ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 12, Nomor 4, Tahun 2023, Halaman 539 - 548
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

PEMODELAN JUMLAH KASUS PNEUMONIA PADA BALITA DI JAWA TIMUR MENGGUNAKAN METODE REGRESI POISSON INVERSE GAUSSIAN DILENGKAPI GUI-R

Krisdiana Nur Utami[1*], Sugito[2], Rukun Santoso[3]
[1,2,3] Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
***e-mail** : krisdiananur@gmail.com

DOI: 10.14710/J.GAUSS.12.4.539-548

Article Info:
Received: 2023-02-14
Accepted: 2024-06-26
Available Online: 2024-03-29

Keywords:
Poisson Inverse Gaussian Regression; Overdipersion; Pneumonia; GUI-R

Abstract: Reducing toddler mortality is one of the desire of sustainable development programs. Modeling count data may be analyzed the usage of Poisson regression.The assumption that must be met in Poisson regression is that the mean and variance values must be equal, often in count data there is a violation of this assumption. This is indicated by the variance value which is greater than the mean value (overdispersion). Poisson Inverse Gaussian (PIG) regression is one form of mixed Poisson regression to model data that experience overdispersion cases. The MLE method is used to estimate the PIG regression parameters and hypothesis testing using the MLTR method. The best model of the PIG regression form is based on the smallest AIC value. The results of hypothesis testing concluded that the percentage of under-fives who received exclusive breast feeding had a significant effect on the number of pneumonia cases among toddler. Data modeling using the PIG regression method in this study is complemented by the creation of a Graphical User Interface (GUI) that can facilitate the process of selecting the best model.

1. PENDAHULUAN

Sustainable Development Goals atau pembangunan berkelanjutan diharapkan mampu tercapai pada waktu mendatang. Target yang harus dicapai pada tahun 2030 yaitu mampu menurunkan angka kematian balita. Pneumonia adalah penyakit pada saluran pernafasan yang menjadi penyebab utama kematian anak usia balita yang kemungkinan akan memberikan dampak negatif di usia dewasa (Riskesdas, 2013). Data jumlah kasus pneumonia merupakan salah satu data cacahan. Data cacahan tidak cocok dimodelkan dengan regresi klasik, sebab akan melanggar asumsi yaitu, bersifat homokedastisitas dan tidak berdistribusi normal (Widiari, 2016). *Generalized Linear Models* atau GLM adalah bentuk perluasan dari model regresi klasik yang biasa disebut sebagai alternatif untuk data yang tidak berdistribusi normal (De Jong dan Heller, 2008). Bentuk dari pemodelan GLM adalah regresi poisson,yang digunakan untuk memodelkan data cacahan yang telah memenuhi asumsi equidispersi.

Data jumlah kasus pneumonia pada balita pada penelitian ini mengalami overdispersi. Keadaan suatu data dengan nilai varians lebih besar daripada *means* akan mengakibatkan kesimpulan yang diperoleh tidak valid serta hilangnya informasi penting karena *underestimate* pada pendugaan standar error. Suatu data yang mengalami overdispersi tidak dapat diselesaikan dengan model regresi Poisson. Cara mengatasi overdispersi adalah membentuk beberapa pemodelan yang merupakan kombinasi dari distribusi poisson dengan beberapa distribusi baik diskret maupun kontinu (*mixed poisson distribution*). Model regresi *mixed* poisson sangat berguna dalam mengatasi overdispersi (Dean *et al*., 1989). Salah satu *mixed poisson distribution* yang digunakan dalam penelitian adalah distribusi Poisson Inverse Gaussian (PIG) dengan random efek distribusi inverse gaussian.

Penelitian ini dikembangkan dengan mengidentifikasi faktor-faktor yang memiliki pengaruh terhadap jumlah kasus pneumonia pada balita di Provinsi Jawa Timur berdasarkan prinsip *backward elimination*. Bertujuan untuk mendapatkan model dan mampu menentukan faktor yang diduga berpengaruh secara signifikan terhadap jumlah kasus pneumonia pada balita di Jawa Timur.

2. TINJAUAN PUSTAKA

Pneumonia merupakan salah satu penyakit yang menyerang saluran pernapasan yang disebabkan oleh bakteri, jamur, dan virus. *Streptococcus pneumoniae* dan *Haemophilus influenzae* merupakan bakteri penyebab pneumonia khususnya menyerang usia balita (Sari dan Cahyati, 2019). Balita yang pernah terjangkit pneumonia akan memiliki konsekuensi jangka panjang yang akan kambuh di usia dewasa, seperti gangguan fungsi paru-paru dan saluran pernafasan lainnya. Pneumonia masih menjadi masalah kesehatan masyarakat yang belum teratasi (Riskesdas, 2013). Pasien penderita pneumonia memiliki jaringan paru-paru (alveoli) yang berisi nanah dan cairan, yang membuat pernapasan menjadi menyakitkan dan mengakibatkan penderita kekurangan oksigen. Pneumonia menimbulkan jumlah kematian yang tinggi, karena sedikitnya perhatian yang diberikan untuk penyakit tersebut (Said, 2010). Sepanjang tahun 2021 Provinsi Jawa Timur menempati urutan paling atas dengan cakupan penemuan kasus pneumonia usia balita sebesar 3.330.329 dan diperkirakan 148.200 balita yang terjangkit pneuomonia (Kementerian Kesehatan RI, 2021).

