

PERBANDINGAN MODEL REGRESI ZERO-INFLATED POISSON (ZIP) DAN REGRESI ZERO-INFLATED NEGATIVE BINOMIAL (ZINB) TERHADAP KASUS DIFTERI DI PROVINSI JAWA TIMUR TAHUN 2024

Ismachul Diana Masrurroh

Program Studi Matematika, FMIPA, Universitas Negeri Surabaya
e-mail: ismachul.22049@mhs.unesa.ac.id

Danang Ariyanto

Program Studi Matematika, FMIPA, Universitas Negeri Surabaya
e-mail: danangariyanto@unesa.ac.id

Abstrak

Kasus difteri merupakan salah satu permasalahan kesehatan masyarakat yang masih menjadi perhatian di Indonesia, khususnya di Provinsi Jawa Timur. Data kasus difteri tahun 2024 menunjukkan karakteristik berupa banyaknya nilai nol (*excess zeros*) dan adanya *overdispersi*, sehingga model regresi Poisson tidak lagi sesuai untuk digunakan. Oleh karena itu, penelitian ini bertujuan untuk membandingkan model regresi *Zero-Inflated Poisson* (ZIP) dan *Zero-Inflated Negative Binomial* (ZINB) dalam memodelkan kasus difteri serta menentukan model terbaik berdasarkan kriteria *Akaike's Information Criterion* (AIC). Data yang digunakan merupakan data sekunder kasus difteri di 38 Kabupaten/Kota di Provinsi Jawa Timur tahun 2024, dengan variabel prediktor meliputi persentase pemberian ASI eksklusif, persentase balita dipantau pertumbuhan dan perkembangan, persentase pelayanan kesehatan bayi, dan persentase imunisasi DPT-HB-Hib3. Hasil analisis menunjukkan bahwa data memiliki *excess zeros* sebesar 76,32% dan mengalami *overdispersi* dengan nilai dispersi sebesar 1,441615 (>1). Hasil pemodelan menunjukkan bahwa variabel persentase pelayanan kesehatan bayi dan persentase imunisasi DPT-HB-Hib3 berpengaruh signifikan terhadap jumlah kasus difteri. Berdasarkan perbandingan nilai AIC, model Regresi *Zero-Inflated Poisson* (ZIP) menghasilkan nilai AIC yang lebih kecil dibandingkan model Regresi *Zero-Inflated Negative Binomial* (ZINB), sehingga dipilih sebagai model terbaik dalam memodelkan kasus difteri di Provinsi Jawa Timur tahun 2024. Penelitian ini diharapkan dapat menjadi referensi dalam pemilihan model statistik yang tepat untuk data kesehatan dengan karakteristik *excess zeros* dan *overdispersi* guna mendukung pengambilan kebijakan pengendalian penyakit yang lebih akurat.

Kata kunci: Difteri, *Zero-Inflated Poisson* (ZIP), *Zero-Inflated Negative Binomial* (ZINB), *excess zeros*, *overdispersi*, AIC

Abstract

Diphtheria cases remain a public health concern in Indonesia, particularly in East Java Province. Data on diphtheria cases in 2024 exhibit characteristics such as an abundance of zeros (excess zeros) and the presence of overdispersion, rendering the Poisson regression model unsuitable for use. Therefore, this study aims to compare the Zero-Inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB) regression models in modeling diphtheria cases and to determine the best model based on the Akaike's Information Criterion (AIC). The data used consists of secondary data on diphtheria cases in 38 districts/cities in East Java Province in 2024, with predictor variables including the percentage of exclusive breastfeeding, the percentage of toddlers monitored for growth and development, the percentage of infant health services, and the percentage of DPT-HB-Hib3 immunizations. The analysis results showed that the data had an excess of zeros of 76.32% and exhibited overdispersion with a dispersion value of 1.441615 (>1). The modeling results indicate that the variables of the percentage of infant healthcare services and the percentage of DPT-HB-Hib3 immunization have a significant effect on the number of diphtheria cases. Based on a comparison of AIC values, the Zero-Inflated Poisson (ZIP) Regression model yields a smaller AIC value compared to the Zero-Inflated (ZINB) model, so it was selected as the best model for modeling diphtheria cases in East Java Province in 2024. This study is expected to serve as a reference in selecting the appropriate statistical model for health data with characteristics of excess zeros and overdispersion to support more accurate disease control policy-making.

Keywords: *Diphtheria, Zero-Inflated Poisson (ZIP), Zero-Inflated Negative Binomial (ZINB), excess zeros, overdispersion, AIC.*

PENDAHULUAN

Difteri merupakan penyakit infeksi menular yang disebabkan oleh *Corynebacterium diphtheriae*, yakni bakteri penghasil toksin. Penyakit ini ditularkan melalui *droplet* pernapasan saat individu terinfeksi batuk atau bersin, sehingga memungkinkan transmisi langsung antarmanusia. Infeksi dapat terjadi baik pada individu simptomatik maupun asimtomatik, yang tetap berpotensi sebagai sumber penularan. Spektrum klinisnya bervariasi, mulai dari bentuk ringan hingga berat, dengan risiko komplikasi serius dan kemungkinan berakhir dengan kematian. Kelompok yang paling rentan adalah anak-anak yang belum mendapatkan vaksinasi lengkap. Gejala umum pada penyakit ini antara lain demam, sakit tenggorokan, dan pembengkakan kelenjar leher (World Health Organization, 2025).

