

Penghitungan Premi Asuransi Kendaraan Bermotor Menggunakan *Generalized Linear Models* dengan Distribusi Tweedie

Tri Andika Julia Putra¹, Donny C. Lesmana^{2*}, I Gusti Putu Purnaba³

^{1,2,3} Departemen Matematika, Fakultas MIPA, IPB University,
Jalan Meranti, Kampus IPB Dramaga 16680, Jawa Barat, Indonesia

* Penulis Korespondensi. Email: donnylesmana@apps.ipb.ac.id

ABSTRAK

Seorang aktuaris mempunyai tugas penting dalam menentukan harga premi yang sesuai untuk setiap nasabah dengan risiko dan karakteristik yang berbeda. Banyak variabel yang dapat mempengaruhi harga premi. Oleh karena itu, aktuaris harus mengetahui variabel-variabel yang berpengaruh signifikan terhadap premi. Tujuan dari penelitian ini adalah untuk menentukan variabel yang dapat mempengaruhi besaran premi murni menggunakan distribusi campuran dalam menentukan besarnya premi melalui *Generalized Linear Models* (GLM) serta menentukan model harga premi yang sesuai berdasarkan variabel-variabel yang mempengaruhinya. Salah satu analisis statistik yang dapat digunakan untuk memodelkan premi asuransi adalah *Generalized Linear Models*. GLM merupakan perluasan dari model regresi klasik yang dapat mengakomodasi fleksibilitas untuk menggunakan beberapa distribusi data tetapi terbatas pada distribusi keluarga eksponensial. Dalam model GLM, premi diperoleh dengan mengalikan nilai ekspektasi bersyarat dari frekuensi klaim dan biaya klaim. Berdasarkan penelitian yang telah dilakukan diketahui bahwa frekuensi klaim dan besarnya klaim mengikuti distribusi Tweedie. Dari kedua model tersebut diketahui bahwa variabel yang mempengaruhi premi murni adalah jumlah anak, pendapatan per bulan, status pernikahan, pendidikan, pekerjaan, penggunaan kendaraan, besarnya *bluebook* yang dibayarkan, dan jenis kendaraan nasabah. Hal ini menunjukkan bahwa model GLM merupakan model yang representatif dan berguna bagi perusahaan asuransi.

Kata Kunci:

Premi Asuransi; *Generalized Linear Models*; Frekuensi Klaim; Biaya Klaim; Distribusi Tweedie

ABSTRACT

It is an important task for an actuary in determining the appropriate premium price for each customer with different risks and characteristics. Many variables can affect the premium price. Therefore, actuaries must determine the variables that significantly affect the premium. The purpose of this study is to determine the variables that can affect the amount of pure premium using a mixed distribution in determining the amount of premium through *Generalized Linear Models* (GLM) and determine the appropriate premium price model based on the variables that influence it. One of the statistical analyzes that can be used to model insurance premiums is the *Generalized Linear Models*. GLM is an extension of the classic regression model that can accommodate the flexibility of its users to use multiple data distributions but is limited to the exponential family distribution. In the GLM model, the premium is obtained by multiplying the conditional expected value of the frequency of claims and the cost of claims.

Based on the research that has been done, it is known that the frequency of claims and the size of claims follow the Tweedie distribution. From the two models, it is known that the variables affecting the pure premium are the number of children, monthly income, marital status, education, occupation, vehicle use, the number of bluebooks paid, and the type of vehicle from the customer. This shows that the GLM model is a representative and useful model for the insurance company business.

Keywords:

Premiums of Insurance; Generalized Linear Models; Frequency of Claims; Cost of Claim; Tweedie Distribution

Format Sitasi:

T. A. J. Putra, D. C. Lesmana and I. G. P. Purnaba, "Penghitungan Premi Asuransi Kendaraan Bermotor Menggunakan *Generalized Linear Models* dengan Distribusi Tweedie," *Jambura J. Math.*, vol. 3, no. 2, pp. 115-127, 2021

1. Pendahuluan

Asuransi atau pertanggungan adalah satu perjanjian, dengan mana seseorang penanggung mengikatkan diri pada tertanggung dengan menerima suatu premi, untuk memberikan penggantian kepadanya karena suatu kerugian, kerusakan atau kehilangan keuntungan yang diharapkan, yang mungkin akan dideritanya karena suatu peristiwa yang tak tertentu [1]. Peran mendasar dari asuransi adalah memberikan perlindungan finansial, dimana perlindungan yang diberikan berupa benefit yang diperoleh dari premi yang dibayarkan oleh nasabah setiap satuan waktu selama masa tanggungan. Berdasarkan hal tersebut maka menjadi tugas penting bagi seorang aktuaris untuk menentukan harga premi yang sesuai untuk nasabahnya sesuai dengan risiko dan karakteristiknya.

Salah satu metode yang dapat dilakukan dalam penentuan tarif premi adalah dengan mengalikan nilai harapan bersyarat dari frekuensi klaim dengan besarnya klaim, dengan mempertimbangkan karakteristik risiko yang diamati. Klaim adalah sebuah permintaan resmi kepada perusahaan asuransi untuk meminta pembayaran berdasarkan ketentuan perjanjian [2]. Di dalam konteks asuransi, tertanggung harus mengajukan klaim pada pihak penanggung (perusahaan asuransi) pada saat terjadi kegagalan risiko. Namun, sebelum melakukan klaim kepada pihak penanggung, peserta asuransi diwajibkan untuk membayarkan premi. Premi merupakan biaya yang harus dibayarkan peserta asuransi kepada perusahaan asuransi sesuai dengan polis yang disepakati [3].