Distribusi Poisson adalah distribusi diskret yang peluang kejadiannya relatif kecil serta bergantung pada interval waktu tertentu ataupun pada wilayah tertentu (Walpole,1995). Suatu variabel random diskrit Y akan berdistribusi Poisson apabila memiliki bentuk fungsi kepadatan peluang diskrit oleh (Myers, 1990) pada persamaan (1) :

$$f(y,\mu) = \frac{e^{-\mu}\mu^{y}}{y!} \qquad \text{dengan } y = 0, 1, 2, \dots \text{ dan } \mu > 0. \tag{1}$$

Rataan dan variansi pada distribusi poisson mempunyai nilai yang sama ditunjukkan pada persamaan (2) :

$$E(Y) = Var(Y) = \mu \tag{2}$$

Distribusi Inverse Gaussian merupakan salah satu distribusi kontinu yang termasuk ke dalam *Generalized Linier Models* (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 pada Distribusi Inverse Gaussian merupakan variabel non negatif dengan fungsi kepadatan peluang yang dinotasikan sebagai $Y \sim IG(\mu, \sigma^2)$, (Jong dan Heller, 2008) :

$$f(y) = (2\pi y^3 \sigma)^{-0,5} e^{-(y-\mu)^2/(2y\mu^2\sigma^2)}, \; y > 0, \mu > 0, \sigma^2 > 0 \tag{3}$$

Rata-rata dan varian dari distribusi Inverse Gaussian adalah :

$$E(Y) = \mu \text{ dan } Var\,(Y) = \sigma^2 \mu^3 \tag{4}$$

dan σ^2 adalah parameter *disperse*, paramater *disperse* adalah parameter yang menunjukkan keadaan overdispersi pada data.

Overdispersi merupakan keadaan suatu data yang memiliki nilai varians lebih besar dari pada nilai rata-rata $\{Var(Y) > E(Y))\}$. Data yang mengalami overdispersi akan menyebabkaan kesalahan fatal dalam interpretasi model, terutama pada estimasi parameter model karena dapat menaksir *standard error* yang terlalu rendah dan dapat memberikan kesimpulan yang keliru tentang signifikan atau tidaknya parameter regresi yang terlibat.

Pengujian overdispersi pada data dengan statistik uji *The Variance Test* (VT) yang ditunjukkan pada persamaan (5) :

$$VT = \sum_{i=1}^{n} \frac{(y-\bar{y})^2}{\bar{y}} = (n-1)\frac{s^2}{\bar{y}} \quad (5)$$

Nilai *Variance Test* (VT) merupakan indeks disperse. Dengan ketentuan nilai VT yang kurang dari 1 maka data mengalami underdispersi, apabila nilai VT lebih dari 1 maka terjadi overdispersi (Karlis dan Xekalaki, 2000).

Distribusi Poisson Inverse Gaussian merupakan salah satu distibusi *mixed* poisson. Distribusi PIG mempunyai dua parameter yaitu rata-rata (μ) yang disebut juga parameter lokasi dan parameter dispersi (τ) atau disebut parameter bentuk. Fungsi kepadatan peluang distribusi PIG pada persamaan (6) (Zha *et al*., 2014) :

$$P(Y = y\,|\mu,\tau) = \frac{\mu^{y} e^{\frac{1}{\tau}}}{y_i!}\left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}}(2\mu\tau+1)^{-\left(\frac{y-\frac{1}{2}}{2}\right)} K_{y-\frac{1}{2}}\left(\frac{1}{\tau}\sqrt{2\pi\tau+1}\right), y > 0, \mu > 0, \tau > 0 \quad (6)$$

Rumus rataan dari distribusi PIG adalah :

$$E(Y) = \mu \quad (7)$$

Rumus variansi dari distribusi PIG pada persamaan (7):

$$Var(Y) = \mu + \tau\mu^2 \quad (8)$$

Model regresi PIG dapat dituliskan pada persamaan (9):

