ESTIMASI: Journal of Statistics and Its Application
Vol. 6, No. 1, Januari, 2025, Hal. 101-115
DOI: 10.70561/ejsa.v6i1.43164

Pendekatan Zero-Inflated Poisson Inverse Gaussian dalam Pemodelan Kasus Malaria di Puskesmas Kota Makassar

Fauziah Nurhidayah[1], **Raupong**[2], **A. Muthiah Nur Angriany**[3*]
[123]Departemen Statistika, Fakultas MIPA, Universitas Hasanuddin, Makassar, 90245, Indonesia)
*** Corresponding author, email: amuthiah97@gmail.com**

Abstract

Poisson regression is one of the approaches used to model count data. However, this method has an assumption of equidispersion that is not always met in actual data. One problem that often arises is overdispersion, especially when there are excess zeros in the dependent variable. The Mixed Poisson method, namely Zero-Inflated Poisson Inverse Gaussian (ZIPIG) regression is one approach that can be used when there is overdispersion in the data. Parameter estimation in the ZIPIG model is done using the Maximum Likelihood Estimation (MLE) method through Fisher Scoring Algorithm iterations. This study discusses how ZIPIG modeling is used to identify factors that influence the number of malaria cases in Makassar City Health Center in 2021. The results of the analysis show that the independent variables that have a significant effect on the number of malaria cases are the number of family heads with access to proper sanitation facilities (X_1) *and the presence of public places that meet health requirements* (X_2)*.*

Keywords*: Excess Zero, Malaria, Maximum Likelihood Estimation, Overdispersi, Zero-Inflated Poisson.*

Abstrak

Regresi Poisson merupakan salah satu pendekatan yang digunakan untuk memodelkan data cacahan. Namun, metode ini memiliki asumsi equidispersi yang tidak selalu terpenuhi pada data aktual. Salah satu permasalahan yang sering muncul adalah overdispersi, terutama ketika terdapat kelebihan angka nol pada variabel dependen. Metode Mixed Poisson, yaitu regresi Zero-Inflated Poisson Inverse Gaussian (ZIPIG), merupakan salah satu pendekatan yang dapat digunakan ketika terjadi overdispersi pada data. Estimasi parameter pada model ZIPIG dilakukan menggunakan metode Maximum Likelihood Estimation (MLE) melalui iterasi Algoritma Fisher Scoring. Penelitian ini membahas bagaimana pemodelan ZIPIG digunakan untuk mengidentifikasi faktor-faktor yang memengaruhi jumlah kasus malaria di Puskesmas Kota Makassar tahun 2021. Hasil analisis menunjukkan bahwa variabel independen yang berpengaruh signifikan terhadap jumlah kasus malaria adalah jumlah kepala keluarga dengan akses terhadap fasilitas sanitasi yang layak (X_1) dan keberadaan tempat umum yang memenuhi syarat kesehatan (X_2).

Kata Kunci: *Excess Zero*, Malaria, *Maximum Likelihood Estimation*, *Overdispersi*, *Zero-Inflated Poisson.*

1. Pendahuluan

Data cacahan merupakan sekumpulan bilangan bulat non-negatif yang terdiri atas beberapa peristiwa diskrit [1]. Data cacahan sering kali tidak berdistribusi normal, memiliki kemiringan yang besar, dan memiliki nilai ekstrem, serta terkadang mengandung banyak angka nol [2]. Oleh karena itu, pemodelan data cacahan tidak dapat dilakukan dengan Ordinary Least Square (OLS) dan lebih sesuai dengan menggunakan pendekatan Generalized Linear Models (GLMs) yang tidak memerlukan asumsi normalitas dan asumsi heterosketastisitas dalam penerapannya [3]. Salah satu metode

dalam keluarga GLMs yang umum digunakan untuk memodelkan data cacahan dengan kejadian jarang adalah regresi Poisson [4].

Pemodelan data menggunakan Regresi Poisson mensyaratkan terpenuhinya asumsi equidispersi atau nilai variansi dan rataan sama. Namun, pada beberapa data cacahan yang berdistribusi Poisson sering kali ditemukan nilai variansi lebih kecil dari nilai rataan (underdispersi) ataupun nilai variansi lebih besar dari nilai rataan (overdispersi). Overdispersi pada data dapat menyebabkan estimasi standar eror menjadi terlalu kecil (underestimate), yang berpotensi menghasilkan kesimpulan yang tidak akurat [5]. Salah satu penyebab terjadinya overdispersi pada data yaitu tingginya proporsi angka nol pada variabel dependen. Untuk mengatasi permasalahan ini, pemodelan dapat dilakukan dengan pendekatan yang mengombinasikan distribusi diskrit dan distribusi kontinu dalam distribusi Poisson, yang dikenal sebagai metode Mixed Poisson Distribution [4].

Data cacahan yang mengikuti distribusi Poisson dan mengandung banyak angka nol dapat dimodelkan dengan salah satu dari metode Mixed Poisson yang menggabungkan antara distribusi Zero Inflated dan distribusi Poisson atau disebut regresi Zero Inflated Poisson [6]. Sementara itu, untuk memodelkan data cacahan yang memiliki kemiringan yang besar sehingga mengalami overdispersi dapat dilakukan dengan metode Mixed Poisson yang menggabungkan antara distribusi Poisson dan distribusi Inverse Gaussian atau disebut regresi Poisson Inverse Gaussian [7]. Pengembangan selanjutnya dari metode Mixed Poisson Distribution untuk memodelkan data cacahan yang mengikuti distribusi Poisson yang mengandung excess zeros dan mengalami overdispersi yaitu dengan menggabungkan antara Zero Inflated dengan Poisson Inverse Gaussian. Penerapan metode Zero Inflated Poisson Inverse Gaussian (ZIPIG) dalam konsepnya menggunakan model Zero Inflated mengatasi excess zeros dan model Poisson Inverse Gaussian mengatasi masalah overdispersi.

