



PENENTUAN PREMI ASURANSI KESEHATAN DENGAN FAKTOR UNDERWRITING MENGGUNAKAN *GENERALIZED LINEAR MODELS*

Sainal Abidin^{1*}, Sri Dewi Anugrawati², Nurwahidah³

Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Alauddin Makassar^{1,2}
sainal200505@gmail.com*

Abstrak– Salah satu usaha yang dapat dilakukan seseorang dalam mengantisipasi adanya risiko yang akan datang yakni dengan berpartisipasi di suatu asuransi kesehatan. Asuransi kesehatan telah banyak ditemui di Indonesia, menurut data Badan Pusat Statistik (BPS) jumlah penduduk Indonesia yang dalam asuransi kesehatan pada tahun 2018 ialah 208.054.199 jiwa. Penentuan premi asuransi kesehatan berdasarkan frekuensi klaim dan besarnya klaim dapat ditentukan dengan menggunakan Generalized Linear Models (GLM). Penelitian ini merupakan jenis penelitian terapan dengan menggunakan data sekunder berupa data frekuensi klaim dan besaran klaim dari salah satu kumpulan data pribadi biaya medis yang diperoleh dari situs <https://www.kaggle.com>. Jumlah data yang digunakan sebanyak 1338 orang. Hasil dari penelitian yang telah dilakukan menunjukkan bahwa frekuensi klaim dan besar klaim mengikuti distribusi Bernoulli dan gamma. Dari kedua model yang didapatkan diketahui bahwa variabel yang mempengaruhi premi murni adalah usia, BMI, jumlah anak dan riwayat merokok. Hasil yang diperoleh menunjukkan bahwa premi yang didapat memiliki rentang biaya sebesar \$1.305 - \$79,354.48627.

Kata Kunci – Asuransi Kesehatan, Premi Murni, *Generalized Linear Models*.

I. PENDAHULUAN

Kesehatan adalah hal mendasar yang sangat penting dimiliki bagi setiap manusia. Tanpa kesehatan yang baik, manusia akan

sulit melaksanakan kegiatan sehari-hari. Kesehatan merupakan hak paten bagi warga negara, sehingga sudah selayaknya negara mempunyai jaminan kesehatan bagi warganya. Salah satu bentuk jaminan yang diperbolehkan dalam Islam adalah akad tabarru' atau tolong menolong yang banyak digunakan dalam praktik takafful atau asuransi syariah dengan unsur saling menanggung risiko antar peserta asuransi. Salah satu usaha yang dapat dilakukan seseorang dalam mengantisipasi adanya risiko yang akan datang yakni dengan berpartisipasi di suatu asuransi kesehatan. Asuransi kesehatan telah banyak ditemui diberbagai negara, di Indonesia sendiri menurut data Badan Pusat Statistik (BPS) jumlah penduduk Indonesia yang bergabung dalam asuransi kesehatan pada tahun 2018 ialah 208.054.199 jiwa. Bagi perusahaan asuransi, risiko adalah klaim. Tetapi, sebelum peserta asuransi mengajukan klaim kepada pihak penanggung, peserta asuransi harus terlebih dahulu membayar premi. Agar perusahaan asuransi bebas dari kerugian besar yang disebabkan oleh klaim, maka proses underwriting perlu dilakukan. Dengan underwriting perusahaan dapat mengetahui potensi – potensi risiko tersebut yang mungkin terjadi, termasuk seberapa besar risiko yang mampu ditanggung oleh perusahaan baik secara kuantitatif ataupun kualitatif, sehingga faktor underwriting mempengaruhi terhadap penentuan premi asuransi kesehatan. Penentuan premi juga dapat dilakukan dengan