$ln(\mu_i) = \mathbf{x}_i^T\boldsymbol{\beta}$

atau

$$E(Y_i|X_i) = \mu_i = exp(\mathbf{x}_i^T\boldsymbol{\beta}) \quad (9)$$

$$E(Y_i|X_i) = \mu_i = exp(\beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki}) \;, \; i = 1,2,\ldots,n$$

Parameter $\boldsymbol{\beta}$ yang sesuai dengan model regresi PIG pada persamaan (9) dapat diestimasi menggunakan *Maximum Likelihood Estimation.* Model dari regresi PIG dapat dilakukan mengikuti langkah-langkah pada persamaan (10) :

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

$$L(\boldsymbol{\beta};\tau) = \prod_{i=1}^{n}\left\{\frac{e^{x_i^T\beta y_i} e^{\frac{1}{\tau}}}{y_i!}\left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}}(2\mu_i\tau+1)^{-\frac{\left(y_i-\frac{1}{2}\right)}{2}} K_{si}(Z_i)\right\} \quad (10)$$

Fungsi likelihood tersebut diubah bentuk logaritma natural (ln) sehingga pada persamaan (10) dapat dinyatakan dalam persamaan (11) :

$$L(\boldsymbol{\beta};\tau) = \ln L(\boldsymbol{\beta};\tau)$$

$$= \sum_{i=1}^{n} y_i\,\boldsymbol{x}_i^T\boldsymbol{\beta} + \frac{n}{\tau} - ln\left(\sum_{i=1}^{n} y_i\,!\right) + \frac{n}{2}\,ln\left(\frac{2}{\pi}\right) - \frac{n}{2}\ln\tau \quad (11)$$

$$- \sum_{i=1}^{n}\left(\frac{2y_i-1}{4}\right)\ln(2\boldsymbol{x}_i^T\boldsymbol{\beta\tau}+1) + \sum_{i=1}^{n} lnK_{si}(z_i)$$

Fungsi dimaksimumkan dengan bantuan algoritma *Fisher Scoring* oleh Smyth (2002) untuk menentukan penaksir *maksimum likelihood* parameter $\boldsymbol{\theta} = [\boldsymbol{\beta}^T \;\; \tau]^T$ dengan persamaan (12) :

$$\widehat{\boldsymbol{\theta}}_{(r+1)} = \widehat{\boldsymbol{\theta}}_{(r)} + \boldsymbol{I}^{-1}(\widehat{\boldsymbol{\theta}}_{(m)})\boldsymbol{D}(\widehat{\boldsymbol{\theta}}_{(m)}). \quad (12)$$

dengan

$\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}, \hat{\tau})^T$

$\boldsymbol{D}(\hat{\theta}) = \left(\frac{\partial l}{\partial\hat{\tau}}, \frac{\partial l}{\partial\widehat{\boldsymbol{\beta}}^T}\right)^T$

$\boldsymbol{I}(\hat{\theta}_{(m)}) = -E[\boldsymbol{H}(\widehat{\boldsymbol{\theta}}_{(m)})]$

$$\boldsymbol{H}\left(\hat{\theta}_{(m)}\right)_{(k+1)(k+1)} = \begin{bmatrix} \frac{\partial^2 l}{\partial \hat{\tau}^2} & \frac{\partial^2 l}{\partial \hat{\tau} \partial \widehat{\boldsymbol{\beta}}} \\ \frac{\partial^2 l}{\partial \widehat{\boldsymbol{\beta}} \partial \hat{\tau}} & \frac{\partial^2 l}{\partial \widehat{\boldsymbol{\beta}} \partial \widehat{\boldsymbol{\beta}}^T} \end{bmatrix}_{\theta=\theta_m}$$

Sehingga $\boldsymbol{I}\left(\widehat{\boldsymbol{\theta}}_{(\boldsymbol{m})}\right) : -\boldsymbol{E}\begin{bmatrix} \frac{\partial^2 l}{\partial \hat{\tau}^2} & \frac{\partial^2 l}{\partial \hat{\tau} \partial \widehat{\boldsymbol{\beta}}} \\ \frac{\partial^2 l}{\partial \widehat{\boldsymbol{\beta}} \partial \hat{\tau}} & \frac{\partial^2 l}{\partial \widehat{\boldsymbol{\beta}} \partial \widehat{\boldsymbol{\beta}}^T} \end{bmatrix}$

Langkah-langkah pengerjaan dengan menggunakan *Fisher Scoring Algorithm* yaitu :

