PEMODELAN REGRESI *ZERO INFLATED POISSON INVERSE GAUSSIAN* PADA JUMLAH KASUS MALARIA DI PUSKESMAS KOTA MAKASSAR

SKRIPSI

FAUZIAH NURHIDAYAH

H051201016

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR

JUNI 2024

PEMODELAN REGRESI *ZERO INFLATED POISSON INVERSE GAUSSIAN* PADA JUMLAH KASUS MALARIA DI PUSKESMAS KOTA MAKASSAR

SKRIPSI

Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Sains pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

FAUZIAH NURHIDAYAH

H051201016

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

JUNI 2024

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

Pemodelan Regresi *Zero Inflated Poisson Inverse Gaussian* Pada Jumlah Kasus Malaria di Puskesmas Kota Makassar

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun

Makassar, 5 Juni 2024

METERAI TEMPEL 10000 8D29EAKX5137012?7

Fauziah Nurhidayah

NIM. H051201016

PEMODELAN REGRESI *ZERO INFLATED POISSON INVERSE GAUSSIAN* PADA JUMLAH KASUS MALARIA DI PUSKESMAS KOTA MAKASSAR

Disetujui Oleh:

Pembimbing Utama

Drs. Raupong, M.Si.

NIP. 19621015 198810 1 001

Ketua Program Studi

Dr. Anna Islamiyati, S.Si., M.Si.

NIP. 19770808 200501 2 002

Pada 5 Juni 2024

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama	:	Fauziah Nurhidayah
NIM	:	H051201016
Program Studi	:	Statistika
Judul Skripsi	:	Pemodelan Regresi *Zero Inflated Poisson Inverse Gaussian* Pada Jumlah Kasus Malaria di Puskesmas Kota Makassar

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1.	Ketua	:	Drs. Raupong, M.Si.	(...................)
2.	Anggota	:	Dr. Nirwan, M.Si.	(...................)
3.	Anggota	:	Dr. Erna Tri Herdiani, S.Si., M.Si.	(...................)

Ditetapkan di : Makassar

Tanggal : 5 Juni 2024

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji syukur penulis panjatkan kehadirat Allah SWT, karena atas berkat dan rahmat-Nya penulis dapat menyelesaikan skripsi ini. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa Sallam* beserta keluarga dan sahabatnya. Penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelas sarjana pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulisan skripsi ini tidak terlepas dari bantuan dan dukungan dari berbagai pihak sehingga dapat menyelesaikannya dengan segala keterbatasan kemampuan pengetahuan penulis. Oleh karena itu, penulis mengucapkan terima kasih kepada kedua orang tua penulis, Ibunda **Jumariah** dan Ayahanda **Saharuddin, S.E.** yang telah memberikan doa, dukungan, cinta, dan kasih sayang serta atas segala pengorbanan dalam setiap langkah kehidupan penulis. Penghargaan dan ucapan terima kasih dengan penuh keikhlasan dan ketulusan kepada:

1. **Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Anna Islamiyati, S.Si., M.Si.**, selaku Ketua Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin yang senantiasa memberikan arahan dan motivasi dengan penuh kesabaran kepada penulis selama menjadi mahasiswa.
4. **Bapak Drs. Raupong, M.Si.**, selaku Pembimbing Utama yang telah memberikan begitu banyak ilmu dan pengajaran serta telah meluangkan waktu dan pemikirannya hingga penulis mampu menyelesaikan tugas akhir ini.
5. **Bapak Dr. Nirwan, M.Si.** dan **Ibu Dr. Erna Tri Herdiani, S.Si., M.Si**., selaku Tim Penguji yang telah meluangkan waktu untuk memberikan kritik dan saran yang membangun dalam penyempurnaan tugas akhir ini.

6. Segenap **Dosen Pengajar** dan **Staf Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin** yang telah memberikan ilmu kepada penulis selama menempuh pendidikan di Departemen Statistika.
7. Kakak penulis, **Agus Riansyah Utama, Dwi Asmi Amaliah**, dan **Muh. Hajar Aswadi** yang telah memberikan doa, motivasi, dan dukungan bagi penulis.
8. Keluarga besar **Himastat FMIPA Unhas** dan **KM FMIPA Unhas** sebagai tempat penulis menemui banyak pengalaman berharga.
9. Sahabat-sahabat penulis, Safira Saiful, Najwa Isnaini Lagga, Azalia Filadelfia Pagalo, Andi Nurannisa Azzahra, Nurhaliza Rais, Nahdah Azatil Ismah, Sri Rahayu Yusri, Alisha Shafa Azzahrah, Ruslinda, Najlah Fauziah, dan Irma Mulia, yang telah membersamai dan bersedia menjadi tempat berkeluh kesah penulis, terima kasih untuk segala kehangatan dan kebaikan kepada penulis.
10. Teman-teman **POIS20N** terkhusus kepada Maryana Maharani, Taufiqurrahman Sadikin, M. Razy Qarar, Ahmad Ryan Al-aqsha, Izzul Haq, Farhan Ngkal, Azhar Ramadhan, Fadlan Amin, dan Ahmad Mukhlish Mursidin sebagai teman bercerita dalam suka dan duka selama kehidupan perkuliahan penulis. Terima kasih untuk segala cerita dan canda tawa yang telah dibagi bersama.
11. Kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih atas segala bentuk dukungan yang menjadi motivasi bagi penulis dalam menyelesaikan tugas akhir ini.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kekurangan. Akhir kata, semoga tulisan ini dapat memberikan manfaat kepada berbagai pihak.