Penelitian mengenai frekuensi klaim dan besarnya klaim sebagai dasar untuk menentukan besarnya premi asuransi kendaraan bermotor serta penggunaan distribusi tweedie dengan *generalized linear models* telah dilakukan oleh beberapa peneliti, di antaranya, Shi [4] menyatakan bahwa frekuensi klaim secara independen mengikuti sebaran tweedie dan memberikan hasil yang baik dengan menggunakan *generalized linear models*. Frees [5] menyatakan bahwa pada perusahaan asuransi skor klaim yang diperoleh dari frekuensi klaim dan besarnya klaim adalah alat yang berguna untuk menilai risiko kewajiban ketika terjadi kecelakaan pada asuransi kendaraan bermotor. David [6] memberikan gambaran umum dalam penggunaan *generalized linear models* untuk menghitung besarnya premi murni dari hasil perkalian nilai harapan frekuensi klaim dan nilai harapan dari besarnya klaim, dengan distribusi yang digunakan adalah distribusi Poisson untuk frekuensi klaim dan distribusi gamma untuk besarnya klaim. Iyit [7] memperlihatkan penggunaan *generalized linear models* pada bidang yang lain,

yaitu pada penentuan pasokan, transformasi dan konsumsi, impor dan ekspor bahan bakar padat, minyak, gas, listrik, dan energi terbarukan negara-negara Uni Eropa. Garrido [8] memeperlihatkan bahwa asusmsi independensi frekuensi klaim dengan besarnya klaim dapat dilonggarkan dengan pendekatan *generalized linear models* serta memasukan *rating variable* kadalam model sebagai modifikasi model yang digunakan. Fitrianti [9] menyatakan bahwa kesadaran dan kepedulian masyarakat terhadap banyaknya kecelakaan kendaraan sepeda motor merupakan faktor yang mempengaruhi banyaknya kecelakaan seperda motor menggunakan *generalized linear models*. Bonat [10] menyatakan bahwa perluasan berbentuk distribusi Poisson-Tweedie dapat membantu mengatasi dispersi pada data terhitung, dimana dispersi terjadi dikarenakan banyaknya angka nol yang muncul pada data. Sukono [11] memberikan gambaran penggunaan metode bayes untuk menghitung besarnya premi murni dengan menggunakan harapan frekuensi klaim dan nilai harapan dari besarnya klaim, dengan distribusi yang digunakan adalah distribusi Poisson untuk frekuensi klaim dan distribusi gamma untuk besarnya klaim. Debrabant [12] melakukan identifikasi terhadap titik hitam atau titik-titik dimana terjadi kecelakaan yang tidak biasa pada suatu daerah dengan menggunakan model Poisson-tweedie, yang dapat digunakan untuk mengambil keputusan dalam pencegahan kecelakaan di masa depan. Saha [13] menerapkan distribusi Poisson-Tweedie dalam mengembangkan model frekuensi tabrakan, Distribusi Poisson-Tweedie menawarkan kerangka kerja terpadu untuk memodelkan data jumlah yang terlalu menyebar, kurang tersebar, nol-inflasi, spasial, dan membujur, serta beberapa variabel respons dari jenis yang sama atau campuran. Saha juga menyatakan bahwa pada distribusi Tweedie dengan variabel dispersi yang lebih beragam akan menghasilkan hasil yang lebih baik jika menggunakan metode bayes. Li [14] menyatakan bahwa untuk data yang berukuran besar, *generalized linear models* masih bersifat konsisten dalam pemilihan variabel atau dapat dikatakan bahwa *generalized linear models* dapat digunakan untuk data yang berukuran besar.

Pada studi sebelumnya masih menggunakan variabel-variabel penjelas dan distribusi yang lazim digunakan dalam penentuan besarnya premi, sehingga minimnya justifikasi terhadap variabel-variabel yang mempengaruhi besarnya premi. Oleh karena itu, penelitian ini akan menentukan variabel-variabel lain yang berpengaruh terhadap besarnya premi, menggunakan distribusi campuran dalam menentukan besarnya premi melalui Generalized Linear Models (GLM) serta menentukan model harga premi yang sesuai berdasarkan variabel-variabel yang mempengaruhinya. Penghitungan harga premi dilakukan dengan cara mengalikan nilai ekspektasi dari frekuensi klaim dan nilai ekspektasi dari besarnya klaim. Penggunaan GLM dalam penentuan variabel-variabel yang berpengaruh didasari karena pada praktiknya, perusahaan asuransi akan menangani volume data yang besar dan perlu diatasi dengan cara yang seefisien mungkin. Model aktuaria dapat membantu perusahaan asuransi untuk mengatasi data yang sangat banyak [15]. Selain dari itu, GLM merupakan pengembangan dari regresi linier, sehingga memungkinkan pemodelan perilaku non linier tanpa harus memperhatikan beberapa asumsi yang memberatkan sesuai dengan data yang ada [16]. Aspek ini sangat berguna untuk analisis asuransi umum, dikarenakan frekuensi klaim dan besarnya klaim tentunya dapat mengikuti sebaran distribusi yang lebih beragam. Dalam penentuan distribusi dari masing-masing variabel respon seringkali hanya dibatasi pada distribusi diskret untuk variabel frekuensi klaim, dan distribusi kontinu untuk variabel besarnya klaim. Namun pada nyatanya, pada perusahaan asuransi data yang diperoleh tidak hanya terpaut pada dua jenis distribusi tersebut saja, akan tetapi

memungkinkan setiap variabel mengikuti distribusi campuran, yang terdiri atas distribusi diskret dan distribusi kontinu. Dengan demikian, perlu dilakukan pembaruan dengan melakukan uji kecocokan distribusi campuran terhadap variabel respon yang digunakan dan melakukan pemodelan dengan perilaku non linear dalam penentuan besarnya premi, agar tercapainya premi yang sesuai dengan masing-masing karakteristik nasabah.