Data difteri Jatim 2024 menunjukkan karakteristik data yang unik di mana 38 Kabupaten/Kota, terdapat 29 wilayah (76,3%) melaporkan nol kasus (Dinas Kesehatan Provinsi Jawa Timur, 2024). Menurut penelitian yang dilakukan oleh Husada dkk (2024), selama periode 11 tahun terdapat 4.009 kasus. Dalam kasus difteri yang ada di Jawa Timur tercatat lima daerah dengan jumlah kasus terbanyak yang ditemukan meliputi Kota Sidoarjo, Kota Surabaya, Kabupaten Blitar, Kota Malang, dan Kabupaten Malang, yang diperkirakan bahwa sebagian besar korban tidak mendapatkan imunisasi secara lengkap (Husada dkk., 2024). Faktor-faktor lain yang dapat menyebabkan difteri adalah Persentase Cakupan Imunisasi Dasar, Jumlah Balita Gizi Kurang, Imunisasi DPT-HB-Hib3, dan Kepadatan Penduduk (Tasya Azzahra, 2023).

Data yang hanya menunjukkan overdispersi ringan dengan dominasi jumlah nilai nol yang tinggi cenderung lebih efisien dimodelkan menggunakan *Zero-Inflated Poisson* (ZIP). Namun, jika overdispersi tetap terjadi secara signifikan setelah *excess zeros* ditangani, maka model *Zero-Inflated Negative Binomial* (ZINB) menjadi alternatif yang lebih fleksibel karena memiliki parameter dispersi tambahan (Greene, 1994). Model ZIP dan ZINB dipilih bukan hanya karena datanya banyak nol, tapi karena secara epidemiologi, kasus difteri di Jawa Timur memiliki dua status wilayah yakni wilayah dengan kekebalan kelompok (*structural zero*) dan wilayah rentan yang belum terinfeksi (*sampling zero*) (Lambert, 1992).

Penelitian terbaru oleh Salsabila dkk. (2025)

pada kasus pneumonia menunjukkan bahwa ZINB sangat efektif dalam memberikan estimasi parameter yang presisi pada data kesehatan dengan sebaran nol yang tinggi (Salsabila dan Efendi, 2025). Meskipun model ZINB sering dianggap superior dalam menangani overdispersi, kompleksitas model tersebut tidak selalu menjamin kesesuaian terbaik. Studi literatur oleh Al Farisyi (2020) pada kasus tanah longsor di Kota Semarang membuktikan bahwa model *Zero-Inflated Poisson* (ZIP) justru menghasilkan nilai *Akaike's Information Criterion* (AIC) yang lebih kecil dibandingkan ZINB. Temuan ini mengindikasikan bahwa ketika overdispersi pada data sepenuhnya disebabkan oleh tingginya proporsi nilai nol (*excess zeros*), model ZIP yang lebih sederhana dan memiliki stabilitas estimasi yang lebih baik tanpa risiko model statistik memiliki kompleksitas yang berlebihan (*over-fitting*) (Al Farisyi, 2020).

Perbedaan karakteristik data yang mengalami *excess zeros* dan overdispersi menyebabkan pemilihan model antara ZIP dan ZINB masih menjadi pertimbangan dalam penelitian statistik. Ketidaktepatan dalam menentukan model yang sesuai berpotensi menghasilkan estimasi parameter yang kurang akurat, sehingga dapat memengaruhi interpretasi faktor-faktor yang berhubungan dengan kasus difteri. Oleh karena itu, penelitian ini dilakukan untuk membandingkan performa model Regresi *Zero-Inflated Poisson* (ZIP) dan *Zero-Inflated Negative Binomial* (ZINB) pada kasus difteri di Jawa Timur tahun 2024, guna menjamin bahwa strategi pengendalian kasus penyakit didasarkan pada landasan statistik yang paling akurat dan dapat dipertanggungjawabkan secara ilmiah.

KAJIAN TEORI

GENERALIZED LINEAR MODEL

Generalized Linear Model (GLM) adalah suatu pengembangan dari model regresi linear klasik yang digunakan untuk menggambarkan variabel respon yang tidak selalu mengikuti distribusi normal seperti data jumlah (Santi dkk., 2021). Persamaan GLM memiliki bentuk umum sebagai berikut:

$$g(\mu_i) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} = \mathbf{x}_i^T \boldsymbol{\beta} \quad (1)$$

EXCESS ZEROS

Salah satu kendala utama dalam analisis regresi data cacah adalah fenomena kelebihan nol (*excess*

zeros), yaitu kondisi di mana proporsi nilai nol pada variabel respon jauh melebihi ekspektasi distribusi teoritis. (Margaretha dkk., 2019). Perhitungan persentase nol sebagai berikut:

$$\text{Persentase Nol} = \frac{\text{Jumlah obsevasi 0}}{\text{Banyak observasi}} \times 100\% \quad (2)$$

