



## **PEMODELAN KASUS PNEUMONIA BERAT PADA BALITA DI KOTA SURABAYA MENGGUNAKAN *ZERO- INFLATED NEGATIVE BINOMIAL***

**Deswita Nada Fitriani<sup>1)</sup>, A'yunin Sofro<sup>2,\*)</sup>**

<sup>1)</sup>*Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Negeri Surabaya*

<sup>2)</sup>*Program Studi Sains Aktuaria, Fakultas Matematika dan Ilmu Pengetahuan  
Alam, Universitas Negeri Surabaya  
\* email: [ayuninsofro@unesa.ac.id](mailto:ayuninsofro@unesa.ac.id)*

**Abstrak:** Pneumonia berat pada balita merupakan salah satu permasalahan kesehatan masyarakat yang signifikan, khususnya di perkotaan seperti Surabaya, karena dapat meningkatkan risiko morbiditas dan mortalitas balita. Data jumlah kasus pneumonia berat pada balita di Kota Surabaya berbentuk data cacah yang menunjukkan adanya overdispersi serta proporsi nilai nol berlebih. Penelitian ini bertujuan untuk memodelkan jumlah kasus pneumonia berat pada balita di Kota Surabaya menggunakan model *Zero-Inflated Negative Binomial* (ZINB) serta mengidentifikasi faktor-faktor yang memengaruhinya. Analisis awal menggunakan regresi Poisson menunjukkan ketidaksesuaian asumsi akibat overdispersi dan nilai nol berlebih, sehingga pemodelan dilanjutkan menggunakan beberapa pendekatan termasuk regresi ZINB. Pemilihan model dilakukan berdasarkan nilai *Akaike Information Criterion* (AIC). Hasil analisis menunjukkan bahwa model ZINB memiliki nilai AIC paling rendah dan mampu mengakomodasi karakteristik data dengan baik. Pada model ZINB, variabel jumlah rumah tangga dengan akses air bersih terbukti berpengaruh signifikan pada komponen *count* maupun pada komponen logit. Selain itu, variabel *dummy* tahun 2023 juga berpengaruh signifikan pada komponen *count*, yang menunjukkan jumlah kasus pneumonia berat yang lebih rendah dibandingkan dengan tahun 2022. Hasil ini menunjukkan bahwa akses air bersih merupakan faktor lingkungan yang berperan penting dalam menurunkan jumlah kasus pneumonia berat pada balita di Kota Surabaya.

**Kata Kunci:** nol berlebih; overdispersi; pneumonia berat pada balita; *zero-inflated negative binomial*

**Abstract:** Severe pneumonia among children under five remains a significant public health concern, particularly in urban areas such as Surabaya, as it increases the risk of morbidity and mortality in this age group. The data on the number of severe pneumonia cases among children under five in Surabaya City are count data characterized by overdispersion and an excess proportion of zero values. This study aims to model the number of severe pneumonia cases among children under five in Surabaya City using the *Zero-Inflated Negative Binomial* (ZINB) model and to identify factors influencing its occurrence. Preliminary analysis using Poisson regression indicated violations of model assumptions due to overdispersion and excess zeros; therefore, further modeling was conducted using several approaches, including ZINB regression. Model selection was based on the *Akaike Information Criterion* (AIC). The results show that the ZINB model yields the lowest AIC value and adequately accommodates the characteristics of the data. In the ZINB model, the variable representing



*the number of households with access to clean water has a significant effect on both the count and logit components. In addition, the dummy variable for the year 2023 is significant in the count component, indicating a lower number of severe pneumonia cases compared to the year 2022. These findings suggest that access to clean water is an important environmental factor in reducing the number of severe pneumonia cases among children under five in Surabaya City.*

**Keywords:** excess zeros; overdispersion; severe pneumonia among under five children; zero-inflated negative binomial

## PENDAHULUAN

Pneumonia berat pada balita merupakan salah satu bentuk infeksi saluran pernafasan bawah dengan tingkat keparahan tinggi yang dapat menimbulkan dampak kesehatan yang serius hingga kematian. Kelompok balita memiliki kerentanan yang lebih besar karena sistem kekebalan tubuh yang belum berkembang secara optimal (Ayuk & Adaeze C., 2025). Di Indonesia, kejadian pneumonia berat pada balita masih ditemukan di berbagai wilayah dan menunjukkan perbedaan yang cukup nyata antar wilayah (Ananda & Budyana, 2024). Menurut Dinas Kesehatan Kota Surabaya (2024), sebagai wilayah perkotaan dengan kepadatan penduduk dan kondisi lingkungan yang beragam, jumlah kasus pneumonia berat pada balita di Kota Surabaya bervariasi antar kecamatan, di mana pada sebagian wilayah tidak ditemukan kasus pada periode tertentu, sementara wilayah lainnya mencatat jumlah kasus yang relatif lebih tinggi. Pola tersebut menunjukkan bahwa kejadian pneumonia berat pada balita memiliki karakteristik data berupa jumlah kejadian yang relatif jarang dan tidak merata antar wilayah .

Perbedaan jumlah kasus pneumonia berat pada balita berkaitan erat dengan variasi kondisi lingkungan dan layanan kesehatan. Kualitas hunian, akses terhadap air bersih dan sanitasi, serta cakupan layanan kesehatan dasar seperti imunisasi dan suplementasi vitamin A terbukti memengaruhi tingkat kerentanan balita terhadap pneumonia di berbagai negara berpendapatan rendah dan menengah (Solomon et al., 2022). Selain itu, kondisi gizi balita juga berkontribusi terhadap daya tahan tubuh terhadap infeksi (Morales et al., 2024). Kombinasi berbagai faktor tersebut menghasilkan pola kejadian pneumonia berat yang tidak seragam antar wilayah, sehingga hubungan antara faktor lingkungan, gizi, dan layanan kesehatan dengan jumlah kasus pneumonia berat tidak menunjukkan pola yang sama di setiap wilayah (Suchery & Hastono, 2025).