Data jumlah kasus malaria merupakan contoh data cacahan. Malaria merupakan penyakit yang ditularkan oleh hewan yaitu nyamuk Anopheles yang membawa mikroorganisme bernama Plasmodium. Penelitian sebelumnya terhadap faktor yang mempengaruhi jumlah kasus malaria pada beberapa wilayah di Indonesia antara lain oleh Sendow dan Chernovita (2021) bahwa faktor iklim menjadi penyebab terjadinya kasus malaria di Kabupaten Mimika [8]. Penelitian lain oleh Aprilia (2023) bahwa tingkat kepadatan penduduk, tingkat kemiskinan, persentase rumah tangga dengan sanitasi layak, angka kesakitan, persentase keluhan kesehatan, jumlah sarana kesehatan, jumlah tenaga kesehatan, dan persentase tempat umum yang sesuai standar kesehatan memiliki pengaruh signifikan terhadap kasus positif malaria di Jawa Timur [9].

Berdasarkan data Kementerian Kesehatan pada tahun 2021 jumlah wilayah di Indonesia dengan status bebas malaria yaitu 347 kabupaten/kota, jumlah tersebut lebih tinggi daripada tahun 2020 yaitu 318 kabupaten/kota. Provinsi Sulawesi Selatan memiliki total konfirmasi kasus positif malaria sebanyak 914 orang dengan persentase eliminasi malaria sebesar 87.5% dengan 20 Kabupaten/Kota berstatus bebas malaria dan 4

Kabupaten/Kota berstatus endemis rendah. Menurut Peta Endemisitas Malaria Tahun 2021 oleh Kementerian Kesehatan Republik Indonesia, Kota Makassar berstatus eliminasi malaria yang berarti bahwa jumlah kasus malaria di Kota Makassar mengalami penurunan secara signifikan sehingga telah mengurangi penularan yang berkelanjutan [10]. Sebanyak 27 kasus malaria tercatat dari 47 puskesmas di wilayah Kota Makassar [11]. Sementara itu, sesuai dengan salah satu dari tujuan Sustainable Develompment Goals (SDGs), Kementerian Kesehatan menargetkan agar semua wilayah di Indonesia terbebas dari penyakit malaria paling lambat pada tahun 2030 [10]. Oleh karena itu, maka pemodelan jumlah kasus malaria dilakukan untuk melihat faktor apa saja yang dapat mempengaruhi jumlah kasus malaria.

Penelitian sebelumnya menggunakan metode Poisson Inverse Gaussian dilakukan oleh Handarzeni (2022) terhadap kasus tuberkulosis di Sumatera dan Saraiva dkk. (2022) terhadap kasus demam berdarah dengue di Brazil [7][12]. Sementara itu, Nariswari dkk. (2023) melakukan pemodelan kasus stunting menggunakan metode Zero Inflated Poisson [13]. Secara umum, pendugaan parameter pada model Generalized Linear Models (GLMs) dapat dilakukan menggunakan metode Maximum Likelihood Estimation (MLE). Namun, metode ini menghasilkan sistem persamaan yang bersifat implisit dan nonlinier, sehingga diperlukan algoritma Fisher Scoring untuk menyelesaikannya [14]. Oleh karena itu, penelitian ini menggunakan pemodelan Regresi Zero Inflated Poisson Inverse Gaussian untuk menganalisis jumlah kasus malaria di Puskesmas Kota Makassar dengan metode Maksimum Likelihood yang diestimasi melalui iterasi algoritma Fisher Scoring.

2. Material dan Metode

2.1 Sumber dan Variabel Penelitian

Penelitian ini menggunakan data sekunder dari publikasi dengan judul Profil Dinas Kesehatan Kota Makassar Tahun 2021 yang diterbitkan pada Tahun 2022 pada laman website Dinas Kesehatan Kota Makassar. Variable yang digunakan adalah sebagai berikut:

Tabel 1. Variabel Penelitian

Variabel	Keterangan
Y	Jumlah positif malaria
X_1	Jumlah KK dengan akses terhadap fasilitas sanitasi yang layak
X_2	Tempat-tempat umum memenuhi syarat kesehatan
X_3	Jumlah sarana kesehatan
X_4	Jumlah tenaga kesehatan

2.2 Pemodelan Regresi Zero Inflated Poisson Inverse Gaussian

Regresi Zero Inflated Poisson Inverse Gaussian (ZIPIG) merupakan suatu pemodelan regresi gabungan dari metode Zero Inflated Poisson dan metode Poisson