OVERDISPERSI

Parameter dispersi dalam beberapa literatur dinotasikan dengan simbol ϕ (phi) (Mustafidah, 2024), dengan langkah-langkah pengujian overdispersi adalah sebagai berikut:

- $H_0: \phi \leq 1$, suatu data tidak mengalami overdispersi
- $H_1: \phi > 1$, suatu data mengalami overdispersi

Uji statistik untuk menguji nilai overdispersi menggunakan nilai *pearson chi square* dan *deviance* (df) dengan rumus sebagai berikut:

$$\phi = \frac{x^2}{\text{derajat bebas}} \quad (3)$$

Jika nilai *disperse* $\phi > 1$ maka suatu data mengalami overdispersi (Mustafidah dan Sofro, 2024).

REGRESI ZERO-INFLATED POISSON (ZIP)

Model *Zero-Inflated Poisson* (ZIP) adalah model campuran yang diperkenalkan oleh Lambert (1992) untuk menangani data jumlah kejadian (*count data*) dengan *excess zeros*. Berbeda dengan distribusi Poisson standar, model *Zero-Inflated Poisson* (ZIP) mengasumsikan bahwa setiap observasi dapat berasal dari salah satu dari dua proses pembentuk data:

- Keadaan Nol Sempurna (*Structural Zeros*): Keadaan di mana observasi pasti bernilai 0 dengan probabilitas π_i .
- Keadaan Poisson (*Count State*): Keadaan di mana observasi mengikuti distribusi Poisson standar dengan rata-rata μ_i dan probabilitas $(1 - \pi_i)$.

Peluang terjadinya nilai y_i pada variabel respon Y_i didefinisikan sebagai berikut (Lambert, 1992):

$$P(Y = y_i) = \begin{cases} \pi_i + (1 - \pi_i) e^{-\mu_i}, & y_i = 0 \\ (1 - \pi_i) \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, & y_i > 0 \end{cases} \quad (4)$$

Fungsi probabilitas regresi Zero-Inflated Poisson tersebut akan di estimasi menggunakan *Maximum Log-Likelihood Estimation* (MLE) menjadi fungsi sebagai berikut (Saputro dan Qudratullah, 2021):

$$\begin{aligned} \ell(\beta, \gamma) &= \ln L(\beta, \gamma) \\ &= \sum_{y_i=0}^n \ln(e^{\mathbf{x}_i^T \gamma} + e^{-e^{\mathbf{x}_i^T \beta}}) \\ &\quad - \sum_{i=1}^n \ln(1 + e^{\mathbf{x}_i^T \gamma}) \\ &\quad + \sum_{y_i>0} (y_i \mathbf{x}_i^T \beta - e^{\mathbf{x}_i^T \beta} - \ln(y_i!)) \end{aligned} \quad (5)$$

REGRESI ZERO-INFLATED NEGATIVE BINOMIAL (ZINB)

Penelitian yang dilakukan oleh Hilbe (2011), bahwa *Zero-Inflated Negative Binomial* (ZINB) adalah perluasan dari distribusi *Negative Binomial* yang digunakan untuk menangani data cacah dengan overdispersi dan data kelebihan nol (*excess zeros*) (Hilbe, 2011). Dalam hal ini Model ZINB diasumsikan menjadi dua model berbeda yakni:

- Model *Negative Binomial State* yaitu model dengan probabilitas $(1 - \pi_i)$ dan berdistribusi *negative binomial* dengan mean (μ_i).
- Model *Zero Inflated State* yaitu model dengan probabilitas π_i yang hanya menghasilkan observasi bernilai nol.

Fungsi probabilitas untuk ZINB sebagai berikut (Saputro dan Qudratullah, 2021):

$$P(y_i) = \begin{cases} \pi_i + (1 - \pi_i)(1 + \phi\mu_i)^{-\frac{1}{\phi}}, & y_i = 0 \\ (1 - \pi_i) \frac{\Gamma(y_i + \frac{1}{\phi})}{\Gamma(\frac{1}{\phi}) y_i!} \left(\frac{\phi\mu_i}{1 + \phi\mu_i}\right)^{y_i} (1 + \phi\mu_i)^{-\frac{1}{\phi}}, & y_i > 0 \end{cases} \quad (6)$$

Fungsi probabilitas regresi Zero-Inflated Poisson tersebut akan di estimasi menggunakan *Maximum Log-Likelihood Estimation* (MLE) menjadi fungsi sebagai berikut (Saputro dan Qudratullah, 2021):

$$\begin{aligned} \ell(\theta|y_i) &= \sum_{i=1}^n \left[I_{(y_i=0)} \ln \left(\frac{e^{\mathbf{x}_i^T \gamma}}{1 + e^{\mathbf{x}_i^T \gamma}} + \left(\frac{1}{1 + e^{\mathbf{x}_i^T \gamma}} \right) (1 + \phi\mu_i)^{-\frac{1}{\phi}} \right) \right. \\ &\quad \left. + \sum_{y_i>0} \ln \left[\frac{1}{1 + e^{\mathbf{x}_i^T \gamma}} \frac{\Gamma(y_i + \frac{1}{\phi})}{\Gamma(\frac{1}{\phi}) y_i!} \left(\frac{\phi\mu_i}{1 + \phi\mu_i} \right)^{y_i} (1 + \phi\mu_i)^{-\frac{1}{\phi}} \right] \right] \end{aligned} \quad (7)$$