Adanya variasi kejadian pneumonia berat pada balita tersebut menunjukkan bahwa data yang dihasilkan memiliki karakteristik sebagai data cacah dengan frekuensi kejadian yang relatif rendah, proporsi nilai nol yang tinggi, serta variasi yang cukup besar antar wilayah (Schober & Vetter, 2021). Karakteristik tersebut memerlukan penggunaan pendekatan analisis yang sesuai dengan sifat data, karena model statistik yang tidak sesuai dengan struktur data berpotensi menghasilkan hasil analisis yang kurang akurat (Cauchemez et al., 2019). Dalam konteks ini, kerangka *Generalized Linear Model* (GLM) banyak digunakan dalam analisis epidemiologi



karena mampu menangani variabel respons yang tidak mengikuti distribusi normal dan dapat dimodelkan melalui distribusi dari keluarga eksponensial yang sesuai dengan fungsi penghubung yang tepat (Maya Santi & Wiyono, 2021). Model Poisson umumnya digunakan sebagai pendekatan awal dalam pemodelan data cacah, namun pada data kejadian penyakit, asumsi equidispersi pada model ini sering kali tidak terpenuhi, sehingga mendorong penggunaan model *Negative Binomial* sebagai alternatif yang lebih sesuai. Ketika data juga menunjukkan proporsi nilai nol yang tinggi, pendekatan *zero-inflated* diperlukan untuk menggambarkan proses pembentukan data secara lebih tepat (Hu et al., 2011). Menurut Lambert (1992), model *Zero-Inflated Poisson* (ZIP) cocok digunakan ketika data memiliki banyak nol tetapi tidak menunjukkan overdispersi yang kuat, sedangkan *Zero-Inflated Negative Binomial* (ZINB) lebih sesuai diterapkan pada data yang memiliki kelebihan nol sekaligus overdispersi (J. H. Lee et al., 2012).

Beberapa penelitian terdahulu telah menerapkan model ZINB untuk memodelkan data kesehatan yang memiliki proporsi nol tinggi dan variasi antar wilayah yang besar. Penelitian oleh Astuti et al., (2023) menggunakan ZINB untuk memodelkan jumlah kematian balita akibat pneumonia di Provinsi Jawa Timur ketika banyak wilayah tidak mencatat kasus sehingga struktur data mengandung banyak nol dan variabilitas antar wilayah tinggi, sehingga ZINB lebih sesuai dibandingkan model tanpa mekanisme *zero-inflation*. Penelitian serupa juga ditunjukkan oleh Amaliana et al., (2018) pada kasus tetanus neonatorum di Jawa Timur yang menunjukkan bahwa model ZINB efektif dalam menangani data kesehatan dengan nol berlebih dan overdispersi. Selain itu, penelitian oleh Wahyu Salsabila, Efendi, et al., (2025) menunjukkan bahwa regresi ZINB dengan pendekatan *Maximum Likelihood Estimation* (MLE) mampu menghasilkan estimasi parameter yang lebih stabil pada data pneumonia balita dengan proporsi nol yang besar, meskipun masih memiliki keterbatasan terkait ukuran sampel serta cakupan variabel prediktor yang digunakan. Namun demikian, sebagian besar penelitian sebelumnya dilakukan pada skala wilayah yang luas, seperti tingkat provinsi, dan belum secara khusus mengkaji jumlah kasus pneumonia berat pada balita dengan unit analisis tingkat kecamatan, khususnya di wilayah perkotaan besar. Selain itu, penelitian yang menerapkan model ZINB dengan mempertimbangkan perbedaan jumlah balita antar wilayah dalam analisis masih terbatas. Kondisi ini menunjukkan adanya kekosongan riset yang menjadi dasar dilakukannya penelitian ini. Penelitian ini bertujuan untuk memodelkan jumlah kasus pneumonia berat pada balita di Kota Surabaya menggunakan model ZINB serta mengidentifikasi faktor-faktor yang memengaruhi perbedaan jumlah kasus pneumonia berat pada balita di Kota Surabaya.

## METODE PENELITIAN

Penelitian ini merupakan penelitian terapan dengan pendekatan pemodelan statistik untuk menganalisis jumlah kasus pneumonia berat pada balita di Kota Surabaya. Data yang digunakan berupa data sekunder yang bersumber dari Profil Kesehatan Kota Surabaya periode 2022–2024, dengan unit analisis sebanyak 31 kecamatan. Seluruh variabel yang digunakan dalam penelitian ini disajikan secara rinci pada Tabel 1 berikut:



**Tabel 1.** Data Penelitian

Variabel	Kode	Deskripsi
Variabel Respon	y	Jumlah kasus pneumonia berat pada balita
	$X_1$	Jumlah rumah sehat
	$X_2$	Jumlah bayi yang menerima imunisasi dasar lengkap
Variabel Prediktor	$X_3$	Jumlah balita yang menerima vitamin A
	$X_4$	Jumlah rumah tangga dengan akses air bersih
	$X_5$	Jumlah balita dengan status gizi buruk
Variabel Kategorik	D 2023	Variabel kategorik tahun 2023
	D 2024	Variabel kategorik tahun 2024
Offset	n	Jumlah balita (populasi balita berisiko)

Variabel kategorik tahun dimodelkan dalam bentuk variabel *dummy*, yaitu D2023 dan D2024, dengan tahun 2022 digunakan sebagai kategori referensi. Penggunaan variabel *dummy* berupa tahun bertujuan untuk menangkap perbedaan jumlah kasus pneumonia berat antar tahun pengamatan. Sedangkan variabel *offset* berupa jumlah balita (populasi balita berisiko) digunakan untuk memperhitungkan perbedaan ukuran populasi balita antar kecamatan.

Analisis data dilakukan menggunakan perangkat lunak *R Studio* dengan tahapan sebagai berikut:

1. Melakukan penginputan dan pemeriksaan data untuk memastikan kelengkapan serta konsistensi nilai pengamatan.
2. Melakukan analisis statistik deskriptif untuk menggambarkan karakteristik data.
3. Melakukan uji *excess zero* untuk mengetahui apakah data memiliki nol berlebih.
4. Melakukan uji multikolinearitas antara variabel prediktor. Uji multikolinearitas dilakukan untuk memastikan tidak adanya hubungan linier yang kuat antar variabel prediktor dalam model regresi (Shrestha, 2020). Multikolinearitas dapat menyebabkan ketidakstabilan estimasi parameter (Kim, 2019). Deteksi multikolinearitas dilakukan menggunakan *Variance Inflation Factor* (VIF) (Kyriazos & Poga, 2023), yang didefinisikan sebagai:

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

dengan  $R^2$  merupakan koefisien determinasi hasil regresi variabel prediktor terhadap prediktor lainnya. Nilai VIF lebih dari 10 mengindikasikan adanya multikolinearitas, sedangkan nilai  $VIF < 10$  menunjukkan bahwa model tidak mengalami masalah multikolinearitas dan layak digunakan untuk analisis lanjutan (Kyriazos & Poga, 2023).

