Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)
Volume 04, No. 3 (2015), hal 323-328

ESTIMASI PARAMETER MODEL REGRESI *ZERO ADJUSTED INVERSE GAUSSIAN* (ZAIG) UNTUK MENENTUKAN BESAR KLAIM

Nurul Huda, Dadan Kusnandar, Mariatul Kiftiah

INTISARI

Model regresi Zero Adjusted Inverse Gaussian (ZAIG) merupakan gabungan dari Distribusi diskrit Bernoulli dan Distribusi kontinu Inverse Gaussian. Model regresi ZAIG adalah model regresi yang dapat menangani kasus terjadi dan tidak terjadinya klaim, karena kasus tersebut berdistribusi diskrit dan kontinu. Dari kasus terjadi dan tidak terjadinya klaim tersebut dapat ditentukan perkiraan probabilitas klaim dan besar klaim pada suatu perusahaan asuransi. Pada penelitian ini, digunakan data biaya klaim pada perusahaan asuransi auto PT. Asuransi Umum Bumiputera Muda 1967 (BUMIDA) tahun 2013-2014. Data biaya klaim yang digunakan merupakan variabel respon yang berdistribusi Inverse Gaussian. Dengan mengestimasi parameter pada data tersebut, dapat diketahui faktor-faktor yang mempengaruhi probabilitas klaim dan besar klaim. Estimasi parameter dilakukan dengan cara menggunakan metode Maximum Likelihood Estimation (MLE) yang dilakukan secara iteratif dengan metode iterasi Newton-Raphson. Hasil estimasi yang diperoleh kemudian dipilih dengan melakukan pemilihan model terbaik. Hasil dari model terbaik menunjukkan bahwa pemilihan model terbaik tidak mempengaruhi probabilitas terjadinya klaim, namun berpengaruh terhadap besarnya biaya klaim yang akan dikeluarkan oleh perusahaan terhadap peserta asuransi. Dari probabilitas klaim dan besarnya klaim yang diperoleh, dapat ditentukan ekspetasi besar klaim dari suatu kendaraan yang diasuransikan.

Kata kunci: *Probabilitas klaim, besar klaim, ZAIG, MLE*

PENDAHULUAN

Suatu perusahaan asuransi harus menggunakan manajemen risiko sebagai faktor terjadinya suatu kejadian [1]. Untuk menangani klaim dari suatu kejadian, perusahaan dapat menggunakan metode prediksi. Oleh karena itu, perusahaan harus menemukan cara untuk memprediksi klaim dan biaya premi yang tepat untuk menutupi risiko. Perusahaan asuransi juga harus mempertimbangkan masalah prediksi terhadap persaingan pasar asuransi, karena memungkinkan perusahaan asuransi untuk mendapatkan keuntungan dengan menawarkan premi yang lebih rendah atau dengan menghindari terjadinya pembayaran yang tidak sesuai dengan kesepakatan awal. Dengan memperkirakan probabilitas klaim dan besar klaim perusahaan dapat menggunakan perkiraan ini untuk menawarkan diskon premi agar premi yang ditawarkan lebih rendah atau tidak menawarkan diskon premi pada klien apabila kemungkinan terjadinya klaim dari kendaraan yang diasuransikan cukup besar.

Beberapa peneliti menyarankan *Generalized Linear Model* (GLM) untuk model biaya klaim sebagai fungsi dari faktor risiko pada asuransi seperti yang terdapat pada Ismail dan Jemain (2009) ataupun Mccullagh dan Nelder (1989), dikarenakan variabel respon yang diamati pada data asuransi berbentuk *non negatif* [2]. Distribusi *Inverse Gaussian* adalah salah satu distribusi yang mengikuti GLM dengan memuat nilai variabel respon kontinu *non negatif* dan merupakan distribusi keluarga eksponensial sehingga model GLM dapat menyelesaikan permasalahan. Pada penelitian ini, digunakan model regresi *Zero Adjusted Inverse Gaussian* yang merupakan model regresi gabungan dari Distribusi *Bernoulli* dan Distribusi *Inverse Gaussian*. Model regresi ini dapat digunakan pada kasus klaim bernilai nol (tidak terjadi klaim) dan kasus terjadinya klaim, dimana masing-masing kasus berturut-turut merupakan distribusi probabilitas diskrit dan distribusi probabilitas kontinu.