Makassar, 5 Juni 2024

Fauziah Nurhidayah

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIK

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini :

Nama	:	Fauziah Nurhidayah
NIM	:	H051201016
Program Studi	:	Statistika
Departemen	:	Statistika
Fakultas	:	Matematika dan Ilmu Pengetahuan Alam
Jenis Karya	:	Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-ekslusif** (***Non-exclusive Royalty-Free Right***) atas tugas akhir saya yang berjudul:

"Pemodelan Regresi *Zero Inflated Poisson Inverse Gaussian* Pada Jumlah Kasus Malaria di Puskesmas Kota Makassar"

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalihkan-media/format-kan, mengelola dalam bentuk pangkalan data (database), merawat dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.
Dibuat di Makassar tanggal 5 Juni 2024

takan,

rhidayah)

ABSTRAK

Regresi *Poisson* merupakan salah satu metode yang dapat digunakan untuk memodelkan data cacahan. Pemodelan regresi *Poisson* memiliki beberapa asumsi yang harus dipenuhi, salah satunya yaitu *equidispersi*. Namun, pada beberapa kondisi terdapat data cacahan yang mengalami *overdispersi*. Terjadinya *overdispersi* dapat disebabkan oleh kelebihan angka nol pada variabel dependen, sehingga pemodelan dapat dilakukan dengan salah satu dari metode *Mixed Poisson* yaitu regresi *Zero Inflated Poisson Inverse Gaussian* (ZIPIG). Estimasi parameter pada model *Zero Inflated Poisson Inverse Gaussian* (ZIPIG) dilakukan dengan metode *Maximum Likelihood Estimation* (MLE) dengan iterasi Algoritma *Fisher Scoring*. Penelitian ini diterapkan pada data jumlah kasus malaria di Puskesmas Kota Makassar Tahun 2021 yang memberikan hasil bahwa variabel independen yang memiliki pengaruh signifikan adalah Jumlah KK dengan akses terhadap fasilitas sanitasi yang layak (X_1) dan Tempat-tempat umum memenuhi syarat kesehatan (X_2).

Kata kunci : *Excess Zero*, Malaria, MLE, *Overdispersi*, Regresi ZIPIG

ABSTRACT

Poisson regression is a method that can be used to model count data. Poisson regression modeling has several assumptions that must be met, one of which is equidispersion. However, in some conditions, there is count data that experiences overdispersion. The occurrence of overdispersion can be caused by excess zeros in the dependent variable, so that modeling can use Zero Inflated Poisson Inverse Gaussian (ZIPIG). Parameter estimation in the Zero Inflated Poisson Inverse Gaussian (ZIPIG) model using the Maximum Likelihood Estimation (MLE) with Fisher Scoring algorithm iteration. This research was applied to data on malaria cases in the Community Health Centers in the city of Makassar in 2021. The result of the research show that the independent variabels that have a significant influence are number of families with proper sanitation facility access (X_1) *dan public place with qualified health* (X_2).

Keyword *: Excess Zero, Malaria, MLE, Overdispersion, ZIPIG Regression*

DAFTAR ISI

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Data cacahan ialah sekumpulan bilangan bulat non-negatif yang terdiri atas beberapa peristiwa diskrit (Feng dkk., 2020). Data cacahan sering kali tidak berdistribusi normal, memiliki kemiringan yang besar, dan memiliki nilai ekstrem serta terkadang mengandung banyak angka nol (Green, 2021). Oleh karena itu, pemodelan tidak dapat dilakukan dengan *Ordinary Least Square* (OLS) sehingga pemodelan dilakukan menggunakan *Generalized Linear Models* (GLMs) yang tidak memerlukan asumsi normalitas dan asumsi heterosketastisitas dalam penerapannya (Ikhsani dkk., 2023). Data cacahan dengan kasus jarang sering dimodelkan menggunakan salah satu dari keluarga GLMs yaitu Regresi *Poisson* (Rahayu, 2020).

Pemodelan data menggunakan Regresi *Poisson* mensyaratkan dipenuhinya asumsi *equidispersi* atau nilai variansi dan rataan sama. Akan tetapi, 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*). Apabila terjadi *overdispersi* pada data, maka estimasi standar eror akan mengalami *underestimate* sehingga kesimpulan yang dihasilkan menjadi tidak akurat (Chan dkk., 2022). Salah satu penyebab terjadinya *overdispersi* pada data yaitu kelebihan angka nol pada variabel dependen. Pemodelan pada data yang mengalami *overdispersi* dapat dilakukan menggunakan kombinasi dari distribusi diskrit dan distribusi kontinu dengan distribusi poisson atau disebut metode *Mixed Poisson Distribution* (Rahayu, 2020).

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* (Fávero dkk., 2021). 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* (Saraiva dkk., 2022). 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. 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.

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 atau dengan kata lain terjadinya kasus malaria pada wilayah tersebut tidak terjadi karena penularan lokal melainkan penderita terjangkit setelah mengunjungi wilayah dengan status endemis malaria ...nkes RI, 2021). Sebanyak 27 kasus malaria tercatat dari 47 puskesmas di ...h Kota Makassar (Dinkes Kota Makassar, 2022). 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 (Kemenkes RI, 2021). 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. Sementara itu, Nariswari dkk. (2023) melakukan pemodelan kasus stunting menggunakan metode *Zero Inflated Poisson*. Secara umum, melakukan pendugaan parameter pada model GLMs dapat menggunakan metode *Maximum Likelihood Estimation* (MLE), namun metode tersebut menghasilkan sistem persamaan implisit dan non linier sehingga Algoritma *Fisher Scoring* digunakan untuk menyelesaikan fungsi tersebut (Darsyab & Ramadhan, 2022). Dengan demikian, penelitian ini menggunakan metode pemodelan dengan Regresi *Zero Inflated Poisson Inverse Gaussian* pada jumlah kasus malaria di Puskesmas Kota Makassar.