2. Metode

Pada bagian ini dipaparkan langkah-langkah yang dilakukan dalam pelaksanaan penelitian. Penelitian diawali dengan melakukan rekapitulasi data dari variabel respon serta variabel penjelas yang akan digunakan. Setelah data dari kedua variabel diperoleh terlebih dahulu akan dilakukan pengujian distribusi untuk masing-masing variabel respon yang digunakan, dimana dalam penentuan distribusi yang mengikuti data variabel respon akan diperkuat dengan melakukan uji kesesuaian distribusi. Berdasarkan informasi mengenai distribusi data variabel respon akan dilakukan pemodelan untuk masing-masing variabel respon. Model yang dihasilkan dari proses pembentukan model selanjutnya akan ditentukan estimasi parameter dari masing-masing model, lalu diuji signifikansi parameter dari masing-masing model. Setelah model lengkap dengan setiap parameternya, maka model dari besarnya premi dapat ditentukan. Proses penelitian dilanjutkan dengan memberikan interpretasi dari model besarnya premi yang diperoleh dengan tujuan untuk menarik kesimpulan dari penelitian yang telah dilakukan.

2.1. Generalized Linear Models (GLM)

Analisis regresi linier merupakan pemodelan dengan pendekatan secara statistik untuk menentukan hubungan linier antara variabel satu dengan variabel lainnya, yaitu variabel respon (variabel yang dipengaruhi) dengan variabel penjelas (variabel penyebab). Pendekatan regresi secara teori mendapatkan hasil yang baik ketika variabel respon berdistribusi normal dan varians data (keragaman data) bernilai konstan. Generalized Linear Models (GLM) dapat mengatasi ketika variabel respon tidak berdistribusi normal dan keragaman data bernilai konstan [16]. GLM didefinisikan sebagai perluasan dari regresi linier dengan menggunakan distribusi keluarga eksponensial. Tujuan dari model GLM ini adalah untuk mengestimasi variabel respon (Y) yang bergantung pada penjelasan dari variabel-variabel penjelas (X).

Variabel pengamatan Y yang berdistribusi keluarga eksponensial memiliki fungsi peluang sebagai berikut [17]:

$$f(y|\theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right), \quad y \in S \quad (1)$$

dengan y adalah variabel respon, θ adalah parameter kanonik, ϕ adalah parameter skala, dan S adalah himpunan bagian dari himpunan bilangan asli (\mathbb{N}) atau bilangan real (\mathbb{R}). Sedangkan $b(\theta)$ dan $c(y, \phi)$ adalah suatu fungsi yang diketahui. Pada distribusi keluarga eksponensial berlaku: $E(y) = b'(\theta)$ dan $Var(y) = \phi b''(\theta)$ [17].

GLM memiliki tujuan yang sama dengan regresi linier pada umumnya, yaitu menentukan nilai harapan bersyarat dari variabel respon dengan menggunakan data pengamatan yang sudah ada. Dalam hal ini, akan ditentukan parameter $\beta_1, \beta_2, \dots, \beta_n$ melalui fungsi hubung dari nilai rata-rata variabel penjelas (μ_i), yang dapat dituliskan sebagai kombinasi linier dari variabel penjelas x_i seperti berikut:

$$g(\mu_i) = \beta_0 + \sum_{j=1}^n \beta_j x_{ij} = \mathbf{x}_i^T \boldsymbol{\beta} = \eta_i, \quad i = 1, 2, \dots, m \tag{2}$$

dengan $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_n \end{pmatrix}$ adalah vektor kolom yang berukuran $((n + 1) \times 1)$, $\mathbf{x}_i = \begin{pmatrix} 1 \\ x_{i1} \\ \dots \\ x_{in} \end{pmatrix}$ adalah vektor kolom yang berukuran $((n + 1) \times 1)$, n adalah banyaknya variabel penjelas, m adalah banyaknya amatan. Fungsi yang menghubungkan prediktor linier η_i dengan nilai μ_i adalah fungsi g , dengan g adalah fungsi yang monoton dan terturunkan. Pada GLM (g) disebut sebagai fungsi hubung. Pada penelitian ini, fungsi hubung yang digunakan adalah fungsi hubung kanonik g , yaitu fungsi yang memenuhi persamaan $g(\mu_i) = \theta_i$ [17].

2.2. Distribusi Tweedie

Pada distribusi keluarga eksponensial juga mencakup distribusi yang merupakan campuran komponen diskret dan kontinu. Tweedie (1984) menyatakan bahwa dalam pemodelan klaim asuransi, distribusi campuran yang paling banyak digunakan adalah distribusi Tweedie. Distribusi Tweedie berguna untuk memodelkan variabel respon $y \geq 0$ [18]. Variabel respon yang memiliki distribusi Tweedie yang termasuk dalam keluarga eksponensial dapat dimodelkan dengan menggunakan pendekatan GLM. Varian dari variabel respon terdistribusi Tweedie adalah

$$Var(Y) = \phi \mu^p \tag{3}$$

dengan $E(Y) = \mu$ adalah *mean* dari variabel respon, ϕ adalah parameter dispersi dan p adalah parameter daya varian ekstra di atas *mean* dalam varian variabel respon. Keluarga Tweedie dikenal sebagai kelas distribusi campuran [19]. Secara spesifik nilai parameter p pada persamaan 3 diberikan dalam Tabel 1.

