

## REGRESI ZERO INFLATED POISSON UNTUK PEMODELAN ANGKA POSITIF PENYAKIT MALARIA DI JAWA TIMUR

**Ajeng Diva Aprilia**

Jurusan Matematika, FMIPA, Universitas Negeri Surabaya

e-mail : [ajeng.19013@mhs.unesa.ac.id](mailto:ajeng.19013@mhs.unesa.ac.id)

**A'yunin Sofro**

Jurusan Matematika, FMIPA, Universitas Negeri Surabaya

Penulis Korespondensi : [ayuninsofro@unesa.ac.id](mailto:ayuninsofro@unesa.ac.id)

### Abstrak

Jumlah kasus positif malaria pada tahun 2020 – 2022 di Jawa Timur merupakan data *count* dengan adanya banyak nilai nol dan ditemukan keadaan overdispersi . Penelitian ini bertujuan untuk mengatasi masalah pada regresi *Poisson* yaitu keadaan overdispersi dengan regresi *Zero Inflated Poisson* (ZIP). Hasil penelitian menunjukkan model ZIP lebih baik daripada model regresi *Poisson*. Pemodelan data dengan regresi ZIP melakukan estimasi parameter menggunakan *Maximum Likelihood Estimation*, pengujian kesesuaian model dengan uji G dan uji Wald untuk pengujian signifikansi parameter. Model regresi ZIP untuk data kasus positif malaria adalah :  $\log(\mu_i) = 3,695 + 0,0057X_1 - 0,0569X_2 - 0,0085X_3 - 0,0619X_4 + 0,0604X_5 - 0,0118X_6 - 0,012X_8$  dan  $\text{logit}(\omega_i) = -5,5185 + 0,0039X_2 - 0,592X_6$ . Hasil dari pengujian menunjukkan bahwa tingkat kepadatan penduduk, tingkat kemiskinan, persentase rumah tangga dengan keadaan sanitasi layak, angka kesakitan atau *morbidity rate*, persentase keluhan kesehatan, jumlah sarana kesehatan, jumlah tenaga kesehatan, dan persentase tempat umum yang sesuai standar kesehatan secara signifikan berpengaruh terhadap kasus positif malaria di Jawa Timur.

**Kata Kunci:** Malaria, *excess zeros*, overdispersi, *Zero Inflated Poisson*

### Abstract

The number of confirmed cases of malaria in East Java from 2020 to 2022 is a count data with numerous zero values and is exhibiting overdispersion. With Zero Inflated Poisson (ZIP) regression, this work seeks to solve the overdispersion issue that plagues Poisson regression. The ZIP model is superior to the Poisson regression model, according to the results. Data modeling with ZIP regression carries out model fit testing with the G test, parameter significance testing with the Wald test, and parameter estimation using Maximum Likelihood Estimation. For case data of malaria positivity, the ZIP regression model in this research is :  $\log(\mu_i) = 3,695 + 0,0057X_1 - 0,0569X_2 - 0,0085X_3 - 0,0619X_4 + 0,0604X_5 - 0,0118X_6 - 0,012X_8$ , and  $\text{logit}(\omega_i) = -5,5185 + 0,0039X_2 - 0,592X_6$ . The test's findings indicate that factors such as population density, poverty levels, the proportion of households with access to clean water and sanitation, the morbidity rate, the number of medical facilities, the number of health complaints, and the percentage of public spaces that adhere to health standards all significantly influence the number of confirmed cases of malaria in East Java.

**Keywords:** Malaria, *excess zeros*, overdispersion, *Zero Inflated Poisson*.

## 1. PENDAHULUAN

Penyakit menular yang disebarkan melalui gigitan nyamuk jenis *Anopheles* betina dengan membawa genus *Plasmodium* berupa parasit adalah malaria (Santi et al., 2019). Berdasarkan laporan World Malaria Report 2020, Indonesia memegang peringkat negara kedua dengan jumlah kasus malaria tertinggi di Asia Tenggara. Penyakit malaria

tergolong ke dalam salah satu penyebab kematian tertinggi di dunia, salah satunya Indonesia.. Menurut data informasi dari Kementerian Kesehatan, secara keseluruhan pada tahun 2021 total kasus malaria di Indonesia mencapai 94.610 kasus. Dilansir dari Dinas Kesehatan Jawa Timur bahwa di Jawa Timur masih dijumpai kasus malaria yang masih tinggi, dimana kasus malaria mencapai ribuan kasus yaitu sebesar 15.529 kasus.

Kasus positif malaria pada beberapa daerah di Jawa Timur tidak menyebar secara menyeluruh dan relatif jarang terjadi di beberapa daerah lainnya sehingga terdapat banyak nilai nol yang menyebabkan overdispersi. Sehubungan dengan hal tersebut, diperlukan suatu analisis yang sesuai untuk mengidentifikasi faktor - faktor yang memberikan pengaruh pada jumlah kasus positif malaria di Jawa Timur. Suatu teknik statistika untuk analisis data *count* untuk suatu kejadian dalam suatu kurun waktu yang jarang terjadi tertentu adalah regresi *Poisson* (Gupta, 2004). Data *count* merupakan tipe data statistik yang pengamatannya hanya mengambil nilai bilangan bulat non negatif dan nilai tersebut muncul dari perhitungan bukan pemeringkatan (Obubu, 2018). Suatu kondisi pada regresi *Poisson* yang disebut *equidispersi* terjadi ketika nilai mean dan varians bernilai sama pada variabel respon (Bal, 2020). Namun dalam penerapannya masih dijumpai kondisi *overdispersi*, yang terjadi ketika nilai varians melebihi nilai rata - rata. Adanya *overdispersi* tersebut dapat disebabkan oleh data pada variabel respon yang memiliki sejumlah besar nilai nol (Yau, 2001).