Penelitian ini bertujuan untuk mengkaji cara estimasi parameter pada kasus terjadi dan tidak terjadinya klaim serta memodelkan data dengan variabel respon yang berdistribusi *Inverse Gaussian.* Data yang digunakan pada penelitian ini adalah data biaya klaim asuransi PT. Asuransi Umum

Bumiputera Muda 1967 (BUMIDA) pada tahun 2013-2014 berdasarkan usia kendaraan, kapasitas kendaraan, dan merk kendaraan. Langkah-langkah yang digunakan dalam memodelkan data yaitu terlebih dahulu dilakukan uji kelayakan data untuk mengetahui apakah data tersebut berdistribusi *Inverse Gaussian.* Apabila terbukti, maka data tersebut layak untuk digunakan oleh model, jika tidak maka digunakan data berikutnya hingga data layak untuk digunakan. Selanjutnya mengestimasi parameter pada model regresi ZAIG dengan metode MLE, yaitu membentuk fungsi *likelihood* kemudian mencari *log-likelihood.* Setelah didapat hasil estimasi, dilakukan kembali pemilihan model terbaik untuk mengetahui variabel penjelas yang mempengaruhi model. Kemudian hasil pemilihan model terbaik tersebut dimodelkan terhadap $\hat{\pi}$ untuk probabilitas klaim dan $\hat{\mu}$ untuk biaya klaim, sehingga didapat probabilitas klaim dan biaya klaimnya.

DISTRIBUSI *BERNOULLI*

Bernoulli merupakan distribusi variabel acak diskrit yang terdiri dari dua kemungkinan hasil yang berkomplementer. Dua kemungkinan hasil dari distribusi *Bernoulli* biasanya dikodekan sebagai 0 atau 1. Ketika $y = 1$ dapat disebut sebagai sukses, $y = 0$ sebagai kegagalan. Untuk $f(1) = \pi$ dan $f(0) = 1 - \pi$ dimana $0 < \pi < 1$ [3]. Penerapannya pada asuransi yaitu klaim atau tidak ada klaim. Fungsi probabilitas dapat didefinisikan sebagai berikut:

$$f(y) = \pi^y (1-\pi)^{1-y}, \qquad y = 0,1$$

dimana y dianggap sebagai variabel biner yang menunjukkan terjadinya klaim paling sedikit satu kali dan π sebagai probabilitas setidaknya satu klaim pada suatu polis. Nilai harapan dan ragamnya yaitu $E(Y) = \pi$ dan $Var(Y) = \pi(1-\pi)$, dengan

$$y_i = \begin{cases} (1-\pi), & y_i = 0 \\ \pi, & y_i > 0 \end{cases}$$

dimana y_i merupakan variabel acak untuk rata-rata besar klaim pada suatu polis i.

DISTRIBUSI *INVERSE GAUSSIAN*

Inverse Gaussian merupakan salah satu distribusi kontinu yang masuk ke dalam GLM dengan variabel respon berdistribusi *Inverse Gaussian*. Karena *Inverse Gaussian* merupakan salah satu distribusi GLM, maka komponen utama distribusi *Inverse Gaussian* yaitu komponen random, komponen sistematik dan fungsi link. Variabel respon Y_i pada Distribusi *Inverse Gaussian* merupakan variabel *non negatif* dengan fungsi kepadatan peluang yang dinotasikan sebagai $y \sim IG(\mu, \sigma^2)$ yaitu:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi y^3}} exp\left\{-\frac{1}{2y}\left(\frac{y-\mu}{\mu\sigma}\right)^2\right\}, y > 0 \tag{1}$$

dengan y : variabel acak

$f(y)$: fungsi kepadatan peluang distribusi *Inverse Gaussian*

μ, σ : parameter pada variabel respon

dengan ekspetasi dan variannya adalah $E(Y) = \mu$, $Var(Y) = \sigma^2\mu^3$ [3].