Tabel 1. Kasus khusus distribusi Tweedie

p	Distribusi
$p = 0$	Normal
$p = 1$	Poisson
$1 < p < 2$	Tweedie
$p = 2$	Gamma
$p = 3$	Inverse Gaussian

Sumber: [20]

Distribusi Tweedie untuk $1 < p < 2$ adalah kasus khusus dari keluarga Tweedie dan kurang terkenal dibandingkan dengan distribusi lain [21]. Fungsi kepekatan peluang distribusi Tweedie sebagai berikut:

$$f_S(y) = \exp \left[\frac{-1}{\phi} \left(\frac{\mu^{2-p}}{2-p} + \frac{y}{(p-1)\mu^{p-1}} \right) + c(y, \phi) \right] \quad (4)$$

dengan $c(y, \phi)$ tidak memiliki bentuk analitik. Berdasarkan persamaan di atas diperoleh juga nilai harapan dan varian sebagai berikut:

$$E(Y) = \mu \quad \text{dan} \quad Var(Y) = \phi\mu^p \quad [22] [5]. \quad (5)$$

2.3. Penduga Kemungkinan Maksimum

Variabel pengamatan Y yang berdistribusi keluarga eksponensial memiliki fungsi peluang $f(y|\theta, \phi)$, sehingga diperoleh fungsi kemungkinannya sebagai berikut [17]:

$$L(\theta, \phi) = \prod_{i=1}^n f(y_i|\theta, \phi) = \prod_{i=1}^n \exp \left(\frac{y_i\theta - b(\theta)}{\phi} + c(y_i, \phi) \right) \quad (6)$$

serta fungsi *log-likelihood*nya sebagai berikut:

$$LL(\theta, \phi) = \ln(L(\theta, \phi)) = \frac{n(\bar{y}\theta - b(\theta))}{\phi} + \sum_{i=1}^n \ln c(y_i, \phi). \quad (7)$$

Penduga kemungkinan maksimum dari θ dapat diperoleh dengan mencari θ dengan $b'(\theta) \equiv \mu$ sama dengan rata-rata sampel \bar{y} . Jadi untuk setiap distribusi keluarga eksponensial berlaku $\hat{\mu} = \bar{y}$ [17].

2.4. Penghitungan Premi Murni Agregat

Premi murni adalah nilai harapan biaya klaim tahunan yang dinyatakan oleh pemegang polis dan diperoleh dengan mengalikan nilai harapan frekuensi klaim dengan nilai harapan besarnya klaim. Perkiraan frekuensi klaim dengan besarnya klaim dapat dinyatakan sebagai berikut:

$$E \left[\sum_{i=1}^N C_i \right] = E[N] \times E[C_i] \quad (8)$$

untuk besarnya klaim (C_1, C_2, \dots, C_N) yang saling bebas terhadap banyaknya klaim (N) [8].

3. Hasil dan Pembahasan

3.1. Implementasi Data

Pada penelitian ini data yang digunakan adalah data dari salah satu perusahaan asuransi di Amerika Serikat. Data diperoleh dari situs kaggle.com, yang diunduh pada

tanggal 22 Oktober 2020. Data berisikan mengenai hal-hal yang berkaitan dengan nasabah, dengan jumlah data sebanyak 10.302. Namun, ada sebanyak 1663 data tersensor. Data tersensor adalah data yang diperoleh dari amatan yang tidak secara utuh [23], sehingga analisis hanya dilakukan terhadap 8639 data. Variabel-variabel yang dianalisis pada penelitian ini adalah jumlah anak nasabah, umur nasabah, tahun menjadi nasabah, pendapatan per bulan, status pernikahan (2 kategori), jenis kelamin (2 kategori), pendidikan (5 kategori), pekerjaan (9 kategori), lama waktu kendaraan digunakan, penggunaan kendaraan (2 kategori), pungutan legal *bluebook*, umur kendaraan, mobil merah (2 kategori), dan tipe kendaraan (6 kategori).

3.2. Pemodelan Frekuensi Klaim

Data frekuensi klaim diduga mengikuti salah satu distribusi diskret, akan tetapi setelah dilakukan *fitting distribution* terhadap distribusi diskret dengan menggunakan uji Anderson-Darling(A-D), data tidak mengikuti salah satu distribusi diskret. Berdasarkan hal tersebut dilakukan *fitting distribution* terhadap distribusi campuran Tweedie, dengan bantuan fungsi "*tweedie.profile*" pada *software* R. Setelah dilakukan pengujian distribusi terhadap data frekuensi klaim dengan bantuan *software* R, diperoleh nilai p yaitu 1,132653. Hal ini menunjukkan bahwa nilai p berada di antara 1 dan 2, atau dapat dikatakan bahwa variabel respon frekuensi klaim mengikuti sebaran Tweedie. Selanjutnya dilakukan penentuan fungsi hubung terbaik, fungsi hubung terbaik adalah fungsi hubung dengan hasil AIC terkecil. Berdasarkan uji yang dilakukan diperoleh bahwa fungsi hubung *log* merupakan fungsi hubung terbaik dengan nilai AIC = 20270,65.

Dengan menggunakan fungsi hubung *log* maka diperoleh model frekuensi klaim sebagai berikut:

$$\begin{aligned} g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} &\leftrightarrow \ln \mu_i = \mathbf{x}_i^T \boldsymbol{\beta} \\ &\leftrightarrow \mu_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}} \end{aligned} \tag{9}$$

Dari model frekuensi klaim pada persamaan 9 akan diuji kebaikan model. Uji yang digunakan untuk melihat kebaikan model adalah uji *partial likelihood ratio*. Uji dilakukan dengan menggunakan bantuan *software* R, hasil pengujian signifikansi model secara keseluruhan menunjukkan bahwa model signifikan dan layak digunakan karena $p\text{-value} = 2,2 \times 10^{-16} < 0.05$. Sedangkan untuk pengujian secara parsial, dilakukan uji wald. Berdasarkan uji yang dilakukan, variabel jumlah anak nasabah, pendapatan per bulan, status pernikahan, pendidikan, pekerjaan, penggunaan kendaraan, besarnya *bluebook*, dan tipe kendaraan secara signifikan mempengaruhi frekuensi klaim. Estimasi parameter dari masing-masing variabel disajikan pada Tabel 2.

