Codeverse: Journal of Emerging Digital Realities
Vol. 01, No. 2, December 2026, pp. 67~78
ISSN: 3123-3481, DOI: 10.11591/codeverse.v01.i1.pp67-78

Analisis Faktor-faktor yang Memengaruhi Jumlah Kasus Pneumonia di Jawa Timur Tahun 2022 Menggunakan Metode Regresi Poisson

Muhammad Zulfadhli [1*]
[1]Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Article Info

Article history:

Received month dd, yyyy
Revised month dd, yyyy
Accepted month dd, yyyy

Keywords:

Clustering
Pemilih pemula
Pemilu 2024
Stratified Two-Stage
Swing voter

ABSTRACT

Penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi jumlah kasus pneumonia pada anak di Jawa Timur menggunakan pendekatan statistik regresi Poisson. Variabel independen yang digunakan meliputi persentase Bayi Berat Badan Lahir Rendah (BBRL), jumlah balita yang mengalami batuk atau kesukaran bernapas, dan persentase penderita COVID-19. Data diperoleh dari Badan Pusat Statistik (BPS) dan Kementerian Kesehatan RI dengan cakupan seluruh kabupaten/kota di Jawa Timur sebanyak 38 wilayah. Berdasarkan hasil analisis regresi Poisson, ditemukan bahwa variabel BBRL dan jumlah balita batuk/kesukaran bernapas berpengaruh signifikan terhadap kasus pneumonia, sementara persentase COVID-19 tidak signifikan. Selanjutnya dilakukan pemeriksaan multikolinearitas dan overdispersion, yang menunjukkan bahwa model Poisson mengalami overdispersion. Untuk mengatasi hal ini, digunakan model regresi binomial negatif. Model ini memperkuat hasil sebelumnya bahwa jumlah balita batuk/kesukaran bernapas merupakan variabel yang paling signifikan memengaruhi jumlah kasus pneumonia pada anak. Model terbaik dipilih berdasarkan nilai Akaike Information Criterion (AIC), di mana model binomial negatif dengan satu prediktor (balita batuk/kesukaran bernapas) menghasilkan AIC terkecil. Variabel persentase Bayi Berat Badan Rendah (BBRL), jumlah balita batuk atau kesukaran bernapas dan persentase penderita Covid-19 di Jawa Timur Tahun 2022, memiliki nilai VIF kurang dari 10, sehingga dapat disimpulkan bahwa tidak terdapat multikolinearitas antar variabel prediktor.

Corresponding Author:

Muhammad Zulfadhli
Department of Statistics, Faculty of Science and Data Analytics
Institut Teknologi Sepuluh Nopember
60111 Sukolilo, Surabaya, Indonesia
Email: muhammadzulfadhli23@gmail.com

1. PENDAHULUAN

Pneumonia adalah infeksi akut yang menyerang paru-paru, khususnya alveoli kantong udara kecil tempat pertukaran oksigen dan karbon dioksida. Pneumonia sendiri dipengaruhi dari beberapa faktor, yaitu faktor internal dan eksternal. Dimana faktor internal dilihat dari usia, imunisasi, status gizi, pemberian ASI eksklusif, Berat Badan Lahir (BBL). Sedangkan faktor eksternal yaitu batuk, dan covid. Beberapa faktor yang dapat mempengaruhi jumlah kasus pneumonia di antaranya adalah status gizi buruk, imunisasi tidak lengkap, sanitasi, paparan asap rokok dan polusi udara, dan infeksi virus lainnya, salah satunya Covid-19. Untuk mengetahui apakah hubungan faktor-faktor tersebut dengan jumlah kasus pneumonia signifikan, diperlukan analisis statistik yang tepat. Salah satu metode yang dapat digunakan adalah analisis regresi Poisson, karena jumlah kasus pneumonia berbentuk count atau hitungan, sehingga analisis ini sesuai untuk memodelkan jumlah kejadian yang jarang terjadi.