MODEL REGRESI *ZERO ADJUSTED INVERSE GAUSSIAN*

Gabungan dari distribusi diskrit dan distribusi kontinu pada Distribusi *Bernoulli* dan Distribusi *Inverse Gaussian* mengakibatkan model ini disebut sebagai model regresi *Zero Adjusted Inverse Gaussian* (ZAIG) [4]. ZAIG merupakan model regresi alternatif yang dapat digunakan untuk mencari probabilitas klaim pada besaran klaim yang bernilai nol dan positif. Fungsi probabilitas ZAIG adalah:

$$f(y) = \begin{cases} (1-\pi), & y = 0 \\ \pi . g(y), & y > 0 \end{cases}$$

Karena distribusi kontinu yang digunakan adalah *Inverse Gaussian* maka fungsi probabilitas ZAIG dapat ditulis sebagai berikut:

$$f(y) = \begin{cases} (1-\pi), & y = 0 \\ \pi . \dfrac{1}{\sigma\sqrt{2\pi y_i^3}} exp\left\{-\dfrac{1}{2y_i}\left(\dfrac{y_i-\mu}{\mu\sigma}\right)^2\right\}, & y > 0 \end{cases} \quad (2)$$

dan ekspetasinya yaitu:

$$E(Y) = \pi\mu \quad (3)$$

dengan π adalah probabilitas klaim dan μ sebagai besarnya klaim.

Model regresi ZAIG merupakan gabungan dari distribusi *Bernoulli* dan *Inverse Gaussian*, maka fungsi *link logit* dan *log* dapat digunakan dalam regresi ZAIG. Fungsi *logit* dan *log* disebut sebagai fungsi link karena menghubungkan π dan μ dengan fungsi linear $X_i'\beta$, sehingga model regresi ZAIG dapat ditulis dalam bentuk

$$E(Y_i) = \pi = \frac{e^{\pi}}{1+e^{\pi}}, \text{ dengan } g(\pi) = \ln\frac{\pi}{(1-\pi)}$$

$$E(Y_i) = \mu = e^{\mu}, \text{ dengan } g(\mu) = \ln\mu$$

Untuk mendapatkan estimator parameter masing-masing pada π dan μ, dibentuk persamaan estimasi regresi ZAIG yaitu:

$$\hat{\pi} = \frac{e^{X_i'\beta_\pi}}{1+e^{X_i'\beta_\pi}} \quad (4)$$

$$\hat{\mu} = e^{X_i'\beta_\mu} \quad (5)$$

dengan $\hat{\pi}$: estimator parameter π

$\hat{\mu}$: estimator parameter μ

$i = 1,2,\ldots,n$ dan β merupakan parameter yang tidak diketahui dalam model dan harus diestimasi.

ESTIMASI PARAMETER

Estimasi parameter dilakukan untuk mengetahui peranan masing-masing variabel penjelas terhadap variabel respon yang berpotensi menjadi faktor resiko. Nilai estimasi parameter dapat dicari menggunakan metode *Maximum Likelihood Estimation* (MLE) dengan mengetahui fungsi kepadatan peluang regresi ZAIG. Pada model regresi ZAIG, β_π dan β_μ adalah parameter yang akan diestimasi menggunakan metode MLE. Estimasi parameter β_π dan β_μ pada model regresi ZAIG dapat diperoleh dengan memaksimumkan fungsi *likelihood*. Diketahui bahwa fungsi *likelihood* sebagai berikut:

$$L(\theta) = \prod_{i=1}^{n} f(y_i;\theta) \quad (6)$$

Kemudian ditentukan fungsi *likelihood* dari model regresi ZAIG, dengan mengasumsikan $y_1, y_2, \ldots, y_n$ merupakan variabel acak ZAIG yang saling bebas. Sehingga didapat fungsi *likelihood* pada Persamaan (6) menjadi:

$$\begin{aligned} L(\beta_\mu, \beta_\pi) &= \prod_{i=1}^{n} f(y_i;\mu,\pi) \\ &= \prod_{i=1}^{n}(1-\pi)\prod_{i=1}^{n}\pi \cdot \frac{1}{\sigma\sqrt{2\pi y_i{}^3}} exp\left\{-\frac{1}{2y_i}\left(\frac{y_i-\mu}{\mu\sigma}\right)^2\right\} \end{aligned}$$

atau fungsi *log-likelihood*:

$$\ln L(\beta_\mu, \beta_\pi) = \sum_{i=1}^{n}\ln(1-\pi) + \sum_{i=1}^{n}\ln\frac{\pi}{\sigma\sqrt{2\pi y_i{}^3}} + \sum_{i=1}^{n}\left[-\left\{\frac{1}{2y_i}\left(\frac{y_i-\mu}{\mu\sigma}\right)^2\right\}\right]$$

HASIL DAN ANALISIS DATA

Data yang digunakan pada penelitian ini adalah data biaya klaim asuransi auto PT. Asuransi Umum Bumiputera Muda 1967 (BUMIDA) Pontianak. Data yang digunakan dalam penelitian ini merupakan data biaya klaim asuransi pada tahun 2013-2014. Dari 169 data auto yang diasuransikan pada perusahaan ini, tercatat 154 kendaraan yang telah mengajukan klaim. Dari data yang dilampirkan terdapat tiga faktor yang digunakan sebagai variabel pada penelitian ini, yaitu merk, usia dan kapasitas dari setiap kendaraan. Ketiga faktor tersebut juga memiliki subfaktor-subfaktor, yaitu:

Merk (M) : Chevrolet (m_1), Daihatsu (m_2), Suzuki (m_3), Honda (m_4), Mitsubishi (m_5), Nissan (m_6),Ford (m_7) dan Toyota (m_8).

Usia (U) : 2th (u_1), 3th (u_2), 4th (u_3), 5th (u_4), 6th (u_5), 7th (u_6) dan 1th (u_7).

Kapasitas (K) : 1200cc (k_1), 1300cc (k_2), 1400cc (k_3), 1500cc (k_4), 1600cc (k_5), 1800cc (k_6), 2000cc (k_7), 2400cc (k_8), 2500cc (k_9), 998cc (k_{10}) dan 2405 (k_{11}).

Untuk mendapatkan nilai estimasi, data tersebut diestimasi menggunakan MLE melalui program R. Hasil dari estimasi tersebut ditunjukan dalam Tabel 1.

Tabel 1. Hasil Estimasi parameter menggunakan program R

Variabel	Parameter			
	β_μ		β_π	
	Estimasi	p value	Estimasi	p value
D_{m_1}	16.05	<2e-16	-2.39	0.231
D_{m_2}	14.65	<2e-16	-1.60	0.342
D_{m_3}	15.85	<2e-16	-1.42	0.423
D_{m_4}	14.55	<2e-16	-1.92	0.339
D_{m_5}	14.52	<2e-16	-2.34	0.231
D_{m_6}	14.37	<2e-16	-14.32	0.976
D_{m_7}	14.71	<2e-16	0.92	0.657
D_{m_8}	15.17	<2e-16	-3.19	0.103
D_{u_1}	-0.53	0.193	0.11	0.932
D_{u_2}	-0.24	0.489	-0.78	0.604
D_{u_3}	-0.53	0.143	-1.46	0.345
D_{u_4}	-0.47	0.200	-1.55	0.344
D_{u_5}	-0.76	0.054	-13.57	0.977
D_{u_6}	0.25	0.691	0.05	0.974
D_{k_1}	-0.18	0.456	0.16	0.912
D_{k_2}	1.36	0.007	1.00	0.592
D_{k_3}	0.53	0.128	0.82	0.572
D_{k_4}	0.88	0.004	-1.46	0.391
D_{k_5}	0.76	0.254	1.82	0.242
D_{k_6}	0.31	0.387	0.35	0.838
D_{k_7}	0.97	0.009	-0.48	0.763
D_{k_8}	0.56	0.243	1.95	0.238
D_{k_9}	0.60	0.332	-14.52	0.983
$D_{k_{10}}$	-0.67	0.016	-12.87	0.985