1.2 Rumusan Masalah

Rumusan masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut:

1. Bagaimana bentuk model jumlah kasus malaria di Puskesmas Kota Makassar dengan menggunakan regresi *Zero Inflated Poisson Inverse Gaussian* (ZIPIG)?
2. Faktor-faktor apa saja yang dapat mempengaruhi jumlah kasus malaria di Kota Makassar?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Variabel dependen yang digunakan yaitu data jumlah positif malaria berdasarkan pelaporan puskesmas pada setiap kecamatan di Kota Makassar pada tahun 2021
2. Variabel dependen yang digunakan berdistribusi poisson

3. Variabel dependen yang digunakan mengandung excess zeros yang menyebabkan terjadinya *overdispersi*

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mendapatkan model jumlah kasus malaria di Puskesmas Kota Makassar dengan menggunakan regresi *Zero Inflated Poisson Inverse Gaussian* (ZIPIG).
2. Untuk menentukan faktor-faktor yang mempengaruhi jumlah kasus malaria di Kota Makassar.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat antara lain sebagai berikut:

1. Menambah dan mengembangkan pengetahuan mengenai pembentukan model regresi *Zero Inflated Poisson Inverse Gaussian* (ZIPIG)
2. Menjadi referensi tambahan untuk penelitian selanjutnya.

BAB II
TINJAUAN PUSTAKA

2.1 Regresi Poisson

Distribusi Poisson merupakan suatu distribusi untuk peristiwa yang memiliki peluang kejadian kecil, di mana kejadian bergantung pada selang waktu atau daerah tertentu dengan hasil pengamatan berupa variabel diskrit dan antar variabel independen saling bebas. Fungsi peluang untuk data berdistribusi *Poisson* bergantung pada parameter tunggal, yaitu rataan μ (Darsyab & Ramadhan, 2022). Fungsi peluang dari distribusi Poisson bila variabel dependen y_i merupakan bilangan bulat non-negatif sebagai berikut (Lukman, 2021):

$$f(y;\mu) = \frac{\exp(-\mu)\,\mu^{y}}{y!} \qquad \text{untuk } y = 0, 1, 2, \ldots \text{ dan } \mu > 0 \tag{2.1}$$

Rataan dan variansi dalam Regresi *Poisson* bernilai sama atau dapat dituliskan sebagai berikut:

$$E(Y) = Var(Y) = \mu \tag{2.2}$$

Terdapat tiga komponen dalam *Generalized Linear Models* (GLMs), yaitu;

1. Komponen random yang terdiri atas variabel respon Y dengan nilai observasi yang saling bebas satu sama lainnya
2. Komponen sistematis yang terdiri dari komponen model linier umum yang menghubungkan vektor $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_n)^T$ dengan sekumpulan variabel dependen melalui model linier $\boldsymbol{\eta} = \boldsymbol{X}^T\boldsymbol{\beta}$ dimana $\boldsymbol{X}^T$ adalah matriks yang berisi nilai variabel independen untuk n buah pengamatan dan $\boldsymbol{\beta}$ adalah vektor dari parameter-parameter di dalam model
3. Komponen fungsi *link* yang menghubungkan komponen random dan komponen sistematis melalui persamaan linier dengan $\eta_i = g(\cdot)$.

Misalkan y_i merupakan variabel acak dengan $i = 1, 2, \ldots, n$, dimana n adalah banyaknya data dan y_i berdistribusi Poisson dengan fungsi kepadatan peluang seperti pada Persamaan (2.1) untuk $\mu > 0$ dengan μ_i merupakan rataan dari variabel dependen Y, maka fungsi peluang Poisson termasuk keluarga ekponensial ditulis dengan linier (Firmansyah dkk., 2020):

$$f(y_i;\mu) = \exp(y_i \ln(\mu_i) - \mu_i - \ln(y_i!)) \tag{2.3}$$

Fungsi hubung yang digunakan adalah logaritma natural sehingga menghasilkan hubungan *log-linear* antara variabel dependen dan variabel independen (Fitrial & Fatikhurrizqi, 2020). Dengan demikian, model regresi *Poisson* adalah sebagai berikut (Lukman dkk., 2021):

$$\eta = \ln(\mu_i) = \exp(\boldsymbol{X}_i^T \boldsymbol{\beta}) \tag{2.4}$$

dengan $\boldsymbol{X}_i^T$ merupakan vektor variabel independen berukuran $n \times p$ yang menjelaskan variabel independen dan $\boldsymbol{\beta}$ merupakan vektor berukuran $p \times 1$ berisi parameter regresi (Firmansyah dkk., 2020).

Pemeriksaan distribusi dari suatu data dapat dilakukan dengan uji *Kolmogorov-Smirnov* dengan hipotesis sebagai berikut:

H_0: variabel dependen berdistribusi poisson

H_1: variabel dependen tidak berdistribusi poisson

Memiliki statistik uji sebagai berikut:

$$D_{hitung} = maks|F_n(Y) - F_0(Y)| \tag{2.5}$$

dengan:

$F_n(Y)$	=	fungsi distribusi kumulatif dibawah hipotesis H_0
$F_0(Y)$	=	perbandingan antara fungsi distribusi kumulatif yang diamati dengan jumlah sampel

Adapun kriteria pengujiannya yaitu H_0 diterima apabila $D_{hitung} < D_{N;\alpha}$ yang artinya variabel dependen berdistribusi poisson.

