

ANALISIS REGRESI POISSON INVERSE GAUSSIAN UNTUK MENENTUKAN FAKTOR-FAKTOR YANG MEMPENGARUHI JUMLAH KASUS DEMAM BERDARAH *DENGUE* (DBD)

Muhammad Marizal^{1*}, Veny Alvionita²

Prodi Matematika, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. HR. Soebrantas No. 155 Simpang Baru, Panam, Pekanbaru, 28293
e-mail: m.marizal@uin-suska.ac.id

diterima: 6 Desember 2022; direvisi: 30 Mei 2023; disetujui: 30 Mei 2023

ABSTRAK

Salah satu data cacahan yaitu jumlah kasus DBD dimana untuk memodelkannya data tersebut dapat dilakukan dengan menggunakan regresi poisson. Regresi poisson mempunyai asumsi yang harus dipenuhi yaitu nilai rata-rata dan variansinya harus sama atau disebut (equidispersi). Namun pada penelitian ini data mengalami pelanggaran yang mana nilai variansinya lebih besar dari nilai rata-ratanya atau disebut (overdispersi). Untuk mengatasi hal tersebut dengan menggunakan metode mixed poisson yaitu regresi poisson inverse gaussian untuk data cacahan yang mengalami overdispersi dan memiliki fungsi likelihood yang close form yaitu parameternya diketahui, sehingga banyak sekali peneliti yang menggunakan model ini. Jumlah kasus DBD di Provinsi Riau adalah data cacahan yang mengalami kasus overdispersi. Oleh karena itu untuk memodelkan jumlah kasus DBD dengan menggunakan regresi poisson inverse gaussian. Berdasarkan model yang digunakan tersebut menunjukkan faktor-faktor yang mempengaruhi jumlah kasus DBD diprovinsi riau pada tahun 2019 yaitu jumlah sarana Kesehatan.

Kata Kunci: Demam Berdarah Dengue, *Overdispersi*, Regresi Poisson Inverse Gaussian

POISSON INVERSE GAUSSIAN REGRESSION ANALYSIS TO DETERMINE FACTORS AFFECTING THE NUMBER OF DENGUE FEVER CASES

ABSTRACT

One of the census data is the number of dengue cases where to model the data can be done using poisson regression. Poisson regression has an assumptions that must be met, namely the average value dan variance must be the same or colled (equidispersion). But, in this study, the data experienced a violation in ehich the variance value was greater than the average value or colled (overdispersion). To overcome thie problem by using the mixed poisson method, namely poisson inverse gaussian regression for count data that is overdispersed and has a likelihood function that is cloce form, that is the parameters are know, so many researchers use this model. The number of DBD cases in Riau Province is counting data that has overdispersion. So therefore, to model the number of cases of DBD using poisson inverse gaussian regression. Based on the model used, it shows the factors that influence the number of dengue cases in Riau Province in 2019, namely the number of health facilities.

Keywords: Dengue Fever. *Overdispersion*, Poisson Inverse Gaussian Regression

PENDAHULUAN

Penyakit Demam Berdarah *Dengue* (DBD) yaitu suatu penyakit yang ditimbulkan dari virus *Dengue* dan disebarkan oleh nyamuk *Aedes Aegypty* lewat gigitannya dan telah terinfeksi oleh virus *Dengue* tersebut. Masa inkubasi pada

virus *Dengue* dalam diri manusia sekitar 3 sampai 14 hari sebelum munculnya gelajagejala klinis, sedangkan virus *Dengue* pada tubuh nyamuk berkisar 8 sampai 10 hari.

Pada tahun 1952 penyakit demam berdarah *Dengue* (DBD) ditemukan pertama kali di Manila (Filipina). Kemudian

menyebarkan ke negara lain seperti Thailand, Vietnam, Malaysia dan juga Indonesia. Penyakit ini muncul pertama kali di Indonesia pada tahun 1968 tepatnya di Kota Surabaya. Kemudian penyakit DBD ini terus meningkat dan menyebar keseluruh wilayah yang ada di Indonesia. Penyakit DBD ini hampir setiap tahun selalu ditemukan diseluruh Indonesia khususnya pada saat musim penghujan. Selain faktor iklim dan kondisi lingkungan, ada faktor yang menunjukkan bahwa kasus DBD ini berhubungan dengan mobilitas dan kepadatan penduduk serta perilaku masyarakat (Budijanto, 2019).