Salah satu metode untuk menangani kondisi terlalu banyak nol atau *excess zeros* dalam variabel respon yaitu regresi *Zero Inflated Poisson*. Dilakukan analisis pada penelitian ini terhadap kasus positif malaria di Jawa Timur selama tiga tahun terakhir, yaitu tahun 2020 sampai dengan 2022 pada 38 kabupaten/kota dengan menggunakan dan membandingkan antara dua metode. Data dalam penelitian ini menerapkan penggunaan data panel yang diartikan sebagai gabungan antara data silang (*cross section*) dan data deret waktu (*time series*) atau dikenal dengan data panel. Pada data deret waktu mencakup lebih dari satu periode, yaitu tiga tahun. Data penelitian terdiri dari beberapa variabel prediktor dalam jangka waktu tertentu pada 38 kabupaten/kota yang disebut *cross section*.

Pemodelan Jumlah Kasus Malaria di Indonesia Menggunakan *Generalized Linear Model* oleh Santi et al. (2021) adalah contoh relevan dari penelitian sebelumnya yang menggunakan GLM. "Model Regresi *Zero Inflated Poisson* Pada Data *Overdispersion*" adalah judul lain yang berhubungan dengan *Zero Inflated Poisson*. Penelitian tersebut menunjukkan bahwa ZIP penanganan yang tepat jika digunakan dalam kasus kasus *overdispersi* yang

diakibatkan oleh kondisi *excess zeros* pada variabel respon.

Berdasarkan uraian tersebut dan ketersediaan data, maka faktor - faktor yang dianalisis adalah tingkat kepadatan penduduk, tingkat kemiskinan, persentase rumah tangga dengan keadaan sanitasi layak, angka kesakitan atau *morbidity rate*, persentase keluhan kesehatan, jumlah sarana kesehatan, jumlah tenaga kesehatan dan persentase tempat umum yang sesuai standar kesehatan. Model regresi *Poisson* dan *Zero Inflated Poisson* (ZIP) digunakan dalam penelitian ini untuk membandingkan model mana yang lebih efektif dalam mengidentifikasi variabel yang secara signifikan mempengaruhi bertambahnya kasus positif malaria.

## 2. KAJIAN TEORI

### REGRESI POISSON

Salah satu penerapan *Generalized Linear Model* pada data dengan variabel responnya diasumsikan mengikuti distribusi *Poisson* adalah regresi *Poisson* (Agresti, 2015). Komponen acak, komponen sistematis, dan fungsi penghubung adalah tiga komponen dari GLM yang didefinisikan sebagai berikut :

- Komponen yang mengidentifikasi distribusi peluang dari variabel respon  $y_i = y_1, y_2, \dots, y_n$  dengan mean  $E(y_i) = \mu_1, \mu_2, \dots, \mu_n$  merupakan komponen acak. Distribusi dari pengamatan  $y_i$  merupakan salah satu jenis dari keluarga eksponensial.
- Variabel - variabel penjelas  $x_1, x_2, \dots, x_n$  yang menghasilkan model  $x_i\beta = \beta_0 + \sum_{i=1}^n x_i\beta_i$  adalah bagian dari komponen sistematis.
- Link function* (fungsi penghubung) merupakan komponen ketiga yang digunakan untuk membuat model :

$$\eta_i = g(\mu_i) \quad i = 1, 2, \dots, n$$

*Generalized Linear Model* standar yang diterapkan pada data count dimana mean dari variabel respon  $\mu_i$  dihubungkan oleh *link function* dengan penduga linear  $\eta_i$  adalah model regresi *Poisson* dan diformulasikan sebagaimana persamaan (2.1).

$$g(\mu_i) = \ln \mu_i = \eta_i$$

$$\ln(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij}$$

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} \quad (2.1)$$

dengan  $j = 1, 2, 3, \dots$  merupakan indeks dari variabel - variabel prediktor.

$$\ln(\mu_i) = x_i\beta$$

$$\mu_i = \exp(x_i\beta)$$

Persamaan untuk model regresi *Poisson* dengan penghubung log dapat dituliskan sebagaimana persamaan (2.2) (Cahyandari, 2014).

$$g(\mu_i) = \log \mu_i = x_i\beta \quad (2.2)$$

dengan,

$\mu_i$  : rata - rata dari jumlah peristiwa selama periode tertentu pada observasi ke- $i$

$x_i$  : vektor dari variabel prediktor

$\beta$  : vektor dari parameter regresi *Poisson*

Apabila terdapat suatu kondisi dimana data variabel respon pada regresi *Poisson* memiliki banyak nilai nol atau *excess zeros*, model *Zero Inflated Poisson* adalah model yang lebih baik untuk digunakan.