UJI MULTIKOLINEARITAS

Multikolinieritas dalam data dapat menggunakan perhitungan nilai *Variance Inflation Factor* (VIF). Jika nilai VIF lebih dari 10, maka dapat dikatakan bahwa terjadi multikolinieritas yang tinggi (Montgomery dkk., 2021). Nilai VIF dapat dinyatakan sebagai berikut:

$$VIF_j = \frac{1}{1 - R_j^2}, j = 1, 2, 3, \dots, p \quad (8)$$

Melalui pengujian kriteria sebagai berikut:

- Nilai $VIF > 10$ maka H_0 ditolak, artinya terdapat multikolinearitas
- Nilai $VIF \leq 10$ maka H_0 gagal menolak, artinya tidak terdapat multikolinearitas (Montgomery dkk., 2021)

UJI SIMULTAN

Pada pengujian simultan menggunakan uji *likelihood* ratio yang digunakan untuk mengetahui apakah variabel bebas mempengaruhi variabel terikat secara signifikan (Agresti, 2015). Hipotesis yang digunakan:

1. $H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$
2. H_1 : paling sedikit ada satu $\beta_j \neq 0$ dimana $j=1,2,\dots,p$

Statistik Uji G dinyatakan sebagai berikut:

$$G = -2 \ln \left[\frac{L_0}{L_1} \right] = -2(\ln L_0 - \ln L_1) \quad (9)$$

UJI PARSIAL

Uji parsial bertujuan untuk menguji signifikansi pengaruh variabel prediktor secara individu terhadap variabel respon, baik pada komponen model *count* maupun komponen *zero-inflation* (Agresti, 2015),

Statistik uji Wald (Z) didefinisikan sebagai rasio antara nilai estimasi parameter dengan galat bakunya (*standard error*). Untuk menguji hipotesis :

1. $H_0: \beta_j = 0$ (variabel prediktor ke- j tidak signifikan)
2. $H_1: \beta_j \neq 0$ (variabel prediktor ke- j signifikan)

dengan menggunakan statistik uji Wald sebagai berikut (Agresti, 2015):

$$Z = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (10)$$

METODE

DATA PENELITIAN

Data yang digunakan dalam kajian penelitian ini merupakan data sekunder mengenai kasus difteri di Provinsi Jawa Timur tahun 2024. Ruang lingkup dalam penelitian ini meliputi 29 Kabupaten dan 9 Kota yang ada di Provinsi Jawa Timur. Variabel respon pada penelitian ini adalah Jumlah Kasus Difteri di Provinsi Jawa Timur tahun 2024. Sedangkan, variabel prediktor pada penelitian ini meliputi persentase pemberian ASI eksklusif,

persentase balita dipantau pertumbuhan dan perkembangan, persentase pelayanan kesehatan bayi di puskesmas dan persentase pemberian Imunisasi DPT-Hb-Hib3.

TAHAPAN PENELITIAN

Adapun tahapan penelitian sebagai berikut:

1. Menentukan rumusan masalah, tujuan, manfaat penelitian dan batasan masalah hingga literatur ruang lingkup penelitian.
2. Mengumpulkan data penelitian yang berdasarkan data sekunder dari Badan Pusat Statistik (BPS) Jatim dan Dinas Kesehatan Provinsi Jawa Timur.
3. Menganalisis deskriptif variabel respon dan variabel prediktor.
4. Melakukan Uji Multikolinieritas terhadap variabel prediktor.
5. Melakukan Uji overdispersi dan excess zeros pada variabel respon.
6. Membangun model ZIP dan ZINB dengan estimasi parameter MLE, serta melakukan uji simultan dan uji parsial untuk mengetahui ketepatan model dan variabel prediktor yang signifikan terhadap model.
7. Membandingkan model terbaik dengan menggunakan nilai AIC terkecil sebagai model terbaik dalam penelitian ini.

HASIL DAN PEMBAHASAN

ANALISIS DESKRIPTIF

Analisis deskriptif menyajikan ukuran statistik seperti nilai rata-rata (*mean*), varians, simpangan baku (standar deviasi), nilai maksimum (*max*), dan nilai minimum (*min*) untuk memberikan gambaran umum mengenai pola dan sebaran data. Berikut disajikan hasil analisis deskriptif dari variabel-variabel dalam penelitian ini:

Tabel 1 Analisis Deskriptif

Var	Mean	Varians	sd.	Max	Min
y	0,4211	0,791	0,889	3	0
x_1	82,00	74,8	8,65	95,60	63,80
x_2	91,13	104	10,2	100,00	60,20
x_3	83,06	143	12,0	100,00	62,80
x_4	79,27	138	11,9	100,00	60,10

