Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)
Volume xx, No. x (2016), hal xx-xx

ANALISIS JUMLAH KLAIM ASURANSI KENDARAAN BERMOTOR RODA EMPAT MENGGUNAKAN MODEL *GENERALIZED POISSON REGRESSION* (GPR)

Cahya Septia, Dadan Kusnandar, Hendra Perdana

INTISARI

Model regresi Poisson digunakan untuk menganalisis data diskrit. Model regresi Poisson memiliki asumsi bahwa nilai mean dan variansnya sama (equidispersi). Namun pada penelitian ini ditemukan nilai variansi lebih besar daripada nilai rata-rata yang dikenal dengan overdispersi. Overdispersi akan menyebabkan nilai standar error menjadi underestimate. Oleh karena itu perlu dilakukan pendekatan dengan model regresi yang lebih sesuai dalam hal ini digunakan model Generalized Poisson Regression (GPR). Model GPR dapat digunakan pada data baik dalam keadaan equidispersi, underdispersi maupun overdispersi. Penelitian ini diaplikasikan pada data jumlah klaim asuransi kendaraan bermotor roda empat di PT. BUMIDA Kalimantan Barat tahun 2015. Estimasi parameter menggunakan Maximum Likelihood Estimation. Setelah parameter diestimasi didapat faktor-faktor yang mempengaruhi jumlah klaim asuransi. Hasil estimasi yang diperoleh kemudian dipilih dengan melakukan pemilihan model terbaik berdasarkan nilai AIC, BIC yang paling kecil. Dari keempat model yang diestimasi diperoleh model terbaik adalah model GPR dengan semua rating factors.

***Kata kunci**: Regresi Poisson, Overdispersi, Maximum Likelihood Estimation*

PENDAHULUAN

Asuransi sebagai sebuah mekanisme saling menanggung bagi sebuah perlindungan, merupakan langkah yang tepat untuk seseorang dalam membagi atau mengalihkan resiko, karena asuransi menjawab kebutuhan rasa aman bagi setiap orang [1]. Berdasarkan jenis usahanya, asuransi terdiri dari asuransi kerugian (*non-life insurance*), asuransi jiwa (*life insurance*), dan reasuransi. Asuransi kerugian menurut UU No. 2 Tahun 1992 adalah usaha yang memberikan jasa-jasa dalam penanggungan risiko atas kerugian, kehilangan manfaat, dan tanggung jawab hukum kepada pihak ketiga yang timbul dari peristiwa tersebut. Asuransi jiwa adalah suatu asuransi yang bertujuan untuk memberikan proteksi terhadap orang per individu dan atau per kelompok (keluarga) atas kerugian finansial tak terduga. Sedangkan reasuransi adalah asuransi yang memberikan jasa dalam asuransi ulang terhadap resiko yang dihadapi oleh perusahaan asuransi kerugian atau perusahaan asuransi jiwa [2]. Jenis asuransi kerugian diantaranya asuransi kebakaran, pengangkutan, dan asuransi kendaraan. Asuransi kebakaran memberikan proteksi terhadap kerugian akibat kebakaran rumah, kantor, hotel. Sedangkan asuransi pengangkutan memberikan proteksi selama masa pengangkutan barang lewat darat, laut maupun udara. Jenis asuransi kerugian lainnya adalah asuransi kendaraan dan asuransi kecelakaan yang memberikan proteksi terhadap kerugian akibat kecelakaan, kerusakan, dan kehilangan [2]. Asuransi kendaraan merupakan salah satu asuransi yang banyak diminati, karena asuransi kendaraan adalah suatu bentuk asuransi yang menjamin kerugian, kerusakan, dan kehilangan atas kendaraan bermotor akibat terjadinya risiko yang menimpa obyek penanggung [1].