Karena variabel independennya berupa data kualitatif maka data tersebut dikuantitatifkan dengan mengunakan variabel dummy. Banyaknya varibel dummy sama dengan banyaknya faktor dikurang satu. Sehingga faktor-faktor tersebut menjadi:

Merk (D_m) : Chevrolet (D_{m_1}), Daihatsu (D_{m_2}), Suzuki (D_{m_3}), Honda (D_{m_4}), Mitsubishi (D_{m_5}), Nissan (D_{m_6}),Ford (D_{m_7}) dan Toyota (D_{m_8}).

Usia (D_u) : 2th (D_{u_1}), 3th (D_{u_2}), 4th (D_{u_3}), 5th (D_{u_4}), 6th (D_{u_5}) dan 7th (D_{u_6}).

Kapasitas (D_k) : 1200cc (D_{k_1}), 1300cc (D_{k_2}), 1400cc (D_{k_3}), 1500cc (D_{k_4}), 1600cc (D_{k_5}), 1800cc (D_{k_6}), 2000cc (D_{k_7}), 2400cc (D_{k_8}), 2500cc (D_{k_9}) dan 998cc ($D_{k_{10}}$).

berdasarkan faktor-faktor dari Tabel 1 yang berupa variabel dummy, dapat ditulis model untuk probabilitas klaim yaitu:

$$\hat{\pi} = \frac{e^{X'\beta_\pi}}{1+e^{X'\beta_\pi}}$$
$$= \frac{e^{\beta_0+\beta_{\pi_{m_1}}D_{m_1}+\beta_{\pi_{m_2}}D_{m_2}+\cdots+\beta_{\pi_{m_7}}D_{m_7}+\beta_{\pi_{u_1}}D_{u_1}+\beta_{\pi_{u_2}}D_{u_2}+\cdots+\beta_{\pi_{u_5}}D_{u_5}+\beta_{\pi_{k_1}}D_{k_1}+\beta_{\pi_{k_2}}D_{k_2}+\cdots+\beta_{\pi_{k_9}}D_{k_9}}}{1+e^{\beta_0+\beta_{\pi_{m_1}}D_{m_1}+\beta_{\pi_{m_2}}D_{m_2}+\cdots+\beta_{\pi_{m_7}}D_{m_7}+\beta_{\pi_{u_1}}D_{u_1}+\beta_{\pi_{u_2}}D_{u_2}+\cdots+\beta_{\pi_{u_5}}D_{u_5}+\beta_{\pi_{k_1}}D_{k_1}+\beta_{\pi_{k_2}}D_{k_2}+\cdots+\beta_{\pi_{k_9}}D_{k_9}}}$$

dan model untuk besar klaim yaitu:

$$\hat{\mu} = e^{X'\beta_\mu}$$
$$= e^{\beta_0+\beta_{\mu_{m_1}}D_{m_1}+\beta_{\mu_{m_2}}D_{m_2}+\cdots+\beta_{\mu_{m_7}}D_{m_7}+\beta_{\mu_{u_1}}D_{u_1}+\beta_{\mu_{u_2}}D_{u_2}+\cdots+\beta_{\mu_{u_5}}D_{u_5}+\beta_{\mu_{k_1}}D_{k_1}+\beta_{\mu_{k_2}}D_{k_2}+\cdots+\beta_{\mu_{k_9}}D_{k_9}}$$

Setelah dilakukan estimasi parameter, terdapat beberapa variabel pada parameter β_μ yang signifikan, yaitu untuk variabel Merk. Sedangkan untuk parameter β_π tidak terdapat variabel yang signifikan terhadap parameter. Secara umum dan menurut perusahaan asuransi, faktor Usia dan Kapasitas kendaraan sebenarnya mempengaruhi terjadinya klaim dan besar klaim karena semakin bertambahnya usia suatu kendaraan maka semakin besar peluang terjadinya klaim dan besarnya klaim. Akan tetapi, pada penelitian ini variabel yang tidak signifikan mempengaruhi model tetap digunakan karena variabel usia dan kapasitas kendaraan juga berpengaruh terhadap peluang terjadinya klaim dan besar klaim.