2.2 Regresi Poisson Inverse Gaussian

Distribusi *Poisson Inverse Gaussian* merupakan salah satu bentuk dari distribusi *Mixed Poisson* antara distribusi *Poisson* dan distribusi *Inverse Gaussian*. Terdapat dua parameter yang menentukan distribusi ini, yaitu rataan (μ) dan parameter *dispersi* (τ) (Handarzeni, 2022). Adapun fungsi kepadatan peluang untuk distribusi *Inverse Gaussian* adalah sebagai berikut (Vasconcelos dkk., 2023):

$$f(y;\mu) = \frac{1}{\sqrt{2\pi\sigma^2 y^3}} \exp\left(-\frac{(y-\mu)^2}{2\mu^2\sigma^2 y}\right) \tag{2.6}$$

Berdasarkan Persamaan (2.1) dan (2.6) kemudian diperoleh probabilitas usi *Poisson Inverse Gaussian* dapat dihitung sebagai berikut (Saraiva dkk.,

$$P(Y=y|\mu)=\frac{\mu^y}{y!}\exp\left(\frac{1}{\tau}\right)\left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}}(1+2\mu\tau)^{-\left(\frac{y-\frac{1}{2}}{2}\right)}K_{y-\frac{1}{2}}\left(\frac{1}{\tau}\sqrt{2\mu\tau+1}\right) \tag{2.7}$$

dengan $K_{y-\frac{1}{2}}\left(\frac{1}{\tau}\sqrt{2\mu\tau+1}\right)$ adalah fungsi Bessel modifikasi jenis ketiga yaitu:

$$K_{y-\frac{1}{2}}\left(-\frac{1}{\tau}\sqrt{(2\mu\tau+1)}\right)=\left(\frac{\pi}{2\left(\frac{1}{\tau}\sqrt{2\mu\tau+1}\right)}\right)^{y-\frac{1}{2}}\left(\exp\left(-\frac{1}{\tau}\sqrt{(2\mu\tau+1)}\right)\right)$$

Rataan untuk distribusi *Poisson Inverse Gaussian* yaitu:

$$E(Y)=\mu \tag{2.8}$$

sedangkan, variansi untuk distribusi *Poisson Inverse Gaussian* yaitu (Phang dkk., 2021):

$$Var(Y)=\mu+\tau\mu^2 \tag{2.9}$$

Model Regresi *Poisson Inverse Gaussian* adalah sebagai berikut (Handarzeni, 2022):

$$\mu_i=\exp\left(\boldsymbol{X}_i^T\boldsymbol{\beta}\right);\ i=1,2,\dots,n \tag{2.10}$$

dimana,

$\boldsymbol{X}_i^T$: $[\mathbf{1}\quad x_{1i}\quad x_{2i}\quad \dots\quad x_{ki}]$ sebagai vektor variabel independen ke-j dengan $j=1,2,\dots,k$ pada pengamatan ke-$i=1,2,\dots,n$ yang merupakan matriks berukuran $n\times p$

$\boldsymbol{\beta}$: $[\beta_0\quad \beta_1\quad \beta_2\quad \dots\quad \beta_k]^T$ sebagai vektor koefisien regresi

2.3 Regresi Zero Inflated Poisson

Banyaknya angka nol atau *excess zero* dengan proporsi lebih dari 50% dapat menyebabkan pemodelan dengan menggunakan regresi poisson menjadi tidak tepat (Famoye & Singh, 2006). Kondisi tersebut juga merupakan salah satu penyebab nilai variansi menjadi lebih besar dari nilai rataan atau disebut *overdispersi*. Pemodelan dengan regresi *Zero Inflated Poisson* (ZIP) merupakan salah satu e analisis yang dapat dilakukan. Model ZIP terdiri atas dua bagian (*state*) nodel *zero state* dan *Poisson state*. Model *zero state* atau disebut juga model logit terjadi dengan probabilitas p_i dan hanya menghasilkan observasi yang bernilai nol. Sementara model *Poisson State* terjadi dengan probabilitas $(1 - p_i)$ dan berdistribusi poisson dengan mean μ_i (Nariswari dkk., 2023). Apabila sebuah data mengikut distribusi *Poisson*, maka *Zero Inflated Poisson* (ZIP) diberikan oleh:

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i)\,\mathrm{P}(Y = 0|\mu_i) & ;\ y_i = 0 \\ (1 - p_i)P(Y = y|\mu_i) & ;\ y_i > 0 \end{cases} \tag{2.11}$$

Berdasarkan Persamaan (2.1) dan Persamaan (2.11) kemudian dibentuk fungsi kepadatan peluang untuk $y = 0$ dan $y > 1$ sebagai berikut (Feng, 2021):

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i)\exp(\mu_i) & ;\ y_i = 0 \\ (1 - p_i)\dfrac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!} & ;\ y_i > 0 \end{cases} \tag{2.12}$$

dimana

$$p_i = \frac{\exp(\boldsymbol{X}_i^T\gamma)}{1 + \exp(\boldsymbol{X}_i^T\gamma)} \qquad 1 - p = \frac{1}{1 + \exp(\boldsymbol{X}_i^T\gamma)}$$

Regresi *Zero Inflated Poisson* merupakan model regresi yang memiliki dua parameter yaitu parameter distribusi *Poisson* μ dan parameter proporsi *zero* p dengan rataan untuk X didefinisikan dengan:

$$E(Y) = (1 - p_i)\mu_i \tag{2.13}$$

sedangkan, variansi yaitu (Park & Jun, 2023):

$$Var(Y) = (1 - p_i)(\mu_i + \mu_i^2 p_i) \tag{2.14}$$

Model gabungan untuk μ dan p adalah sebagai berikut (Azagi, 2022):

$$ln(\mu_i) = \boldsymbol{X}_i^T\boldsymbol{\beta} \tag{2.15}$$

$$\begin{aligned} logit(p_i) &= \ln\left(\frac{p_i}{1 - p_i}\right) \\ &= \ln\left(\frac{\exp(\boldsymbol{X}_i^T\gamma)}{1 + \exp(\boldsymbol{X}_i^T\gamma)} \times \frac{1 + \exp(\boldsymbol{X}_i^T\gamma)}{1}\right) \\ &= \ln(\exp(\boldsymbol{X}_i^T\gamma)) \\ &= \boldsymbol{X}_i^T\gamma \end{aligned} \tag{2.16}$$

dengan $\boldsymbol{X}_i^T$ merupakan matriks variabel independen, $\boldsymbol{\beta}$ dan γ merupakan vektor parameter yang akan ditaksir, serta p merupakan probabilitas observasi bernilai nol.