Analisis regresi adalah suatu metode yang digunakan untuk menganalisis hubungan antara variabel respon dan variabel prediktor. Umumnya analisis regresi digunakan untuk menganalisis data variabel respon yang merupakan data kontinu. Namun pada aplikasinya, analisis yang digunakan dapat berupa data yang variabel responnya merupakan data diskrit. Salah satu model regresi yang digunakan untuk regresi Poisson [3]. Ada asumsi yang harus dipenuhi dalam model regresi Poisson yaitu nilai variansi dan nilai mean bernilai sama. Namun dari beberapa penelitian, banyak ditemukan nilai variansi lebih besar dari nilai mean. Ini menunjukkan terjadinya overdispersi pada data diskrit tersebut. Adanya fenomena tersebut mengakibatkan dugaan model regresi Poisson menjadi kurang tepat karena mengakibatkan standar eror yang cukup besar. Ada beberapa metode yang bisa digunakan dalam penanganan overdispersi pada data diskrit diantaranya regresi Binomial Negatif, *Generalized Poisson Regression* (GPR), dan *Quasi-Likelihood* [4].

Penelitian ini bertujuan untuk membentuk model GPR dari data klaim asuransi kendaraan bermotor roda empat di PT. BUMIDA Kalimantan Barat tahun 2015. Data tersebut dibentuk kedalam *rating factors* dan *classes factors*. Variabel responnya adalah jumlah klaim kendaraan bermotor roda empat tahun 2015, sedangkan *rating factors* terdiri dari harga pertanggungan, usia kendaraan, lokasi penggunaan, dan jenis kendaraan. Langkah-langkah analisisnya dimulai dengan pengambilan data, dilanjutkan dengan pemeriksaan variabel respon dan variabel prediktor, membentuk model regresi Poisson, pemeriksaan overdispersi. Jika data mengalami overdispersi maka dilanjutkan dengan membentuk model GPR dan mengestimasi model tersebut menggunakan *Maximum Likelihood Estimation* (MLE). Model terbaik ditentukan berdasarkan nilai *Akaike Information Criteria* (AIC) dan *Bayesian Schwartz Information Criteria* (BIC) yang paling kecil.

DISTRIBUSI POISSON

Distribusi Poisson merupakan salah satu distribusi peluang bagi variabel acak diskrit. Pada data yang variabel prediktornya berbentuk *rating factors* fungsi kepadatan probabilitasnya didefinisikan oleh persamaan: [5]

$$f(Y_i|x_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, y = 0,1,2,\ldots \text{dan } i = 1,2,\ldots,n \tag{1}$$

dengan: $E(Y_i|x_i) = Var(Y_i|x_i) = \mu_i,\ i = 1,2,\ldots,n$

$y_i =$ banyaknya klaim yang terjadi

MODEL REGRESI POISSON

Model Regresi Poisson merupakan *Generalized Linear Model* (GLM) yang data dari variabel responnya diasumsikan berdistribusi Poisson. Model regresi Poisson ditunjukkan sebagai berikut:

$$\mu_i = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p)$$

dengan:

$\mu_i =$ jumlah rata-rata pada data

$\beta_i =$ parameter yang diestimasi

$x_p =$ variabel prediktor dari data

Model Regresi Poisson merupakan model regresi nonlinear yang digunakan untuk menganalisis data diskrit. Langkah pertama untuk mengestimasi parameter regresi Poisson adalah menentukan fungsi *likelihood*. Berdasarkan Persamaan (1) fungsi *likelihood* tersebut ditunjukkan oleh Persamaan (2).

$$L(\beta; y_i) = \prod_{i=1}^{n} f_i\,(y_i;\beta)$$

$$= \frac{\left[\prod_{i=1}^{n} \mu_i^{y_i}\right]\exp\left[-\sum_{i=1}^{n}\mu_i\right]}{\prod_{i=1}^{n} y_i!} \tag{2}$$

dengan $\mu_i = \exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)$

Kemudian dibentuk fungsi *log-likelihood* nya yang ditunjukkan oleh Persamaan (3).