5. Memodelkan data menggunakan Regresi Poisson sebagai pendekatan awal (*baseline*). Regresi Poisson merupakan bagian dari kerangka *Generalized Linear Model* (GLM) yang digunakan ketika variabel respons berbentuk data cacah dan diasumsikan memenuhi kondisi equidispersi. Model ini mencakup variabel prediktor numerik, variabel kategorik tahun pengamatan dalam bentuk variabel *dummy*, serta *offset* logaritma jumlah balita untuk

memperhitungkan perbedaan besaran populasi berisiko antar kecamatan, model regresi Poisson dinyatakan sebagai:

$$\ln(\mu_i) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \sum_{k=1}^K \gamma_k D_{ik} + \ln(E_i), \quad i = 1, 2, \dots, n \quad (2)$$

di mana  $x_{ij}$  adalah variabel prediktor numerik ke- $j$ ,  $D_{ik}$  merupakan variabel *dummy* untuk kategori tahun pengamatan ke- $k$ , dan  $E_i$  menyatakan besaran eksposur (jumlah balita) pada pengamatan ke- $i$ . Koefisien  $\beta_j$  dan  $\gamma_k$  masing-masing menunjukkan pengaruh variabel prediktor dan perbedaan antar tahun terhadap jumlah kasus pneumonia berat, sedangkan *offset*  $\ln(E_i)$  dimasukkan dengan koefisien tetap satu.

6. Melakukan uji overdispersi untuk menilai apakah regresi Poisson memenuhi asumsi equidispersi, yaitu kesamaan antara nilai harapan dan varians respon (W. Lee et al., 2023). Pengujian dilakukan menggunakan rasio dispersi ( $\hat{\phi}$ ), yaitu perbandingan antara statistik *Pearson's Chi-Square* dan derajat bebas residual (df), yang dirumuskan sebagai:

$$\hat{\phi} = \frac{\sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}}{df} \quad (3)$$

Nilai  $\hat{\phi}$  yang mendekati 1 menunjukkan tidak adanya overdispersi, sedangkan nilai  $\hat{\phi}$  yang jauh lebih besar dari 1 menunjukkan adanya overdispersi sehingga Regresi Poisson tidak sesuai dan perlu model alternatif lain.

7. Melakukan pemodelan alternatif menggunakan ZIP untuk mengatasi permasalahan *excess zero*. Menurut Feng (2020), nilai nol pada data dapat berasal dari dua mekanisme yang berbeda, yang diakomodasi dalam model ZIP dengan memisahkan komponen *count* yang mengikuti distribusi Poisson dan komponen *zero-inflation* yang memodelkan pembentukan nilai nol struktural. Pada komponen *count*, nilai harapan dinyatakan sebagai:

$$\ln(\mu_i) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \sum_{k=1}^K \gamma_k D_{ik} + \ln(E_i), \quad i = 1, 2, \dots, n \quad (4)$$

Sementara itu, probabilitas terjadinya nol struktural dimodelkan menggunakan regresi logistik sebagai berikut:

$$\text{logit}(\pi_i) = \gamma_0 + \sum_{j=1}^p \gamma_j z_{ij} \quad (5)$$

8. Melakukan pemodelan alternatif menggunakan *Negative Binomial* untuk mengatasi permasalahan overdispersi (Jhwueng & Wu, 2023). Secara umum, model *Negative Binomial* dituliskan sebagai:

$$\ln(\mu_i) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \sum_{k=1}^K \gamma_k D_{ik} + \ln(E_i), \quad i = 1, 2, \dots, n \quad (6)$$

Regresi *Negative Binomial* memiliki parameter dispersi yang memungkinkan varians respons lebih besar dari nilai harapan, sehingga lebih sesuai untuk data *count* yang mengalami overdispersi.

9. Melakukan pemodelan alternatif menggunakan ZINB untuk menangani data *count* yang menunjukkan overdispersi dan proporsi nol yang berlebih (Almulhim et al., 2025). Model ZINB memisahkan proses pembentukan data menjadi dua bagian, yaitu mekanisme yang secara khusus menghasilkan nilai nol dan mekanisme pembentukan data *count* yang mengikuti distribusi *Negative Binomial* dengan nilai harapan  $\mu_i$ . Komponen *count* dimodelkan menggunakan fungsi penghubung log dengan struktur nilai harapan sebagai berikut:

$$\ln(\mu_i) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \sum_{k=1}^K \gamma_k D_{ik} + \ln(E_i), \quad i = 1, 2, \dots, n \quad (7)$$

Selanjutnya, peluang terjadinya nol struktural dimodelkan melalui regresi logistik dengan fungsi logit sebagai berikut:

$$\text{logit}(\pi_i) = \gamma_0 + \sum_{j=1}^p \gamma_j z_{ij} \quad (8)$$

10. Melakukan pemilihan model dengan membandingkan nilai *Akaike Information Criterion* (AIC) sebagai kriteria utama, di mana model dengan nilai AIC terkecil akan dipilih sebagai model yang paling sesuai. Selain itu, rasio dispersi digunakan sebagai kriteria pendukung untuk mengevaluasi kesesuaian asumsi model terhadap karakteristik data.

Rumus AIC didefinisikan sebagai:

$$AIC = -2 \ln L(\hat{\Omega}) + 2k \quad (9)$$

Di mana  $\ln L(\hat{\Omega})$  merupakan nilai maksimum fungsi *likelihood* dan  $k$  merupakan jumlah parameter yang diestimasi dalam model.