Berdasarkan penjelasan Dinas Kesehatan Riau kasus Demam Berdarah *Dengue* (DBD) pada tahun 2018 dan 2019 mengalami peningkatan sebesar 59.9 dari 100.000 penduduk dari 12 per 100.000 penduduk dari sasaran yang telah ditentukan yaitu > 49 per 100.000 penduduk. Dengan demikian hampir semua Kabupaten atau Kota mengalami peningkatan perkara DBD.

Regresi poisson digunakan untuk memodelkan suatu data cacahan seperti data jumlah kasus Demam Berdarah *Dengue* (DBD) (Adiatma, 2021). Model regresi poisson ini mempunyai asumsi yang wajib dipenuhi yaitu nilai rata-rata dan variansinya wajib sama besar yang mana hal ini dikatakan *Equidispersi*. Tetapi jika nilai varians diatas rata-rata dapat dikatakan terjadinya *Overdispersi*. Sedangkan, jika nilai varians dibawah rata-rata hal ini disebut *Underdispersi* (Sundari, 2021). Namun pada penelitian ini asumsi tidak terpenuhi sehingga data mengalami pelanggaran asumsi yaitu data mengalami *Overdispersi*, hal ini mengakibatkan *Underestimate* pada estimasi standar eror untuk memperoleh hasil kesimpulannya. Maka untuk menangani hal tersebut dilakukan penggabungan model distribusi poisson dengan distribusi diskrit atau kontinu, yang disebut *mixed poisson distribution*. Peneliti menggunakan metode dari *Mixed poisson* yaitu Regresi Poisson *Inverse Gaussian* (PIG) (Ramadhan, 2017). Beberapa penelitian menggunakan model ini untuk menggambarkan terjadinya penyakit.

Ramadhan (2017) menggunakan metode ini dengan penelitian penyakit kusta di Provinsi Sulawesi Tenggara. Adapun Herindrawati (2017) melakukan pemodelan dengan menggunakan metode ini dengan kasus HIV di Provinsi Sulawesi Selatan. Namun belum terdapat penelitian yang menggunakan metode ini untuk menganalisis penyakit DBD di Provinsi Riau. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi jumlah kasus DBD di Provinsi Riau dengan menggunakan Regresi Poisson Inverse Gaussian (PIG).

BAHAN DAN METODE

Data yang digunakan pada penelitian ini yaitu data sekunder. Data diambil dari Dinas Kesehatan Provinsi Riau dan Badan Pusat Statistik (BPS) tahun 2019 dengan kasus Demam Berdarah *Dengue* (DBD). Metode dari *Mixed Poisson Distribution* ditentukan berdasarkan parameternya, terdapat dua parameter yaitu rata-rata (μ) dan parameter *dispersi* (τ). Metode tersebut dinamakan distribusi poisson inverse gaussian. Adapun parameternya yaitu sebagai berikut.

$$P(Y = y|\mu) = \frac{\mu^y e^{\frac{1}{\tau}}}{y!} \left(\frac{2}{\mu\tau}\right)^{\frac{1}{2}} (2\mu\tau + 1)^{\left(\frac{y-\frac{1}{2}}{2}\right)} K_{y-\frac{1}{2}}\left(\frac{1}{\tau}\sqrt{2\mu\tau + 1}\right)$$

Adapun rata-rata dan variansi pada distribusi poisson inversse gaussian sebagai berikut:

$$\begin{aligned} E(Y) &= E\{E(Y|\mu\nu)\} = E(\mu\nu) = \mu \\ \text{Var}(Y) &= \text{Var}\{E(Y|\mu\nu)\} \\ &\quad + E\{\text{Var}(Y|\mu\nu)\} \\ &= \mu + \tau\mu^2 \end{aligned}$$

Model dari regresi poisson inverse gaussian yaitu sebagai berikut:

$$\begin{aligned} P(Y = y|x_i; \beta; \tau) &= \left\{ \frac{e^{x_i^T \beta y i e^{\frac{1}{\tau}}}}{y_i!} \left(\frac{2}{\mu\tau}\right)^{\frac{1}{2}} (2e^{x_i^T \beta} \tau \right. \right. \\ &\quad \left. \left. + 1\right)^{-\frac{(y_i-\frac{1}{2})}{2}} K_{y-\frac{1}{2}}\left(\frac{1}{\tau}\sqrt{2\mu\tau + 1}\right) \right\} \end{aligned}$$