$$\ln L(\boldsymbol{\beta}; y_i) = \sum_{i=1}^{n} y_i\,(\boldsymbol{x}_i^T\boldsymbol{\beta}) - \sum_{i=1}^{n} \exp(\boldsymbol{x}_i^T\boldsymbol{\beta}) - \sum_{i=1}^{n} \ln y_i! \quad (3)$$

GENERALIZED POISSON REGRESSION (GPR)

Model GPR merupakan suatu model regresi yang digunakan untuk menganalisis hubungan antara sebuah variabel dependen yang berupa data diskrit dengan satu atau beberapa variabel independen. Model GPR dapat digunakan baik dalam keadaan equidispersi, underdispersi, atau overdispersi. Fungsi kepadatan peluang distribusi *Generalized Poisson Regression* (GPR) adalah [6]:

$$P_r(Y_i = y_i) = \left(\frac{\mu_i}{1+\alpha\mu_i}\right)^{y_i} \frac{(1+\alpha y_i)^{y_i-1}}{y_i!} \exp\left(-\frac{\mu_i(1+\alpha y_i)}{1+\alpha y_i}\right), y = 0,1,2,... \quad (4)$$

dimana mean adalah $E(Y_i|\boldsymbol{x}_i) = \mu_i = \exp(\boldsymbol{x}_i^T\boldsymbol{\beta})$ dan variansi adalah $Var(Y_i|\boldsymbol{x}_i) = \mu_i(1+\alpha\mu_i)^2$ dengan α adalah parameter disperse. Fungsi kepadatan peluang akan berdistribusi Poisson jika $\alpha = 0$ sehingga nilai mean sama dengan variansi, yaitu $E(Y_i|\boldsymbol{x}_i) = Var(Y_i|\boldsymbol{x}_i)$. Jika $\alpha > 0$ maka hal ini menunjukkan terjadinya overdispersi. Jika diasumsikan bahwa

$$E(y_i) = \mu(\boldsymbol{x}_i, \boldsymbol{\beta}) = \exp(\boldsymbol{x}_i^T\boldsymbol{\beta})$$

dengan $\boldsymbol{x}_i$ adalah vektor yang berukuran p×1 yang menjelaskan variabel independen dan $\boldsymbol{\beta}$ adalah vektor yang berukuran p×1 merupakan parameter regresi, maka fungsi *log-likelihood* dari model GPR adalah

$$\begin{aligned} L(\beta,\alpha,y_i) &= \prod_{i=1}^{n} f(\beta,\alpha,y_i) \\ &= \prod_{i=1}^{n}\left[\left(\frac{\mu_i}{1+\alpha\mu_i}\right)^{y_i} \frac{(1+\alpha y_i)^{y_i-1}}{y_i!} \exp\left(-\frac{\mu_i(1+\alpha y_i)}{1+\alpha\mu_i}\right)\right] \end{aligned} \quad (5)$$

$$\begin{aligned} \ln L(\beta,\alpha,y_i) &= \ln\left[\prod_{i=1}^{n}\left(\frac{\mu_i}{1+\alpha\mu_i}\right)^{y_i} \prod_{i=1}^{n}\frac{(1+\alpha y_i)^{y_i-1}}{y_i!} \prod_{i=1}^{n}\exp\left(-\frac{\mu_i(1+\alpha y_i)}{1+\alpha\mu_i}\right)\right] \\ &= \sum_{i=1}^{n} y_i \ln\frac{\mu_i}{1+\alpha\mu_i} + \sum_{i=1}^{n}(y_i-1)\ln(1+\alpha y_i) - \sum_{i=1}^{n}\ln\exp\left(\frac{\mu_i(1+\alpha y_i)}{1+\alpha\mu_i}\right) - \sum_{i=1}^{n}\ln y_i! \end{aligned} \quad (6)$$