### REGRESI ZERO INFLATED POISSON

Salah satu penerapan penggunaan metode statistika yang bertujuan untuk menangani kondisi terlalu banyak nol atau *excess zeros* dalam variabel respon disebut regresi *Zero Inflated Poisson* (Hu, 2021). Model regresi *Poisson* dikembangkan untuk menganalisis adanya pengaruh variabel prediktor terhadap variabel respon yang memiliki banyak nilai nol yang dikenal sebagai regresi *Zero Inflated Poisson* (ZIP). Keadaan variabel respon yang terlalu banyak nilai nol disebut dengan *excess zeros* ditunjukkan dengan kejadian yang jarang terjadi (*rare case*) dan hanya terjadi di beberapa daerah pengamatan.

Untuk setiap pengamatan pada variabel respon  $y_i$  yang saling bebas, dalam pengamatan nilai nol dapat terjadi dalam dua keadaan (*state*) yang berbeda untuk setiap pengamatan. *Zero state* merupakan keadaan pertama dan *Poisson state* merupakan keadaan kedua (Jansakul, 2002). Sebagai hasilnya  $y_i$  mengikuti distribusi *Zero Inflated Poisson* sesuai dengan yang ditunjukkan pada persamaan (2.3).

$$y_i \sim \begin{cases} 0, & \text{dengan probabilitas } \omega_i \\ \text{poisson}(\mu_i), & \text{dengan probabilitas } (1 - \omega_i) \end{cases} \quad (2.3)$$

Maka fungsi probabilitasnya dapat ditunjukkan pada persamaan 2.4 (Yau, 2001).

$$P(y_i) = \begin{cases} \omega_i + (1 - \omega_i)e^{-\mu_i}, & \text{untuk } y_i = 0 \\ \frac{(1 - \omega_i)e^{-\mu_i}\mu_i^{y_i}}{y_i!}, & \text{untuk } y_i > 0, 0 \leq \omega_i \leq 1 \end{cases} \quad (2.4)$$

Sehingga dengan parameter  $\mu_i$  dan  $\omega_i$ , maka model regresi *Zero Inflated Poisson* ditunjukkan pada persamaan 2.5 (Lambert, 1992).

$$\log(\mu_i) = x_i\beta \text{ dan } \logit(\omega_i) = \log \frac{\omega_i}{1 - \omega_i} = x_i\gamma \quad (2.5)$$

dengan,  $x_i$  merupakan vektor dari variabel - variabel predictor, variabel - variabel yang mempengaruhi *Poisson state* diwakili oleh vektor dari parameter regresi  $\beta$ , sedangkan  $\gamma$  merupakan vektor dari parameter regresi yang ditaksir mewakili variabel yang mempengaruhi *zero state*.

### ESTIMASI PARAMETER

Pada regresi *Zero Inflated Poisson*, dalam tahap estimasi parameter pada penelitian ini menerapkan *Maximum Likelihood Estimation* (MLE). Dari model yang terbentuk, fungsi *likelihood* dengan  $\beta$  dan  $\gamma$  merupakan parameter yang akan diestimasi adalah :

$$L(\beta, \gamma) = \begin{cases} \prod_{i=1}^n \frac{e^{x_i\gamma} + \exp(-e^{x_i\beta})}{1 + e^{x_i\gamma}}, & \text{untuk } y_i = 0 \\ \prod_{i=1}^n \frac{\frac{1}{1 + e^{x_i\gamma}} \exp(-e^{x_i\beta}) (e^{x_i\beta})^{y_i}}{y_i!}, & \text{untuk } y_i > 0 \end{cases} \quad (2.6)$$

Selanjutnya dibuat fungsi log *likelihood* dari persamaan (2.6) sehingga diperoleh :

$$\log L(\beta, \gamma) = \begin{cases} \sum_{i=1}^n \log(e^{x_i\gamma} + \exp(-e^{x_i\beta})) - \sum_{i=1}^n \log(1 + e^{x_i\gamma}), & \text{untuk } y_i = 0 \\ \sum_{i=1}^n ((x_i\beta)y_i - e^{x_i\beta}) - \sum_{i=1}^n \log(1 + e^{x_i\gamma}) - \sum_{i=1}^n y_i!, & \text{untuk } y_i > 0 \end{cases} \quad (2.7)$$

Sehingga total fungsi log *likelihood* dapat diberikan oleh  $L_T = \log L(\beta, \gamma)_{y_i=0} + \log L(\beta, \gamma)_{y_i>0}$ , dan diperoleh :

$$L_T = \sum_{y_i=0} \log(\exp(x_i\gamma) + \exp(-e^{x_i\beta})) - \sum_{y_i=0} \log(1 + \exp x_i\gamma) + \sum_{y_i>0} (x_i\beta y_i - \exp(x_i\beta)) - \sum_{y_i>0} \log(1 + \exp x_i\gamma) - \sum_{y_i>0} (y_i!) \quad (2.8)$$

### UJI SIGNIFIKANSI PARAMETER SIMULTAN

Tahap untuk mengukur secara keseluruhan atau simultan adanya pengaruh dari variabel prediktor terhadap variabel respon adalah uji signifikansi parameter menggunakan uji G.