mengalikan nilai harapan bersyarat dari frekuensi klaim dan besarnya klaim, dengan mempertimbangkan faktor underwriting atau karakteristik risiko yang diamati. Penentuan premi asuransi kesehatan berdasarkan frekuensi klaim dan besarnya klaim dapat ditentukan dengan menggunakan Generalized Linear Models (GLM). Supriatma dkk [1] melakukan penelitian tentang menentukan harga premi asuransi menggunakan *Generalized Linear Models* dengan model distribusi yang digunakan untuk besar klaim adalah distribusi gamma dan banyak klaim distribusi binomial dengan $\alpha = 5\%$. Putra, dkk [2] meneliti tentang *Generalized Linear Models* dalam menghitung premi asuransi kendaraan bermotor dimana frekuensi dan besarnya klaim berdistribusi tweedie dan variabel yang mempengaruhi premi murni adalah faktor *underwriting*. Rohmaniah and Danardono [3] meneliti perhitungan premi model dua tahunan dengan factor *underwriting* menggunakan *Generalized Linear Models*. Hasil menunjukkan faktor *Underwriting* yang berpengaruh ialah usia, status peminum alcohol, penyakit jantung dan diabetes Berdasarkan penelitian yang telah dilakukan sebelumnya, metode *Generalized Linear Models* dapat digunakan dalam menentukan premi asuransi berdasarkan faktor *underwriting*. Faktor risiko dapat dijadikan sebagai faktor *underwriting* apabila objektif dan mudah dikendalikan, kemudian nilainya berdampak dan dapat digunakan untuk menentukan premi asuransi kesehatan.

II. HASIL DAN PEMBAHASAN

Langkah-langkah penentuan premi asuransi dengan menggunakan Generalized Linear Models yaitu :

A. Data Penelitian

Data pada penelitian ini adalah data dari salah satu kumpulan data pribadi biaya medis yang diperoleh dari situs kaggle.com. Data ini berjumlah sebanyak 1338. Variabel – variabel yang digunakan pada penelitian ini adalah usia, jenis kelamin (2 kategori), BMI (Body Mass Index), langka kaki, jumlah anak,

merokok (2 kategori), charges dan klaim asuransi.

B. Uji Distribusi Data

1. Uji distribusi Frekuensi Klaim

Frekuensi klaim pada data penelitian ini berbentuk kategori di mana 0 didefinisikan sebagai tidak mengajukan klaim dan 1 didefinisikan mengajukan klaim. Berikut adalah Tabel 1 dari frekuensi klaim.

Tabel 1: Data Frekuensi Klaim

Frekuensi klaim	0	1
Banyak Peserta	555	783

Data frekuensi klaim di atas diduga mengikuti salah satu distribusi diskrit, yaitu distribusi Bernoulli yang dibuktikan dengan analisa tes statistic uji *Chi-Square* dengan hipotesis sebagai berikut [4] :

H_0 : Data variabel respon mengikuti distribusi Bernoulli

H_1 : Data variabel respon tidak mengikuti distribusi Bernoulli

Pada taraf signifikansi/tingkat keyakinan $\alpha = 5\%$, dengan kriteria pengujian yaitu nilai $X^2 < X^2_{(1-\alpha);dk}$ maka gagal menolak H_0 . Setelah dilakukan uji *Chi-Square* dengan bantuan *software R*, diperoleh nilai $X^2 = 2.17250076348203e - 13$. Nilai $X^2_{(1-\alpha);dk}$ dengan df 1 dan $\alpha = 0.05$ adalah 3.8415. Karena $X^2 < X^2_{(1-\alpha);dk}$ maka gagal menolak H_0 . Berdasarkan hasil uji dapat disimpulkan bahwa variabel respon frekuensi klaim mengikuti distribusi Bernoulli.

2. Uji distribusi Besar Klaim

Data besar klaim pada penelitan ini diduga mengikuti salah satu distribusi kontinu, yaitu distribusi gamma yang dibuktikan dengan analisis *fitting distribution* terhadap distribusi kontinu dengan melihat nilai AIC terkecil. Berikut hasil analisis *fitting distribution* dengan nilai AIC.