Inverse Gaussian. Model ZIP berperan untuk mengatasi excess zeros dan model PIG mengatasi overdispersi pada data. Terdapat tiga parameter dalam model ZIPIG yaitu mean μ, dispersi τ, dan zero inflated p. Fungsi kepadatan peluang untuk model ZIPIG ketika $y = 0$ dapat dituliskan:

$$P(Y = 0|\mu, \tau, p) = p + (1 - p)\exp\left(\frac{1}{\tau}\right)\left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}}(2\mu\tau + 1)^{\frac{1}{4}} \left(\frac{\pi}{2\left(\frac{1}{\tau}\sqrt{2\mu\tau+1}\right)}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{\tau}\sqrt{(2\mu\tau + 1}\right) \quad (1)$$

sedangkan untuk $y = 1, 2, \ldots$ dituliskan sebagai berikut:

$$P(Y = y|\mu, \tau, p) = (1 - p)\frac{\mu^y}{y!}\exp\left(\frac{1}{\tau}\right)\left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}}(2\mu\tau + 1)^{-\frac{\left(y-\frac{1}{2}\right)}{2}} \left(\frac{\pi}{2\left(\frac{1}{\tau}\sqrt{2\mu\tau+1}\right)}\right)^{y-\frac{1}{2}} \exp\left(-\frac{1}{\tau}\sqrt{(2\mu\tau + 1}\right) \quad (2)$$

dengan $p = \frac{\exp(-\gamma \boldsymbol{X}_i^T \boldsymbol{\beta})}{1+\exp(-\gamma \boldsymbol{X}_i^T \boldsymbol{\beta})}$ dan $1 - p = \frac{1}{1+\exp(-\gamma \boldsymbol{X}_i^T \boldsymbol{\beta})}$

adapun model poisson state dan mode zero inflated dapat dilihat pada persamaan (3) dan persmaan (4).

$$\ln(\mu) = \boldsymbol{X}_i^T \boldsymbol{\beta} \quad (3)$$

$$\begin{aligned} logit(p) &= \ln\left(\frac{p}{1-p}\right) \\ &= -\gamma \boldsymbol{X}_i^T \boldsymbol{\beta} \end{aligned} \quad (4)$$

dengan mendistribusikan Persamaan (3) dan Persamaan (4) pada Persamaan (1) dan Persamaan (2) maka diperoleh fungsi peluang untuk model regresi ZIPIG ketika $y = 0$ adalah:

$$P(Y = 0|\boldsymbol{\beta}, \gamma, \tau) = \frac{exp(-\gamma \boldsymbol{X}_i^T \boldsymbol{\beta})}{1+exp(-\gamma \boldsymbol{X}_i^T \boldsymbol{\beta})} + \frac{1}{1+exp(-\gamma \boldsymbol{X}_i^T \boldsymbol{\beta})} exp\left(\frac{1}{\tau}\right)\left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}} (2\, exp(\boldsymbol{X}_i^T \boldsymbol{\beta})\, \tau + 1)^{\frac{1}{4}} \left(\frac{\pi}{2\left(\frac{1}{\tau}\sqrt{2\, exp(\boldsymbol{X}_i^T \boldsymbol{\beta})\tau+1}\right)}\right)^{\frac{1}{2}} exp\left(-\frac{1}{\tau}\sqrt{2\, exp(\boldsymbol{X}_i^T \boldsymbol{\beta})\, \tau + 1}\right) \quad (5)$$

dan untuk $y = 1, 2, \ldots$ dituliskan sebagai berikut:

$$P(Y = y|\boldsymbol{\beta}, \gamma, \tau) = \frac{1}{1+exp(-\gamma \boldsymbol{X}_i^T \boldsymbol{\beta})} \left(\frac{\left(exp(\boldsymbol{X}_i^T \boldsymbol{\beta})\right)^{y_i} exp\left(\frac{1}{\tau}\right)}{y_i!}\right)\left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}} (2\, exp(\boldsymbol{X}_i^T \boldsymbol{\beta})\, \tau + 1)^{-\frac{\left(y_i-\frac{1}{2}\right)}{2}}$$

$$\left(\frac{\pi}{2\left(\frac{1}{\tau}\sqrt{2\,exp(\boldsymbol{X}_i^T\boldsymbol{\beta})\tau+1+1}\right)}\right)^{y_i-\frac{1}{2}}$$
$$exp\left(-\frac{1}{\tau}\sqrt{2\,exp(\boldsymbol{X}_i^T\boldsymbol{\beta})\,\tau+1+1}\right) \quad (6)$$

2.2 Pengujian Asumsi Regresi Zero Inflated Poisson Inverse Gaussian

Dalam pemodelan ZIPIG, terdapat beberapa asumsi yang harus diuji, yaitu distribusi variabel dependen, multikolinearitas, dan overdispersi. Langkah pertama adalah menguji apakah variabel dependen mengikuti distribusi Poisson dengan menggunakan uji Kolmogorov-Smirnov. Hipotesis yang diuji dalam pengujian ini adalah:

H_0 : variable dependen berdistribusi poisson

H_1 : variable dependen tidak berdistribusi poisson

Statistik uji dalam pengujian asumsi distribusi Poisson dapat dihitung menggunakan rumus berikut:

$$D_{hitung} = maks\ |F_n(Y) - F_0(Y)|$$

dengan $F_n(Y)$ adalah fungsi distribusi kumulatif dibawah hipotesis dan $F_0(Y)$ adalah perbandingan antara fungsi distribusi kumulatif yang diamati. Hipotesis nol diterima (H_0) diterima apabila $D_{hitung} < D_{N;\alpha}$yang artinya variabel dependen berdistribusi poisson**.**

Selain itu, dalam pemodelan regresi, asumsi multikolinearitas harus diuji untuk memastikan bahwa tidak terdapat hubungan linear yang sangat tinggi antara variabel independen. Multikolinearitas dapat menyebabkan ketidakstabilan dalam taksiran regresi dan menghasilkan koefisien regresi yang besar. Untuk mendeteksi multikolinearitas, digunakan Variance Inflation Factor (VIF) yang dihitung menggunakan Persamaan 7 yaitu:

$$VIF = \frac{1}{1 - R_j^2} \quad (7)$$

dengan R_j^2 merupakan koefisien determinasi ke-j yang diperoleh dengan meregresikan variabel dependen [15]. Kriteria pengujian yaitu H_0 diterima apabila $VIF < 10$ yang berarti tidak terjadi multikolinearitas antar variabel independen yang digunakan.