Penerapan uji statistik G adalah seperti pada persamaan (2.9) (Agresti, 2015).

$$G = -2 \ln \left( \frac{L(1)}{L(2)} \right) = -2 [\ln L(1) - \ln L(2)] \sim \chi^2_{(p;\infty)} \quad (2.9)$$

Kriteria pengujiannya adalah jika  $G > \chi^2_{(p;\infty)}$  maka dapat diartikan bahwa variabel prediktor

berpengaruh secara signifikan, dimana  $p$  merupakan banyaknya variabel prediktor yang terdapat pada model.

### UJI SIGNIFIKANSI PARAMETER PARSIAL

Tahap selanjutnya yaitu uji secara parsial untuk mengukur adanya pengaruh dari masing – masing variabel prediktor terhadap variabel respon yang dapat diketahui dengan melakukan uji Wald. Penerapan uji Wald adalah seperti pada persamaan (2.10) berikut :

$$W = \frac{\beta_j^2}{(SE(\beta_j))^2} \sim Z \quad (2.10)$$

dengan :  $SE(\beta_j) = \sqrt{Var(\beta_j)}$

$SE(\beta_j)$  merupakan standar error *Maximum Likelihood Estimation* (MLE).

Kriteria pengujiannya adalah jika  $|W| > Z_{\alpha/2}$  atau  $p - \text{value} < \alpha$  dengan  $\alpha$  adalah taraf signifikan, dapat disimpulkan bahwa parameter signifikan.

### PEMILIHAN MODEL TERBAIK DENGAN AIC

Dalam pemilihan model yang lebih tepat digunakan, dilakukan dengan melihat perbandingan antara model regresi *Poisson* dan regresi *Zero Inflated Poisson* mana yang lebih baik. Dengan melihat nilai AIC dari kedua model dengan rumus perhitungannya seperti pada persamaan (2.11) berikut :

$$AIC = -2 \ln L(\beta) + 2p \quad (2.11)$$

Diketahui bahwa  $L(\beta)$  menyimbolkan nilai likelihood dan jumlah parameter sejumlah  $p$ .

## 3. METODE

### DATA PENELITIAN

Data penelitian didapatkan dari dua sumber, yaitu BPS dan Publikasi Dinas Kesehatan pada Provinsi Jawa Timur tahun 2020 sampai dengan 2022 pada 38 kabupaten/kota. Dalam penelitian ini jumlah kasus positif malaria ditentukan menjadi variabel respon (Y) dan terdapat delapan variabel prediktor (X), yaitu tingkat kepadatan penduduk (X1), tingkat kemiskinan (X2), persentase rumah tangga dengan keadaan sanitasi layak (X3), angka kesakitan atau *morbidity rate* (X4), persentase keluhan kesehatan (X5), jumlah sarana kesehatan (X6), jumlah tenaga kesehatan (X7), dan persentase tempat umum yang sesuai standar kesehatan (X8).

### TAHAP PENELITIAN

Tahapan yang ditetapkan sebagai acuan dalam penelitian agar terstruktur adalah :

1. Studi literatur terkait materi yang berkaitan dengan penelitian dan menentukan batasan masalah sehingga ruang lingkup penelitian lebih jelas.
2. Pengumpulan data, data diambil berupa data sekunder pada Badan Pusat Statistik dan Publikasi Dinas Kesehatan Jawa Timur tahun 2020, 2021, dan 2022.
3. Pengolahan data dan penentuan variabel respon serta variabel prediktor.
4. Pemodelan dengan menerapkan regresi *Poisson* dan pengecekan overdispersi.
5. Pemodelan untuk menangani kasus overdispersi menggunakan regresi *Zero Inflated Poisson* dengan estimasi parameter menggunakan *Maximum Likelihood*.
6. Pengujian signifikansi parameter simultan dan parsial untuk menguji faktor – faktor yang berpengaruh terhadap kasus positif malaria.
7. Pemilihan model yang lebih tepat dengan menggunakan nilai AIC.
8. Menginterpretasi model dari hasil *odds ratio*.
9. Membuat kesimpulan.

## 4. HASIL DAN PEMBAHASAN

### STATISTIK DESKRIPTIF

Dalam penelitian ini, jumlah kasus positif malaria sebagai variabel respon (Y) dan terdapat delapan variabel prediktor (X). Statistik deskriptif yang diperoleh dari data penelitian di Jawa Timur selama 3 tahun (2020, 2021, 2022) dapat ditunjukkan pada tabel 1 berikut ini :

Tabel 1. Nilai Statistik Deskriptif

Variabel	Mean	Min	Max	Std. deviasi	Varians
Y	10,06	0	211	22,99	39,05
X1	19,75	2,95	85,95	21,75	34,95
X2	10,88	1,39	23,76	4,64	5,87
X3	81,25	39,44	98,71	13,25	12,97
X4	14,8	5,11	40,09	5,3	7,64
X5	32,17	13,56	57,25	8,87	21,4
X6	33,53	7	121	18,21	24,5
X7	545,35	79	8733	1101,29	489,3
X8	68,47	17,5	96	13,69	13,85

Setelah diperoleh statistik deskriptif tersebut, pada variabel respon (Y) menghasilkan nilai varians 39,05 yang melebihi dari nilai rataannya 10,06 yang menunjukkan bahwa data mengalami overdispersi.

