variabel LHR. Dengan transformasi logaritma variabel LHR output yang diberikan oleh fungsi glm.nb dari perangkat lunak R adalah sebagai berikut.

Coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.42665	1.71063	-0.249	0.8030
Panjang	0.07448	0.03131	2.379	0.0174 *
log(lhrt)	0.38224	0.15864	2.410	0.0160 *
Null deviance	: 71.362 on 56 degrees of freedom			
Residual deviance	: 60.939 on 54 degrees of freedom			
AIC	: 571.31			



Gambar 3. (a) Nilai prediksi vs. residu untuk kecelakaan total, panjang ruas dan volume lalu lintas dengan regresi binomial negatif, (b) *Normal quantile plot* untuk kecelakaan total, panjang ruas dan volume lalu lintas dengan regresi binomial negatif

Sehingga model yang dihasilkan adalah:
 $Ln(Total) = -0.427 + 0.075(Panjang) + 0.382 * Ln(LHR)$.
 Variabel panjang ruas dan $Ln(LHR)$ signifikan pada level 95 persen. Nilai *deviance* dibagi dengan derajat kebebasan mendekati satu.

Gambar 3(a) yang merupakan plot *deviance residual* terhadap nilai prediksi menunjukkan gejala varian konstan, dimana sebagian besar titik terletak pada rentang yang konstan disekitar nol. *Normal quantile plot* pada

Gambar 3(b) menunjukkan penyimpangan yang kecil terhadap distribusi normal.

KESIMPULAN

Formula *accident rate* yang selama ini digunakan kurang valid, karena hubungan antara jumlah kecelakaan dengan LHR maupun panjang ruas pada kenyataannya tidak bersifat linier. Model regresi linier dengan distribusi normal tidak memadai untuk memprediksi kecelakaan, karena dari data yang digunakan

dalam penelitian ini ditemukan bahwa tingkat kecelakaan tidak terbukti mengikuti distribusi normal.

Regresi Poisson lebih mencerminkan karakteristik kejadian kecelakaan, yang diketahui bersifat diskrit, acak (random), dan langka. Namun, penggunaan distribusi Poisson yang mengasumsikan kesamaan *varian* dan *mean* tidak tercermin dalam data; kejadian kecelakaan yang dikaji dalam penelitian ini mengindikasikan gejala *overdispersi*.

Sebagai alternatif dari regresi Poisson, regresi dengan menggunakan distribusi Binomial Negatif paling baik merepresentasikan kejadian kecelakaan dengan adanya gejala *overdispersi* pada data.

DAFTAR PUSTAKA

- Allison, Paul D. 1999. *Logistic Regression Using the SAS System: Theory and Application*. Cary, NC.: SAS Institute Inc.
- Bonneson, J. A., and P. T. McCoy. 1997. "Effect of Median Treatment on Urban Arterial Safety: An Accident Prediction Model," *Highway Research Record* 1581, Highway Research Board, Washington, D.C.
- Brown, H. C. and A. P. Tarko. 1999. "Effects of Access Control on Safety on Urban Arterial Streets," *Highway Research Record* 1665, Highway Research Board, Washington, D.C.
- Devore, Jay L., 1991. *Probability and Statistics for Engineering and the Sciences*. Brooks/Cole Publishing Company, Pacific Grove, California.
- Garber, N.J. and A.A. Ehrhart, 2000. *The effects of speed, flow, and geometric characteristics on crash rates for different types of virginia highways*. Final Report, Virginia Transportation Research Council, VTRC 00-R15.
- Hadi, M., J. Aruldas, L.F. Chow, and J.J.A. Wattleworth, 1995. Estimating safety effects of cross-section design of various highway types using negative binomial regression. *Transport. Res. Record*. 1500:169-176.
- Knuiman, M. W., Council, F.M. and D. W. Reinfurt. 1993. "Association of Median Width and Highway Accident Rates," *Highway Research Record* 1401, Highway Research Board, Washington, D.C.
- Lord, D. 2002. Application of Accident Prediction Models for Computation of Accident Risk on Transportation Network, *Highway Research Records* 1784, Highway Research Board, Washington, D.C.
- Lord, D., and B. N. Persaud. 2000. "Accident Prediction Models With and Without Trend: Application of the Generalized Estimating Equations Procedure," *Highway Research Record* 1717, Highway Research Board, Washington, D.C.
- Maher, M.J. and I. Summersgill, 1996. A comparative methodology for the fitting of predictive accident models. *Accid. Anal. And Prev.*, 28(3): 281-296.
- Miaou, S., Hu, P. S., Wright, T., Rathi, A. K. and S. C. Davis. 1992. "Relationship Between Truck Accidents and Highway Geometric Design: A Poisson Regression Approach," *Highway Research Record* 1376, Highway Research Board, Washington, D.C.
- Neter, J., W. Wasserman, and M.H. Kutner, 1999. *Applied Regression Models*, 2nd ed. Richard D. Irwin, Inc., Boston.
- Petruccelli, Joseph D, B. Nandram and M. Chen. 1999. *Applied Statistics for Engineers and Scientists*. New Jersey; Prentice Hall.
- Qin, X; Ivan, J.N.; Ravishanker, N., & Liu, J. 2005. Hierarchical Bayesian Estimation of Safety Performance Functions for Two-Lane Highways Using Markov chain Monte Carlo Modeling, *Journal of Transportation Engineering*, Vol. 131, No. 5.
- Rice, John A. 1988. *Mathematical Statistics and Data Analysis*. California; Wadsworth & Brooks/Cole Advanced Books & Software.
- Saccomano, F.F. & C. Buyco. 1988. Generalized Loglinear Models of Truck Accident Rates, *Highway Research Record* 1172, Highway Research Board, Washington, D.C.