Beberapa penelitian terkait regresi non-linear yang telah dilakukan oleh peneliti, diantaranya regresi logistik biner (Achmad, et al., 2025; Otok, et al., 2026), regresi logistik multinomial (Saida, et al., 2022), regresi logistik nonparametrik (Zulfadhli, et al., 2025; Zulfadhli et al., 2025). Regresi *Poisson* merupakan suatu model regresi non-linear yang digunakan untuk menganalisis data berdistribusi *poisson* dimana variabel respon berbentuk data *count* dan nilainya adalah integer tidak negatif (Esra, 2023). Selain itu, dilakukan pemeriksaan multikolinearitas untuk mengetahui adanya korelasi antar variabel prediktor. Namun, model regresi *poisson* seringkali terjadi *overdispersion* yakni nilai varians variabel respon lebih besar dibandingan nilai rata-ratanya. Padahal, salah satu asumsi utama pada model regresi *poisson* adalah terjadinya *equidispersion* yaitu nilai variansi dari variabel respon harus sama dengan nilai rata-ratanya. Salah satu cara untuk mengatasi *overdispersion* adalah dengan menggunakan model regresi binomial negatif (Cahyandari, 2014). Dalam menentukan hubungan antara variabel respon dan prediktor juga dibutuhkan model regresi terbaik. Salah satu metode untuk memilih model regresi terbaik adalah menggunakan nilai *Akaike Information Criterion* (AIC) yang ditemukan oleh Akaike dan Schwarz. Model regresi terbaik adalah model dengan nilai AIC terkecil (Rachmillah, 2015).

Pada penelitian ini, data yang digunakan didapatkan dari *website* resmi BPS yang selanjutnya dianalisis mengenai pengaruh persentase BBRL (Berat Badan Rendah Lahir), jumlah balita Batuk, dan persentase penderita Covid-19 terhadap jumlah kasus pneumonia pada anak di Jawa Timur Tahun 2022. Data akan dianalisis dengan menggunakan regresi *poisson*. Temuan dari studi ini diharapkan menambah wawasan para pembaca dan dapat menjadi rujukan bagi pembuat kebijakan dalam merancang strategi penanganan kasus pneumonia di Jawa Timur.

2. METODE

Data dalam penelitian ini diperoleh dari *website* resmi Badan Pusat Statistik (BPS) dan Kementerian Kesehatan Republik Indonesia. Populasi yang digunakan adalah 38 Kab/Kota di Jawa Timur. Adapun rincian variable yang digunakan adalah sebagai berikut.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Skala Pengukuran	Satuan
Y	Jumlah Kasus Pneumonia pada Anak di Jawa Timur Tahun 2022	Rasio	Kasus
X_1	Persentase Bayi Berat Badan Rendah (BBRL) di Jawa Timur Tahun 2022	Rasio	Persen
X_2	Jumlah Balita Batuk atau Kesukaran Bernapas di Jawa Timur tahun Tahun 2022	Rasio	Orang
X_3	Persentse Penderita Covid-19 di Jawa Timur Tahun 2022	Rasio	Persen

Langkah analisis yang dilakukan dalam praktikum ini adalah sebagai berikut.

1. Mengumpulkan data jumlah kasus pneumonia di Jawa Timur Tahun 2022 dan faktor-faktor yang diduga mempengaruhinya, melalui *website* resmi Badan Pusat Statistik (BPS) dan Kementerian Kesehatan.
2. Mendeskripsikan karakteristik jumlah kasus pneumonia Jawa Timur Tahun 2022 dan faktor-faktor yang diduga mempengaruhinya, menggunakan statistika deskriptif.
3. Pengujian kasus multikolinearitas berdasarkan kriteria nilai VIF.
4. Memodelkan kasus pneumonia di Jawa Timur Tahun 2022 dengan regresi *Poisson.*
5. Jika Terjadi overdispresi, maka dilanjutkan menganalisis model regresi Binomial Negatif dan *Generalized Poisson Regression* (GPR).
6. Menentukan model terbaik dengan membandingkan nilai AIC pada model regresi *Poisson,* regresi Binomial Negatif dan *Generalized Poisson Regression* (GPR).