### PEMERIKSAAN MULTIKOLINEARITAS

Pengujian multikolinearitas pada jumlah kasus positif malaria tahun 2020 – 2022 untuk pendeteksian adanya hubungan linear antara variabel – variabel prediktor dapat diketahui dari nilai VIF. Nilai VIF terdapat pada tabel 2 berikut :

Tabel 2. Pengujian Multikolinearitas

Variabel (X)	Nilai VIF
X1	1,608
X2	1,553
X3	1,239
X4	3,24
X5	3,359
X6	2,087
X7	2,306
X8	1,231

Merujuk pada tabel 2, terlihat bahwa hasil perhitungan VIF kurang dari 10 sehingga dapat ditunjukkan jika pada setiap variabel prediktor tidak terdapat multikolinearitas.

### MODEL REGRESI POISSON

Penerapan model regresi *Poisson* adalah salah satu analisis yang sesuai untuk mengetahui faktor – faktor apa saja yang berpengaruh terhadap jumlah kasus positif malaria di Jawa Timur selama tahun 2020 – 2022. Berikut hasil estimasi parameter dari model regresi *Poisson* dengan Maximum Likelihood Estimation (MLE).

Tabel 3. Hasil Estimasi Parameter Regresi Poisson

Parameter	Estimasi	Standar Error	Z value	P value
Y	3,835	0,2511	15,272	2e-16
X1	0,0009	0,0015	0,574	0,56602
X2	-0,1194	0,0105	-11,276	2e-16
X3	-0,0045	0,0021	-2,141	0,0322
X4	-0,0649	0,0088	-7,346	2,05e-13
X5	0,06	0,0048	12,4	2e-16
X6	-0,0031	0,0025	-1,228	0,2195
X7	-0,0002	6,944e-05	-3,791	0,00015
X8	-0,0126	0,0019	-6,009	1,87e-09

Seperti halnya hasil perhitungan pada tabel 3, menunjukkan enam variabel – variabel prediktor

meliputi X2, X3, X4, X5, X7, dan X8 yang terhadap model dipengaruhi secara signifikan. Nilai  $p\text{-value} < \alpha = 0,05$  yang digunakan sebagai dasar dalam penentuan adanya pengaruh signifikan. Model yang didapatkan dengan menggunakan regresi *Poisson* yaitu :

$$\mu_i = \exp(3,835 + 0,0009X1 - 0,1194X2 - 0,0045X3 - 0,0649X4 + 0,06X5 - 0,0031X6 - 0,0002X7 - 0,0126X8)$$

Setelah diperoleh model regresi *Poisson* dilakukan pengecekan apakah masih terdapat kondisi overdispersi pada model.

### PEMERIKSAAN OVERDISPERSI

Model regresi *Poisson* menjadi kurang tepat apabila diterapkan dalam memodelkan data jika terjadi overdispersi yang dapat dilihat pada variabel respon. Merasiakan nilai devians dengan derajat bebasnya dan *Pearson Chi-Squared* dengan derajat bebasnya merupakan langkah dalam pemeriksaan overdispersi. Hasil perhitungan dari model regresi *Poisson* memperoleh nilai taksiran dispersi berikut ini:

Tabel 4. Pendeteksian Overdispersi Regresi Poisson

Kriteria	Nilai	Db	Rasio
Devians	1849,2	105	17,611
<i>Pearson's chi-squared</i>	2697,506	105	25,691

Hasil rasio yang diperoleh pada tabel 4 memiliki nilai melebihi satu sehingga mengindikasikan bahwa pada model regresi *Poisson* ditemukan keadaan overdispersi. Hal tersebut mengartikan bahwa penerapan model regresi *Poisson* kurang tepat jika digunakan pada data. Selanjutnya model regresi lain yang dapat digunakan dalam kasus overdispersi, yaitu regresi *Zero Inflated Poisson*.

### MODEL REGRESI ZERO INFLATED POISSON (ZIP)

Apabila model regresi *Poisson* masih dijumpai keadaan overdispersi maka dapat ditangani dengan *Zero Inflated Poisson*. Sama halnya dengan regresi *Poisson*, dalam tahap untuk menghasilkan nilai  $p\text{-value}$  dalam uji parameter dilakukan menggunakan MLE. Hasil estimasi dan pengujian ditunjukkan pada tabel 5 :