Pada Tabel 1 diatas diketahui bahwa Nilai varians sebesar 0,791 lebih besar dibandingkan nilai rata-rata ($0,791 > 0,4211$). Kondisi ini menunjukkan bahwa data mengalami overdispersi. Sedangkan pada variabel prediktor meliputi variabel persentase pemberian ASI eksklusif x_1 memiliki rata-rata sebesar

82,0 dengan simpangan baku 8,65, menunjukkan penyebaran data yang relatif merata antarwilayah. Variabel persentase balita yang dipantau pertumbuhan dan perkembangannya x_2 memiliki rata-rata 91,1 dan simpangan baku 10,2 yang mengindikasikan adanya variasi capaian antarwilayah. Variabel persentase pelayanan kesehatan bayi x_3 memiliki rata-rata 83,1 dengan simpangan baku 12,0, sedangkan variabel persentase imunisasi lanjutan DPT-HB-Hib 3 x_4 memiliki rata-rata 79,3 dengan simpangan baku 11,8. Besarnya simpangan baku pada x_2 , x_3 dan x_4 menunjukkan bahwa capaian indikator kesehatan bayi dan balita di Provinsi Jawa Timur masih bervariasi antar kabupaten/kota.

MULTIKOLINIERITAS

Uji multikolinieritas dapat diketahui melalui indikator nilai *Variance Inflation Factor* (VIF), akan ditunjukkan hasil sebagai berikut:

Tabel 2 Uji Multikolinearitas

Variabel Prediktor	Hasil VIF
x_1	1,324548
x_2	1,500499
x_3	2,902698
x_4	2,946482

Pada Tabel 2, diketahui nilai VIF untuk x_1 hingga x_4 nilainya kurang dari 10. Sehingga, dapat diartikan pada setiap variabel prediktor tidak terjadi multikolinearitas.

EXCESS ZEROS

Pengecekan bisa dilakukan dengan menghitung nilai presentasi nol sebagai berikut:

$$\begin{aligned}\text{Persentase Nol} &= \frac{\text{Jumlah obsevasi 0}}{\text{Banyak observasi}} \times 100\% \\ &= \frac{29}{38} \times 100\% = 76,32\%\end{aligned}$$

Penelitian data yang diperoleh, terdapat 29 dari 38 pengamatan yang bernilai nol atau sebesar 76,32%. Kondisi tersebut menunjukkan adanya indikasi kelebihan nol (*excess zeros*) atau karakteristik data termasuk kedalam (*zero-inflation*).

OVERDISPERSI

Statistik uji overdispersi yang digunakan adalah *deviance* dan *pearson chi-square*. Pengujian overdispersi diringkas pada tabel berikut

Tabel 3 Uji Overdispersi

Nilai Rata-rata Variabel Respon	:0,4211
Nilai Varians	:0,7909
Nilai dispersi	:1,441615

Pada Tabel 3, diperoleh nilai dispersi sebesar 1,441615. Karena nilai dispersi yang diperoleh lebih besar dari satu ($1,441615 > 1$), maka keputusan yang diambil adalah menolak H_0 . Hal tersebut menunjukkan variabel respon (Y) mengalami overdispersi. Oleh karena itu diperlukan model alternatif yang mampu mengakomodasi kondisi overdispersi, seperti model ZIP dan ZINB.

REGRESI ZERO-INFLATED POISSON (ZIP)

Penanganan fenomena kelebihan nol (*excess zeros*) pada data cacahan, digunakan regresi ZIP sebagai pengembangan dari model Poisson. Hasil penaksiran parameter pada regresi *Zero-Inflated Poisson* (ZIP) dengan menggunakan MLE, akan ditunjukkan pada tabel berikut:

Tabel 4 Hasil Regresi ZIP

Prediktor	Estimasi	std. error	p-value
<i>a. Poisson State</i>			
<i>Intercept</i>	-8,33673	6,28613	0,18477
β_1	0,012195	0,03826	0,7349
β_2	0,06709	0,05676	0,2372
β_3	-0,13166	0,04207	0,00175
β_4	-0,13528	0,03960	0,00063
<i>b. Zero-Inflated</i>			
<i>Intercept</i>	6,055	3118,485	0,998
γ_1	-5,524	63,606	0,931
γ_2	-12,357	134,395	0,927
γ_3	24,374	295,518	0,934
γ_4	-7,641	100,050	0,939

Hasil regresi ZIP pada Tabel (4), dapat dibentuk model ZIP sebagai berikut:

a. Model Poisson State

$$\ln(\mu) = -8,33673 + 0,012195x_1 + 0,06709x_2 + 0,13166x_3 - 0,13528x_4$$

b. Model Zero-Inflated State

$$\text{logit}(\pi_i) = 6,055 - 5,524x_1 - 12,357x_2 + 24,374x_3 - 7,641x_4$$

Ringkasan hasil ZIP tersebut menunjukkan variabel yang memiliki koefisien positif akan cenderung meningkatkan jumlah kasus difteri, sedangkan pada variabel yang memiliki koefisien negatif cenderung menurunkan rata-rata jumlah kasus

difteri.