## HASIL DAN PEMBAHASAN

### Hasil

#### Analisis Deskriptif

Tahap awal penelitian ini diawali dengan analisis deskriptif untuk memberikan gambaran umum mengenai hubungan antara variabel respon dan variabel prediktor. Ringkasan statistik deskriptif seluruh variabel penelitian disajikan pada Tabel 2 berikut.

**Tabel 2.** Statistik Deskriptif

	<b>Mean</b>	<b>Median</b>	<b>Min</b>	<b>Q<sub>1</sub></b>	<b>Q<sub>3</sub></b>	<b>Max</b>
<b>Y</b>	3.43	0.00	0.00	0.00	0.00	148.00
<b>Offset</b>	6328	5276	1237	3919	7647	14406
<b>X<sub>1</sub></b>	19091	15954	6181	13251	25069	43330

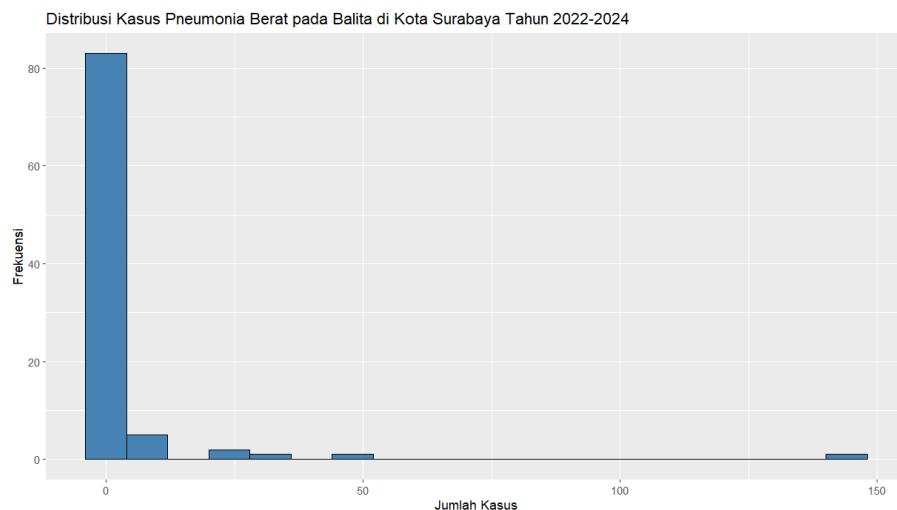


$X_2$	1286	1112	509	835	1495	2793
$X_3$	4871	4339	284	3029	5807	11334
$X_4$	19497	16102	8633	12237	24776	41770
$X_5$	3.742	3.000	0.000	1.000	5.000	25.000

Tabel 2 menunjukkan bahwa jumlah kasus pneumonia berat pada balita memiliki rentang yang cukup lebar, dengan nilai minimum 0 dan maksimum 148 kasus. Nilai rata-rata sebesar 3.43 kasus dan median 0 menunjukkan bahwa sebagian besar wilayah memiliki jumlah kasus yang rendah, dengan variasi yang dipengaruhi oleh beberapa wilayah dengan kasus relatif tinggi. Jumlah balita sebagai *offset* juga bervariasi antar wilayah dengan nilai rata-rata sebesar 6328 balita.

### Pemeriksaan *Excess zero*

Pemeriksaan terhadap keberadaan *excess zero* pada variabel respon dilakukan melalui visualisasi histogram sebagai berikut:



**Gambar 1.** Distribusi Kasus Pneumonia Berat pada Balita di Surabaya

Berdasarkan histogram pada Gambar 1, hasil perhitungan menunjukkan bahwa sebanyak 81.7% data memiliki nilai nol. Proporsi ini melebihi batas umum 50% yang mengindikasikan keberadaan *excess zero*.

### Uji Multikolinearitas

Uji multikolinearitas dilakukan untuk menilai ada tidaknya hubungan linier yang kuat antar variabel prediktor dalam model. Pengujian ini menggunakan nilai *Variance Inflation*



*Factor* (VIF). Hasil uji multikolinearitas untuk seluruh variabel prediktor disajikan pada Tabel 3 berikut.

**Tabel 3.** Uji Multikolinearitas

Variabel Prediktor	VIF
$X_1$	4.403580
$X_2$	6.220354
$X_3$	4.503015
$X_4$	3.818669
$X_5$	1.627665

Tabel 3 menunjukkan bahwa semua variabel prediktor memiliki nilai VIF yang lebih kecil dari 10, sehingga dapat dikatakan bahwa tidak terjadi multikolinearitas pada variabel prediktor.

### Pemodelan Regresi Poisson

Pemodelan regresi Poisson digunakan sebagai model awal (*baseline*) untuk memodelkan jumlah kasus pneumonia berat pada balita. Estimasi parameter yang diperoleh dalam model regresi Poisson disajikan dalam Tabel 4 berikut.

**Tabel 4.** Estimasi Parameter pada Regresi Poisson

	Estimate	Std. Error	z value	Pr(> z )	
<b>(Intercept)</b>	-6.563e+00	1.653e-01	-39.706	< 2e-16	***
$X_1$	-2358e-04	1.666e-05	-14.050	< 2e-16	***
$X_2$	8.327e-04	3.014e-04	2.763	0.00573	**
$X_3$	7.032e-04	6.094e-05	11.540	< 2e-16	***
$X_4$	2.188e-05	1.649e-05	1.327	0.18453	
$X_5$	-3.786e-01	2.832e-02	-13.370	< 2e-16	***
Tahun 2023	-2.692e+00	2.333e-01	-11.542	< 2e-16	***
Tahun 2024	-9.854e-01	1.411e-01	-6.982	2.92e-12	***

Berdasarkan Tabel 4, hasil pemodelan pada tingkat signifikansi 5% menunjukkan bahwa variabel  $X_1$ ,  $X_2$ ,  $X_3$ , dan  $X_5$  berpengaruh signifikan terhadap jumlah kasus pneumonia berat pada balita. Variabel *dummy* tahun pengamatan menunjukkan adanya penurunan jumlah kasus pneumonia berat yang signifikan pada tahun 2023 dan 2024 dibandingkan tahun referensi (2022).