Dari persamaan tersebut maka metode yang digunakan untuk menaksirkan

parameter β pada regresi poisson inverse gaussian yaitu *maximum likelihood*, adapun distribusi poisson inverse gaussian sebagai berikut:

$$L(\beta; \tau) = \prod_{i=1}^n P(Y = y_i | x_i; \beta; \tau)$$

$$l(\beta; \tau) = \prod_{i=1}^n \left\{ \frac{\mu_i^{y_i} e^{-\frac{1}{\tau}}}{y_i!} \left(\frac{2}{\pi\tau} \right)^{\frac{1}{2}} (2\mu_i\tau + 1) + \sum_{i=1}^n \left(\frac{2y_{i-1}}{4} \right) \ln K_{y-\frac{1}{2}} \left(\frac{1}{\tau} \sqrt{2\mu\tau + 1} \right) \right\}$$

Selanjutnya, diubah fungsi *likelihood* kedalam bentuk logaritma natural (ln) dan menghasilkan persamaan berikut:

$$L(\beta; \tau) = \ln L(\beta; \tau)$$

$$= \sum_{i=1}^n y_i x_i^T \beta + \frac{n}{\tau} - \ln \left(\sum_{i=1}^n y_i! \right)$$

$$+ \ln \left(\frac{2}{n} \right) - \left(\frac{n}{2} \right) \ln \tau$$

$$- \sum_{i=1}^n \left(\frac{2y_{i-1}}{4} \right) \ln(2x_i^T \beta + 1)$$

$$+ \sum_{i=1}^n \left(\frac{2y_{i-1}}{4} \right) \ln K_5(z)$$

Pengujian secara simultan digunakan dalam mengetahui pengaruh dari semua variabel penjelas secara simultan terhadap variabel respon dengan hipotesis. Uji G mengikuti distribusi yang bernama chi-square sehingga uji G dibandingkan dengan table chi-square dengan derajat bebasnya. Sedangkan uji parsial digunakan untuk mengetahui pengaruh dari setiap variabel penjelas secara satu persatu terhadap variabel respon dengan hipotesisnya.

HASIL DAN PEMBAHASAN

Dalam melakukan pengujian terhadap data, maka hal terlebih dahulu yang harus dilakukan adalah menganalisis deskriptif pada variabel independen dan variabel dependennya untuk mengetahui masing-masing karakteristik dari variabel yang digunakan pada penelitian. Variabel dependen (y) adalah jumlah kasus DBD,

sementara variabel independent terdiri dari 5 variabel yaitu x_1 rata-rata lama sekolah, x_2 jumlah sarana kesehatan, x_3 kepadatan penduduk per km², x_4 ketinggian wilayah dan x_5 jumlah tenaga kesehatan.

Statistika Dekskriptif

Tabel 1. Statistika Deskriptif

Variabel	Min	Maks	Rata-Rata	Variansi
y	60	1116	390,01	108013,71
x_1	7	11	8,42	1,174
x_2	241	884	534,17	3,721
x_3	38	1808	221,42	2,513
x_4	2	91	18,25	787,841
x_5	12	61	22,25	201,295

Berdasarkan Tabel 3.1, diketahui nilai rata-rata atau *mean* sebesar 390,08 dan nilai variansinya sebesar 108013,719 pada jumlah kasus DBD di Provinsi Riau tahun 2019. Adapun jumlah kasus DBD terbanyak pada tahun 2019 yaitu di Kabupaten Bengkalis sebesar 1116 kasus. Sedangkan jumlah kasus DBD terkecil pada tahun 2019 yaitu di Kabupaten Palalawan sebesar 60 kasus.

Overdispersi

Berdasarkan pada tabel analisis deskriptif didapatkan hasil nilai variansi pada data penelitian lebih besar dari pada rata-rata maka penelitian ini mengalami *overdispersi*. Adapun cara lain untuk melihat adanya kasus *overdispersi* dapat diperoleh dari deviansi dibagi dengan nilai derajat bebas. Dengan menggunakan *package* AER pada *software* R didapatkan nilai devians = 786,25 dan nilai derajat bebas = 6, maka diperoleh hasil nilai *overdispersi* yaitu sebesar 131,042, karena nilai yang didapatkan lebih dari 1 maka dapat disimpulkan bahwa terjadi *overdispersi*.