UJI SIMULTAN PADA ZIP

Pengujian signifikansi parameter secara simultan dilakukan dengan metode *Likelihood Ratio Test* (uji G). Dengan hasil sebagai berikut:

Tabel 5 Hasil Uji Simultan ZIP

	Log-likelihood	Df	<i>p-value</i>
Model penuh	-23,932	10	0,05626

Nilai Uji G pada Tabel 5 sebesar 15,152 dengan df sebesar 10 dan nilai *p-value* sebesar 0,05626 > 0,10. Sehingga, diputuskan menolak H_0 . Maka dari itu, dapat dipastikan terdapat minimal satu variabel prediktor yang signifikan terhadap variabel respon.

UJI PARSIAL PADA ZIP

Pengujian parsial menggunakan uji statistik Wald didapatkan nilai sebagai berikut:

1. Parameter β

Pada Tabel (4) didapatkan bahwa nilai *p-value* dari persentase pelayanan kesehatan bayi (x_3) adalah $0,001752 < \alpha = 0,10$ dan persentase imunisasi lanjutan DPT-HB-Hib 3 (x_4) adalah $0,000635 < \alpha = 0,10$. Berdasarkan hasil perhitungan tersebut didapatkan variabel (x_3) dan (x_4) berpengaruh secara signifikan terhadap jumlah kasus difteri di Provinsi Jawa Timur tahun 2024.

2. Parameter γ

Pada Tabel (4) diketahui bahwa tidak terdapat variabel prediktor yang memiliki nilai *p-value* < 0,10. Maka diputuskan bahwa menerima H_0 , yang berarti pada parameter γ tidak terdapat variabel prediktor yang signifikan terhadap kasus difteri.

INTERPRETASI HASIL ZIP

Pada model *Poisson State* yang dihasilkan, akan diinterpretasikan mengenai dampak setiap variabel prediktor yang memiliki pengaruh signifikan terhadap variabel respon sebagai berikut:

- Nilai koefisien persentase pelayanan kesehatan bayi (x_3) adalah 0,13166, sehingga dapat diinterpretasikan bahwa apabila terjadi peningkatan persentase pelayanan kesehatan bayi 1% dapat menyebabkan

terjadinya peningkatan kasus difteri sebesar $e^{(0,13166)} = 1,1408$ kali.

- Nilai koefisien imunisasi lanjutan DPT-HB-Hib3 (x_4) adalah -0,13528, sehingga dapat diinterpretasikan bahwa apabila terjadi peningkatan persentase imunisasi lanjutan DPT-HB-Hib3 1% dapat menyebabkan terjadinya penurunan kasus difteri sebesar $e^{(-0,13528)} = 0,8734$ kali

Pada hasil dari model *Zero-Inflated State* dihasilkan bahwa pada uji signifikansi parameter secara parsial menunjukkan tidak terdapat variabel prediktor yang signifikan.

REGRESI ZERO-INFLATED NEGATIVE BINOMIAL

Regresi ZINB merupakan salah satu regresi yang diterapkan untuk menganalisis data hitung (*count*) yang mengalami overdispersi dan jumlah nol yang berlebihan (*excess zeros*). Estimasi parameter ZINB dilakukan menggunakan MLE, dengan hasil yang disajikan pada tabel berikut:

Tabel 6 Hasil Regresi ZINB

Prediktor	Estimasi	std. error	<i>p-value</i>
<i>a. Negative Binomial State</i>			
<i>Intercept</i>	-7,66567	6,09063	0,20884
β_1	0,01005	0,03807	0,79181
β_2	0,016173	0,05511	0,262692
β_3	0,13027	0,04198	0,001915
β_4	-0,13284	0,03907	0,000674
<i>b. Zero-Inflated</i>			
<i>Intercept</i>	7,607	9683,71	0,999
γ_1	-6,162	166,025	0,970
γ_2	-13,410	213,669	0,950
γ_3	26,507	619,886	0,966
γ_4	-8,205	197,746	0,967

Hasil regresi ZINB pada Tabel (6), dapat dibentuk model ZIP sebagai berikut:

- Model *Negative Binomial State*

$$\ln(\mu_i) = -7,66567 + 0,01005x_1 + 0,06173x_2 + 0,13027x_3 - 0,13284x_4$$

- Model *Zero-Inflated State*

$$\text{logit}(\pi_i) = 7,607 - 6,162x_1 - 13,410x_2 + 26,507x_3 - 8,205x_4$$

Ringkasan hasil ZINB tersebut menunjukkan

variabel yang memiliki koefisien positif akan cenderung meningkatkan jumlah kasus difteri, sedangkan pada variabel yang memiliki koefisien negatif cenderung menurunkan rata-rata jumlah kasus difteri.

UJI SIMULTAN PADA ZINB

Pengujian signifikansi parameter secara simultan dilakukan dengan metode *Likelihood Ratio Test* (uji G). Dengan hasil sebagai berikut:

Tabel 7 Hasil Uji Simultan ZINB

	Log-likelihood	Df	<i>p-value</i>
Model penuh	-23,926	11	0,05603

Nilai Uji G model ZINB pada Tabel 7 sebesar 15,164 dengan df sebesar 11 serta nilai *p-value* sebesar $0,05603 < 0,10$, maka keputusan yang diambil adalah menolak H_0 . Maka dari itu, dapat dipastikan terdapat minimal satu variabel prediktor yang signifikan terhadap variabel respon.