7. Membuat interpretasi model jumlah pneumonia pada anak di Jawa Timur 2022
8. Membuat kesimpulan dan saran.

2.2 Statistika Deskriptif

Statistika deskriptif adalah metode dalam statistika yang berhubungan dengan pendataan, pengumpulan, penyajian data, dan penyimpulan hasil pengamatan terhadap seluruh kejadian secara kuantitatif. Hasil tersebut dapat dideskripsikan dalam bentuk angka maupun visual dengan metode statistika. Statistika deskriptif bertujuan untuk mengetahui karakteristik suatu kejadian dengan menghitung nilai rata-rata, median, modus, standar deviasi, dan varians dari objek pengamatan (Susanto & Haryono, 2016). Beberapa contoh penelitian yang menggunakan analisis deskriptif (Dirwan, et al., 2025).

2.2.1 Mean (Rata-Rata)

Mean (rata-rata) adalah teknik pendeskripsian kelompok berdasarkan nilai tengah kelompok tersebut dengan membagi jumlah nilai-nilai data dengan jumlah individu. Nilai *mean* yang dihasilkan bahkan bisa digunakan untuk membandingkan data populasi satu dengan data populasi lain (Martias, 2021). Dalam menentukan nilai *mean* data tunggal suatu sampel dapat dinyatakan pada Persamaan 2.1.

$$\bar{X} = \frac{\sum_{i=1}^{N} x_i}{n} \qquad (2.1)$$

Keterangan :
$\bar{X}$ = Rata-rata sampel
x_i = Nilai pada data ke-i
n = Banyak data

2.2.2 Varians

Varians merupakan teknik statistik yang berfungsi untuk menjelaskan homogenitas kelompok data yang menunjukkan sebaran data dekat dengan rataannya. Varians selalu bernilai positif dan nilainya yang kecil menunjukkan bahwa sebaran data dekat dengan rataannya (Martias, 2021). Varians suatu sampel dinotasikan dengan S^2, dinyatakan pada Persamaan 2.2.

$$S^2 = \frac{\sum_{i=1}^{n}(xi-\bar{x})^2}{n-1} \qquad (2.2)$$

Keterangan :
S^2 = Varians sampel
$\bar{X}$ = Rata-rata sampel
x_i = Nilai pada data ke-i
n = Jumlah sampel

2.3 Multikolinearitas

Multikolinearitas adalah suatu kondisi dimana terjadi korelasi antara variabel bebas atau antar variabel bebas tidak bersifat saling bebas. Besaran (*quality*) yang dapat digunakan untuk mendeteksi adanya multikolinearitas adalah faktor inflasi ragam (*Variance Inflation Faktor* / VIF). VIF digunakan sebagai kriteria untuk mendeteksi multikolinearitas pada regresi linear yang melibatkan lebih dari dua variabel bebas. Nilai VIF lebih besar dari 10 mengidentifikasi adanya masalah multikolinearitas yang serius (Sriningsih, 2018). VIF untuk koefisien regresi-j diformulasikan dengan Persamaan 2.3 berikut.

$$VIF = \frac{1}{1-R_j^2} \qquad (2.3)$$

Keterangan :
VIF = Nilai *Variance Inflation Factor*
R_j^2 = Koefisien determinasi

2.4 Regresi Poisson

Regresi *Poisson* merupakan salah satu bentuk analisis regresi yang digunakan untuk model data cacah (count data). Regresi *Poisson* diartikan juga sebagai metode statistik yang digunakan untuk memodelkan hubungan antara variabel independen (variabel penjelas) dengan variabel dependen (variabel respons) yang merupakan data hitung atau data diskrit. Regresi *Poisson* didasarkan pada distribusi *Poisson*, yang merupakan distribusi probabilitas diskrit yang digunakan untuk menggambarkan jumlah kejadian yang jarang terjadi dalam interval waktu atau ruang tertentu (Yulianingsih, 2012). Estimasi parameter dari distribusi *poisson* adalah sebagai berikut.