Dari hasil uji wald yang tertera pada Tabel 2 diperoleh model frekuensi klaim sebagai berikut:

$$\begin{aligned} E[N_i] = \mu_f = \exp &\left(\beta_0 + \beta_1 x_1^i + \beta_4 x_4^i + \beta_5 x_5^i + \beta_7 x_7^i + \sum_{j=1}^9 \beta_{8,j} x_{8,j}^i + \beta_{10} x_{10}^i \right. \\ &\left. + \beta_{11} x_{11}^i + \sum_{j=1}^6 \beta_{12,j} x_{12,j}^i \right) \end{aligned} \tag{10}$$

dengan β_0 adalah *intercept*, dan β_1 adalah parameter dari variabel jumlah anak, x_1^i adalah variabel jumlah anak nasabah ke- i , β_4 adalah parameter dari variabel pendapatan per bulan, x_4^i adalah variabel pendapatan nasabah ke- i (dalam \$), β_5 adalah parameter dari variabel status pernikahan, x_5^i adalah variabel biner untuk status pernikahan nasabah ke- i , dengan status pernikahan sudah menikah bernilai 0 dan untuk status pernikahan belum menikah bernilai 1, β_7 adalah parameter dari variabel pendidikan, x_7^i adalah variabel jenjang pendidikan nasabah ke- i , dengan nasabah yang menempuh jenjang pendidikan magister bernilai 1 dan yang lainnya bernilai 0, $\beta_{8,j}$ adalah parameter dari variabel pekerjaan untuk kategori ke- j , $x_{8,j}^i$ adalah variabel jenis pekerjaan nasabah ke- i , dengan $j = 4, 6$, dan 9 berturut-turut adalah pengacara, *professional*/ahli, dan pekerjaan lainnya bernilai 1, sedangkan untuk nilai j lainnya bernilai 0, β_{10} adalah parameter dari variabel penggunaan kendaraan, x_{10}^i variabel biner untuk jenis penggunaan mobil oleh nasabah ke- i , dengan penggunaan komersial bernilai 0 dan penggunaan pribadi bernilai 1, β_{11} adalah parameter dari variabel besarnya *bluebook*, x_{11}^i adalah variabel besaran *bluebook* yang dibayarkan oleh nasabah ke- i (dalam \$), $\beta_{12,j}$ adalah parameter dari variabel tipe kendaraan untuk kategori ke- j dan $x_{12,j}^i$ adalah variabel tipe mobil dari nasabah ke- i , dengan $j = 2, 4$ dan 6 berturut-turut adalah *panel truck*, mobil *sports*, dan SUV bernilai 1, dan yang lainnya bernilai 0.

Tabel 2. Estimasi parameter model frekuensi klaim

Sumber	β	Std. Error	p-value
<i>Intercept</i>	0,0528166	0,1342234	0,693962
Jumlah anak	0,1186674	0,0290391	4,42e-05
Pendapatan per bulan	-0,0010793	0,0003843	0,004989
Status pernikahan: belum menikah	0,2060592	0,0319151	1,13e-10
Pendidikan: Magister	-0,2036201	0,0957677	0.033516
Pekerjaan: Pengacara	0,2147868	0,0985116	0,029261
Pekerjaan: <i>Professional</i>	0,1326934	0,0661551	0,044910
Pekerjaan: lainnya	0,3698516	0,1029445	0,000329
Penggunaan kendaraan: pribadi	-0,2482142	0,0502653	8,04e-07
<i>Bluebook</i>	-0,0098818	0,0027120	0,000270
Tipe kendaraan: <i>Panel truck</i>	0,1979704	0,0848804	0,019705
Tipe kendaraan: <i>Sports car</i>	0,3915279	0,0678376	8,13e-09
Tipe kendaraan: SUV	0,2379924	0,0585604	4,87e-05

3.3. Pemodelan Besarnya Klaim

Sama halnya dengan frekuensi klaim, data besarnya klaim diduga mengikuti salah satu distribusi kontinu, akan tetapi setelah dilakukan *fitting distribution* terhadap distribusi kontinu dengan menggunakan uji khi-kuadrat, data tidak mengikuti salah satu distribusi kontinu. Berdasarkan hal tersebut dilakukan *fitting distribution* terhadap distribusi campuran Tweedie, dengan bantuan fungsi "*tweedie.profile*" pada *software R*. Setelah dilakukan pengujian distribusi terhadap data frekuensi klaim dengan bantuan *software R*, diperoleh nilai p yaitu 1,757143. Hal ini menunjukkan bahwa variabel respon besarnya klaim mengikuti sebaran Tweedie. Selanjutnya dilakukan penentuan fungsi hubung terbaik, fungsi

hubung terbaik adalah fungsi hubung dengan hasil AIC terkecil. Berdasarkan uji yang dilakukan diperoleh bahwa fungsi hubung *log* merupakan fungsi hubung terbaik dengan nilai AIC = 50428,61.