UJI PARSIAL PADA ZINB

Pengujian parsial menggunakan uji statistik Wald didapatkan nilai sebagai berikut:

1. Parameter β

Pada Tabel (6) didapatkan nilai *p-value* dari persentase pelayanan kesehatan bayi (x_3) adalah $0,001915 < 0,10$ dan persentase imunisasi lanjutan DPT-HB-Hib 3 (x_4) adalah $0,000674 < 0,10$. Karena terdapat variabel prediktor yang memiliki nilai *p-value* $< \alpha$, maka keputusannya adalah menolak H_0 . Sehingga, variabel (x_3) dan (x_4) berpengaruh secara signifikan terhadap kasus difteri di Provinsi Jawa Timur tahun 2024.

2. Parameter γ

Pada Tabel (6) diketahui bahwa tidak terdapat variabel prediktor yang memiliki nilai *p-value* $< 0,10$. Maka diputuskan bahwa menerima H_0 , yang berarti pada parameter γ tidak terdapat variabel prediktor yang signifikan terhadap kasus difteri.

INTERPRETASI HASIL ZINB

Pada model *Negative Binomial State* yang dihasilkan, akan diinterpretasikan mengenai dampak setiap variabel prediktor yang memiliki pengaruh signifikan terhadap variabel respon sebagai berikut:

- Nilai koefisien persentase pelayanan kesehatan bayi (x_3) adalah 0,13027, sehingga dapat diinterpretasikan bahwa apabila terjadi peningkatan persentase pelayanan

kesehatan bayi 1% dapat menyebabkan terjadinya peningkatan kasus difteri sebesar $e^{(0,13027)} = 1,1392$ kali.

- Nilai koefisien imunisasi lanjutan DPT-HB-Hib3 (x_4) adalah -0,13284, sehingga dapat diinterpretasikan bahwa apabila terjadi peningkatan persentase imunisasi lanjutan DPT-HB-Hib3 1% dapat menyebabkan terjadinya penurunan kasus difteri sebesar $e^{(-0,13284)} = 0,8756$ kali.

Pada hasil dari model *Zero-Inflated State* dihasilkan bahwa pada uji signifikansi parameter secara parsial menunjukkan tidak terdapat variabel prediktor yang signifikan.

PEMILIHAN MODEL TERBAIK

Proses pemodelan melibatkan dua jenis regresi, yaitu ZIP dan ZINB. Dari kedua regresi ini, akan dilakukan identifikasi model terbaik berdasarkan nilai AIC dalam tabel berikut:

Tabel 8 Hasil Uji Simultan ZINB

Model Regresi	Nilai AIC
<i>Zero-Inflated Poisson</i> (ZIP)	67,8647
<i>Zero-Inflated Negative Binomial</i> (ZINB)	69,8525

Pada Tabel (8) didapatkan nilai AIC untuk masing-masing model di mana regresi ZIP lebih kecil dibandingkan dengan regresi ZINB yaitu $67,8647 < 69,8525$. Dengan demikian, diperoleh bahwa model regresi ZIP merupakan pemodelan yang lebih baik untuk menganalisis kasus difteri di Provinsi Jawa Timur tahun 2024.

PENUTUP

KESIMPULAN

Hasil Model ZIP dan ZINB dihasilkan sebagai berikut:

a. Model Regresi ZIP

1. Model Poisson State

$$\ln(\mu) = -8,33673 + 0,012195x_1 + 0,06709x_2 + 0,13166x_3 - 0,13528x_4$$

2. Model Zero-Inflated State

$$\text{logit}(\pi_i) = 6,055 - 5,524x_1 - 12,357x_2 + 24,374x_3 - 7,641x_4$$

Hasil pengujian simultan menunjukkan nilai statistik G sebesar 15,152 (df=10) dengan *p-value* sebesar $0,05626 < \alpha = 0,10$, sehingga regresi ZIP dinyatakan layak untuk memodelkan data kasus

difteri. Hal ini mengindikasikan adanya pengaruh signifikan dari minimal satu variabel prediktor terhadap variabel respons. Secara parsial, pada *Poisson State*, variabel persentase pelayanan kesehatan bayi (x_3), dan persentase Imunisasi DPT-Hb-Hib3 di Jawa Timur (x_4). Sebaliknya, tidak ditemukan yang variabel prediktor yang berpengaruh signifikan pada komponen *Zero-Inflated* terhadap kasus difteri di Jawa Timur tahun 2024

b. Model Regresi ZINB

1. Model Negative Binomial State

$$\ln(\mu_i) = -7,66567 + 0,01005x_1 + 0,06173x_2 + 0,13027x_3 - 0,13284x_4$$

2. Model Zero-Inflated State

$$\text{logit}(\pi_i) = 6,055 - 5,524x_1 - 12,357x_2 + 24,374x_3 - 7,641x_4$$

Hasil pengujian simultan menunjukkan nilai statistik G sebesar 15,164 (df=11) dengan *p-value* sebesar $0,05603 < \alpha = 0,10$, sehingga regresi ZINB dinyatakan layak untuk memodelkan data kasus difteri. Hal ini mengindikasikan adanya pengaruh signifikan dari minimal satu variabel prediktor terhadap variabel respons. Secara parsial, pada *Poisson State*, variabel persentase pelayanan kesehatan bayi (x_3), dan persentase Imunisasi DPT-Hb-Hib3 di Jawa Timur (x_4). Sebaliknya, tidak ditemukan yang variabel prediktor yang berpengaruh signifikan pada komponen *Zero-Inflated* terhadap kasus difteri di Jawa Timur tahun 2024.