$$\mu_i = \exp(\beta_0 + \beta_2 X_{i2} \ldots + \beta_0 X_{ik}) \qquad (2.4)$$

β_0 = intersep atau konstanta

$\beta_1 \ldots \beta_k$ = koefisien regresi atau parameter
$x_1 \ldots x_k$ = variabel independen
Uji serentak distribusi *poisson* merupakan uji yang dilakukan memeriksa pengaruh variabel-variabel independen secara bersama-sama terhadap variabel dependen dalam model regresi *Poisson*. Rumus uji serentak pada distribusi *poisson* adalah sebagai berikut.
Hipotesis:
$H_0 : \beta_1 = \beta_2 = \ldots\ldots = \beta_k = 0$
H_1 : paling sedikit terdapat satu j dimana $\beta_j \neq 0$, j = (1, …, k)
Taraf signifikan: α
Daerah Kritis: Tolak H_0 jika $G^2 \geq X^2_{(v,\propto)}$, dimana v adalah banyaknya variabel prediktor
Statistik Uji :

$$G = -2\left[\ln\left(\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right) = 2\left(L(\widehat{\Omega}) - L(\widehat{\omega})\right) \quad (2.5)$$

Keterangan :
G = Statistik uji *deviance*
$L(\widehat{\omega})$ = Fungsi *likelihood* dari model yang diestimasi (model penuh)
$L(\widehat{\Omega})$ = Fungsi *likelihood* dari model nol (model yang tidak memiliki prediktor)
Kemudian uji parsial dari distribusi *poisson* untuk mengetahui pengaruh antar variabel terhadap variabel respon adalah sebagai berikut.
Hipotesis:
$H_0 : \beta_i = 0$ (Pengaruh variabel ke-i tidak signifikan)
$H_1 : \beta_i \neq 0$ (Pengaruh variabel ke-i signifikan)
Taraf signifikan: α
Daerah kritis: Tolak H_0 jika $|Z_{hitung}| \geq Z_{\frac{\alpha}{2})}$ dimana α adalah taraf signifikan dan v adalah derajat bebas.
Statistik uji:

$$Z = \frac{\beta_j}{se\{\beta_j\}} \quad (2.6)$$

Keterangan:
$\hat{\beta}_j$ = Beta ke-j
SE(βj) = Standar error Beta

2.4.1 Estimasi Parameter Model Regresi Poisson

Estimasi model regresi poisson digunakan untuk memodelkan data count (jumlah kejadian) yang bersifat diskret dan mengikuti distribusi Poisson. Misalnya: jumlah kecelakaan per hari, jumlah pelanggan per jam, jumlah cacat produk, dan lain-lain (Agresti, 2015). Fungsi *likelihood* dari regresi Poisson adalah sebagai berikut.

$$lnL(\beta) = ln\left(\prod_{i=1}^{n}\frac{exp(-\mu_i)\mu_i^{y_i}}{y_i!}\right) = \sum_{i=1}^{n} ln\left(\frac{exp(-\mu_i)\mu_i^{y_i}}{y_i!}\right) \quad (2.7)$$

$$lnL(\beta) = \sum_{i=1}^{n}\left(-e^{x_i^T\beta} + y_i\, ln(e^{x_i^T\beta}) - \ln(y_i!)\right) \quad (2.8)$$

$$lnL(\beta) = -\sum_{i=1}^{n}(exp\,(x_i^T\beta)) + \sum_{i=1}^{n} y_i ln\,(exp(x_i^T\beta) - \sum_{i=1}^{n} ln\,(y_i!) \quad (2.9)$$

Turunan pertama fungsi ln *likelihood* ditunjukkan pada Persamaan 2.10 berikut.

$$\frac{\partial\, ln\, L(\beta)}{\partial\beta^T} = -\sum_{i=1}^{n} x_i\, exp\,(x_i^T\beta) + \sum_{i=1}^{n} y_i\, x_i = 0 \quad (2.10)$$

Keterangan :
$L(\beta)$ = *Log-likelihood*
$x_i^T\beta$ = Prediktor linier
Persamaan 2.10 belum menghasilkan solusi yang tepat sehingga perlu diselesaikan menggunakan numerik yaitu dengan iterasi Newton-Raphson sebagai berikut.