Tabel 2: *Fitting distribution* dengan nilai AIC

Distribusi	AIC
Normal	2542.161430
Gamma	2402.165628
Burr	2418.608068
Log Normal	2991.559713

Tabel 2 menunjukkan bahwa distribusi gamma memiliki nilai AIC paling kecil dibandingkan dengan distribusi lainnya, sehingga dapat disimpulkan bahwa variabel respon besar klaim mengikuti distribusi gamma.

C. *Generalized Linear Models (GLM)*

Generalized Linear Model (GLM) merupakan bentuk umum atau general dari Model Linear. Diketahui bahwa vektor \mathbf{y} memiliki n komponen, yang merepresentasikan realisasi dari sebuah matriks respon \mathbf{Y} . Setiap komponen independen dan berdistribusi dengan mean atau $E(\mathbf{Y}) = \mu$. Jika model yang dihasilkan memiliki prediktor \mathbf{X} , dengan beberapa parameter yang tidak diketahui β_1, \dots, β_n maka model tersebut merupakan kombinasi linear $\mu = \sum_j^p \mathbf{x}_j \beta_j$ atau dapat ditulis dalam bentuk matriks sebagai $\mu = \mathbf{x}\beta$. Sebuah model GLM terdiri dari prediktor linier [5].

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}$$

dimana

η_i : *Canonical link*

β_p : Parameter model yang belum diketahui

x_{pi} : variabel independen

D. *Distribusi Keluarga Eksponensial*

Sebuah konsep penting yang mendasari Generalized Linear Model (GLM) adalah distribusi keluarga eksponensial. Anggota keluarga distribusi eksponensial semuanya memiliki fungsi kepadatan probabilitas untuk respon yang teramati y yang dapat ditulis dalam bentuk berikut [6]:

$$f(y; \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right)$$

dimana

$a(\cdot), b(\cdot)$ dan $c(\cdot)$: Fungsi spesifik

θ : Natural parameter lokasi

ϕ : Parameter disperse

Beberapa jenis distribusi yang kerap digunakan dalam GLM dapat dilihat sebagai berikut:

1. *Distribusi Bernoulli*

Distribusi Bernoulli memiliki fungsi peluang sebagai berikut [7]:

$$f(x) = \begin{cases} \pi^x(1 - \pi)^{1-x} & x = 0, 1 \\ 0 & \text{lainnya} \end{cases}$$

dimana

x : variabel acak yang merepresentasikan hasil dari eksperimen tunggal

π : parameter $0 \leq \pi \leq 1$

2. *Distribusi Gamma*

Distribusi Gamma memiliki fungsi peluang sebagai berikut [8]:

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}$$

dimana

$\Gamma(\alpha)$: Fungsi gamma

α : Parameter bentuk (shape parameter) yang merupakan bilangan bulat positif

β : Parameter skala (scale parameter) yang merupakan bilangan riil positif

x : Variable acak yang yang memenuhi $x > 0$

E. *Penentuan Link Function*

Fungsi *link* memberikan penjelasan hubungan linier antara fungsi transformasi dari mean, $g(\mu)$ dengan variable prediktor yang menghubungkan komponen sistematis η terhadap nilai mean μ . Fungsi hubung atau *link function* merupakan nilai harapan dari komponen acak. Fungsi *link* dimodelkan sebagai berikut [9].

$$\eta = g(\mu) \text{ atau } \mu = g^{-1}(\eta) \tag{1}$$

Berikut ini penghubung kanonik (θ) untuk beberapa distribusi [10]

Tabel 3: Link Function

Nama	Fungsi Hubung Kanonik
<i>Identity</i>	$g(\mu) = \mu$
Fungsi Log	$g(\mu) = \ln \mu$
Fungsi Logit	$g(\mu) = \ln\left(\frac{\pi}{1-\pi}\right)$
<i>Negatif Inverse</i>	$g(\mu) = \mu^{-1}$

Link function yang digunakan pada data frekuensi klaim yang berdistribusi Bernoulli adalah link logit. Untuk besar klaim yang berdistribusi gamma memiliki *link function negative inverse*. Namun ketika membuat model dengan fungsi *link negative inverse* kurang efisien digunakan karena

penaksiran parameter yang diperoleh tidak konvergen [1]. Sehingga fungsi link yang dipilih adalah fungsi link log.