Karena $\mu_i = \exp(x_i^T\beta)$ maka

$$\begin{aligned} \ln L(\beta,\alpha,y_i) &= \sum_{i=1}^{n} y_i \ln\frac{\exp(x_i^T\beta)}{1+\alpha\exp(x_i^T\beta)} + \sum_{i=1}^{n}(y_i-1)\ln(1+\alpha y_i) - \sum_{i=1}^{n}\frac{\exp(x_i^T\beta)(1+\alpha y_i)}{1+\alpha\exp(x_i^T\beta)} - \sum_{i=1}^{n}\ln y_i! \\ &= \sum_{i=1}^{n}\left(y_i(x_i^T\beta) - (\ln 1+\alpha\exp(x_i^T\beta))\right) + \sum_{i=1}^{n}(y_i-1)\ln(1+\alpha y_i) - \sum_{i=1}^{n}\frac{\exp(x_i^T\beta)(1+\alpha y_i)}{1+\alpha\exp(x_i^T\beta)} \\ &\quad - \sum_{i=1}^{n}\ln y! \end{aligned} \quad (7)$$

Penduga MLE bagi β dan α diperoleh dengan menurunkan Persamaan (7) terhadap β dan α dan menyamakannya dengan nol. Sehingga diperoleh turunan $\ln L(\beta,\alpha,y_i)$ terhadap β adalah sebagai berikut:

$$\sum_{i=1}^{n}(y_i x_i - \alpha x_i) = \frac{((1+\alpha y_i)\beta\exp(x_i^T\beta))(1+\alpha\exp(x_i^T\beta)) - (\exp(x_i^T\beta)(1+\alpha y_i))(\alpha\beta\exp(x_i^T\beta))}{(1+\alpha\exp(x_i^T\beta))^2} \quad (8)$$

Sedangkan turunan $\ln L(\beta, \alpha, y_i)$ terhadap α adalah sebagai berikut:

$$\sum_{i=1}^{n}\exp(x_i^T\beta)+\sum_{i=1}^{n}\frac{y_i^{\,2}-y_i}{(1+\alpha y_i)}=\sum_{i=1}^{n}\frac{\left(\exp(x_i^T\beta)y_i(1+\alpha\exp(x_i^T\beta))\right)-\left((1-\alpha y_i)^2(\exp(x_i^T\beta))^2\right)}{(1+\alpha\exp(x_i^T\beta))^2} \quad (9)$$

Solusi eksplisit dari Persamaan (8) dan (9) jarang bisa diselesaikan secara analitik dan harus dilakukan secara iteratif. Dalam penelitian ini dilakukan dengan bantuan komputer menggunakan metode *Newton-Raphson.*

APLIKASI NUMERIK

Data yang digunakan dalam penelitian ini adalah data sekunder mengenai data klaim asuransi kendaraan bermotor roda empat yang meliputi data klaim kendaraan bermotor roda empat tahun 2015, harga pertanggungan, usia kendaraan lokasi kejadian dan jenis kendaraan. Data-data tersebut dikelompokkan ke dalam *rating factors* dan *rating classes* sebagai berikut:

1. Harga Pertanggungan : < 100 juta, 100-200 juta, 200-300 juta, 300-400 juta, >400 juta
2. Usia Kendaraan : < 2 tahun, 2-4 tahun, 4-7 tahun, >7 tahun
3. Lokasi Penggunaan : Pontianak dan luar Pontianak
4. Jenis Kendaraan : angkutan penumpang dan angkutan barang

Dari uraian tersebut, terdapat 13 *rating classes* yang akan digunakan sebagai variabel prediktor. Setelah dihitung, didapat nilai mean dan variansi dari jumlah klaim sebagai berikut

$$E(Y)=\bar{y}=\frac{\sum_{i=1}^{80}y_i}{80}=1{,}6125$$

$$Var(Y)=\frac{\sum_{i=1}^{80}(y_i-\bar{y})}{79}=14{,}51883$$

Data tersebut kemudian dianalisis lalu diestimasi dengan menggunakan MLE melalui program R. Hasil estimasi ditunjukkan dalam Tabel 1.