### Uji Overdispersi

Salah satu asumsi utama dalam regresi Poisson adalah terpenuhinya kondisi equidispersi. Hasil uji overdispersi disajikan pada Tabel 5 berikut.



**Tabel 5.** Tes Overdispersi

<b>Komponen</b>	<b>Nilai</b>
<i>Dispersion ratio</i>	37.6915
<i>Pearson's Chi-Squared</i>	3203.7777
<i>p-value</i>	0.000

Berdasarkan hasil pengujian pada Tabel 5, diperoleh nilai *dispersion ratio* sebesar 37.6915, yang jauh lebih besar dari satu. Nilai ini menunjukkan bahwa varians data jumlah kasus pneumonia berat pada balita jauh lebih besar dibandingkan nilai harapannya, sehingga asumsi equidispersi pada regresi Poisson tidak terpenuhi. Selain itu, nilai *Pearson's Chi-Squared* sebesar 3203.7777 dengan *p-value* sebesar 0.000 yang mengindikasikan bahwa overdispersi yang terjadi bersifat signifikan pada tingkat signifikansi 5%.

### Pemodelan Regresi ZIP

Pemodelan selanjutnya dilakukan dengan mempertimbangkan adanya proporsi nilai nol yang berlebih (*excess zero*) pada data jumlah kasus pneumonia berat pada balita. Untuk mengatasi kondisi tersebut, digunakan model ZIP. Hasil estimasi parameter model ZIP disajikan pada Tabel 6 berikut.

**Tabel 6.** Estimasi Parameter pada Regresi ZIP

	<b>Estimate</b>	<b>Std. Error</b>	<b>z value</b>	<b>Pr(&gt; z )</b>	
<b>Model Count</b>					
<b>(Intercept)</b>	-4.075e+00	4.422e-01	-9.217	< 2e-16	***
$X_1$	-2.789e-04	4.074e-05	-6.845	7.64e-12	***
$X_2$	1.074e-03	1.139e-03	0.943	0.346	
$X_3$	1.702e-03	2.790e-04	6.098	1.07e-09	***
$X_4$	-3.253e-04	3.350e-05	-9.709	< 2e-16	***
$X_5$	4.102e-02	6.083e-02	0.674	0.500	
Tahun 2023	-3.591e+00	3.191e-01	-11.255	< 2e-16	**
Tahun 2024	-2.049e-01	2.365e-01	-0.866	0.386	
<b>Model Logit</b>					
<b>(Intercept)</b>	2.388e+00	8.447e-01	2.827	0.0047	***
$X_1$	5.998e-05	8.408e-05	0.713	0.4756	
$X_2$	1.583e-03	1.610e-03	0.983	0.3255	
$X_3$	6.673e-05	3.304e-04	0.202	0.8399	
$X_4$	-2.933e-04	1.159e-04	-2.531	0.0114	*
$X_5$	3.675e-01	1.679e-01	2.189	0.0286	*

Berdasarkan Tabel 6, pada tingkat signifikansi 5%, hasil pemodelan ZIP menunjukkan bahwa pada komponen *count*, variabel  $X_1$ ,  $X_3$ , dan  $X_4$  berpengaruh signifikan terhadap jumlah kasus pneumonia berat pada balita. Selain itu, variabel *dummy* tahun 2023 menunjukkan pengaruh yang signifikan dibandingkan tahun referensi (2022). Pada komponen logit, variabel  $X_4$  dan  $X_5$  berpengaruh signifikan terhadap peluang terjadinya nol struktural.

### **Pemodelan Regresi *Negative Binomial* (NB)**

Selain keberadaan *excess zero*, hasil uji sebelumnya menunjukkan adanya overdispersi pada data jumlah kasus pneumonia berat pada balita. Untuk mengatasi kondisi tersebut, pemodelan selanjutnya dilakukan menggunakan regresi *Negative Binomial* (NB) untuk menangani overdispersi. Hasil estimasi parameter model NB disajikan pada Tabel 7 berikut.

**Tabel 7.** Estimasi Parameter pada Regresi *Negative Binomial*

	<b>Estimate</b>	<b>Std. Error</b>	<b>z value</b>	<b>Pr(&gt; z )</b>	
<b>(Intercept)</b>	-9.415e+00	1.219e+00	-7.726	1.11e-14	***
$X_1$	-6.467e-07	9.942e-05	-0.007	0.99481	
$X_2$	1.216e-03	1.811e-03	0.671	0.50217	
$X_3$	1.067e-03	4.227e-04	2.525	0.01158	***
$X_4$	-1.431e-04	9.475e-05	-1.511	0.13086	
$X_5$	-5.486e-01	1.792e-01	-3.061	0.00221	**
Tahun 2023	-3.722e+00	1.223e+00	-3.043	0.00234	**
Tahun 2024	-8.068e-01	1.029e+00	-0.784	0.43320	

Berdasarkan Tabel 7, pada tingkat signifikansi 5%, variabel  $X_3$  dan  $X_5$  menunjukkan pengaruh signifikan terhadap jumlah kasus pneumonia berat pada balita. Selain itu, variabel *dummy* tahun 2023 juga berpengaruh signifikan dibandingkan tahun referensi (2022).

### **Pemodelan Regresi ZINB**

Pemodelan alternatif selanjutnya dilakukan menggunakan model ZINB sebagai model utama untuk mengatasi karakteristik data yang menunjukkan adanya overdispersi serta proporsi nilai nol yang berlebih (*zero-inflation*). *Zero inflation* mengindikasikan bahwa sebagian nilai nol pada data tidak hanya berasal dari proses kejadian acak, tetapi juga dari nol struktural, yaitu wilayah yang secara struktural memiliki kemungkinan yang lebih besar untuk tidak mengalami kejadian pneumonia berat pada balita.

Model ZINB terdiri atas dua komponen yang menunjukkan dua proses pembentukan data yang berbeda. Komponen *count* memodelkan rata-rata jumlah kasus pneumonia berat pada wilayah yang berpotensi mengalami kejadian kasus menggunakan distribusi *Negative Binomial*, dengan rata-rata kejadian dinyatakan sebagai  $\mu_i$ . Sementara itu, komponen logit (*zero-inflation*) memodelkan peluang nol struktural yang dinyatakan sebagai  $\pi_i$ , yaitu peluang bahwa suatu



wilayah berada pada kondisi dengan jumlah kasus bernilai nol akibat proses struktural tertentu, dan bukan berasal dari proses kejadian kasus. Hasil estimasi parameter model ZINB disajikan pada Tabel 8 berikut.