2.4 Regresi Zero Inflated Poisson Inverse Gaussian

Regresi *Zero Inflated Poisson Inverse Gaussian* merupakan suatu pemodelan 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 $Y \sim ZIPIG(\mu, \tau, p)$ dapat dituliskan dengan (Nur dkk., 2021):

$$P(Y = y|\mu, \tau, p) = \begin{cases} p + (1-p)P(Y = 0|\tau, p) & ; y = 0 \\ (1-p)P(Y = y|\tau, p) & ; y = 1, 2, \ldots \end{cases} \tag{2.17}$$

Berdasarkan Persamaan (2.1) dan Persamaan (2.17) kemudian dibentuk Persamaan (2.18) dan (2.19) sebagai fungsi kepadatan peluang untuk model ZIPIG berikut:

untuk $y = 0$ dituliskan

$$P(Y = 0|\mu, \tau, p) = p + (1-p)\exp\left(\frac{1}{\tau}\right)\left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}}(2\mu\tau + 1)^{\frac{1}{4}} \left(\frac{\pi}{2\left(\frac{1}{\tau}\sqrt{2\mu\tau + 1}\right)}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{\tau}\sqrt{(2\mu\tau + 1}\right) \tag{2.18}$$

dan untuk $y = 1, 2, \ldots$ dituliskan

$$P(Y = y|\mu, \tau, p) = (1-p)\frac{\mu^y}{y!}\exp\left(\frac{1}{\tau}\right)\left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}}(2\mu\tau + 1)^{-\frac{\left(y-\frac{1}{2}\right)}{2}} \left(\frac{\pi}{2\left(\frac{1}{\tau}\sqrt{2\mu\tau + 1}\right)}\right)^{y-\frac{1}{2}} \exp\left(-\frac{1}{\tau}\sqrt{(2\mu\tau + 1}\right) \tag{2.19}$$

dengan

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

Misalkan sebuah variabel respon $Y \sim ZIPIG(\mu, \tau, p)$ maka model ZIPIG dapat dituliskan pada dua komponen model sebagai berikut:

a. Model poisson state

$$\ln(\mu) = \boldsymbol{X}_i^T \boldsymbol{\beta}$$
$$\mu = \exp(\boldsymbol{X}_i^T \boldsymbol{\beta}) \tag{2.20}$$

Model *zero inflated*

$$logit(p) = \ln\left(\frac{p}{1-p}\right)$$

$$logit(p) = \ln\left(\frac{\exp(-\gamma \boldsymbol{X}_i^T\boldsymbol{\beta})}{1+\exp(-\gamma \boldsymbol{X}_i^T\boldsymbol{\beta})} \times \frac{1+\exp(-\gamma \boldsymbol{X}_i^T\boldsymbol{\beta})}{1}\right)$$

$$= \ln(\exp(-\gamma \boldsymbol{X}_i^T\boldsymbol{\gamma}))$$

$$logit(p) = -\gamma \boldsymbol{X}_i^T\boldsymbol{\beta} \tag{2.21}$$

dengan mendistribusikan Persamaan (2.20) dan Persamaan (2.21) pada Persamaan (2.18) dan Persamaan (2.19) maka diperoleh fungsi peluang untuk model regresi ZIPIG sebagai berikut:

untuk $y = 0$ dituliskan

$P(Y = 0|\boldsymbol{\beta}, \gamma, \tau)$

$$= \frac{\exp(-\gamma \boldsymbol{X}_i^T\boldsymbol{\beta})}{1+\exp(-\gamma \boldsymbol{X}_i^T\boldsymbol{\beta})} + \frac{1}{1+\exp(-\gamma \boldsymbol{X}_i^T\boldsymbol{\beta})}\exp\left(\frac{1}{\tau}\right)\left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}}$$
$$(2\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})\,\tau+1)^{\frac{1}{4}}\left(\frac{\pi}{2\left(\frac{1}{\tau}\sqrt{2\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})\,\tau+1}\right)}\right)^{\frac{1}{2}} \tag{2.22}$$
$$\exp\left(-\frac{1}{\tau}\sqrt{2\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})\,\tau+1}\right)$$

untuk $y = 1, 2, \ldots$ dituliskan

$P(Y = y|\boldsymbol{\beta}, \gamma, \tau)$

$$= \frac{1}{1+\exp(-\gamma \boldsymbol{X}_i^T\boldsymbol{\beta})}\left(\frac{(\exp(\boldsymbol{X}_i^T\boldsymbol{\beta}))^{y_i}\exp\left(\frac{1}{\tau}\right)}{y_i!}\right)\left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}}$$
$$(2\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})\,\tau+1)^{-\frac{\left(y_i-\frac{1}{2}\right)}{2}}$$
$$\left(\frac{\pi}{2\left(\frac{1}{\tau}\sqrt{2\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})\,\tau+1+1}\right)}\right)^{y_i-\frac{1}{2}} \tag{2.23}$$
$$\exp\left(-\frac{1}{\tau}\sqrt{2\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})\,\tau+1+1}\right)$$

2.5 Multikolinearitas

Multikolinearitas merupakan hubungan linear yang sangat tinggi pada model regresi pada setiap variabel independennya. Akibat terjadinya multikolinearitas adalah pemakaian metode regresi menjadi kurang tepat karena taksiran regresinya tidak stabil dan koefisien regresinya sangat besar. Salah satu cara yang dapat dilakukan untuk mendeteksi terjadinya gejala multikolinearitas yaitu dengan melihat nilai *Variance Inflation Factor* (VIF) menggunakan persamaan berikut:

$$VIF = \frac{1}{1 - R_j^2} \tag{2.24}$$

dengan

$$R_j^2 = \frac{\sum_{i=1}^{n} \left(\hat{y}_{ij} - \bar{y}\right)^2}{\sum_{i=1}^{n} \left(y_{ij} - \bar{y}\right)^2} \tag{2.25}$$

dimana R_j^2 merupakan koefisien determinasi ke-j yang diperoleh dengan meregresikan variabel dependen (Kyriazos & Poga, 2023). Kriteria pengujian yaitu H_0 diterima apabila $VIF < 10$ yang berarti tidak terjadi multikolinearitas antar variabel independen yang digunakan.