Tabel 1 Hasil Estimasi Parameter GPR dengan semua *rating factors*

	Variabel	Estimasi	SE	*p-value*
D_{11}	Harga 100-200 juta	1,81417	0,14500	0,00000
D_{12}	Harga 200-300juta	0,87807	0,18100	0,00000
D_{13}	Harga 300-400juta	-1,00321	0,44800	0,02500
D_{14}	Harga >400juta	-0,35799	0,32500	0,27100
D_{21}	2-4 tahun	-0,10042	0,22300	0,65200
D_{22}	4-7 tahun	1,12941	0,14500	0,00000
D_{23}	>7 tahun	0,46916	0,17100	0,00600
D_{31}	Pontianak	-0,00023	0,01600	0,98900
D_{41}	Angkutan penumpang	-3,33597	0,44300	0,00000

Berdasarkan Tabel 1 didapat faktor-faktor yang mempengaruhi jumlah klaim adalah:

1. Harga Pertanggungan : Harga 100-200 juta, Harga 200-300 juta dan Harga 300-400 juta
2. Usia Kendaraan : Usia 4-7 tahun dan >7 tahun
3. Jenis Kendaraan : Angkutan penumpang

Berdasarkan Tabel 1 lokasi penggunaan Pontianak tidak signifikan, sehingga *rating factors* untuk lokasi penggunaan dikeluarkan dari model dan didapat hasil estimasinya ditunjukkan pada Tabel 2.

Berdasarkan Tabel 2 didapat faktor-faktor yang mempengaruhi jumlah klaim adalah:

1. Harga Pertanggungan : Harga 100-200 juta, Harga 200-300 juta dan Harga 300-400 juta
2. Usia Kendaraan : Usia 4-7 tahun dan >7 tahun
3. Jenis Kendaraan : Angkutan penumpang

Tabel 2 Hasil Estimasi Parameter GPR tanpa Lokasi

	Variabel	Estimasi	SE	*P value*
D_{11}	Harga 100-200 juta	1,81000	0,14000	0,00000
D_{12}	Harga 200-300juta	0,88000	0,18000	0,00000
D_{13}	Harga 300-400juta	-1,00000	0,45000	0,02500
D_{14}	Harga >400juta	-0,36000	0,33000	0,27100
D_{21}	2-4 tahun	-0,10000	0,22000	0,65200
D_{22}	4-7 tahun	1,13000	0,14000	0,00000
D_{23}	>7 tahun	0,47000	0,17000	0,00600
D_{41}	Angkutan penumpang	-3,34000	0,44000	0,00000

Dari empat estimasi model yang telah didapat, yaitu Regresi Poisson dengan semua *rating factors*, regresi Poisson tanpa lokasi, GPR dengan semua *rating factors* dan GPR tanpa lokasi diperoleh nilai AIC, BIC dan *Log-likelihood* yang ditunjukkan pada Tabel 3.

Tabel 3 Nilai AIC, BIC

Model	AIC	BIC
GP dengan semua *rating factors*	186,3892	210,2090
GP tanpa lokasi	184,3970	205,8352
GPR dengan semua *rating factors*	133,9459	112,5077
GPR tanpa lokasi	135,9457	116,8895

Berdasarkan Tabel 3 diperoleh nilai AIC dan BIC yang paling kecil adalah estimasi pada GPR dengan semua *rating factors*, sehingga diperoleh model *Generalized Poisson Regression* sebagai berikut:

$$\begin{aligned}\mu = \exp(&1{,}81417D_{11} + 0{,}87807D_{12} - 1{,}00321D_{13} - 0{,}35799D_{14}\\ &-0{,}10042D_{21} + 1{,}12941D_{22} + 0{,}46916D_{23} - 0{,}00023D_{31} - 3{,}33597D_{41})\end{aligned}$$