**Tabel 8.** Estimasi Parameter pada Regresi ZINB

	<b>Estimate</b>	<b>Std. Error</b>	<b>z value</b>	<b>Pr(&gt; z )</b>	
<b>Model Count</b>					
<b>(Intercept)</b>	-7.254e+00	1.831e+00	-3.961	7.47e-05	***
$X_1$	2.343e-05	1.333e-04	0.176	0.8605	
$X_2$	2.057e-03	3.439e-03	0.598	0.5497	
$X_3$	1.182e-03	7.921e-04	1.492	0.1357	
$X_4$	-3.372e-04	8.459e-05	-3.986	6.71e-05	***
$X_5$	-1.845e-01	1.981e-01	-0.931	0.3518	
Tahun 2023	-3.341e+00	7.369e-01	-4.534	5.80e-06	***
Tahun 2024	-1.170e+00	5.976e-01	-1.958	0.0502	.
<b>Model Logit</b>					
<b>(Intercept)</b>	1.663e+00	1.052e+00	1.580	0.1141	
$X_1$	9.732e-05	8.597e-05	1.132	0.2576	
$X_2$	1.497e-03	1.828e-03	0.819	0.4128	
$X_3$	2.023e-05	3.453e-04	0.059	0.9533	
$X_4$	-2.708e-04	1.199e-04	-2.258	0.0239	*
$X_5$	3.208e-01	1.871e-01	1.715	0.0864	.

Berdasarkan Tabel 8, pada tingkat signifikansi 5%, hasil estimasi parameter model ZINB menunjukkan bahwa pada komponen model *count*, variabel  $X_4$  (jumlah rumah tangga dengan akses air bersih) menunjukkan pengaruh signifikan dengan arah koefisien negatif, yang menunjukkan bahwa peningkatan variabel ini berkaitan dengan penurunan rata-rata jumlah kasus pneumonia berat pada balita di wilayah yang berpotensi mengalami kejadian kasus. Selain itu, variabel *dummy* tahun 2023 juga berpengaruh signifikan dengan arah koefisien negatif, yang menunjukkan bahwa jumlah kasus pneumonia berat pada tahun tersebut lebih rendah dibandingkan tahun referensi (2022).

Pada komponen model logit (*zero-inflation*), variabel  $X_4$  juga menunjukkan pengaruh signifikan dengan arah negatif terhadap peluang nol struktural. Dalam konteks model ZINB, hasil ini menunjukkan bahwa peningkatan akses air bersih menurunkan probabilitas suatu wilayah berada pada kondisi nol struktural. Dengan kata lain, peningkatan akses air bersih menurunkan peluang suatu wilayah berada pada kondisi nol struktural, sehingga wilayah tersebut tidak sepenuhnya berada pada kelompok wilayah tanpa kejadian pneumonia berat pada balita.



Berdasarkan hasil estimasi parameter yang signifikan, bentuk model ZINB mengacu pada formula (7) dan (8), yaitu terdiri atas model *count* dan model *zero-inflation* dengan *log link function* dinyatakan sebagai berikut:

$$\ln(\mu_i) = -7.254 - 0.0003372x_{4i} - 3.341D_{2023}$$
$$\text{logit}(\pi_i) = 1.663 - 0.0002708X_{4i}$$

### Pemilihan Model Terbaik

Pemilihan model terbaik dilakukan dengan mempertimbangkan nilai *Akaike Information Criterion* (AIC) sebagai kriteria utama, serta rasio dispersi sebagai kriteria pendukung untuk mengevaluasi kesesuaian model terhadap karakteristik data. Hasil perbandingan model ditampilkan pada Tabel 9 berikut.

**Tabel 9.** Perbandingan Model

Model	AIC	Rasio Dispersi
Poisson	1055.7017	37.6915
Negative Binomial	211.7238	0.4801
ZeroInflated Poisson	269.8399	6.9357
Zero-Inflated Negative Binomial	201.5217	0.9726

Berdasarkan Tabel 9, model Poisson menghasilkan nilai AIC tertinggi sebesar 1055.7017, diikuti oleh model ZIP dengan nilai AIC sebesar 269.8399, dan model *Negative Binomial* dengan nilai AIC sebesar 211.7238. Model ZINB menunjukkan nilai AIC paling rendah, yaitu sebesar 201.5217, sehingga dipilih sebagai model terbaik karena memberikan keseimbangan terbaik antara kecocokan model dan kompleksitas parameter. Selain itu, nilai rasio dispersi pada model ZINB sebesar 0.9726 mendekati satu, yang mengindikasikan bahwa variasi data telah terakomodasi secara optimal. Sebaliknya, model Poisson dan ZIP masih menunjukkan indikasi overdispersi, sementara model *Negative Binomial* menghasilkan rasio dispersi di bawah satu, yang mengindikasikan underdispersi sehingga kurang sesuai dengan karakteristik data.

### Pembahasan

Hasil analisis menunjukkan bahwa data kasus pneumonia berat pada balita di Kota Surabaya memiliki proporsi nilai nol yang tinggi, yaitu sebesar 81.7% dan mengalami overdispersi yang kuat, ditunjukkan oleh rasio dispersi regresi Poisson sebesar 37.69. Kondisi tersebut menunjukkan bahwa asumsi equidispersi tidak terpenuhi, sehingga regresi Poisson kurang sesuai untuk memodelkan data ini. Hal tersebut sejalan dengan temuan pada data survei kesehatan lainnya yang menunjukkan bahwa data hitung dengan banyak nol cenderung mengalami overdispersi, sehingga memerlukan model alternatif yang dapat menangani proporsi nol berlebih dan overdispersi (Abu Bakar et al., 2022).