Model regresi ZIP terpilih sebagai model terbaik dalam penelitian ini karena memiliki nilai AIC terkecil (67,8647) dan mampu menangani data kasus difteri di Jawa Timur tahun 2024 secara lebih stabil dibandingkan model ZINB, meski terdapat overdispersi ringan. Melalui model ini, ditemukan dua variabel yang signifikan terhadap kasus difteri, yaitu persentase pelayanan kesehatan bayi (x_3), serta cakupan imunisasi DPT-Hb-Hib3 di Jawa Timur (x_4).

SARAN

Penelitian selanjutnya disarankan untuk melakukan komparasi model selain model *Zero-Inflated* seperti model *Hurdle* atau pendekatan regresi lainnya yang mampu menangani karakteristik *overdispersi* dan *excess zeros*. Selain itu, penambahan variabel prediktor yang lebih merepresentasikan faktor kesehatan, lingkungan, dan sosial ekonomi juga perlu dipertimbangkan agar model dapat menjelaskan variasi kasus secara lebih komprehensif serta mengidentifikasi faktor-faktor yang

berpengaruh signifikan dengan lebih akurat.

DAFTAR PUSTAKA

- Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. John Wiley & Sons, INC, 473.
- Ariyanto, D., Sofro, A. Y., Maulana, D. A., Oktaviarina, A., & Romadhonia, R. W. (2022). Analisis Statistika untuk Penelitian Dasar dengan Menggunakan R Studio. *Mitra Edukasi dan Publikasi*.
- Dinas Kesehatan Provinsi Jawa Timur. (n.d.). *Profil Kesehatan Provinsi Jawa Timur 2024*.
- Dr. Dominicus Husada, dr., DTM&H., MCTM (TP)., S. A. (2024). *Laporan Kasus Difteri di Jawa Timur selama 11 Tahun*. Universitas Airlangga.
- Greene, W. H. (1994). *Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models*. 74(1934), 535–546.
- Hilbe. (2011). *Negative Binomial Regression*.
- Kasim, F., & Hendra. (2023). Pengaruh Pengangguran dan Kemiskinan Terhadap Tindakan Kriminal di Kabupaten Tolitoli Periode 2021. *Jurnal Ilmu Ekonomi (JIE)*, 5(2), 247–261.
- Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1–14. <https://doi.org/10.2307/1269547>
- Margaretha, C. E., Ispriyanti, D., & Widiyari, T. (2019). Pemodelan Regresi Hurdle Poisson Dalam Mengatasi Excess Zeros Untuk Kasus Penyakit Tetanus Neonatorum Pada Neonatal Di Jawa Timur. *Jurnal Gaussian*, 8(3), 389–397. <https://doi.org/10.14710/j.gauss.v8i3.26683>
- Montgomery, d. C., peck, e. A., & vining, g. G. (2021). *Introduction To linier Regression*.
- Mustafidah, M. E. (2024). *MATH unesa*. 12(02), 428–438.
- Salman Al Farisyi, M. (2020). *Perbandingan Regresi Zero Inflated Poisson (ZIP) dan Regresi Zero-Inflated Negative Binomial (ZINB) pada Kasus Tanah Longsor di Kota Semarang*. repository.unimus.ac.id. <http://repository.unimus.ac.id/id/eprint/4185>
- Santi, V. M., Wiyono, A., & Sudarwanto. (2021). Pemodelan Jumlah Kasus Malaria di Indonesia Menggunakan Generalized Linear Model. *Jurnal Statistika Dan Aplikasinya*, 5(1), 112–120. <https://doi.org/10.21009/jsa.05111>
- Saputro, M. I. A., & Qudratullah, M. F. (2021). Estimation of Zero-Inflated Negative Binomial Regression Parameters Using the Maximum Likelihood Method (Case Study: Factors Affecting

Infant Mortality in Wonogiri in 2015). *Proceeding International Conference on Science and Engineering*, 4(February), 240-254.

Tasya Azzahra. (2023). Pemodelan Regresi Hurdle Negative Binomial pada Jumlah Kasus Difteri Jawa Barat 2020. *Jurnal Riset Statistika*, 125-130. <https://doi.org/10.29313/jrs.v3i2.3014>

Wahyu Salsabila, S., & Efendi, A. (2025). Zero Inflated Negative Binomial (ZINB) Regression: Application to the Pneumonia Study and Simulation under Several Scenarios. *Matematika Murni Dan Aplikasi*, 10(1), 457-468. <http://dx.doi.org/10.18860/ca.v10i1.32499>

World Bank Group, O. (2025). *Diphtheria*. World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/diphtheria>