Pemodelan Regresi Poisson Inverse Gaussian

Setelah dilakukan analisis data dengan regresi poisson *inverse gaussian* (PIG) diperoleh beberapa kombinasi model regresi poisson inverse gaussian (PIG) yang konvergen, kemudian ditentukan model terbaiknya. Berikut empat kemungkinan model yang konvergen diantaranya sebagai berikut:

$$\begin{aligned} \mu &= \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5) \\ \mu &= \exp(\beta_0 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5) \\ \mu &= \exp(\beta_0 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4) \\ \mu &= \exp(\beta_0 + \beta_2 x_2 + \beta_3 x_3) \end{aligned}$$

Sebelum menentukan model terbaik dilakukan pengujian estimasi parameter serta pengujian parameter. Berikut adalah tabel estimasi parameter model-model yang

kemungkinan menjadi model terbaik. berikut merupakan tabel dari estimasi parameter model regresi poisson *inverse gaussian* (PIG).

Dari Tabel 2, didapatkan hasil estimasi parameter dari keempat model *Regresi Poisson Inverse Gaussian* (PIG). Dari keempat model berikut selanjutnya melakukan pengujian hipotesis untuk mengetahui pengaruh antara variabel dependen dan variabel independen.

Tabel 2. Estimasi Parameter Model Regresi Poisson Inverse Gaussian

Variabel	x_1, x_2, x_3, x_4, x_5	x_2, x_3, x_4, x_5	x_2, x_3, x_4	x_2, x_3
β_0	17,043206	9,091159	0,19365	2,0203936
β_1	-0,305054	-	-	-
β_2	-0,738549	0,215596	0,6014	0,5069325
β_3	0,008408	0,004327	-0,00047124	-0,0013151
β_4	0,001652	-0,004367	-0,0000007794	-
β_5	-0,009463	0,052036	-	-
τ	5,1566	7,4997	-0,9147	-0,8432

Pengujian Hipotesis

Terdapat kemungkinan model yang cocok pada model regresi poisson *inverse gaussian* dilakukan pengujian secara simultan atau bersama-sama. Adapun hipotesisnya sebagai berikut:

- a. $H_0: \beta_1 = \beta_2 = \dots \beta_k$ (variabel independen secara simultan tidak berpengaruh terhadap variabel dependen)

- b. H_1 : Minimal satu terdapat $\beta_i \neq 0$ dimana $i = 1, 2, \dots, k$ (variabel independen secara simultan berpengaruh terhadap variabel dependen)

Kriteria penolakan apabila $G > X^2_{(a,v)}$ maka tolak H_0 . Berikut diperoleh hasil dari pengujian secara simultan.

Tabel 3. Pengujian Parameter Secara Simultan

Variabel Dari Model	Statistik G	Df	$X^2_{(a,v)}$	Keputusan
x_1, x_2, x_3, x_4, x_5	174,6128	6	12,59159	Tolak H_0
x_2, x_3, x_4, x_5	171,8096	7	14,06714	Tolak H_0
x_2, x_3, x_4	157,142	8	15,50731	Tolak H_0
x_2, x_3	157,7182	9	16,91898	Tolak H_0

Berdasarkan Tabel 3, dinyatakan bahwa seluruh kemungkinan model diperoleh hasil Statistik $G > X^2_{(a,v)}$, maka keputusan yang diperoleh yaitu tolak H_0 , yang mana terdapat minimal satu parameter yang berpengaruh signifikan terhadap model tersebut. Selanjutnya untuk mengetahui pengaruh variabel prediktor yang signifikan pada model, maka dilanjutkan dengan pengujian parameter secara parsial atau individu.

Pengujian Parsial

Pengujian parsial dilakukan untuk mengetahui pengaruh signifikan antara variabel independen dengan variabel dependen, dilihat dari kriteria penolakan jika nilai $Z > \chi^2_{(a,v)}$ atau taraf signifikan (α) > P -value. Adapun hipotesis pengujian secara parsial yaitu sebagai berikut.

- a. $H_0: \beta_i = 0$ (variabel independen tidak berpengaruh terhadap variabel dependen)

- b. $H_1: \beta_i \neq 0$; dimana $i = 1,2,3, \dots, p$.
 (variabel independen berpengaruh terhadap variabel dependen)
- Adapun kriteria penolakan apabila nilai $Z > \chi^2_{(\alpha, v)}$ atau taraf signifikan (α) > P-value. Berikut hasil pengujian secara parsial.