Tabel 5. Hasil MLE dengan ZIP

Parameter	Estimasi	Standar Error	Z value	P value
$\beta_0$	3,695	0,2624	14,078	2e-16
$\beta_1$	0,0057	0,00178	3,211	0,0013
$\beta_2$	-0,0569	0,0136	-4,179	2,93e-05
$\beta_3$	-0,0085	0,0021	-3,968	7,25e-05
$\beta_4$	-0,0619	0,008	-7,702	1,34e-14
$\beta_5$	0,0604	0,0047	12,805	2e-16
$\beta_6$	-0,0118	0,0026	-4,54	5,63e-06
$\beta_7$	0,00005	6,533e-05	0,772	0,44
$\beta_8$	-0,012	0,0019	-6,025	1,69e-09
$\gamma_0$	-5,5185	2,999	-1,840	0,0658
$\gamma_1$	0,0254	0,0196	1,298	0,1942
$\gamma_2$	0,0039	0,0922	4,076	4,58e-05
$\gamma_3$	0,3759	0,0276	0,144	0,8851
$\gamma_4$	-0,0591	0,0912	-0,649	0,5164
$\gamma_5$	0,0547	0,0562	0,973	0,3306
$\gamma_6$	-0,592	0,0296	-1,998	0,0457
$\gamma_7$	0,0012	0,0007	1,759	0,0785
$\gamma_8$	-0,0107	0,0224	-0,481	0,6308

Berdasarkan tabel 5 maka dapat terbentuk model regresi *Zero Inflated Poisson* dalam fungsi log dan fungsi logit berikut ini :

$$\log(\mu_i) = 3,695 + 0,0057X_1 - 0,0569X_2 - 0,0085X_3 - 0,0619X_4 + 0,0604X_5 - 0,0118X_6 + 0,00005X_7 - 0,012X_8$$

dan

$$\text{logit}(\omega_i) = -5,5185 + 0,0254X_1 + 0,0039X_2 + 0,3759X_3 - 0,0591X_4 + 0,0547X_5 - 0,592X_6 + 0,0012X_7 - 0,0107X_8$$

#### UJI SIGNIFIKANSI PARAMETER SIMULTAN

Tahap selanjutnya yang dilakukan guna menentukan layak tidaknya model regresi *Zero Inflated Poisson* dapat digunakan dalam menunjukkan hubungan dari variabel prediktor dan variabel respon, yaitu dilakukan dengan uji signifikansi model. Uji G digunakan untuk menilai signifikansi model secara simultan. Hasil pengujian tersebut menghasilkan nilai statistik uji G dengan perhitungan  $-2 \log \text{likelihood}$  sebesar 1860,8 sehingga  $G = 1860,8 > \chi^2_{0,05;8} = 15,51$ . Berdasarkan keputusan yang telah dibuat untuk model regresi ZIP, jumlah kasus positif malaria pada tahun 2020 –

2022 dapat dipengaruhi oleh faktor – faktor seperti tingkat kepadatan penduduk (X1), tingkat kemiskinan (X2), persentase rumah tangga dengan keadaan sanitasi layak (X3), angka kesakitan atau *morbidity rate* (X4), persentase keluhan kesehatan (X5), jumlah sarana kesehatan (X6), jumlah tenaga kesehatan (X7), dan persentase tempat umum yang sesuai standar kesehatan (X8).

#### UJI SIGNIFIKANSI PARAMETER PARSIAL

Setelah melakukan uji secara simultan maka selanjutnya dilakukan uji Wald yang digunakan untuk uji signifikansi pada setiap parameter model regresi ZIP. Merujuk pada tabel 5 nilai statistik uji Wald diperoleh dengan membagi nilai estimasi dengan standar error atau dapat dilihat pada nilai *Z - value*. Ketika uji Wald untuk model  $\log(\mu_i)$  digunakan untuk menentukan signifikansi setiap parameter model regresi, dapat dilihat bahwa tingkat kepadatan penduduk (X1), tingkat kemiskinan (X2), persentase jumlah rumah tangga yang memiliki akses sanitasi layak (X3), angka kesakitan atau *morbidity rate* (X4), persentase keluhan kesehatan (X5), jumlah sarana kesehatan (X6), dan persentase tempat umum yang sesuai standar kesehatan (X8) yang signifikan pada  $\alpha = 0,05$  oleh karena itu, variabel jumlah tenaga kesehatan (X7) tidak signifikan. Sedangkan untuk model  $\text{logit}(\omega_i)$  diperoleh bahwa hanya ada dua variabel, yaitu persentase penduduk miskin (X2) dan jumlah sarana kesehatan (X6) yang signifikan pada  $\alpha = 0,05$ .

#### PEMILIHAN MODEL TERBAIK

Penentuan model yang paling baik dari dua model yang ada dalam penelitian, yaitu antara model regresi *Poisson* dan regresi *Zero Inflated Poisson*. Dari kedua model tersebut kemudian dilakukan penentuan model terbaik yang dapat dipilih dari nilai AIC berikut :

Tabel 6 Nilai AIC Pada Model

Model	AIC
Model Regresi <i>Poisson</i>	2194,676
Model Regresi <i>Zero Inflated Poisson</i>	1896,896

Berdasarkan nilai AIC yang didapatkan seperti pada tabel 6, nilai AIC yang didapatkan dari model *Zero Inflated Poisson* bernilai lebih kecil. Oleh karena itu, dapat dipilih model regresi ZIP yang lebih baik dalam pemodelan data jumlah kasus positif malaria

lebih baik daripada regresi *Poisson*. Setelah model regresi ZIP dipilih sebagai model terbaik dilakukan interpretasi model ZIP.