Selain pengujian multikolinearitas, asumsi overdispersi perlu diuji dalam model ZIPIG. Overdispersi dapat diperiksa dengan membandingkan nilai deviance terhadap derajat bebas (db). Pengujian overdispersi dilakukan menggunakan persamaan:

$$VT = \frac{D}{db} \quad (8)$$

$$D = 2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right)\right] \tag{9}$$

dengan

D : Nilai devians
VT : The Variance test (parameter dispersi)
y_i : Nilai variabel independen dari pengamatan ke-i
db : Derajat bebas
$\hat{\mu}_i$: Penduga rataan

Kriteria pengujian yaitu H_0 ditolak apabila $VT > 1$ yang artinya nilai dispersi lebih besar dari 1, maka terjadi overdispersi [16]. Overdispersi menunjukkan bahwa model Poisson standar tidak cukup untuk menjelaskan variasi dalam data, sehingga diperlukan model yang lebih fleksibel seperti ZIPIG. Dengan demikian, pengujian asumsi ini menjadi langkah penting dalam memastikan bahwa model yang digunakan dapat menghasilkan estimasi yang valid dan akurat.

2.3 Estimasi Parameter

Estimasi parameter untuk regresi ZIPIG dapat dilakukan dengan metode MLE. Model ZIPIG memiliki dua model yang saling bebas, yaitu untuk variabel respon bernilai nol ($Y_i = 0$) dan model variabel respon bernilai lebih dari nol ($Y_i = 1, 2, \dots$). Fungsi likelihood model ZIPIG dapat ditulisakan sebagai berikut:

$$L(\boldsymbol{\beta}, \gamma, \tau) = \big(L_1(\boldsymbol{\beta}, \gamma, \tau)\big)\big(L_2(\boldsymbol{\beta}, \gamma, \tau)\big) \tag{10}$$

dengan

$$L_1(\boldsymbol{\beta}, \gamma, \tau) = \prod_{\substack{i=1 \\ y_i=0}}^{n} \frac{1}{1+\exp(-\gamma \boldsymbol{X}_i^T \boldsymbol{\beta})}\left[\exp(-\gamma \boldsymbol{X}_i^T \boldsymbol{\beta}) + \exp\left(\frac{1}{\tau} - \frac{1}{\tau}\sqrt{(2(\exp(\boldsymbol{X}_i^T \boldsymbol{\beta}))\tau) + 1}\right)\right] \tag{11}$$

$$L_2(\boldsymbol{\beta}, \gamma, \tau) = \prod_{\substack{i=1 \\ y_i>0}}^{n} \left(\frac{1}{1+\exp(-\gamma \boldsymbol{X}_i^T \boldsymbol{\beta})}\right)\left(\frac{(\exp(-\boldsymbol{X}_i^T \boldsymbol{\beta}))^{y_i} \exp\left(\frac{1}{\tau}\right)}{y_i!}\right) \left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}} [2\exp(\boldsymbol{X}_i^T \boldsymbol{\beta})\tau + 1]^{-\left(\frac{y_i - \frac{1}{2}}{2}\right)} \left(\frac{\pi}{2\left(\frac{1}{\tau}\sqrt{(2(\exp(\boldsymbol{X}^T \boldsymbol{\beta}))\tau) + 1}\right)}\right)^{y_i - \frac{1}{2}} \tag{12}$$

$$\exp\left(-\frac{1}{\tau}\sqrt{(2(\exp(\boldsymbol{X}^T\boldsymbol{\beta}))\tau)+1}\right)$$

Fungsi likelihood ini kemudian diubah menjadi bentuk log-likelihood:

$$l = \ln L(\boldsymbol{\beta},\gamma,\tau) = \ln L_1(\boldsymbol{\beta},\gamma,\tau) + \ln L_2(\boldsymbol{\beta},\gamma,\tau) \quad (13)$$

Karena turunan fungsi likelihood ini menghasilkan bentuk implisit, estimasi parameter dilakukan secara numerik menggunakan algoritma Fisher Scoring. Proses iterasi dalam algoritma Fisher Scoring diberikan oleh:

$$\widehat{\boldsymbol{\theta}}_{(m+1)} = \widehat{\boldsymbol{\theta}}_{(m)} + \boldsymbol{I}^{-1}\left(\widehat{\boldsymbol{\theta}}_{(m)}\right)\boldsymbol{D}\left(\widehat{\boldsymbol{\theta}}_{(m)}\right) \quad (14)$$

dengan:

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

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

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

Iterasi dimulai dengan estimasi awal parameter menggunakan model regresi linier berganda, kemudian memperbarui nilai parameter sampai konvergen berdasarkan kriteria $||\widehat{\boldsymbol{\theta}}_{(m+1)} - \widehat{\boldsymbol{\theta}}_{(m)}|| \leq \epsilon$, dengan ϵ adalah nilai yang sangat kecil mendekati nol.