Dengan menggunakan fungsi hubung *log* maka diperoleh model frekuensi klaim sebagai berikut:

$$\begin{aligned}
 g(\mu_i) &= \mathbf{x}_i^T \boldsymbol{\gamma} \leftrightarrow \ln \mu_i = \mathbf{x}_i^T \boldsymbol{\gamma} \\
 &\leftrightarrow \mu_i = e^{\mathbf{x}_i^T \boldsymbol{\gamma}}
 \end{aligned}
 \tag{11}$$

Dari model frekuensi klaim pada persamaan 10 akan diuji kebaikan model. Uji yang digunakan untuk melihat kebaikan model adalah uji *partial likelihood rasio*. Uji dilakukan dengan menggunakan bantuan *software R*, hasil pengujian signifikansi model secara keseluruhan menunjukkan bahwa model signifikan dan layak digunakan karena $p\text{-value} = 2,67 \times 10^{-9} < 0,05$. Sedangkan untuk pengujian secara parsial, dilakukan uji wald. Berdasarkan uji yang dilakukan, variabel jumlah anak nasabah, status pernikahan, besarnya *bluebook*, dan tipe kendaraan secara signifikan mempengaruhi besarnya klaim, yaitu variabel dengan nilai *p-value* kurang dari 0,05. Estimasi parameter dari masing-masing variabel disajikan pada Tabel 3.

Tabel 3. Estimasi parameter model besarnya klaim

Sumber	$\boldsymbol{\gamma}$	<i>Std. Error</i>	<i>p-value</i>
<i>Intercept</i>	3,3582030	0,4727858	1,32e-12
Jumlah anak	0,2849244	0,1041721	0,00625
Status menikah: belum menikah	0,3546937	0,1133211	0,00175
<i>Bluebook</i>	-0,0286070	0,0095318	0,00270
Tipe kendaraan: <i>Panel truk</i>	0,8903729	0,3032908	0,00334
Tipe kendaraan: <i>Pickup</i>	0,5714183	0,1845274	0,00196
Tipe kendaraan: <i>Sports car</i>	0,5255360	0,2411837	0,02936
Tipe kendaraan: Van	0,6916264	0,2366763	0,00348
Tipe kendaraan: SUV	0,5538445	0,2015019	0,00600

Dari hasil uji wald yang tertera pada Tabel 3 diperoleh model frekuensi klaim sebagai berikut:

$$E[C_{ik}] = \mu_c = \exp \left(\gamma_0 + \gamma_1 x_1^i + \gamma_5 x_5^i + \gamma_{11} x_{11}^i + \sum_{j=1}^6 \gamma_{12,j} x_{12,j}^i \right)
 \tag{12}$$

dengan γ_0 adalah *intercept*, dan γ_1 adalah parameter dari variabel jumlah anak, x_1^i adalah variabel jumlah anak nasabah ke-*i*, γ_5 adalah parameter dari variabel status pernikahan, x_5^i adalah variabel biner untuk status pernikahan nasabah ke-*i*, dengan status pernikahan sudah menikah bernilai 0 dan untuk status pernikahan belum menikah bernilai 1, γ_{11} adalah parameter dari variabel besarnya *bluebook*, x_{11}^i adalah variabel besaran *bluebook* yang dibayarkan oleh nasabah ke-*i* (dalam \$), $\gamma_{12,j}$ adalah parameter dari variabel tipe kendaraan untuk kategori ke-*j* dan $x_{12,j}^i$ adalah variabel tipe mobil dari nasabah ke-*i*, dengan $j = 2,3,4,5$ dan 6 berturut-turut adalah *panel truck*, *pickup*, mobil *sports*, van dan SUV bernilai 1, dan yang lainnya bernilai 0.

3.4. Penghitungan Premi Murni Agregat

Penghitungan premi murni diperoleh dengan mengalikan nilai harapan dari frekuensi klaim dengan nilai harapan besarnya klaim, mengikuti persamaan berikut:

$$\begin{aligned}
 E \left[\sum_{k=1}^{N_i} C_{ik} \right] &= E[N_i] \times E[C_{ik}] \\
 &= \exp(\beta^t x) \times \exp(\gamma^t x) \\
 &= \exp((\beta^t + \gamma^t)x) \tag{13}
 \end{aligned}$$

Estimasi parameter untuk premi murni disajikan pada Tabel 4.

Tabel 4. Estimasi parameter model premi murni

Sumber	β	γ	$\beta + \gamma$
<i>Intercept</i>	0,0528166	3,3582030	3,4110196
Jumlah anak	0,1186674	0,2849244	0,4035918
Pendapatan per bulan	-0,0010793	-	-0,0010793
Status menikah: belum menikah	0,2060592	0,3546937	0,5607529
Pendidikan: Magister	-0,2036201	-	-0,2036201
Pekerjaan: Pengacara	0,2147868	-	0,2147868
Pekerjaan: <i>Professional</i>	0,1326934	-	0,1326934
Pekerjaan: lainnya	0,3698516	-	0,3698516
Penggunaan kendaraan: pribadi	-0,2482142	-	-0,2482142
<i>Bluebook</i>	-0,0098818	-0,0286070	-0,0384888
Tipe kendaraan: <i>Panel truck</i>	0,1979704	0,8903729	1,0883833
Tipe kendaraan: <i>Pickup</i>	-	0,5714183	0,5714183
Tipe kendaraan: <i>Sports car</i>	0,3915279	0,5255360	0,9170639
Tipe kendaraan: Van	-	0,6916264	0,6916264
Tipe kendaraan: SUV	0,2379924	0,5538445	0,7918369