### INTERPRETASI HASIL

Model regresi ZIP yang diperoleh berdasarkan hasil uji signifikansi parameter regresi ZIP untuk model log adalah :

$$\log(\mu_i) = 3,695 + 0,0057X_1 - 0,0569X_2 - 0,0085X_3 - 0,0619X_4 + 0,0604X_5 - 0,0118X_6 - 0,012X_8$$

Dari model log yang diperoleh maka dapat dilakukan interpretasi model untuk menjelaskan pengaruh dari setiap variabel prediktor yang signifikan sebagai berikut :

1. Nilai koefisien untuk  $X_1$  adalah 0,0057. Dapat dijelaskan bahwa jika terjadi peningkatan kepadatan penduduk sebesar 1% akan mejadikan jumlah kasus positif malaria meningkat sebesar  $\exp(0,0057)=1,005 \approx 1$  kali.
2. Nilai koefisien  $X_2$  adalah -0,0569, dapat diinterpretasikan bahwa terjadi peningkatan jumlah kasus positif malaria sebesar  $\exp(-0,0569) = 0,9446 \approx 1$  kali yang disebabkan karena terjadi peningkatan persentase penduduk miskin sebesar 1%.
3. Nilai koefisien  $X_3$  adalah -0,0085, dapat diinterpretasikan bahwa terjadi peningkatan jumlah kasus positif malaria sebesar  $\exp(-0,0085) = 0,9915 \approx 1$  kali yang disebabkan karena terjadi peningkatan persentase rumah tangga dengan keadaan sanitasi layak sebesar 1%.
4. Nilai koefisien  $X_4$  adalah -0,0619, dapat diinterpretasikan bahwa apabila terjadi peningkatan sebesar 1% angka kesakitan atau *morbidity rate* yang dapat menyebabkan meningkatnya jumlah kasus positif malaria sebesar  $\exp(-0,0619) = 0,9399 \approx 1$  kali.
5. Nilai koefisien  $X_5$  adalah 0,0604, dapat diinterpretasikan bahwa jika terjadi peningkatan sebesar 1% pada persentase keluhan kesehatan akan mengakibatkan meningkatnya jumlah kasus positif malaria sebesar  $\exp(0,0604) = 1,066 \approx 1$  kali.
6. Nilai koefisien  $X_6$  adalah -0,0118, dapat diinterpretasikan bahwa jika terjadi peningkatan sebesar 1% pada jumlah sarana kesehatan menyebabkan meningkatnya jumlah kasus

positif malaria sebesar  $\exp(-0,0118) = 0,9882 \approx 1$  kali.

7. Nilai koefisien  $X_8$  adalah -0,012, dapat diinterpretasikan bahwa apabila terjadi peningkatan sebesar 1% pada persentase tempat umum yang sesuai standar kesehatan meningkatkan jumlah kasus positif malaria sebesar  $\exp(-0,012) = 0,988 \approx 1$  kali.

Model regresi ZIP yang diperoleh berdasarkan hasil uji signifikansi parameter regresi ZIP untuk model logit adalah :

$$\text{logit}(\omega_i) = -5,5185 + 0,0039X_2 - 0,592X_6$$

Dari model logit yang diperoleh maka dapat dilakukan interpretasi model untuk menjelaskan pengaruh dari setiap variabel prediktor yang signifikan sebagai berikut :

1. Nilai koefisien untuk  $X_2$  adalah 0,0039. Dapat diinterpretasikan bahwa jika terjadi peningkatan sebesar 1% pada persentase penduduk miskin akan berpengaruh terhadap meningkatnya jumlah kasus positif malaria sebesar  $100(\exp(0,0039) - 1) = 0,39\%$ .
2. Nilai koefisien untuk  $X_6$  adalah -0,592. Dapat diinterpretasikan bahwa jika terjadi peningkatan sebesar 1% pada persentase penduduk miskin akan berakibat dalam penurunan jumlah kasus positif malaria sebesar  $100(\exp(-0,592) - 1) = 44,67\%$ .

### 5. PENUTUP KESIMPULAN

Berdasarkan temuan penelitian ini, kesimpulan yang dapat diambil yaitu pemodelan jumlah kasus positif malaria dapat menggunakan regresi *Poisson*, namun dalam pelaksanaannya ditemukan masalah overdispersi sehingga model tidak memberikan hasil yang baik. Dari hasil pengujian *Pearson Chi Squared* yang dilakukan  $G = 1860,8 > \chi^2_{0,05,8} = 15,51$ , menunjukkan bahwa lebih tepat dan layak dalam penerapan model regresi *Zero Inflated Poisson* untuk data kasus positif malaria di Jawa Timur tahun 2020 - 2022. Dari hasil nilai AIC yang diperoleh, terlihat bahwa regresi *Zero Inflated Poisson* adalah metode yang lebih baik digunakan dalam data jumlah kasus positif malaria. Berdasarkan hasil pengujian dalam penelitian, terdapat tujuh variabel prediktor yang memberikan pengaruh secara signifikan terhadap jumlah kasus positif malaria,

yaitu tingkat kepadatan penduduk (X1), tingkat kemiskinan (X2), persentase rumah tangga dengan keadaan sanitasi layak (X3), angka kesakitan atau *morbidity rate* (X4), persentase keluhan kesehatan (X5), jumlah sarana kesehatan (X6), dan persentase tempat umum yang sesuai standar kesehatan (X8).