Tabel 4. Pengujian Parameter Secara Parsial

Variabel dari model	Pr (> t)					Parameter signifikan
	x_1	x_2	x_3	x_4	x_5	
x_1, x_2, x_3, x_4, x_5	0,344	0,693	0,555	0,635	0,787	-
x_2, x_3, x_4, x_5	-	0,920	0,734	0,617	0,758	-
x_2, x_3, x_4	-	0,0446	0,5157	0,9853	-	β_2
x_2, x_3	-	0,0222	0,1911	-	-	β_2

Berdasarkan Tabel 4, diatas menunjukan bahwa model yang mempengaruhi ada pada model ketiga dan keempat dengan parameter β_2 . Karena model ketiga dan keempat memenuhi kriteria yang mana nilai $p - value < a$ (0.05) maka keputusan yang diperoleh Tolak H_0 . Sehingga β_2 yang mempengaruhi jumlah kasus DBD diprovinsi Riau merupa kan variabel independen x_2 (jumlah sarana kesehatan).

Pemilihan Model Terbaik

Metode *Akaike Information Criteria* (AIC) digunakan untuk menentukan model terbaik. Adapun hasil dari AIC pada model regresi poisson *inverse gaussian*.

Tabel 5. AIC dari Model Regresi Poisson *Inverse Gaussian*

Variabel dari model	AIC
x_1, x_2, x_3, x_4, x_5	188,6128
x_2, x_3, x_4, x_5	183,8096
x_2, x_3, x_4	167,142
x_2, x_3	165,7182

Berdasarkan Tabel 5, diatas bahwa diketahui model yang memiliki nilai AIC terkecil yaitu 165,7182 dengan variabel x_2, x_3 . Maka dari itu model terbaik yaitu pada model keempat. Berikut merupakan hasil dari penaksiran parameter Model Regresi Poisson *Inverse Gaussian*.

Tabel 6. Penaksiran Parameter Model Regresi Poisson *Inverse Gaussian*

Parameter	Taksiran	Stand.Error	Z _{hitung}	P-value
β_0	2,0203	1,4683	1,376	0,2061
β_2	0,5069	0,1791	2,830	0,0222
β_3	-0,0013	0,0009	-1,428	0,1911
τ	-0,8432	0,4680	-1,802	0,109

Berdasarkan Tabel 6, merupakan pendugaan parameter model regresi poisson *inverse gaussian*. Dimana nilai $p - value$ pada jumlah sarana kesehatan (x_2) sebesar 0.0222 kurang dari $a = 0.05$ sehingga tolak H_0 . Maka dari itu jumlah sarana kesehatan (x_2) berpengaruh signifikan terhadap jumlah kasus DBD. Adapun model regresi poisson *inverse gaussian* yaitu sebagai berikut:

$$\mu = \exp(2,0203 + 0,5069 x_2)$$

Berdasarkan dari model diatas, dapat diinterpretasikan pada model regresi poisson *inverse gaussian* bahwa untuk setiap penambahan 1rasio sarana kesehatan maka

akan melipatgandakan rata-rata DBD di Provinsi Riau sebesar $\exp(0,5069) = 1,6601$ kali dari rata-rata variabel respon jika variabel lainnya tetap. Dapat diartikan bahwa semakin banyak sarana kesehatan maka penularan kasus DBD di Provinsi Riau semakin meningkat dari rata-rata semula dengan syarat variabel lain dianggap tetap.

KESIMPULAN

Pada penelitian ini diperoleh hasil analisis regresi poisson *inverse gaussian* bahwa faktor yang mempengaruhi jumlah kasus DBD di Provinsi Riau tahun 2019

secara signifikan yaitu jumlah sarana kesehatan.

DAFTAR PUSTAKA

- Adiatma, P (2018) .“Analisis Faktor-Faktor yang Mempengaruhi Penyebaran Penyakit Demam Berdarah Dengue (Dbd) di Provinsi Jawa Tengah dengan Metode Spatial Autoregressive Model dan Spatial Durbin Model,” *Indonesian Journal of Applied Statistics*, vol. I, no. 1, pp. 1–13,
- Budijanto, D. (2019). “*Profil Kesehatan Indonesia 2019*”.
- Sundari, M. (2021). “Penanganan Overdispersi Pada Regresi Poisson (Studi Kasus : Pengaruh Faktor Iklim Terhadap Jumlah Penderita Penyakit Demam Berdarah di Kota Bogor),” *Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 2, no. 1, pp. 1–9.
- Ramadhan, M. 2017 “Pemodelan Jumlah Kasus Penyakit Kusta Di Provinsi Sulawesi Tenggara Menggunakan Regresi Poisson Inverse Gaussian,” pp. 1–15.
- Herindrawati, A. (2017). “Pemodelan Regresi Poisson Inverse Gaussian Studi Kasus : Jumlah Kasus Baru HIV di Provinsi Jawa Tengah Tahun 2015,” *Sains dan Seni ITS*, vol. 6, no. 1, pp. 143–148.