Berdasarkan faktor-faktor yang mempengaruhi probabilitas klaim pada Tabel 1, perusahaan asuransi dapat menetukan probabilitas dan besar klaim suatu kendaraan yang akan diasuransikan. Sebagai contohnya, jika seseorang mendaftarkan mobilnya yang bermerk Honda dengan kapasitas 2400cc dan berusia 2th. Maka dapat ditentukan probabilitas klaim dan besar klaim yang akan terjadi yaitu:

$$\hat{\pi} = \frac{e^{\beta_{\pi_{m_4}}D_{m_4}+\beta_{\pi_{u_1}}D_{u_1}+\beta_{\pi_{k_8}}D_{k_9}}}{1+e^{\beta_{\pi_{m_4}}D_{m_4}+\beta_{\pi_{u_1}}D_{u_1}+\beta_{\pi_{k_8}}D_{k_9}}}$$
$$= \frac{e^{(-1,92)+0,11+1,95}}{1+e^{(-1,92)+0,11+1,95}}$$
$$= 0,53$$

dan

$$\hat{\mu} = e^{\beta_{\mu_{m_4}}D_{m_4}+\beta_{\mu_{u_1}}D_{u_1}+\beta_{\mu_{k_8}}D_{k_8}}$$
$$= e^{14,55+(-0,53)+0,56}$$
$$= 2.147.897,46$$

Dari perhitungan tersebut didapat besarnya klaim sebesar Rp2.147.897,46 dengan probabilitas terjadinya klaim sebesar 0,53 menggunakan Persamaan (4) dan (5). Kemudian didapat ekspetasi besarnya klaim dari data terjadi dan tidak terjadinya klaim untuk peserta asuransi yang bermerk Honda, kapasitas 2400cc dan berusia 2th adalah sebesar Rp1.138.385,65.

PENUTUP

Berdasarkan pembahasan yang telah dipaparkan, dapat diambil kesimpulan bahwa model regresi ZAIG dapat digunakan untuk menentukan probabilitas klaim dan besar klaim. Berdasarkan data biaya klaim yang telah diestimasi, diketahui bahwa dari tiga faktor dengan total 24 variabel yang digunakan pada penelitian ini, semua variabel mempengaruhi probabilitas klaim dan besar klaim. Adapun faktor-faktor yang mempengaruhi besar klaim tersebut yaitu Merk kendaraan, Usia kendaraan dan Kapasitas kendaraan. Dari faktor-faktor yang mempengaruhi tersebut, suatu perusahaan asuransi dapat memperkirakan besarnya klaim dan probabilitas terjadinya klaim dari suatu kendaraan yang akan diasuransikan pada perusahaannya.

DAFTAR PUSTAKA

[1]. Meulbroek, L.K. The Efificiency of Equity-linked Compensation: Understanding the Full Cost of Awarding Executive Stock Options. *Financial Management*. 2001; 30(2):5-30.

[2]. Hogg R.V, Klugman S.A, *Loss Distributions: 1st Ed*, John Wiley and Sons. New York: 2009.

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

[4]. Bortoluzzo A.B, Claro D.P, Caetano M.A.L, Artes R, Estimating Total Claim Size in the Auto Insurance Industry: a Comparison between Tweedie and Zero-Adjusted Inverse Gaussian Distribution. *Brazilian Administration Review, Curitiba*. 2011; 8:37-47.

NURUL HUDA : Jurusan Matematika FMIPA UNTAN, Pontianak, Nurulhuda0311@gmail.com

DADAN KUSNANDAR : Jurusan Matematika FMIPA UNTAN, Pontianak, dkusnand@yahoo.com

MARIATUL KIFTIAH : Jurusan Matematika FMIPA UNTAN, Pontianak, kiftiahmariatul@ymail.com