## SARAN

Pemodelan terhadap kasus positif malaria tidak terbatas pada model regresi *Poisson* dan regresi *Zero Inflated Poisson* saja. Penanganan masalah overdispersi tidak hanya terbatas pada regresi *Zero Inflated Poisson* terdapat beberapa metode lain yang apabila diterapkan pada data mungkin diperoleh hasil yang lebih maksimal. Pada penelitian berikutnya dengan materi yang serupa dan menyesuaikan perkembangan dalam dunia kesehatan juga diharapkan dapat mempertimbangkan untuk menambahkan variabel prediktor yang lebih berkaitan dengan faktor kesehatan manusia yang kemungkinan lebih berpengaruh terhadap data kasus positif penyakit sehingga dapat memodelkan data menjadi lebih baik lagi dan mengetahui faktor lainnya yang berpengaruh.

## DAFTAR PUSTAKA

- Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. Hoboken, New Jersey: John Wiley & Sons.
- Achmad, N., Payu, M.R.F., dan Rahim, Y. (2022). Pemodelan Pneumonia Berat Menggunakan Regresi Zero Inflated Negative Binomial di Gorontalo. *Euler : Jurnal Ilmiah Matematika, Sains, dan Teknologi*, 10 (1) : 45 - 53. <https://doi.org/10.34312/euler.v10i1.13990>
- BPS Provinsi Jawa Timur. (2020). *Provinsi Jawa Timur Dalam Angka (Jawa Timur Province in Figures) 2020*. Surabaya : BPS Provinsi Jawa Timur.
- BPS Provinsi Jawa Timur. (2021). *Provinsi Jawa Timur Dalam Angka (Jawa Timur Province in Figures) 2021*. Surabaya : BPS Provinsi Jawa Timur.
- BPS Provinsi Jawa Timur. (2022). *Provinsi Jawa Timur Dalam Angka (Jawa Timur Province in Figures) 2022*. Surabaya : BPS Provinsi Jawa Timur.
- Bal, S. & Sodoudi, S. (2020). Modelling and Prediction of Dengue occurrences in Kolkata, India, based on climate factors. *International Journal of Biometeorology*. <https://doi.org/10.1007/s00484-020-01918-9>.
- Cahyandari, R. (2014). Pengujian Overdispersi Pada Model Regresi *Poisson*. *Statistika*, 14(2) : 69 - 76.
- Dinas Kesehatan Provinsi Jawa Timur. (2020). *Profil Kesehatan 2020*. [WWW.DINKES.JATIMPROV.GO.ID](http://WWW.DINKES.JATIMPROV.GO.ID)
- Dinas Kesehatan Provinsi Jawa Timur. (2021). *Profil Kesehatan 2021*. [WWW.DINKES.JATIMPROV.GO.ID](http://WWW.DINKES.JATIMPROV.GO.ID)
- Dinas Kesehatan Provinsi Jawa Timur. (2022). *Profil Kesehatan 2022*. [WWW.DINKES.JATIMPROV.GO.ID](http://WWW.DINKES.JATIMPROV.GO.ID)
- Gupta, P.L and Gupta, R.C. (2004). Modelling Count Data by Random Effect *Poisson* Model. *The india Journal of statistic*. pp 1 - 18.
- Hu, Qiuyan and Liu, Liu. (2021). Weighted Score test based EWMA control charts for Zero-Inflated *Poisson* Models. *Computers & Industrial Engineering*, 152.
- Jansakul, N. & J.P. Hinde. (2002). Score Tests for Zero-Inflated *Poisson* Models. *Computational Statistics & data Analysis*, 40 : 75 - 96.
- Kementerian Kesehatan RI. (2021). *Profil Kesehatan Indonesia 2021*. Jakarta : Kementerian Kesehatan Republik Indonesia.
- Kusuma, W., Komalasari, D., dan Hadijati, M. (2013). Model Regresi Zero Inflated *Poisson* pada Data *Overdispersion*. *Jurnal Matematika* 3 (2) : 71 - 85.
- Lambert, D. (1992). Zero-Inflated *Poisson* Regression with an Application to Defects in Manufacturing. *Technometrics*, 34(1) : 1 - 14.
- Obubu, M et al. (2018). "Modelling Count Data; A Generalized Linear Model Framework", *American Journal of Mathematics and Statistics*, 8(6).
- Santi, Vera M., Abi Wiyono, Sudarwanto, 2021. "Pemodelan Jumlah Kasus Malaria di Indoensia Menggunakan Generalized Linear Model". *Jurnal Statistika dan Aplikasinya*, 5 (1).
- Yau, K.K.W. and Lee, A.H. (2001). "Zero-inflated *Poisson* regression with random effect to evaluate an occupational injury prevention programme". *Statist. Med.*, 20: 2907-2920. <https://doi.org/10.1002/sim.860>