F. Model Frekuensi Klaim dan Besar Klaim

Link function untuk data frekuensi klaim menggunakan link logit. Setiap parameter natural η_i dapat didefinisikan sebagai nilai mean μ_i dari distribusi keluarga eksponensial atau $\mu_i = x_i^T \beta$, $i = 1, 2, \dots, n$. Karena $\eta_i = x_i^T \beta$ sehingga model frekuensi klaim menjadi

$$\pi = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}} \quad (2)$$

Model yang digunakan pada data besar klaim yaitu model log linear. Jika variabel yang mempengaruhi variabel respon Y adalah X_1, X_2, \dots, X_k , maka nilai μ untuk Y pada observasi ke- i menurut model log linear adalah sebagai berikut [11].

$$\begin{aligned} g(\mu_i) &= \ln(\mu_i) = \beta_0 + \beta_1 x_1 \\ &\quad + \beta_2 x_2 + \dots \\ &\quad + \beta_i x_i \\ \mu_i &= \exp(\beta_0 \\ &\quad + \beta_1 x_1 \\ &\quad + \beta_2 x_2 + \dots \\ &\quad + \beta_i x_i) \\ \mu_i &= \exp(x_i^T \beta) \end{aligned} \quad (3)$$

G. Maximum Likelihood Estimation (MLE)

Metode *Maximum Likelihood Estimation* adalah metode pendugaan yang memaksimumkan fungsi *likelihood*. Adapun fungsi *Likelihood* (θ) sebagai berikut [12].

$$L(\theta, \phi; y) = \prod_{i=1}^n f(y_i; \theta, \phi)$$

Untuk mempermudah perhitungan secara matematis, umumnya digunakan fungsi *log-likelihood*.

$$\ln(L(\theta, \phi; y)) = l(\theta, \phi; y)$$

$$= \sum_{i=1}^n \ln f(y_i; \theta, \phi)$$

H. Uji Signifikansi Parameter

1. Uji Serentak

Uji serentak atau uji *Likelihood Ratio Test* untuk model frekuensi klaim dan besar

klaim diberikan hipotesis sebagai berikut

[13]:

$H_0 : \beta_1 = \beta_2 = \dots = \beta_i = 0$ (tidak terdapat pengaruh antara variabel independen terhadap variabel dependet)

$H_1 : \text{minimal ada satu } \beta_i \neq 0$ (terdapat pengaruh antara variabel independen dan variabel dependen)

Dengan nilai kritis tolak H_0 apabila $P - Value < \alpha$. Berdasarkan analisis yang telah dilakukan untuk kedua model tersebut diperoleh nilai $p - value$ uji serentak adalah $2.2e - 16$, dengan demikian dapat dilihat bahwa $p - value < \alpha$ sehingga H_0 ditolak [14]. Hal ini berarti bahwa paling sedikit ada satu variabel independen yang berpengaruh secara serentak terhadap variabel dependen pada model frekuensi klaim dan besar klaim.

2. Uji Parsial

Uji parsial untuk model frekuensi klaim dilakukan dengan hipotesis berikut [15]:

$H_0 : \beta_i = 0$ (tidak ada pengaruh antara masing-masing variabel independen terhadap variabel dependen)

$H_1 : \beta_i \neq 0$ (ada pengaruh antara masing-masing variabel independen terhadap variabel dependen)