2.4 Pengujian Parameter

Pengujian parameter dalam model Regresi ZIPIG dilakukan untuk mengevaluasi pengaruh variabel independen terhadap variabel dependen. Pengujian ini mencakup uji simultan terhadap parameter serta uji parsial terhadap parameter $\beta, \tau,$ dan γ.

a. Uji Simultan

Pengujian secara simultan dilakukan dalam penelitian ini menggunakan hipotesis sebagai berikut:

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

H_1: minimal ada satu $\beta_j \neq 0, j = 1, 2, \ldots, k$

Statistik uji yang digunakan yaitu:

$$G = -2\ln\left(\frac{\ln L(\widehat{\omega})}{\ln L(\widehat{\Omega})}\right) \sim \chi^2_{\alpha;v} \quad (15)$$

dengan Ω adalah ukuran statistik likelihood ratio yang dibentuk dengan menentukan himpunan di bawah populasi yaitu $\Omega = (\boldsymbol{\beta},\gamma,\tau)$ dan ω adalah himpunan parameter di bawah H_0 benar yaitu $\omega = (\beta_0, \gamma_0, \tau_0)$. Pada Ω fungsi likelihood untuk model penuh yang melibatkan seluruh variabel independen $L(\Omega)$, sedangkan pada ω dibentuk fungsi likelihood untuk model tidak melibatkan variabel independen $L(\omega)$.

Kriteria pengujian adalah H_0 ditolak apabila $G > \chi^2_{\alpha,v}$. artinya, secara bersama-sama variabel independen berpengaruh terhadap variabel dependen.

b. Uji Parsial

Pengujian parsial dilakukan untuk mengevaluasi signifikansi masing-masing parameter.

Hipotesis yang diuji adalah:

untuk parameter β_j: $H_0: \beta_j = 0$ Vs $H_1: \beta_j \neq 0$ dengan $j = 1, 2, \ldots, k$

untuk parameter τ: $H_0: \tau = 0$ Vs $H_1: \tau \neq 0$

Untuk parameter γ: $H_0: \gamma = 1$ Vs $H_1: \gamma \neq 1$

Statistik uji yang digunakan untuk melihat signifikan parameter dari setiap parameter adalah:

$$t = \frac{\hat{\theta}_j}{SE(\hat{\theta}_j)} \tag{16}$$

dengan $\hat{\theta}_j$ merupakan nilai estimasi parameter untuk $\boldsymbol{\beta}$, τ, dan γ $SE(\hat{\theta}_j)$ merupakan standar eror dari $\hat{\theta}_j$. Kriteria pengujian yaitu H_0 ditolak jika $|t_{hitung}| > t_{\alpha;db}$ atau $p - value < \alpha$. Artinya variabel independen berpengaruh secara signifikan terhadap variabel dependen.

2.5 Pemilihan Model Terbaik

Kriteria pemilihan model terbaik dapat dilihat melalui nilai Akaike Information Criterion (AIC) yang ditentukan oleh Akaike. Metode Nilai AIC dapat dihitung dengan persamaan berikut:

$$AIC = -2 \ln L(\hat{\boldsymbol{\theta}}) + 2k \tag{17}$$

dengan L merupakan nilai maksimum dari fungsi likelihood dan k merupakan banyak parameter yang diestimasi dalam model. Jika dalam sebuah set data terdapat beberapa model, maka model yang lebih baik adalah model dengan AIC kecil [17].

Pemilihan model terbaik juga perlu memperhatikan signifikansi parameter dispersi τ yang menunjukkan bahwa data tidak memenuhi kondisi equidispersi atau mengalami kondisi overdispersi [18]. Selain itu, pemilihan model terbaik juga perlu memperhatikan signifikansi dari parameter zero inflated γ yang menunjukkan bahwa data sesuai untuk model zero inflated [19].

3. Hasil dan Diskusi

3.1 Pengujian Asumsi Data

Hasil uji kecocokan distribusi Poisson menggunakan uji Kolmogorov-Smirnov menunjakkan bahwa nilai $D_{maks} = 0{,}0285$ lebih kecil dari nilai $D_{(47;0,05)} = 0{,}19420$.

Berdasarkan nilai tersebut maka H_0 diterima yang berarti bahwa variabel Jumlah Kasus Malaria di Puskesmas Kota Makassar berdistribusi Poisson. Selain itu, hasil pengujian asumsi multikolinearitas menunjukkan bahwa nilai Variance Inflation Factor (VIF) untuk masing-masing variabel independen dapat dilihat pada Tabel 2 berikut:

Tabel 2. Uji Multikolinearitas

	X_1	X_2	X_3	X_4
VIF	1,5216	1,6032	1,1025	1,0950

Berdasarkan hasil tersebut, nilai VIF pada setiap variabel independen lebih kecil dari 10 yang berarti bahwa tidak terjadi multikolinearitas antar variabel independen yang digunakan dalam penelitian. Selain itu, uji overdispersi menghasilkan nilai Variance Test (VT) sebesar 1,1173. Karena nilai VT lebih besar dari satu, maka terdapat indikasi overdispersi pada data jumlah kasus malaria di Puskesmas Kota Makassar

3.2 Pendugaan Model Regresi Zero Inflated Poisson Inverse Gaussian

Pemodelan data jumlah kasus malaria di Puskesmas Kota Makassar menggunakan ZIPIG membentuk empat kombinasi kemungkinan model berdasarkan kriteria pemilihan model terbaik,,yaitu:

$\hat{\mu}_i = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4)$
$\hat{\mu}_i = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_4)$
$\hat{\mu}_i = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$
$\hat{\mu}_i = \exp(\beta_0 + \beta_2 x_2)$

Nilai dugaan parameter pada model ZIPIG tersebut diperoleh dengan menggunakan metode MLE adalah sebagai berikut:

Tabel 3. Nilai Dugaan Parameter Dalam Model

Parameter	Model 1	Model 2	Model 3	Model 4
	X_1, X_2, X_3, X_4	X_1, X_2, X_4	X_1, X_2	X_2
β_0	2.3118	-1.0812	0.7238	0.1591
β_1	0.0013	7×10^{-4}	7×10^{-4}	
β_2	-0.4424	-0.2139	-0.2165	-0.0341
β_3	-3.3509			
β_4	0.0083	0.0757		
τ	-0.2113	3.7200	3.4790	1.2620
γ	-1.4910	-1.2200	-1.3940	-36

Tabel 3 menunjukkan nilai dugaan parameter untuk setiap model yang konvergen. Setelah diperoleh nilai dugaan parameter, pengujian parameter dilakukan secara simultan dan parsial untuk masing-masing model guna mengevaluasi kesesuaian model.

3.3 Pengujian Parameter

3.3.1 Uji Simultan

Pengujian secara simultan terhadap parameter pada model ZIPIG dilakukan dengan menentukan fungsi likelihood model penuh $L(\Omega)$ dan model tidak melibatkan variabel independen $L(\omega)$. Hipotesis pengujiannya adalah sebagai berikut:

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

H_1: minimal ada satu $\beta_j \neq 0, j = 1, 2, 3, 4$

hasil uji simultan yang merangkum pengujian terhadap parameter dapat dilihat pada Tabel 4 berikut:

Tabel 4. Hasi Pengujian Simultan

Model	Variabel	Statistik G	db	$\chi^2_{\alpha;db}$	Keputusan
1	X_1, X_2, X_3, X_4	188,6	40	55.76	Tolak H_0
2	X_1, X_2, X_4	123,5	41	56.94	Tolak H_0
3	X_1, X_2	123,2	42	58.12	Tolak H_0
4	X_2	100,5	43	59.30	Tolak H_0

Tabel 4 menunjukkan bahwa nilai Statistik G dalam setiap model dengan tingkat signifikansi $\alpha = 0.05$ lebih besar dari nilai $\chi^2_{\alpha;db}$ sehingga H_0 ditolak. Artinya dalam setiap model terdapat minimal satu parameter yang tidak nol atau setiap variabel pada masing-masing model minimal terdapat satu variabel independen yang berpengaruh secara signifikan terhadap jumlah kasus malaria.

3.3.2 Uji Parsial

Pengujian secara parsial terhadap parameter dilakukan untuk mengetahui bagaimana variabel independen mempengaruhi variabel dependen dalam model. Hipotesis pengujian secara simultan adalah sebagai berikut:
Untuk parameter β_j:

$H_0: \beta_j = 0$

$H_1: \beta_j \neq 0$ dengan $j = 1, 2, 3, 4$

Untuk parameter τ:

$H_0: \tau = 0$

$H_1: \tau \neq 0$

Untuk parameter γ:

$H_0: \gamma = 1$

$H_1: \gamma \neq 1$

hasil uji parsial memberikan informasi yang lebih mendalam mengenai kontribusi setiap parameter dalam model ZIPIG. Tabel 5 dan Tabel 6 memuat ringkasan hasil uji parsial, yang mencakup nilai koefisien regresi, nilai τ, nilai λ, serta signifikansi setiap parameter yang diuji.

Tabel 5. Uji Parsial Model 1 dan Model 2

Parameter	Model 1		Model 2	
	X_1, X_2, X_3, X_4		X_1, X_2, X_4	
	Nilai Dugaan	$p-value$	**Nilai Dugaan**	$p-value$
β_0	2.3118	0.0180	-1.0812	0.6589
β_1	0.0013^*	2×10^{-16}	$7 \times 10^{-4^*}$	0.0153
β_2	-0.4424^*	2×10^{-16}	-0.2139^*	0.0072
β_3	-3.3509^*	2×10^{-16}		
β_4	0.0083	0.7350	0.0757	0.3426
τ	-0.2113^*	2×10^{-16}	3.7200^*	3×10^{-8}
γ	-1.4910^*	0.0084	-1.2200^*	0.0310

Keterangan: * (Signifikan)

Tabel 5. menunjukkan bahwa parameter yang secara parsial berpengaruh signifikan pada Model 1 ialah parameter yang berkaitan dengan variabel Jumlah KK dengan akses terhadap fasilitas sanitasi yang layak (X_1), Tempat-tempat umum memenuhi syarat kesehatan (X_2), dan Jumlah tenaga kesehatan (X_3) serta parameter dispersi dan parameter *zero inflated*. Parameter yang secara parsial berpengaruh signifikan pada Model 2 ialah parameter yang berkaitan dengan variabel Jumlah KK dengan akses terhadap fasilitas sanitasi yang layak (X_1) dan Tempat-tempat umum memenuhi syarat kesehatan (X_2) serta parameter dispersi dan parameter *zero inflated*.

Tabel 6. Uji Parsial Model 3 dan Model 4

Parameter	Model 3		Model 4	
	X_1, X_2		X_2	
	Nilai Dugaan	$p-value$	**Nilai Dugaan**	$p-value$
β_0	0.7238	0.5145	0.1591	0.8100
β_1	$7 \times 10^{-4^*}$	0.0117		
β_2	-0.2165^*	0.0028	-0.0341	0.1900
β_3				
β_4				
τ	3.4790^*	3×10^{-8}	1.2620^*	0.0290
γ	-1.3940^*	0.0240	-36.0000	1.0000

Keterangan: * (Signifikan)

Tabel 6 menunjukkan bahwa parameter yang secara parsial berpengaruh signifikan pada Model 3 ialah parameter yang berkaitan dengan variabel jumlah KK dengan akses terhadap fasilitas sanitasi yang layak (X_1) dan Tempat-tempat umum memenuhi syarat

kesehatan (X_2) serta parameter dispersi dan parameter zero inflated. Parameter yang secara parsial berpengaruh signifikan pada Model 4 ialah parameter dispersi.