1. Menentukan vektor awal parameter $\widehat{\boldsymbol{\theta}}_0$ dengan mengasumsikan data memenuhi model regresi linier berganda (McCullagh dan Nelder, 1989) :
   $\boldsymbol{Y}^* = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i \,; i = 1,2,..,n$

   dengan metode kuadarat terkecil diperoleh:
   $\widehat{\boldsymbol{\beta}}_0 = (\mathbf{X^T X})^{-1}\mathbf{X^T Y}^*$
2. Membentuk vektor gradien $\boldsymbol{D}\left(\widehat{\boldsymbol{\theta}}_{(0)}\right)$.
3. Membentuk matriks hessian $\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}_{(0)}\right)$.
4. Membentuk matrik informasi Fisher $\boldsymbol{I}\left(\widehat{\boldsymbol{\theta}}_{(0)}\right)$.
5. Memasukkan nilai $\widehat{\boldsymbol{\theta}}_0$ sehingga diperoleh vektor gradien $\boldsymbol{D}\left(\widehat{\boldsymbol{\theta}}_{(0)}\right)$ dan matriks hessian $\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}_{(0)}\right)$.
6. Memulai dari m = 0 dilakukan iterasi pada $\widehat{\boldsymbol{\theta}}_{(r+1)} = \widehat{\boldsymbol{\theta}}_{(r)} + \mathbf{I}^{-1}\left(\widehat{\boldsymbol{\theta}}_{(m)}\right)\boldsymbol{D}\left(\widehat{\boldsymbol{\theta}}_{(m)}\right)$, nilai $(\widehat{\boldsymbol{\theta}}_{(m)})$ merupakan sekumpulan penaksir parameter yang konvergen saat iterasi ke-m.
7. Jika belum diperoleh penaksiran parameter yang konvergen saat iterasi ke-m, maka dilanjutkan kembali ke langkah 5 sampai iterasi ke- m+1. Itersai akan berhenti apabila nilai dari $\left\|\widehat{\boldsymbol{\theta}}_{(m+1)} - \widehat{\boldsymbol{\theta}}_{(m)}\right\| \leq \varepsilon\, , dan\; \varepsilon > 0$.

Pengujian parameter model regresi PIG dilakukan secara serentak dengan bantuan metode *Maximum Likelihood Ratio Test* (MLRT) atau berdasarkan nilai statistik uji G (Mc. Cullagh dan Nelder, 1989). Hipotesis pengujian secara serentak yaitu :

H_0: $= \beta_1 = \beta_2 \ldots = \beta_k = 0$

H_1: minimal ada satu $\beta_i \neq 0$, dengan i = 1, 2, …, k.

Statistik Uji :

$$G_{hitung} = -2\ln\left(\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right) \tag{13}$$

dengan:

$L(\widehat{\omega})$: model likelihood tanpa variabel prediktor

$L(\widehat{\Omega})$: model likelihood dengan variabel prediktor

Nilai G tersebut mengikuti distribusi *chi square* dengan derajat bebas v. Kriteria pengujian parameternya adalah menolak H_0 ketika nilai $G_{hitung} > \chi^2(\alpha, v)$, maka minimal terdapat satu parameter yang diduga mempunyai pengaruh yang signifikan.

Hasil dari penaksiran parameter model regresi PIG belum tentu menunjukkan adanya pengaruh yang signifikan terhadap model secara parsial, maka perlu dilakukan uji parsial pada parameter β dan parameter τ. Hipotesis pengujian secara parsial untuk masing-masing parameter yaitu :

Parameter β

Hipotesis :
H_0: β_i = 0, i = 1, 2, …,k (variabel prediktor ke-i tidak berpengaruh terhadap variabel respon)
H_1: $\beta_i \neq$ 0, i = 1, 2, …,k (variabel prediktor ke-i berpengaruh terhadap variabel respon)
Statistik uji parsial parameter β Statistik uji parsial terhadap parameter β menurut Nakaya *et al*. 2005 adalah :

$$t_{hitung} = \frac{\hat{\beta}_i}{Se(\hat{\beta}_i)} \quad (14)$$

Kriteria pengujian adalah tolak H_0 apabila $|t_{hitung}| > \boldsymbol{t(\frac{\alpha}{2}, n-k)}$, atau $p - value < \alpha$, dengan α adalah tingkat signifikasi yang digunakan, $Se(\hat{\beta}_i)=\sqrt{\widehat{var}(\hat{\beta}_i)}$ dengan $\widehat{var}(\beta_i)$ diperoleh dari elemen diagonal ke (m+1) dari matriks $\boldsymbol{I(\hat{\theta}_{(m)})} = -\boldsymbol{E}\left((\boldsymbol{H}^{-1}(\hat{\boldsymbol{\theta}})\right)$.