Berdasarkan faktor-faktor tersebut, jika nasabah mendaftarkan mobilnya dengan harga pertanggungan 100-200 juta, usia kendaraan 4-7 tahun, lokasi penggunaan di Pontianak dan mobil berjenis angkutan penumpang maka dapat ditentukan banyaknya klaim yang akan terjadi, yaitu:

$$\begin{aligned}\mu_i &= \exp(\beta_0 + \beta_{11}D_{11} + \beta_{22}D_{22} + \beta_{31}D_{31} + \beta_{41}D_{41})\\ &= \exp(1{,}81417 + 1{,}12941 - 0{,}00023 - 3{,}33597)\\ &= \exp(-0{,}39267)\\ &= 0{,}675643 \approx 1\end{aligned}$$

Didapat rata-rata klaim yang terjadi sebanyak 1 kali.

Walaupun pada GPR dengan semua *rating factor*s ada *rating factor* yang tidak signifikan, namun tetap digunakan karena model tersebut merupakan model terbaik. Selain itu, dicari nilai ekspektasi dan variansi dari model terbaik sebagai bentuk penanganan. Nilai α yang didapat sebesar $-0{,}05$ sehingga nilai variansinya sebagai berikut:

$$\begin{aligned}Var &= \mu(1+\alpha\mu)^2\\ &= 1{,}6125(1 + (-0{,}05)1{,}6125)^2\\ &= 1{,}362957\end{aligned}$$

Dari nilai variansi pada regresi Poisson diperoleh 14,51883 dan pada GPR nilai variansinya sebesar 1,362975 dan tidak terlalu jauh dengan nilai mean yaitu sebesar 1,6125 sehingga menunjukkan bahwa overdispersi teratasi

PENUTUP

Berdasarkan pembahasan yang telah dipaparkan, dapat diambil kesimpulan bahwa data yang digunakan mengalami overdispersi sehingga model regresi Poisson tidak tepat digunakan. Untuk mengatasi overdispersi digunakan model GPR. Dari empat estimasi model yang telah didapat, yaitu regresi Poisson dengan semua *rating factors*, regresi Poisson tanpa lokasi, GPR dengan semua *rating factors* dan GPR tanpa lokasi diperoleh nilai AIC dan BIC. Model terbaik dilihat dari nilai AIC dan BIC yang paling kecil, dan diperoleh model terbaik adalah GPR dengan semua *rating factors*. Dari model terbaik dicari nilai ekspektasi dan variansi untuk menunjukkan overdispersi teratasi dan didapat nilai variansi sebesar 1,6975 sehingga overdispersi teratasi.

DAFTAR PUSTAKA

[1]. Sastrawidjaja M.S. *Aspek-Aspek Hukum Asuransi dan Surat Berharga*. Bandung: Alumni; 2003.

[2]. Abdulkadir M. *Hukum Asuransi Indonesia*. Bandung: Citra Aditya Bakti; 2006.

[3]. Simarmata R.T. Penanganan Overdispersi Pada Model Regresi Poisson Menggunakan Model Regresi Binomial Negatif. *Jurnal Matematika*. 2011; 4: 95-104.

[4]. Jong P.D, Heller G.Z. *Generalized Linear Models for Insurance Data*. Cambridge: Cambridge University Press; 2008.

[5]. Kusnandar D. *Metode Statistik dan Aplikasinya dengan MINITAB dan Excel*. Yogyakarta: Madyan Press; 2004.

[6]. Hosmer D.W, Lemeshow S. *Applied Logistic Regression*. 2nd Edition. New York: John Wiley and Sons. Inc; 2000.

CAHYA SEPTIA : FMIPA UNTAN, Pontianak, cahyaseptia09@yahoo.com
DADAN KUSNANDAR : FMIPA UNTAN, Pontianak, dkusnand@untan.ac.id
HENDRA PERDANA : FMIPA UNTAN, Pontianak, hendra.perdana@math.untan.ac.id