1. Menentukan nilai taksiran awal parameter $\hat{\beta}_{(0)}$ yang biasanya diperoleh dengan metode *Ordinary Least Square* (OLS) sebagai berikut.

3 $\hat{\beta}_{(0)} = (X^TX)^{-1}(X^Ty)$ (2.11)

2. Membentuk vektor gradient g sebagai berikut.

$$g^T(\beta_{(m)})_{px1} = \frac{\partial lnL(\beta)}{\partial\beta_0}, \frac{\partial lnL(\beta)}{\partial\beta_1}, \ldots, \frac{\partial lnL(\beta)}{\partial\beta_k} \quad (2.12)$$

3. Membentuk matriks Hessian H sebagai berikut.

$$H(\beta_{(m)})_{(k+1)(k+1)} = \begin{bmatrix} \frac{\partial^2 lnL(\beta)}{\partial {\beta_0}^2} & \frac{\partial^2 lnL(\beta)}{\partial \beta_0 \beta_1} & \cdots & \frac{\partial^2 lnL(\beta)}{\partial \beta_0 \beta_k} \\ \cdots & \frac{\partial^2 lnL(\beta)}{\partial {\beta_1}^2} & \cdots & \frac{\partial^2 lnL(\beta)}{\partial \beta_1 \beta_k} \\ & & \ddots & \vdots \\ simetris & & \cdots & \frac{\partial^2 lnL(\beta)}{\partial {\beta_k}^2} \end{bmatrix} \quad (2.13)$$

4. Memasukkan nilai $\hat{\beta}_{(0)}$ ke dalam elemen-elemen vektor g dan matriks H, sehingga diperoleh vektor g$\hat{\beta}_{(0)}$ dan matriks H$\hat{\beta}_{(0)}$.

5. Melakukan iterasi pada persamaan $\hat{\beta}_{(m+1)} = \hat{\beta}_{(m)} - H^{-1}_{(m)g(m)}$ mulai dari m=0, dengan nilai $\hat{\beta}_{(m)}$ adalah kumpulan penaksir parameter yang konvergen pada iterasi ke-m.6. Lanjutkan iterasi hingga m=m+1 dan akan berhenti pada keadaan konvergen yaitu pada saat $\|\beta_{(m+1)} - \beta_{(m)}\|$ lebih kecil atau sama dengan ε, ε adalah bilangan yang sangat kecil.

2.4.2 Pengujian Parameter Model Regresi Poisson

Parameter yang dihasilkan dari serangkaian proses penaksiran belum tentu memberi pengaruh yang signifikan terhadap variabel respon. Untuk menguji apakah parameter model memiliki pengaruh yang signifikan perlu dilakukan pengujian terhadap parameter model tersebut. Pengujian signifikasi parameter model regresi poisson terdiri dari uji serentak dan uji parsial (Agresti, 2015). Pengujian signifikansi parameter model regresi Poisson secara serentak dilakukan dengan menggunakan metode *Maximum Likelihood Ratio Test* (MLRT) dengan hipotesis sebagai berikut.

Hipotesis :
$H_0 : \beta_1 = \beta_2 = ... = \beta_p = 0$
H_1 : minimal ada satu $\beta_k \neq 0$; k = 1, 2, ..., p
Statistik uji :

$$G^2 = -2ln\Lambda = -2ln\left(\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right) \quad (2.14)$$

Keterangan :
G^2 = Statistik uji *likelihood rasio*
Λ = Rasio *likelihood* antara dua model
$L(\widehat{\omega})$ = Nilai *likelihood* dari model yang disederhanakan
$L(\widehat{\Omega})$ = Nilai *likelihood* dari model penuh
Daerah kritis : Tolak H_0 jika nilai G^2 lebih dari $\chi^2_{(\alpha,p)}$.