3.4 Pemilihan Model Terbaik

Pemilihan model terbaik dalam penelitian ini berdasarkan pada nilai AIC terkecil serta adanya parameter dispersi dan parameter zero inflated yang signifikan dalam model. Hasil perhitungan adalah sebagai berikut:

Tabel 7. Kriteria Pemilihan Model Terbaik

Model	AIC	t		γ	
		Nilai Dugaan	$p-value$	Nilai Dugaan	$p-value$
1	202.60	-0.2113^*	2×10^{-16}	-1.4910^*	0.0084
2	135.5	3.7200^*	3×10^{-8}	-1.2200^*	0.0310
3	133.2	3.4790^*	3×10^{-8}	-1.3940^*	0.0240
4	108.5	1.2620^*	0.0290	-36.0000	1.0000

Keterangan: * (Signifikan)

Kriteria pemilihan model terbaik yang ditunjukkan dalam Tabel 7 menunjukkan bahwa Model 4 memiliki nilai AIC yang lebih kecil dibandingkan dengan model lainnya serta parameter dispersi berpengaruh signifikan, namun parameter zero inflated pada model tersebut tidak berpengaruh signifikan, sehingga tidak memenuhi kriteria pemilihan model terbaik. Berdasarkan kriteria pemilihan model terbaik, Model 3 dipilih karena memiliki nilai AIC yang lebih kecil dibandingkan Model 1 dan Model 2 serta parameter dispersi dan parameter zero inflated pada model tersebut berpengaruh signifikan, sehingga memberikan kesimpulan sebagai berikut:

a. Model Poisson State

$$\hat{\mu}_i = \exp(0.7238 + 0.0007x_{1i} - 0.2165x_{2i}), i = 1, 2, \dots, 47$$

Berdasarkan model tersebut dapat diperoleh dari model tersebut yaitu sebagai berikut. Setiap penambahan Jumlah KK dengan akses terhadap fasilitas sanitasi yang layak (X_1) maka akan meningkatkan rata-rata jumlah kasus malaria sebesar $\exp(0.0007) = 1.0007$ apabila variabel independen lainnya tidak dilibatkan dalam model. Hasil tersebut bertolak belakang dengan landasan teoritis, dimana tersedianya akses terhadap fasilitas sanitasi yang layak dapat mencegah perkembangbiakan nyamuk Anopheles yang menjadi vektor penyakit malaria. Hal tersebut dapat dipengaruhi oleh beberapa alasan seperti keterbatasan dalam pengumpulan dan rentangan data yang digunakan hanya data satu tahun sehingga interpretasi pada model ini dapat disimpulkan bahwa dapat terjadi peningkatan ataupun penurunan kasus malaria akibat jumlah KK dengan akses sanitasi yang layak dalam kurun waktu tertentu. Selanjutnya, setiap penambahan Tempat-tempat umum memenuhi syarat kesehatan (X_2) maka akan menurunkan rata-rata jumlah kasus malaria sebesar $\exp(0.2165) = 0.0853$ apabila

variabel independen lainnya tidak dilibatkan dalam model. Model tersebut juga memberikan kesimpulan bahwa secara bersama-sama setiap penambahan Jumlah KK dengan akses terhadap fasilitas sanitasi yang layak (X_1) dan Tempat-tempat umum memenuhi syarat kesehatan (X_2) maka akan menurunkan jumlah rata-rata jumlah kasus malaria sebesar 0.3226.

b. Model logit

$$logit(\hat{p}_i) = 1.3940(0.7238 + 0.0007x_{1i} - 0.2165x_{2i})$$
$$= 1.0090 + 0.0010x_{1i} - 0.3018x_{2i}$$

dengan $i = 1, 2, \dots, 47$

Berdasarkan model tersebut diperoleh kesimpulan bahwa setiap penambahan Jumlah KK dengan akses terhadap fasilitas sanitasi yang layak (X_1) maka akan meningkatkan peluang terjadinya kasus malaria sebesar 0.0010 apabila variabel lain dianggap konstan. Setiap penambahan Tempat-tempat umum memenuhi syarat kesehatan (X_2) maka akan menurunkan peluang terjadinya kasus malaria sebesar 0.3018 apabila variabel lain dianggap konstan. Selanjutnya, secara bersama-sama setiap penambahan Jumlah KK dengan akses terhadap fasilitas sanitasi yang layak (X_1) dan Tempat-tempat umum memenuhi syarat kesehatan (X_2) maka akan menurunkan peluang terjadinya kasus malaria sebesar 0.3008.

4. Kesimpulan

1. Model Regresi *Zero Inflated Poisson Inverse Gaussian* pada data jumlah kasus malaria di Puskesmas Kota Makassar pada Tahun 2021 adalah

   a. Model *Poisson State*

   $$\hat{\mu}_i = \exp(0{,}7238 + 0{,}0007x_{1i} - 0{,}2165x_{2i})$$

   b. Model logit

   $$logit(\hat{p}_i) = 1{,}3940(0{,}7238 + 0{,}0007x_{1i} - 0{,}2165x_{2i})$$
   $$= 1{,}0090 + 0{,}0010x_{1i} - 0{,}3018x_{2i}$$

2. Faktor-faktor yang mempengaruhi jumlah kasus malaria di Puskesmas Kota Makassar pada Tahun 2021 adalah Jumlah KK dengan akses terhadap fasilitas sanitasi yang layak (X_1) dan Tempat-tempat umum memenuhi syarat kesehatan (X_2).