Tabel 4: Hasil Uji Parsial Model Frekuensi Klaim

Parameter	Estimate	Statistic	p-value
Intercept	-8.86	-8.42	3.62E-17
X1	0.0309	5.06	4.28e- 7
X2	-0.0215	-0.136	8.92e- 1
X3	0.285	11	2.31E-28
X4	0.0000786	1.6	1.09e- 1
X5	-1.44	-15.1	2.81E-51
X6	4.38	11.9	1.40E-32

Sumber : Hasil Analisis Program R

Berdasarkan Tabel 4 dengan menggunakan kepercayaan 95% maka variabel independent yang signifikan ($p - value < \alpha$) terhadap variabel dependent ialah X1, X3, X5 dan X6. Setelah didapatkan penaksiran parameternya yang hanya memperhatikan level signifikan pada selang

kepercayaan 95% didapatkan model akhir frekuensi klaim sebagai berikut.

$$\pi = \frac{e^{-8.86+0.0309X_1+0.285X_3-1.44X_5+4.38X_6}}{1+e^{-8.86+0.0309X_1+0.285X_3-1.44X_5+4.38X_6}} \quad (4)$$

dengan :

- X_1 = Usia
- X_3 = BMI
- X_5 = Jumlah Anak
- X_6 = Merokok

Uji parsial untuk model besar klaim dilakukan dengan hipotesis berikut :

$H_0 : \beta_j = 0$ (tidak ada pengaruh antara masing-masing variabel independen terhadap variabel dependen)

$H_1 : \beta_j \neq 0$ (ada pengaruh antara masing-masing variabel independen terhadap variabel dependen)

Tabel 5: Hasil Uji Parsial Model Frekuensi Klaim

Parameter	Estimate	Statistic	p-value
Intercept	7.28	37.0	4.63e-207
X1	0.0287	21.1	1.70e- 85
X2	-0.0572	-1.52	1.29e- 1
X3	0.0136	3.13	21.78e-3
X4	0.00000563	0.495	6.21e-1
X5	0.0842	5.38	8.63e- 8
X6	1.5	29.9	5.60e-151

Sumber : Hasil Analisis Program R

Berdasarkan Tabel 4 menggunakan kepercayaan 95% maka variabel independent yang signifikan ($p\text{-value} < \alpha$) terhadap variabel dependent iyalah X1, X3, X5 dan X6, maka model akhir besar klaim dapat dituliskan sebagai berikut

$$\mu = e^{7.28+0.0287X_1+0.0136X_3+0.0842X_5+1.5X_6} \quad (5)$$

dengan :

- X_1 = Usia
- X_3 = BMI
- X_5 = Jumlah Anak
- X_6 = Merokok

I. Menentukan Premi Murni Menggunakan Model Compound

Perhitungan premi murni diperoleh dengan mengalikan nilai harapan dari

frekuensi klaim dengan nilai harapan besar klaim, mengikuti persamaan berikut [4]:

$$E[S] = E[N] \times E[X] \quad (6)$$

dengan

- $E[S]$: Harapan biaya klaim
- $E[N]$: Harapan frekuensi klaim
- $E[X]$: Harapan besar klaim

Model dari $E[N]$ dan $E[X]$ telah didapatkan sesuai pada persamaan (4) dan (5) sehingga persamaan (6) menjadi

$$E[S] = \frac{e^{-8.86+0.0309X_1+0.285X_3-1.44X_5+4.38X_6}}{1+e^{-8.86+0.0309X_1+0.285X_3-1.44X_5+4.38X_6}} \times (e^{7.28+0.0287X_1+0.0136X_3+0.0842X_5+1.5X_6})$$

$$E[S] = \frac{e^{-1.58+0.0596X_1+0.2986X_3-1.3558X_5+5.88X_6}}{1 + e^{-8.86+0.0309X_1+0.285X_3-1.44X_5+4.38X_6}} \quad (7)$$

Persamaan (7) merupakan model dari premi yang akan digunakan dalam menghitung nilai harapan biaya klaim tahunan.