2.6 Overdispersi

Dalam model regresi poisson terdapat beberapa asumsi yang harus dipenuhi. Salah satu dari asumsi tersebut adalah asumsi kesamaan rataan dan variansi yang disebut *equidispersi*. Nilai variansi yang melebihi nilai rataan disebut *overdispersi*. Salah satu cara untuk melakukan pemeriksaan *overdispersi* yaitu dengan menggunakan nilai *deviance* dibagi dengan derajat bebas. Pengujian *overdispersi* dengan menggunakan persamaan berikut:

$$VT = \frac{D}{db} \tag{2.26}$$

$$D = 2 \sum_{i=1}^{n} \left[y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) \right] \tag{2.27}$$

dimana

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 (C.R. & Yanti, 2021).

2.7 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 $Y_i = 1, 2, \ldots$. Persamaan likelihood untuk model ZIPIG dapat dituliskan sebagai berikut:

$$L(\boldsymbol{\beta}, \gamma, \tau) = \left(\prod_{\substack{i=1\\ y_i=0}}^{n} P(Y_i = y_i | \boldsymbol{X_i}; \boldsymbol{\beta}, \gamma, \tau)\right)\left(\prod_{\substack{i=1\\ y_i>0}}^{n} P(Y_i = y_i | \boldsymbol{X_i}; \boldsymbol{\beta}, \gamma, \tau)\right)$$

$$L(\boldsymbol{\beta}, \gamma, \tau) = \big(L_1(\boldsymbol{\beta}, \gamma, \tau)\big)\big(L_2(\boldsymbol{\beta}, \gamma, \tau)\big) \tag{2.28}$$

dengan

$$L_1(\boldsymbol{\beta}, \gamma, \tau) = \prod_{\substack{i=1\\ y_i=0}}^{n} \frac{1}{1 + \exp(-\gamma \boldsymbol{X}_i^T \boldsymbol{\beta})}\Bigg[\exp(-\gamma \boldsymbol{X}_i^T \boldsymbol{\beta}) + \exp\left(\frac{1}{\tau} - \frac{1}{\tau}\sqrt{(2(\exp(\boldsymbol{X}_i^T \boldsymbol{\beta}))\tau) + 1}\right)\Bigg] \tag{2.29}$$

$$L_2(\boldsymbol{\beta}, \gamma, \tau) = \prod_{\substack{i=1\\ y_i>0}}^{n} \left(\frac{1}{1 + \exp(-\gamma \boldsymbol{X}_i^T \boldsymbol{\beta})}\right)\left(\frac{(\exp(-\boldsymbol{X}_i^T \boldsymbol{\beta}))^{y_i} \exp\left(\frac{1}{\tau}\right)}{y_i!}\right) \tag{2.30}$$

$$\left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}} [2\exp(\boldsymbol{X}_i^T \boldsymbol{\beta})\tau + 1]^{-\left(\frac{y_i - \frac{1}{2}}{2}\right)}$$

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

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

Kedua fungsi likelihood tersebut kemudian diubah menjadi bentuk logaritma natural (ln) sebagai berikut:

$$l = \ln L(\boldsymbol{\beta}, \gamma, \tau)$$

$$l = \ln L_1(\boldsymbol{\beta}, \gamma, \tau) + \ln L_2(\boldsymbol{\beta}, \gamma, \tau) \tag{2.31}$$

dengan

$$l_1 = \ln L_1\,(\boldsymbol{\beta}, \gamma, \tau)$$

$$l_1 = \sum_{i=1}^{n} \ln\left(\exp(-\gamma \boldsymbol{X}_i^T \boldsymbol{\beta}) + \exp\left(\frac{1}{\tau} - \frac{1}{\tau}\sqrt{(2(\exp(\boldsymbol{X}_i^T \boldsymbol{\beta}))\tau) + 1}\right)\right)$$
$$- \sum_{i=1}^{n} \ln(1 + \exp(-\gamma \boldsymbol{X}_i^T \boldsymbol{\beta})) \tag{2.32}$$

dan

$$l_2 = \ln L_2\,(\boldsymbol{\beta}, \gamma, \tau)$$

$$l_2 = \sum_{i=1}^{n} (1 - \mathrm{y_i})(\ln 2 - \ln \pi - \ln \tau) + \sum_{i=1}^{n} \mathrm{y_i} \mathbf{X}_\mathrm{i}^\mathrm{T} \boldsymbol{\beta} + \frac{\mathrm{n}}{\tau}$$
$$- \sum_{i=1}^{n} \ln\big(1 + \exp\big(-\gamma \mathbf{X}_\mathbf{i}^\mathbf{T} \boldsymbol{\beta}\big)\big) + \sum_{i=1}^{n} \ln(\mathrm{y_i}!)$$
$$+ \sum_{i=1}^{n} \ln\left(\frac{1}{2} - \mathrm{y_i}\right) \ln\big(2\big(\exp\big(\mathbf{X}_\mathbf{i}^\mathbf{T} \boldsymbol{\beta}\big)\big)\tau + 1\big)$$
$$- \sum_{i=1}^{n} \left(-\frac{1}{\tau}\sqrt{2(\exp(\boldsymbol{X}_i^T \boldsymbol{\beta}))\tau + 1}\right) \tag{2.33}$$