Berdasarkan karakteristik data tersebut, diperlukan pemodelan alternatif yang lebih sesuai. Pemodelan alternatif pertama dilakukan menggunakan model ZIP\*) untuk menangani *excess zero*, namun masih belum bisa menangani kondisi overdispersi yang signifikan. Selanjutnya dilakukan pemodelan alternatif menggunakan Regresi *Negative Binomial* untuk mengatasi overdispersi, namun belum mengatasi keberadaan nilai nol yang berlebih. Untuk mengatasi kedua permasalahan tersebut, digunakan model ZINB. Model ini memisahkan proses terbentuknya nilai nol melalui komponen *zero-inflation* dan memodelkan jumlah kejadian menggunakan distribusi *Negative Binomial*, sehingga mampu menangani proporsi nilai nol berlebih sekaligus overdispersi pada data (Ahmad et al., 2025).

Sebagai dasar pemilihan model terbaik, dilakukan perbandingan beberapa model alternatif, yaitu Poisson, *Negative Binomial*, ZIP, dan ZINB menggunakan kriteria *Akaike Information Criterion* (AIC). Model Poisson menghasilkan nilai AIC tertinggi sebesar 1055.7017, diikuti oleh model ZIP sebesar 269.8399 dan model *Negative Binomial* sebesar 211.7238. Model ZINB menunjukkan nilai AIC paling rendah, yaitu sebesar 201.5217, yang menunjukkan keseimbangan terbaik antara kecocokan model dan kompleksitas parameter. Selain itu, rasio dispersi model ZINB sebesar 0.9726 mendekati satu, sehingga menunjukkan bahwa variasi data telah teratasi dengan baik. Oleh karena itu, model ZINB dipilih sebagai model yang paling sesuai untuk memodelkan kasus pneumonia berat pada balita di Kota Surabaya yang mengandung overdispersi dan proporsi nilai nol berlebih.

Hasil estimasi parameter model ZINB menunjukkan bahwa pada komponen model *count*, hanya variabel  $X_4$  (rumah tangga dengan akses air bersih) dan *dummy* tahun 2023 yang berpengaruh signifikan pada tingkat signifikansi 5 %. Koefisien negatif pada variabel  $X_4$  menunjukkan bahwa semakin banyak rumah tangga yang memiliki akses air bersih, semakin rendah rata-rata jumlah kasus pneumonia berat pada balita di Surabaya. Penelitian sebelumnya juga menemukan bahwa akses air bersih dan sanitasi yang layak berkaitan dengan penurunan risiko infeksi saluran pernapasan, karena air yang bersih dan kondisi sanitasi yang baik dapat mengurangi paparan patogen penyebab infeksi saluran pernapasan dan memperkuat respon imun anak (Ashraf et al., 2020). Koefisien negatif pada *dummy* tahun 2023 menunjukkan bahwa rata-rata jumlah kasus pneumonia berat pada tahun tersebut lebih rendah dibandingkan tahun referensi 2022.

Pada komponen logit (*zero-inflation*), variabel  $X_4$  (rumah tangga dengan akses air bersih) juga menunjukkan pengaruh signifikan dengan koefisien negatif terhadap peluang suatu wilayah berada pada kondisi nol struktural. Dalam model ZINB, komponen logit memodelkan probabilitas suatu wilayah termasuk ke dalam kelompok nol struktural, yaitu kondisi di mana tidak terjadi kasus pneumonia berat pada balita karena faktor struktural tertentu, dan bukan hanya disebabkan oleh proses kejadian yang bersifat acak. Koefisien negatif pada variabel  $X_4$  menunjukkan bahwa peningkatan akses air bersih menurunkan peluang suatu wilayah berada pada kelompok nol struktural. Dengan demikian, kecamatan dengan tingkat akses air bersih yang lebih baik cenderung tidak sepenuhnya berada pada kondisi tanpa risiko kejadian pneumonia



berat secara struktural, tetapi tetap termasuk dalam kelompok wilayah yang memiliki potensi kejadian kasus pada tingkat yang rendah. Hasil ini menunjukkan bahwa akses air bersih tidak hanya berperan dalam menurunkan rata-rata jumlah kasus, tetapi juga membedakan karakteristik risiko struktural kejadian pneumonia berat pada tingkat wilayah pengamatan.

## SIMPULAN

Berdasarkan hasil analisis, data jumlah kasus pneumonia berat pada balita di Kota Surabaya menunjukkan adanya overdispersi dan nilai nol berlebih, sehingga regresi Poisson tidak sesuai untuk digunakan. Untuk mengatasi karakteristik data tersebut, dilakukan pemodelan alternatif, yaitu ZIP, *Negative Binomial* (NB), dan ZINB. Berdasarkan kriteria AIC, model ZINB memiliki nilai AIC paling rendah sehingga ditetapkan sebagai model yang paling sesuai dalam memodelkan jumlah kasus pneumonia berat pada balita di Kota Surabaya. Pada tingkat signifikansi 5%, hasil estimasi parameter model ZINB menunjukkan bahwa hanya variabel  $X_4$  (jumlah rumah tangga dengan akses air bersih) yang berpengaruh signifikan baik terhadap jumlah kasus pneumonia berat pada komponen *count* maupun terhadap peluang terjadinya nol struktural pada komponen logit. Selain itu, variabel *dummy* tahun 2023 berpengaruh signifikan pada komponen *count*. Secara keseluruhan, hasil penelitian ini memberikan bahan pertimbangan bagi perencanaan kebijakan kesehatan wilayah yang berfokus pada peningkatan akses air bersih sebagai upaya pencegahan pneumonia berat pada balita. Selain itu, hasil penelitian ini menunjukkan bahwa model ZINB relevan digunakan pada studi kesehatan masyarakat yang melibatkan data kejadian penyakit dengan proporsi nol yang tinggi dan variasi data yang besar.