Parameter τ
Hipotesis :
H_0: τ = 0 (variabel respon tidak mengalami overdispersi)
H_1: τ ≠ 0 (variabel respon mengalami overdispersi)
Statistik uji parsial parameter τ :

$$t_{hitung} = \frac{\hat{\tau}}{Se(\hat{\tau})} \quad (15)$$

Kriteria pengujian adalah tolak H_0 apabila $|t_{hitung}| > \boldsymbol{t(\frac{\alpha}{2}, n-k)}$, atau $p - value < \alpha$, dengan α adalah tingkat signifikasi yang digunakan, $Se(\hat{\tau})=\sqrt{\widehat{var}(\hat{\tau})}$ dengan $\widehat{var}(\hat{\tau})$ diperoleh dari elemen diagonal ke (m+1) dari matriks $\boldsymbol{I(\hat{\theta}_{(m)})} = -\boldsymbol{E}\left((\boldsymbol{H}^{-1}(\hat{\boldsymbol{\theta}})\right)$.

Kriteria yang digunakan untuk mendapatkan model terbaik dapat didasarkan pada perolehan nilai *Akaike Information Criterion* (AIC). Nilai AIC dapat dirumuskan dalam persamaan (16):

$$AIC = 2k - 2\ln\left(L(\hat{\theta})\right) \quad (16)$$

dengan $L(\hat{\theta})$ adalah estimasi fungsi *likelihood* dan k merupakan banyaknya parameter yang ditaksir dalam suatu model regresi. Jika terdapat banyak model dalam suatu data penelitian, model dengan nilai AIC terkecil adalah model terbaik (Akaike, 1978).

Tidak adanya gejala multikolinearitas antara variabel prediktor merupakan prasyarat untuk terciptanya model regresi yang baik. Pendekteksian adanya kasus multikolinieritas dapat diketahui yaitu :

1. Jika koefisien korelasi Pearson (r_{ij}) antar variabel prediktor lebih tinggi dari 0,95 maka terdapat korelasi antar variabel tersebut.
2. Nilai VIF (*Varian Inflation Factor*) lebih besar dari 10 menunjukkan adanya multikolinieritas antar variabel prediktor. Nilai VIF dinyatakan dalam persamaan (17):

$$VIF_j = \frac{1}{\left(1-R_j^2\right)} \quad (17)$$

Nilai R_j^2 merupakan koefisien determinasi antara satu variabel prediktor dengan variabel prediktor lainnya (Hocking, 1996).

3. METODE PENELITIAN

Data sekunder dari profil Dinas Kesehatan Jawa Timur tahun 2021 pada website https://dinkes.jatimprov.go.id merupakan jenis data yang digunakan dalam penelitian ini. Terdapat 38 kabupaten dan kota di Jawa Timur sebagai jumlah dari observasi. Variabel penelitian disajikan dalam Tabel 1 :

Tabel 1. Variabel Penelitian

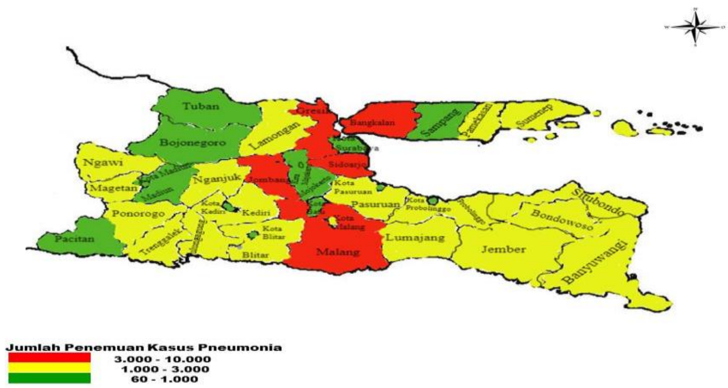
Variabel	Keterangan
Y	Jumlah Kasus Pneumonia
X_1	Persentase Balita Kurang Gizi
X_2	Persentase Pemberian ASI Ekslusif
X_3	Persentase Pemberian Imunisasi Campak
X_4	Persentase Pemberian Imunisasi DPT HB-HIB
X_5	Persentase Cakupan Pelayanan Kesehatan Balita

Pada penelitian ini penulis menambahkan tampilan Grapical User Interface (GUI) R dari paket Shiny sebagai alat pengolah data. Langkah dalam analisis data yaitu :

1. Mempersiapkan aplikasi GUI R yang sesuai dengan keperluan analisis.
2. Mengumpulkan data dan berbagai sumber informasi ilmiah yang mendukung proses penelitian.
3. Mendeskripsikan karakteristik data jumlah kasus pneumonia dan faktor-faktor yang mempengaruhinya.
4. Pemeriksaan multikolinieritas berdasarkan nilai koefisien korelasi dan nilai VIF.
5. Pengujian overdispersi.
6. Pemodelan dengan regresi Poisson Inverse Gaussian (PIG).
7. Mencari model terbaik berdasarkan kriteria nilai AIC terkecil.
8. Membuat interpretasi hasil pemodelan.
9. Menyusun kesimpulan berdasarkan analisis yang telah dilakukan.