III. KESIMPULAN

Distribusi yang cocok untuk data frekuensi klaim adalah distribusi Bernoulli dengan *link function* menggunakan link logit, sedangkan untuk data besar klaim distribusi yang cocok adalah distribusi gamma dengan *link function* menggunakan link log. Berdasarkan hasil yang telah diperoleh menunjukkan bahwa faktor *underwriting* yang berpengaruh signifikan terhadap penentuan premi adalah usia, BMI, jumlah anak dan merokok. Dengan demikian, untuk menentukan premi nasabah baru dapat memperhatikan faktor *underwriting* tersebut dalam menentukan besaran premi dengan menggunakan model GLM sesuai pada Persamaan (7).

REFERENSI

[1] A. Supriatma, Riaman, Sudradjat and S. Tari, "Generalized Linear Models (GLM) untuk Data Asuransi dalam Menentukan Harga Premi," *Seminar Nasional Pendidikan Matematika*, pp. 276-281, 2017.

[2] T. A. J. Putra, D. C. Lesmana and I. G. P. Purnaba, "Perhitungan Premi Asuransi

- Kendaraan Bermotor Menggunakan Generalized Linear Models dengan Distribusi Tweedie," *Jambura Journal of Mathematics*, p. 118, 2021.
- [3] S. A. Rohmaniah and Danardono, "Perhitungan Harga Premi Model Dua Tahunan dengan Faktor Underwriting Menggunakan Generalized Linear Models," *Konferensi Nasional Penelitian Matematika dan Pembelajarannya II*, p. 124, 2017.
- [4] C. K. Waha, A. J. Rindengan and T. Manurung, "Model Distribusi Data Klaim Asuransi Mobil untuk Menentukan Premi Murni," *Jurnal Matematika dan Aplikasi*, p. 110, 2019.
- [5] A. A., *Foundations Linear Generalized Linear Models*, Cambridge: John Wiley & Sons, Inc., 2015.
- [6] I. N. Wardani, "Estimasi Parameter Model Regresi Probit Multinomial dengan Metode Maximum Likelihood," *Mathunse*, vol. Volume 8 No. 2, p. 209, 2020.
- [7] A. Rofiqoh, *Pengembangan Grafik-P Menggunakan Ekspansi Cornish-Fisher*, Surabaya, 2016.
- [8] J. G. Z. and H. P. D., *Generalized Linear Models for Insurance Data*, Cambridge: Cambridge University Press, 2008.
- [9] J. Garrido, C. Genest and J. Schulz, "Generalized Linear Models for Dependent Frequency and Severity of Insurance Claims," *Insurance: Mathematics and Economics*, p. 206, 2016.
- [10] J. Zahro, R. E. Caraka and R. Herliansyah, *Aplikasi Generalized Linear Model Pada R*, Yogyakarta: Innosain, 2018.
- [11] K. Faizah, *Regresi Gamma dan Regresi Lognormal untuk memodelkan Biomassa Klorofil- α di Sungai Danube Eropa*, Surabaya, 2018.
- [12] "Estimasi Parameter pada Model Poisson Generalized Autoregressive Moving Average (GARMA) dengan Algoritma IRLS Studi Kasus: Peramalan Jumlah Kecelakaan di Jalan Tol Surabaya-Gempol," *SAINS DAN SENI ITS*, Vols. Vol. 8, No. 1, p. A18, 2019.
- [13] M. K. Kotimah and S. P. Wulandari, "Model Regresi Logistik Biner Stratifikasi Pada Partisipasi Ekonomi Perempuan Di Provinsi Jawa Timur," *ITS : Jurnal Sains dan Seni POMITS*, p. h.2, 2014.
- [14] D. E. Putri, *Pemodelan Geographically Weighted Gamma Regression (GWGR), Studi Kasus: Pencemaran Sungai di Kota Surabaya Tahun 2013*, Surabaya, 2016.
- [15] Y. A. Tampil, "Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Samratulangi Manado," *Unsrat : Jurnal Matematika dan Aplikasi*, p. h. 58, 2018.