Berdasarkan estimasi parameter premi murni yang dipaparkan pada Tabel 4 diperoleh model premi murni sebagai berikut:

$$\begin{aligned}
 E \left[\sum_{k=1}^{N_i} C_{ik} \right] &= \exp \left((\beta_0 + \gamma_0) + (\beta_1 + \gamma_1)x_1^i + \beta_4 x_4^i + (\beta_5 + \gamma_5)x_5^i + \beta_7 x_7^i \right. \\
 &\quad \left. + \sum_{j=1}^9 \beta_{8,j} x_{8,j}^i + \beta_{10} x_{10}^i + (\beta_{11} + \gamma_{11})x_{11}^i \right. \\
 &\quad \left. + \sum_{j=1}^6 (\beta_{12,j} + \gamma_{12,j})x_{12,j}^i \right) \tag{12}
 \end{aligned}$$

dengan $(\beta_0 + \gamma_0)$ adalah *intercept*, dan $(\beta_1 + \gamma_1)$ adalah parameter dari variabel jumlah anak, x_1^i adalah variabel jumlah anak nasabah ke- i , β_4 adalah parameter dari variabel pendapatan per bulan, x_4^i adalah variabel pendapatan nasabah ke- i (dalam \$), β_5 adalah parameter dari variabel status pernikahan, x_5^i adalah variabel biner untuk status pernikahan nasabah ke- i , dengan status pernikahan sudah menikah bernilai 0 dan untuk status pernikahan belum menikah bernilai 1, β_7 adalah parameter dari variabel

pendidikan, x_7^i adalah variabel jenjang pendidikan nasabah ke- i , dengan nasabah yang menempuh jenjang pendidikan magister bernilai 1 dan yang lainnya bernilai 0, $\beta_{8,j}$ adalah parameter dari variabel pekerjaan untuk kategori ke- j , $x_{8,j}^i$ adalah variabel jenis pekerjaan nasabah ke- i , dengan $j = 4, 6, \text{ dan } 9$ berturut-turut adalah pengacara, *professional/ahli*, dan pekerjaan lainnya bernilai 1, sedangkan untuk nilai j lainnya bernilai 0, β_{10} adalah parameter dari variabel penggunaan kendaraan, x_{10}^i variabel biner untuk jenis penggunaan mobil oleh nasabah ke- i , dengan penggunaan komersial bernilai 0 dan penggunaan pribadi bernilai 1, $(\beta_{11} + \gamma_{11})$ adalah parameter dari variabel besarnya *bluebook*, x_{11}^i adalah variabel besaran *bluebook* yang dibayarkan oleh nasabah ke- i (dalam \$), $(\beta_{12,j} + \gamma_{12,j})$ adalah parameter dari variabel tipe kendaraan untuk kategori ke- j dan $x_{12,j}^i$ adalah variabel tipe mobil dari nasabah ke- i , dengan $j = 2, 3, 4, 5$ dan 6 berturut-turut adalah *panel truck*, *pickup*, mobil *sports*, van dan SUV bernilai 1, dan yang lainnya bernilai 0.

Dari persamaan 12 dan Tabel 4 dapat dilihat bahwa jumlah anak nasabah, nasabah yang belum menikah, pekerjaan dari nasabah dan tipe kendaraan nasabah akan mengakibatkan penambahan besarnya premi yang harus dibayarkan, dengan besarnya penambahan premi masing-masing kategori adalah $\exp(0,4035918)$ dari jumlah anak nasabah, jika nasabah belum menikah maka penambahan besarnya premi sebesar $\exp(0,5607529)$, pada kategori pekerjaan, jika nasabah bekerja sebagai pengacara, *professional*, atau pekerjaan lainnya akan terjadi penambahan masing-masing sebesar $\exp(0,2147868)$, $\exp(0,1326934)$ atau $\exp(0,3698516)$. Sedangkan pada kategori tipe kendaraan nasabah, akan terjadi penambahan besarnya premi untuk tipe *panel truck* sebesar $\exp(1,0883833)$, *pickup* sebesar $\exp(0,5714183)$, *sports car* sebesar $\exp(0,9170639)$, Van sebesar $\exp(0,6916264)$ dan SUV sebesar $\exp(0,7918369)$. Disisi lain variabel lainnya seperti pendapatan per bulan, pendidikan, penggunaan kendaraan dan besarnya *bluebook* yang dibayarkan nasabah akan mengakibatkan penurunan besarnya premi yang akan dibayarkan. Besarnya penurunan dari masing-masing variabel dapat dilihat dari masing-masing kategori variabel, dimana besarnya penurunan premi berdasarkan pendapatan perbulan adalah sebesar $\exp(-0,0010793)$, untuk nasabah dengan pendidikan magister akan mengalami penurunan sebesar $\exp(-0,2036201)$, untuk nasabah yang menggunakan kendaraannya sebagai kendaraan pribadi akan mengalami penurunan sebesar $\exp(-0,2482142)$ dan setiap pembayaran 1 satuan pembayaran *bluebook* oleh nasabah akan mengalami penurunan sebesar $\exp(-0,0384888)$, dengan setiap satuan penambahan atau penurunan besarnya premi adalah dollar (\$).

4. Kesimpulan

Premi murni dari asuransi kendaraan bermotor dipengaruhi oleh beberapa faktor risiko, diantaranya yaitu seperti jumlah anak, pendapatan per bulan, status pernikahan, pendidikan, pekerjaan, penggunaan kendaraan, besarnya *bluebook* yang dibayarkan serta tipe kendaraan dari nasabah. Dengan demikian, untuk menentukan premi nasabah yang baru, perusahaan asuransi dapat memperhatikan faktor risiko tersebut dalam menentukan besaran premi yang akan diberikan dengan sistem penghitungan besaran premi dapat ditentukan melalui model GLM, mengikuti Persamaan 12.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih yang sebesar-besarnya atas pendanaan penelitian

ini kepada Direktorat Penelitian dan Pengabdian kepada Masyarakat, Direktorat Jenderal Pendidikan Tinggi Kementerian Pendidikan Nasional Republik Indonesia, melalui Hibah Penelitian Tesis Nomor 2822/IT3.L1/PN/2020.