4. HASIL DAN PEMBAHASAN

Pada akhir tahun 2021, sebaran kasus pneumonia di Provinsi Jawa Timur sebanyak 61.345 kasus per kabupaten atau kota. Seluruh kabupaten dan kota di Jawa Timur telah terjangkit pneumonia, dengan jumlah kasus terbanyak ditemukan di Kabupaten Sidoarjo 9.308 kasus dan terendah di Kabupaten Pacitan sebanyak 63 kasus. Gambar 1 menggambarkan sebaran jumlah kasus pneumonia pada balita di Jawa Timur pada tahun 2021.



Gambar 1. Persebaran Jumlah Kasus Pneumonia pada Balita di Jawa Timur tahun 2021

Hasil dari statistika deskriptif data dapat dilihat pada Tabel 2.

Tabel 2. Statistika Deskriptif Variabel Respon dan Variabel Prediktor

	N	Minimun	Maximum	Mean	Std. Deviation	Variance
Y	38	63,0	9308,0	1614.34	1771.908	3139659.420
X_1	38	0,8	14,8	7,650	3,2337	10,457
X_2	38	42,1	92,2	72,812	14,3211	205,094
X_3	38	13,9	100,6	83,784	18,3446	336,525
X_4	38	29,8	141,1	77,958	18,6929	349,425
X_5	38	35,6	102,6	87,195	13,1086	171,835

Tabel 2 menunjukkan bahwa rata-rata jumlah kasus pneumonia pada balita di Provinsi Jawa Timur pada tahun 2021 terdapat 1614.34 ≈ 1615 kasus. Jumlah kasus tertinggi sebanyak 9308 kasus, dan terendah sebanyak 63 kasus. Langkah selanjutnya perlu dilakukan pemeriksaan multikolinieritas terlebih dahulu. Hasil pemeriksaan multikolinieritas berdasarkan nilai koefisien korelasi pada Tabel 3.

Tabel 3. Koefisien Korelasi Variabel Respon dan Variabel Prediktor

	X_1	X_2	X_3	X_4
X_2	-0,099			
X_3	0,019	0,3450		
X_4	-0,009	0,0543	0,5162	
X_5	-0,014	0,1593	0,0351	0,1495

Nilai koefisien korelasi Pearson (r_{ij}) semua variabel prediktor memiliki koefisien korelasi Pearson yang kurang dari 0,95 berarti tidak terdapat kasus multikolinieritas.
Cara lain untuk mengetahui adanya kasus multikolinieritas adalah berdasarkan nilai VIF. Nilai VIF dari setiap variabel prediktor yang menunjukkan nilai kurang dari 10, maka disimpulkan tidak terdapat kasus multikolinieritas.

Tabel 4. Hasil Uji Multikolinieritas

Variabel	X_1	X_2	X_3	X_4	X_5
VIF	1,015167	1,220860	1,610095	1,446846	1,061856

Regresi Poisson Inverse Gaussian merupakan model yang cocok diterapkan pada data yang memiliki distribusi sangat menceng dan mengalami overdipersi atau data yang memiliki nilai varians *(variance)* lebih besar dari rata-rata (*means).*

Tabel 5. Uji Overdispersi

$\bar{y}$	s	VT
1614.34	1771.908	71959,65

Pengujian overdispersi berdasarkan Tabel 5 diperoleh nilai (VT) sebesar 71959,65. Hasil VT yang didapatkan melebihi nilai 1, sehingga pemodelan dengan regresi Poisson Inverse Gaussian dapat dilakukan.

Metode yang digunakan untuk menentukan model terbaik adalah *backward elimination* yang didasarkan pada nilai *Akaike Information Criterion* (AIC) dengan beberapa variabel prediktor yang ada pada penelitian ini akan dieliminasi untuk mendapatkan nilai AIC terkecil. Lima variabel prediktor yang digunakan menghasilkan tiga kombinasi model yang sudah konvergen, selanjutnya dicari model terbaiknya. Tiga kombinasi model regresi Poisson Inverse Gaussian yang sudah konvergen adalah :

Tabel 6. Nilai AIC pada Kemungkinan Model regresi Poisson Inverse Gaussian

Variabel dari model	AIC
$X_1, X_2, X_3 X_4, X_5$	647,9986
X_1, X_2, X_3, X_5	646,0374
$\boldsymbol{X_1, X_2, X_5}$	**645,2935**

Nilai AIC dari model regresi Poisson Inverse Gaussian berdasarkan Tabel 6 dapat dilihat bahwa model ketiga merupakan model yang memiliki nilai AIC terkecil yaitu 645,2935 dengan variabel yang terkandung adalah X_1, X_2, X_5. Oleh karena itu model ketiga yang dipilih sebagai model terbaik.