Fungsi likelihood yang diperoleh kemudian diturunkan terhadap parameter $\boldsymbol{\beta}$, τ, dan γ kemudian disamakan dengan nol. Hasil dari turunan tersebut menghasilkan bentuk yang implisit sehingga estimasi parameter tidak dapat dilakukan secara analitik. Oleh karena itu, estimasi parameter akan dilakukan dengan menggunakan metode iterasi dengan algoritma *Fisher Scoring* (Saraiva dkk., 2022). Pendugaan maksimum terhadap parameter dapat diperoleh menggunakan iterasi sebagai berikut (Darsyab & Ramadhan, 2022):

$$\widehat{\boldsymbol{\theta}}_{(m+1)} = \widehat{\boldsymbol{\theta}}_{(m)} + \boldsymbol{I}^{-1}\big(\widehat{\boldsymbol{\theta}}_{(m)}\big)\boldsymbol{D}\big(\widehat{\boldsymbol{\theta}}_{(m)}\big) \tag{2.34}$$

dimana,

$\hat{\boldsymbol{\beta}}^T, \hat{\tau}, \hat{\gamma})^T$

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

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

$$\boldsymbol{H}(\widehat{\boldsymbol{\theta}}_{(m)}) = \begin{bmatrix} \frac{\partial^2 l}{\partial \hat{\gamma}^2} & \frac{\partial^2 l}{\partial \hat{\gamma} \partial \hat{\tau}} & \frac{\partial^2 l}{\partial \hat{\gamma} \partial \widehat{\boldsymbol{\beta}}} \\ \frac{\partial^2 l}{\partial \hat{\tau} \partial \hat{\gamma}} & \frac{\partial^2 l}{\partial \hat{\tau}^2} & \frac{\partial^2 l}{\partial \hat{\tau} \partial \widehat{\boldsymbol{\beta}}} \\ \frac{\partial^2 l}{\partial \widehat{\boldsymbol{\beta}} \partial \hat{\gamma}} & \frac{\partial^2 l}{\partial \widehat{\boldsymbol{\beta}} \partial \hat{\tau}} & \frac{\partial^2 l}{\partial \widehat{\boldsymbol{\beta}} \partial \widehat{\boldsymbol{\beta}}^T} \end{bmatrix}$$

Algoritma *Fisher Scoring* merupakan metode iterasi yang dapat diuraikan sebagai berikut:

a. Menentukan vektor awal parameter $\boldsymbol{\theta_0}$ dengan data diasumsikan memenuhi model regresi linier berganda yaitu:

   $\boldsymbol{Y}^* = \beta_{0_{(0)}} + \beta_1 x_{i1_{(0)}} + \cdots + \beta_k x_{ik_{(0)}} + \varepsilon_i$

   dimana $i = 1, 2, \ldots, n$ dan menggunakan metode kuadrat terkecil yaitu dengan

   $\widehat{\boldsymbol{\beta}}_{(0)} = (\boldsymbol{X^T X})^{-1} \boldsymbol{X^T Y}^*$

b. Membentuk vektor gradien $\boldsymbol{D}(\widehat{\boldsymbol{\theta}})$**,** matriks hessian $\boldsymbol{H}(\widehat{\boldsymbol{\theta}}_{(0)})$, dan matriks informasi *Fisher* $\boldsymbol{I}(\widehat{\boldsymbol{\theta}}_{(0)})$

c. Memasukkan nilai $\widehat{\boldsymbol{\theta}}_{(0)}$ sehingga diperoleh vektor gradien $\boldsymbol{D}(\widehat{\boldsymbol{\theta}}_{(0)})$ dan matriks hessian $\boldsymbol{H}(\widehat{\boldsymbol{\theta}}_{(0)})$

d. Iterasi dimulai dari $m = 0$ pada Persamaan (2.33) dengan nilai $\widehat{\boldsymbol{\theta}}_{(m)}$ ialah sekumpulan penduga parameter yang konvergen saat iterasi ke-m.

e. Apabila saat iterasi ke-m belum memperoleh nilai dugaan parameter yang konvergen, maka iterasi kembali pada langkah (c.) hingga iterasi ke-m+1. Iterasi akan berhenti ketika nilai $\parallel \widehat{\boldsymbol{\theta}}_{(m+1)} - \widehat{\boldsymbol{\theta}}_{(m)} \parallel \leq \varepsilon$, dimana ε adalah bilangan yang sangat kecil mendekati nol.

2.8 Pengujian Parameter

Pengujian parameter pada model Regresi *Zero Inflated Poisson Inverse Gaussian* dilakukan dengan pengujian hipotesis secara simultan pada parameter $\boldsymbol{\beta}$ engujian parsial pada parameter $\boldsymbol{\beta}$, τ, dan γ. Pengujian parameter dilakukan mengetahui pengaruh variabel independen terhadap variabel dependen.

2.8.1. 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:

$$G = -2\ln\left(\frac{\ln L(\widehat{\omega})}{\ln L(\widehat{\Omega})}\right) \sim \chi^2_{\alpha;v} \tag{2.35}$$

dimana 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 $L(\omega)$. Kedua fungsi *likelihood* tersebut kemudian diubah ke dalam bentuk *ln* sehingga membentuk persamaan sebagai berikut:

$$\begin{aligned}
l(\widehat{\Omega}) &= \ln L\,(\widehat{\Omega}) \\
&= \sum_{i=1}^{n} \ln\Bigg(\exp(-\hat{\gamma}\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}) \\
&\qquad + \exp\left(\frac{1}{\hat{\tau}} - \frac{1}{\hat{\tau}}\sqrt{\left(2\left(\exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}\right)\right)\hat{\tau}\right) + 1}\right)\Bigg) \\
&\qquad - \sum_{i=1}^{n} \ln\left(1 + \exp(-\hat{\gamma}\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}})\right) \\
&\qquad + \sum_{i=1}^{n} (1 - y_i)(\ln 2 - \ln \pi - \ln \hat{\tau}) \\
&\qquad + \sum_{\substack{i=1 \\ y_i > 0}}^{n} y_i\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}} + \frac{n}{\hat{\tau}} - \sum_{i=1}^{n} \ln\left(1 + \exp\left(-\hat{\gamma}\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}\right)\right) \\
&\qquad + \sum_{i=1}^{n} \ln(y_i!) \\
&\qquad + \sum_{i=1}^{n} \ln\left(\frac{1}{2} - y_i\right)\ln\left(2\left(\exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}\right)\right)\hat{\tau} + 1\right) \\
&\qquad - \sum_{i=1}^{n}\left(-\frac{1}{\hat{\tau}}\sqrt{2\left(\exp\left(\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}\right)\right)\hat{\tau} + 1}\right)
\end{aligned} \tag{2.36}$$