Referensi

- [1] B. H. dan H. B. U. A. M. Agung-RI, [KUHD] *Kitab Undang-undang Hukum Dagang Tahun 2009 Bab 9 Pasal 246 Tentang Asuransi atau Pertanggungan Seumurnya*. Bhuna Ilmu Populer, 2017.
- [2] S. Asmussen and M. Steffensen, *Probability Theory and Stochastic Modelling: Risk and insurance*. Switzerland: Springer, 2020.
- [3] P. R. Kongstvedt, *Health Insurance and Managed Care*, 5 th. Burlington: Jones And Bartlett Learning, 2020.
- [4] P. Shi, X. Feng, and A. Ivantsova, "Dependent frequency-severity modeling of insurance claims," *Insur. Math. Econ.*, vol. 64, pp. 417–428, 2015, doi: 10.1016/j.insmatheco.2015.07.006.
- [5] E. Frees, G. Lee, and L. Yang, "Multivariate Frequency-Severity Regression Models in Insurance," *Risks*, vol. 4, no. 1, p. 4, 2016, doi: 10.3390/risks4010004.
- [6] M. David, "Auto insurance premium calculation using generalized linear models," *Procedia Econ. Financ.*, vol. 20, no. 15, pp. 147–156, 2015, doi: 10.1016/S2212-5671(15)00059-3.
- [7] N. Iyit, H. Yonar, and A. Genc, "Generalized Linear Models for European Union Countries Energy Data," *Acta Phys. Pol. A*, vol. 130, no. 1, 2016, doi: 10.12693/APhysPolA.130.397.
- [8] J. Garrido, C. Genest, and J. Schulz, "Generalized linear models for dependent frequency and severity of insurance claims," *Insur. Math. Econ.*, vol. 70, pp. 205–215, 2016, doi: 10.1016/j.insmatheco.2016.06.006.
- [9] H. Fitrianti, "Model Faktor-Faktor Banyaknya Kecelakaan Lalu Lintas Pada Kendaraan Sepeda Motor Dengan Pendekatan Generalized Linear Model," *Magistra*, vol. 4, no. 2, pp. 94–104, 2017.
- [10] W. H. Bonat, B. Jorgensen, C. Kokonendji, and J. Hinde, "Extended Poisson – Tweedie : Properties and Regression Models for Count Data," *Stat. Modelling*, pp. 1–26, 2017.
- [11] S. Sukono, R. Riaman, E. Lesmana, R. Wulandari, H. Napitupulu, and S. Supian, "Model estimation of claim risk and premium for motor vehicle insurance by using Bayesian method," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 300, no. 1, 2018, doi: 10.1088/1757-899X/300/1/012027.
- [12] B. Debrabant, U. Halekoh, W. H. Bonat, D. L. Hansen, J. Hjelmberg, and J. Lauritsen, "Identifying traffic accident black spots with Poisson-Tweedie models," *Accid. Anal. Prev.*, vol. 111, pp. 147–154, 2018, doi: 10.1016/j.aap.2017.11.021.
- [13] D. Saha, P. Alluri, E. Dumbaugh, and A. Gan, "Application of the Poisson-Tweedie distribution in analyzing crash frequency data," *Accid. Anal. Prev.*, vol. 137, no. January, p. 105456, 2020, doi: 10.1016/j.aap.2020.105456.

- [14] N. Li, X. Peng, E. Kawaguchi, M. A. Suchard, and G. Li, "A scalable surrogate L0 sparse regression method for generalized linear models with applications to large scale data," *J. Stat. Plan. Inference*, vol. 213, pp. 262–281, 2020, doi: 10.1016/j.jspi.2020.12.001.
- [15] S. A. Klugman, H. H. Panjer, and G. E. Willmot, *Loss Models From Data to Decisions*, 5 th., vol. 6, no. 1. Hoboken, New Jersey: John Wiley & Sons, Inc., 2019.
- [16] A. Agresti, *Foundations Linear Generalized Linear Models*. Cambridge: John Wiley & Sons, Inc., 2015.
- [17] G. Z. Jong, P D, Heller, *Generalized Linear Models for Insurance Data*. Cambridge: Cambridge University Press, 2008.
- [18] M. Tweedie, "An Index Which Distinguishes Between Some Important Exponential Families. The Statistics: Applications and New Directions," 1984.
- [19] U. Simsekli, A. T. Cemgil, and B. Ermiş, "Learning Mixed Divergences in Coupled Matrix and Tensor Factorization Models," in *The Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 2120–2124.
- [20] O. Ozgur and N. Iyit, "Modelling the US Diabetes Mortality Rates via Generalized Linear Model with the Tweedie Distribution," *Int. J. Sci. Res.*, vol. 7, no. 2, pp. 1326–1334, 2018, doi: 10.21275/ART2018368.
- [21] A. A. Khalin and E. B. Postnikov, "A wavelet-based approach to revealing the Tweedie distribution type in sparse data," *Phys. A Stat. Mech. its Appl.*, vol. 553, p. 124653, 2020, doi: 10.1016/j.physa.2020.124653.
- [22] E. W. Frees, *Regression Modeling with Actuarial and Financial Applications*. Cambridge: Cambridge University Press, 2010.
- [23] E. T. Lee and J. W. Wang, *Statistical Methods for Survival Data Analysis*, 4 rd. Hoboken, New Jersey: John Wiley & Sons, Inc., 2013.



This article is an open-access article distributed under the terms and conditions of the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). Editorial of JJoM: Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B.J. Habibie, Moutong, Tilongkabila, Kabupaten Bone Bolango, Provinsi Gorontalo 96119, Indonesia.