Tabel 7. Penaksiran Parameter Model Regresi PIG pada Jumlah Kasus Pneumonia di Jawa Timur

Parameter	Taksiran	Standart Error	$\mathbf{t}_{hitung}$	***P-Value***
β_0	7,76881	1,37295	5,658	0,00000264
β_1	-0,07912	0,05287	-1,497	0,1440
β_2	-0,02229	0,01079	-2,066	0,0468
β_5	-0,0203	0,01359	1,474	0,1499
τ	0,4588	0,3107	1,477	0,149

Penaksiran parameter model regresi berdasarkan Tabel 7 diperoleh informasi bahwa persentase balita yang mendapatkan ASI ekslusif (X_2) adalah variabel prediktor yang berpengaruh secara signifikan terhadap kasus pneumonia pada balita. Nilai *p-value* dari (X_2) sebesar 0,0468 yang berarti kurang dari $\alpha = 0{,}05$ sehingga H_0 ditolak. Nilai *p-value* pada parameter τ yang kurang dari $\alpha = 0{,}05$ dan menunjukkan bahwa data mengalami kasus overdispersi.

Langkah selanjutnya adalah melakukan pengujian parameter model regresi Poisson Inverse Gaussian. Pengujian parameter secara serentak dilakukan pada kemungkinan model yang sesuai dengan model regresi PIG. Hasil pengujian parameter secara serentak seperti yang disajikan dalam Tabel 7:

Tabel 8. Pengujian Parameter Regresi PIG Secara Serentak

Variabel dari model	Statistik G	db	$\boldsymbol{\chi^2_{(\alpha,db)}}$	Keputusan
X_1, X_2, X_5	635.2935	35	49,8018	Tolak H_0

Nilai statistik uji G yang lebih besar dari nilai $\chi^2_{(\alpha,db)}$ untuk semua model yang mungkin, menunjukkan bahwa setidaknya terdapat satu parameter memiliki pengaruh yang signifikan terhadap model.

Pengujian parameter secara parsial dapat dilihat dari nilai *p-value* dan t-hitung. Kriteria penolakan pengujian secara parsial adalah apabilla nilai p - $value < \alpha$ atau $|t_{hit}| > t(\frac{\alpha}{2}, n-k-1)=$ 2,036 maka H_0 ditolak. Berdasarkan nilai *p-value* pada Tabel 6 menunjukkan bahwa variabel yang signifikan dalam model adalah β_2. Sehingga model regresi Poisson Inverse Gaussian (PIG) yang terbentuk adalah $\hat{\mu}$=exp$(7{,}76881 - 0{,}07912X_2 - 0{,}02229 - 0{,}0203X_5)$. Intepretasi dari model regresi yang terbentuk adalah nilai konstanta (β_0) sebesar 7,76881 menyatakan jika semua variabel prediktor tidak diperhatikan maka rata-rata kasus pneumonia balita di Provinsi Jawa Timur adalah sebesar exp(7,76881)= 2365,65 atau kurang lebih 2366 kasus. Parameter β_1 sebesar -0,07912 menyatakan bahwa setiap penambahan 1% persentase balita kurang gizi maka akan sebanding dengan penurunan rata-rata kasus penumonia (Y) sebesar exp$(-0{,}07912)$ = 0,92329 kali dari rata-rata semula dengan variabel lainnya tetap. Parameter β_2 sebesar -0,02229 menyatakan bahwa setiap penambahan 1% persentase balita yang mendapatkan ASI ekslusif (X_2) maka akan sebanding dengan penurunan rata-rata kasus penumonia (Y) sebesar exp$(-0{,}02229)$ = 0,97795 kali dari rata-rata semula dengan variabel lainnya tetap. Parameter β_5 sebesar -0,0203 menyatakan bahwa setiap penambahan 1% persentase cakupan pelayanan kesehatan balita (X_5) maka akan sebanding dengan penurunan rata-rata kasus penumonia (Y) sebesar exp$(-0{,}0203)$ = 0,9799 kali dari rata-rata semula dengan variabel lainnya tetap.

Penelitian ini dilengkapi aplikasi *Graphical User Interface* (GUI) R. Pembuatan aplikasi GUI R digunakan untuk mempermudah pemodelan menggunakan metode regresi Poisson Inverse Gaussian dengan cara menginput data. Pembuatan aplikasi GUI R diperlukan *package* R salah satunya yaitu *Shiny*. Dua komponen utama dalam pembuatan aplikasi GUI R terdiri dari *User Interface* (UI) dan *Server*. Proses pengolahan data yang dijalankan oleh GUI akan didefinisikan dalam *server*, input yang telah dimasukkan akan dipanggil menggunakan identitas yang telah dibuat dalam objek UI. Tampilan GUI regresi Poisson Inverse Gaussian yang telah dijalankan seperti pada Gambar 2.