## DAFTAR PUSTAKA

- Abu Bakar, N. S., Ab Hamid, J., Mohd Nor Sham, M. S. J., Sham, M. N., & Jailani, A. S. (2022). Count data models for outpatient health services utilisation. *BMC Medical Research Methodology*, 22(1). <https://doi.org/10.1186/s12874-022-01733-3>
- Ahmad, M., Amin, K., Ali, A., & Ahmad, R. W. (2025). A Comparative Evaluation of Poisson, Negative Binomial, and Zero-Inflated Models for Count Data. *Dialogue Social Science Review (DSSR)*. [www.thedssr.com](http://www.thedssr.com)
- Almulhim, F. A., Nagy, M., Hammad, A. T., Mansi, A. H., Mekiso, G. T., & El-Raouf, M. M. A. (2025). New two parameter hybrid estimator for zero inflated negative binomial regression models. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-06116-4>
- AMALIANA, L., SA'ADAH, U., & WARDHANI, N. W. S. (2018). PERFORMA PROPORSI ZERO-INFLATION PADA REGRESI ZERO-INFLATED NEGATIVE BINOMIAL (STUDI KASUS: DATA TETANUS NEONATORUM DI JAWA TIMUR). *E-Jurnal Matematika*, 7(1), 41. <https://doi.org/10.24843/mtk.2018.v07.i01.p183>
- Ananda, I. P., & Budyana. (2024). *Determinants of Pneumonia Status in Under-Five Children in Indonesia*.
- Ashraf, S., Islam, M., Unicomb, L., Rahman, M., Winch, P. J., Arnold, B. F., Benjamin-Chung, J., Ram, P. K., Colford, J. M., & Luby, S. P. (2020). Effect of improved water quality, sanitation, hygiene and nutrition interventions on respiratory illness in young children in Rural Bangladesh: A multi-arm cluster-randomized controlled trial. *American Journal of*



- Tropical Medicine and Hygiene*, 102(5), 1124–1130. <https://doi.org/10.4269/AJTMH.19-0769>
- Astuti, C. C., Puka, A. O. B., & Wiguna, A. (2023). ZERO-INFLATED NEGATIVE BINOMIAL MODELING IN INFANT DEATH CASE DUE TO PNEUMONIA IN EAST JAVA PROVINCE. *Barekeng*, 17(4), 1835–1844. <https://doi.org/10.30598/barekengvol17iss4pp1835-1844>
- Ayuk, & Adaeze C. (2025). Complications and long-term impact of early life pneumonia. In *Pediatric Pulmonology* (Vol. 60, Issue S1, pp. S51–S53). John Wiley and Sons Inc. <https://doi.org/10.1002/ppul.27299>
- Cauchemez, S., Hoze, N., Cousien, A., Nikolay, B., & ten bosch, Q. (2019). How Modelling Can Enhance the Analysis of Imperfect Epidemic Data. In *Trends in Parasitology* (Vol. 35, Issue 5, pp. 369–379). Elsevier Ltd. <https://doi.org/10.1016/j.pt.2019.01.009>
- Dinas Kesehatan Kota Surabaya. (2024). *Profil Kesehatan Kota Surabaya Tahun 2022–2024*.
- Feng, C. (2020). Zero-inflated models for adjusting varying exposures: a cautionary note on the pitfalls of using offset. *Journal of Applied Statistics*, 49(1), 1–23. <https://doi.org/10.1080/02664763.2020.1796943>
- Hu, M. C., Pavlicova, M., & Nunes, E. V. (2011). Zero-inflated and hurdle models of count data with extra zeros: Examples from an HIV-risk reduction intervention trial. *American Journal of Drug and Alcohol Abuse*, 37(5), 367–375. <https://doi.org/10.3109/00952990.2011.597280>
- Jhhueng, D. C., & Wu, C. Y. (2023). A Novel Phylogenetic Negative Binomial Regression Model for Count-Dependent Variables. *Biology*, 12(8). <https://doi.org/10.3390/biology12081148>
- Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, 72(6), 558–569. <https://doi.org/10.4097/kja.19087>
- Kyriazos, T., & Poga, M. (2023). Dealing with Multicollinearity in Factor Analysis: The Problem, Detections, and Solutions. *Open Journal of Statistics*, 13(03), 404–424. <https://doi.org/10.4236/ojs.2023.133020>
- Lee, J. H., Han, G., Fulp, W. J., & Giuliano, A. R. (2012). Analysis of overdispersed count data: Application to the Human Papillomavirus Infection in Men (HIM) Study. *Epidemiology and Infection*, 140(6), 1087–1094. <https://doi.org/10.1017/S095026881100166X>
- Lee, W., Kim, J., & Lee, D. (2023). Revisiting the analysis pipeline for overdispersed Poisson and binomial data. In *Journal of Applied Statistics* (Vol. 50, Issue 7, pp. 1455–1476). Taylor and Francis Ltd. <https://doi.org/10.1080/02664763.2022.2026897>
- Maya Santi, V., & Wiyono, A. (2021). Pemodelan Jumlah Kasus Malaria di Indonesia Menggunakan Generalized Linear Model. *Jurnal Statistika Dan Aplikasinya*, 5(1).
- Morales, F., Montserrat-de la Paz, S., Leon, M. J., & Rivero-Pino, F. (2024). Effects of Malnutrition on the Immune System and Infection and the Role of Nutritional Strategies Regarding Improvements in Children's Health Status: A Literature Review. In *Nutrients* (Vol. 16, Issue 1). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/nu16010001>
- Schober, P., & Vetter, T. R. (2021). *Statistical Minute ANESTHESIA & ANALGESIA Count Data in Medical Research: Poisson Regression and Negative Binomial Regression*. [www.anesthesia-analgesia.org](http://www.anesthesia-analgesia.org)



- Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*, 8(2), 39–42. <https://doi.org/10.12691/ajams-8-2-1>
- Solomon, Y., Kofole, Z., Fantaye, T., & Ejigu, S. (2022). Prevalence of pneumonia and its determinant factors among under-five children in Gamo Zone, southern Ethiopia, 2021. *Frontiers in Pediatrics*, 10. <https://doi.org/10.3389/fped.2022.1017386>
- Suchery, P. L., & Hastono, S. P. (2025). Spatial Analysis of Pneumonia Distribution in Children Under Five in West Java: Relationships with Individual and Environmental Determinants. *Jurnal Ilmu Kesehatan Masyarakat*.
- Wahyu Salsabila, S., Efendi, A., & Nurjannah. (2025). Zero Inflated Negative Binomial (ZINB) Regression: Application to the Pneumonia Study and Simulation under Several Scenarios. *CAUCHY - Jurnal Matematika Murni Dan Aplikasi*, 10(1), 457–468. <https://doi.org/10.18860/ca.v10i1.32499>