Daftar Pustaka

[1] Feng, C., Li, L., & Sadeghpour, A. A comparison of residual diagnosis tools for diagnosing regression models for count data. *BMC Medical Research Methodology*, 20(1), 2020.

[2] Green, J. A. Too many zeros and/or highly skewed? A tutorial on modelling health behaviour as count data with Poisson and negative binomial regression. *Health Psychology and Behavioral Medicine*, 9(1), 436–455, 2021.

[3] Ikhsani, N., Kalondeng, A., & Ilyas, N. Pemodelan Regresi Bivariate Poisson Inverse Gaussian pada Kasus Kematian Ibu dan Neonatal di Sulawesi Selatan. *Estimasi: Journal of Statistics and Its Application*, 4(1), 2721–379, 2023.

[4] Rahayu, A. Model-Model Regresi Untuk Mengatasi Masalah Overdispersi Pada Regresi Poisson. *Journal Peqguruang: Conference Series*, 2(1), 1–5, 2020.

[5] Chan, J. S. K., Choy, S. T. B., Makov, U., Shamir, A., & Shapovalov, V. Variable Selection Algorithm for a Mixture of Poisson Regression for Handling Overdispersion in Claims Frequency Modeling Using Telematics Car Driving Data. Risks, 10(4), 2022.

[6] Fávero, L. P., de Freitas Souza, R., Belfiore, P., Corrêa, H. L., & Haddad, M. F. C. Count Data Regression Analysis: Concepts, Overdispersion Detection, Zero-inflation Identification, and Applications with R. *Practical Assessment, Research and Evaluation*, 26, 1–22, 2021.

[7] Saraiva, E. F., Vigas, V. P., Flesch, M. V., Gannon, M., & Pereira, C. A. de B. Modeling Overdispersed Dengue Data via Poisson Inverse Gaussian Regression Model: A Case Study in the City of Campo Grande, MS, Brazil. *Entropy*, 24(9), 1–16, 2022.

[8] Sendow, Jeivy Mirechell. Analisis Pengaruh Cuaca Terhadap Persebaran Penyakit Malaria di Kabupaten Mimika Menggunakan Sistem Informasi Geografis. *Diss*, 2021.

[9] Aprilia, A. D. Regresi Zero Inflated Poisson Untuk Pemodelan Angka Positif Penyakit Malaria Di Jawa Timur. *Jurnal Ilmiah Matematika Volume 11* No 02 E-ISSN : 2716-506, 11(2), 139–146, 2023.

[10] Kemenkes RI. Profil Kesehatan Indonesia 2021. Kementerian Kesehatan Republik Indonesia, 2021.

[11] Dinkes Kota Makassar. *Profil Dinas Kesehatan Kota Makassar Tahun 2021* (Dinas Kesehatan Kota Makassar, Ed.). Dinas Kesehatan Kota Makassar, 2022.

[12] Handarzeni, S. A. Modeling of Tuberculosis Cases in Sumatra Region using Poisson Inverse Gaussian Regression. *Journal of Statistics and Data Science*, 1(2), 36–43, 2022.

[13] Nariswari, R., Widhiyanthi, A. A., Arifin, S., & Yudistira, I. G. A. A. Zero Inflated Poisson Regression: A Solution of Overdispersion in Stunting Data. *AIP Conference Proceedings*, 2975(1), 1–8, 2023.

[14] Darsyab, Muh. Y., & Ramadhan, M. N. Pemodelan Jumlah Kasus Penyakit Kusta di Provinsi Sulawesi Tenggara Menggunakan Metode Regresi Poisson Inverse Gaussian. *Jurnal Litbang Edusaintech* , 3(1), 11–24, 2022.

[15] Kyriazos, T., & Poga, M. Dealing with Multicollinearity in Factor Analysis: The Problem, Detections, and Solutions. Open Journal of Statistics, 13(03), 404–424, 2023.

[16] C.R., M. D., & Yanti, T. S. Regresi Poisson Invers Gaussian (PIG) untuk Pemodelan Jumlah Kasus Pneumonia pada Balita di Provinsi Jawa Tengah Tahun 2019. *Jurnal Riset Statistika*, 1(1), 143–151, 2021.

[17] Wahyuni, S. T., Utami, T. W., & Darsyah, Moh. Y. Pemodelan Generalized Additive Model For Location, Scale, and Shape (Gamlss) Dengan Pemulusan Locally Estimated Scatterplot Smoothing (Loess) pada Kasus Hiv/Aids Di Jawa Timur. *Jurnal Litbang Edusaintech*, 2(1), 18–26, 2021

[18] Wang, W., & Famoye, F. Modeling household fertility decisions with generalized Poisson regression. *Journal of Population Economics*, 10(3), 273–283, 1997

[19] Anggreainy, M. S., Illyasu, A. M., Musyaffa, H., & Kansil, F. H. Analysis of Factors Influencing the COVID-19 Mortality Rate in Indonesia using Zero Inflated Negative Binomial Model. *International Journal of Advanced Computer Science and Application*s, 13(4), 728–734, 2022.