dan

$$
\begin{aligned}
l(\widehat{\omega}) &= \ln L\,(\widehat{\omega}) \\
&= \sum_{i=1}^{n} \ln\left(\exp(-\hat{\gamma}_0) + \exp\left(\frac{1}{\tau} - \frac{1}{\tau}\sqrt{\left(2\left(\exp\left(\hat{\beta}_0\right)\right)\hat{\tau}\right) + 1}\right)\right) \\
&\qquad - \sum_{i=1}^{n} \ln(1 + \exp(-\hat{\gamma}_0)) \\
&\qquad + \sum_{i=1}^{n} (1 - y_i)(\ln 2 - \ln \pi - \ln \hat{\tau}) \\
&\qquad + \sum_{i=1}^{n} y_i \hat{\beta}_0 + \frac{n}{\tau} - \sum_{i=1}^{n} \ln\left(1 + \exp\left(-\gamma \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}\right)\right) \\
&\qquad + \sum_{i=1}^{n} \ln(y_i!) \\
&\qquad + \sum_{i=1}^{n} \ln\left(\frac{1}{2} - y_i\right) \ln\left(2\left(\exp\left(\hat{\beta}_0\right)\right)\hat{\tau} + 1\right) \\
&\qquad - \sum_{i=1}^{n} \left(-\frac{1}{\tau}\sqrt{2\left(\exp\left(\hat{\beta}_0\right)\right)\hat{\tau} + 1}\right)
\end{aligned} \tag{2.37}
$$

dengan kriteria dan kesimpulan pengujian H_0 ditolak apabila $G > \chi^2_{\alpha,v}$. Artinya, secara bersama-sama variabel independen berpengaruh terhadap variabel dependen.

2.8.2. Uji Parsial

Pengujian secara parsial dalam penelitian ini menggunakan hipotesis sebagai berikut:

Untuk parameter β_j:

$H_0: \beta_j = 0$

$H_1: \beta_j \neq 0$ dengan $j = 1, 2, \dots, k$

Untuk parameter τ:

$H_0: \tau = 0$

$H_1: \tau \neq 0$

Untuk parameter γ:

$H_0 : \gamma = 1$

$H_1 : \gamma \neq 1$

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

$$t = \frac{\hat{\theta}_j}{SE(\hat{\theta}_j)} \tag{2.38}$$

dimana $\hat{\theta}_j$ merupakan nilai estimasi parameter untuk $\boldsymbol{\beta}, \tau$, 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.

2.9 Kriteria Pemilihan Model Terbaik

Kriteria pemilihan model terbaik dapat dilihat melalui nilai *Akaike Information Criterion* (AIC) yang ditentukan oleh Akaike. Metode AIC didasarkan pada metode *Maximum Likelihood Estimation* (MLE). Nilai AIC dapat dihitung dengan persamaan berikut:

$$AIC = -2 \ln L(\hat{\boldsymbol{\theta}}) + 2k \tag{2.39}$$

dimana 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 (Wahyuni dkk., 2021).

Pemilihan model terbaik juga perlu memperhatikan signifikansi parameter dispersi τ yang menunjukkan bahwa data tidak memenuhi kondisi *equidispersi* atau mengalami kondisi *overdispersi* (Wang & Famoye, 1997). Selain itu, pemilihan model terbaik juga perlu memperhatikan signifikansi dari parameter *zero inflated* γ yang menunjukkan bahwa data sesuai untuk model *zero inflated* (Anggreainy dkk., 2022).

2.10 Penyakit Malaria

Malaria merupakan penyakit yang disebabkan mikroorganisme bernama *odium* dan bersifat menular melalui vektor yaitu nyamuk *Anopheles*.

Berdasarkan status endemisnya, sebagian besar wilayah bagian barat Indonesia berstatus endemis tinggi. Indonesia berada di peringkat kedua dengan jumlah kasus malaria tertinggi di Asia Tenggara (Kemenkes RI, 2021). Penelitian Santi dkk. (2021) menyebutkan bahwa akses sanitasi yang layak menjadi salah satu penyebab yang berpengaruh secara signifikan terhadap jumlah kasus malaria di Indonesia. Penelitian Sendow & Chernovita (2021) menerangkan bahwa nyamuk *Anopheles* berkembang biak di tempat dengan genangan air yang memadai, sehingga kondisi iklim atau dalam hal ini curah hujan, kelembaban, dan suhu memiliki pengaruh terhadap jumlah kasus malaria di Kabupaten Mimika, Papua. Penelitian Aprilia (2023) menunjukkan hasil 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.

Infeksi Plasmodium pada manusia dapat disebabkan oleh satu dari lebih dari empat jenis Plasmodium yaitu *Plasmodium Falciparum, Plasmodium Malariae, Plasmodium Vivax,* dan *Plasmodium Ovale*. Jenis Plasmodium tersebut dapat memberikan pengaruh terhadap gejala yang ditimbulkan serta pengobatannya. Pengobatan malaria menggunakan *Artemicin-based Combination Therapy* (ACT) yang diberikan pada 24 jam pertama pasien panas dan obat harus diminum habis. Pencegahan penyakit malaria dapat dilakukan dengan Pembersihan Sarang Nyamuk (PSN) dengan menghindari gigitan nyamuk yakni menggunakan kain kelambu atau penggunaan obat *Chloroquine* serta dengan memberikan pengetahuan terkait dengan penyakit malaria (Kemenkes, 2022).