Gambar 2. Tampilan Awal GUI R Pemodelan Regresi Poisson Inverse Gaussian

5. KESIMPULAN

Lima daerah dengan jumlah kasus pneumonia pada balita terbanyak di Provinsi Jawa Timur tahun 2021 yaitu Kabupaten Sidoarjo, Kabupaten Malang, Kabupaten Gresik, Kabupaten Jombang, serta Kabupaten Bangkalan, sebagaimana ditentukan dari analisis dan pembahasan pemodelan kasus pneumonia pada balita menggunakan regresi Poisson Inverse Gaussian. Dengan range jumlah kasus pneumonia tertinggi berkisar antara 3000 - 10.000 kasus pada wilayah tersebut. Terbukti jumlah kasus pneumonia di Provinsi Jawa Timur pada tahun 2021 berlebihan.Sebab itu penelitian ini dilakukan dengan menggunakan metode regresi PIG. Sehingga diperoleh hasil sebagai berikut untuk model regresi Poisson Inverse Gaussian pada jumlah kasus pneumonia pada usia balita yaitu $\hat{\mu}$=exp$(7{,}76881 - 0{,}07912X_1 - 0{,}02229X_2 - 0{,}0203X_5)$. Persentase balita yang mendapatkan ASI ekslusif (X_2) merupakan faktor yang berpengaruh signifikan terhadap jumlah kasus pneumonia pada balita di Provinsi Jawa Timur pada tahun 2021.

DAFTAR PUSTAKA

Akaike, H. (1978). *A Bayesian Analysis of The Minimum AIC Procedure*. Annals of the Institute of Statistical Mathematics, Part A Hal. 914

Dean, C., Lawless, J. F. dan Willmot, G.E. (1989). *A Mixed Poisson-inverse Gaussian Regression Model. The Canadian Journal of Statistics*, Vol. 17, No. 2, hal. 171-181.

De Jong, P. dan Heller, G.Z. (2008), "*Generalized Linear Model for Insurance Data"*, 1st edition, Cambridge University, Press, New York.

Dinas Kesehatan Provinsi Jawa Timur. (2022). *Profil Kesehatan Provinsi Jawa Timur 2021*.

Karlis, D. dan Xekalaki, E. (2000). A Simulation Comparison of Several Procedures for Testing the Poisson Assumption. The Statistician. Vol. 49, No. 3, hal. 355-382.

Kementerian Kesehatan RI. (2021). *Profil Kesehatan Indonesia 2021*. Kementrian Kesehatan Republik Indonesia. https://pusdatin.kemkes.go.id/

McCullagh, P., dan J. A. Nelder, (1989), Generalized Linear Models, 2nd Ed., Chapman and Hall, New York.

Myers. R. H., (1990). *Classical and Modern Regression with Applicaton.* Boston: PWS-KENT Publishing Company.

Nakaya, T., Fotheringham, A. S., Brunsdon, C., dan Charlton, M. (2005). Geographically weighted Poisson regression for disease association mapping.*Statistics In medicine*. *24*(17), 2695-271.

Said, M. (2010). *Pengendalian Pneumonia Anak Balita dalam Rangka Pencapaian MDG4*. Buletin Jendela Epidemilogi. 3: 16-21.

Sari, M.P. dan Cahyati, W. H. (2019). *Tren Pneumonia Balita di Kota Semarang Tahun 2012-2018.* Jurnal HIGEIA. 3(3): 407-416.

Smyth, G. K. (2002). *Optimation*. In Shaarawi, A.H.E., & Piegorsch, W.W. (Ed). *Encyclopedia of Environmetrics*, Vol. 3, (pp. 1481-1487), Chicester: John Wiley & Sons

Riset Kesehatan Dasar (Riskesdas). (2013). Badan Penelitian dan Pengembangan Kesehatan Kementerian RI tahun 2013. Diakses 2 September 2022.

Walpole, R.E. (1995). *Pengantar Metode Statistika.* Ahli Bahasa: Ir. Bambang Sumantri, Jakarta: PT Gramedia Pustaka Utama.

Widiari, S, M. (2016). *Penaksiran Parameter dan Statistik Uji dalam Model Regresi Poisson Inverse Gaussian (PIG).* Tesis. Institut Teknologi Sepuluh Nopember.

Zha, L., Lord, D. dan Zou, Y. (2014), " *The Poisson Inverse Gaussian (PIG) Generalized Linear Regression Model for Analyzing Motor Vehicle Crash Data". Journal of Transportation Safety and Security*. DOI: 10.1080/19439962.2014.977502.