Pengujian signifikansi parameter model regresi Poisson secara parsial dilakukan dengan dengan hipotesis sebagai berikut
Hipotesis :
$H_0 : \beta_k = 0$; k = 1, 2, ..., p
$H_1 : \beta_k \neq 0$; k = 1, 2, ..., p
Statistik uji :

$$Z = \frac{\hat{\beta}_k}{s\hat{e}(\hat{\beta}_k)} \quad (2.15)$$

Keterangan :
Z = Statistik uji parameter parsial
$s\hat{e}(\hat{\beta}_k)$ = Statistik uji parameter $\hat{\beta}_k$
Daerah kritis : Tolak H_0 jika nilai │Z│lebih dari $Z\alpha/_2$

2.4.3 Kebaikan Model

Uji kebaikan model (goodness-of-fit test) digunakan untuk mengevaluasi apakah model regresi Poisson yang dibangun sesuai dengan data yang diamati. Artinya, kita ingin tahu seberapa baik model menjelaskan variabilitas dalam data. Koefisien determinasi (R^2) mempunyai *range* antara 0 sampai 1. Semakin besar nilai R^2 (mendekati 1) maka berarti pengaruh variabel bebas secara serentak dianggap kuat dan apabila (R^2) mendekati nol maka pengaruh variabel bebas terhadap variabel terikat serentak adalah lemah (Dobson, 2018). Koefisien determinasi (R^2) dapat dihitung dengan rumus berikut.

$$R^2 = \frac{JKR}{JKT} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (2.16)$$

Keterangan :
R^2 = Koefisien determinasi
y_i = Data respon ke-i
$\bar{y}$ = Rata-rata variabel respon

$\hat{y}_i$ = Estimasi respon ke-i

2.5 Pemeriksaan Overdispersi

Overdispersi merupakan suatu keadaan dimana variansi data yang diamati lebih besar daripada meannya. *Overdispersi* sering ditemukan dalam data diskrit, seperti jumlah kejadian, kecelakaan, atau kegagalan dan dapat disebabkan oleh adanya sumberkeragaman yang tidak teramati, adanya pencilan pada data. *Overdispersi* dapat diartikna juga sebagai fenomena Dimana variansi data yang diamati lebih besar dari pada ekspektasinya (Cahyandari, 2014).

Pengujian overdispersi adalah sebagai berikut.

$$\theta = \frac{nilai\ ddeviance}{df} \quad (2.17)$$

Keterangan :

θ = Dispersi

Df = *degree of Freedom*

2.6 Analisis Regresi Binomial Negatif

Analisis Regresi Binomial Negatif adalah salah satu bentuk regresi untuk data count (jumlah kejadian) yang digunakan ketika terjadi overdispersion dalam data yaitu ketika varians lebih besar dari mean, sehingga model regresi Poisson tidak lagi cocok (Dobson, 2018). Fungsi massa probabilitas regresi binomial negatif ditunjukkan pada Persamaan 2.18 berikut.

$$f(y_i; \mu_i; \theta) = \frac{\Gamma(y+1/\theta)}{\Gamma(1/\theta)y!}\left(\frac{1}{1+\theta\mu}\right)^{1/\theta}\left(\frac{\theta}{1+\theta\mu}\right)^{y};\ y = 1,2,\dots,n \quad (2.18)$$

Keterangan :

y = Variabel respon

μ = Ekspektasi prediktor linier

θ = Parameter dispersi

Pada Persamaan 2.18, kondisi overdispersi ditunjukkan dengan nilai q > 1. Model regresi binomial negatif dinyatakan pada Persamaan 2.19 berikut.

$$\mu_i = \exp\,(x_i^T\beta) = exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}) \quad (2.19)$$

Keterangan :

μ_i = Rata-rata jumlah kejadian yang terjadi dalam interval waktu tertentu

$\exp(\beta_j)$ = Rasio jumlah kejadian yang diharapkan

β_0 = Intersep

β_1 = Koefisien regresi

X_k = Variabel independen

2.6.1 Estimasi Parameter Model Regresi Binomial Negatif

Estimasi parameter regresi binomial negatif adalah perluasan dari regresi Poisson yang digunakan ketika terjadi overdispersion (varians data lebih besar daripada rata-ratanya). Estimasi parameter dalam model ini bertujuan untuk mendapatkan nilai koefisien regresi (β) dan parameter overdispersi (α) (Dobson, 2018). Persamaan *log-likelihood* untuk binomial negatif ditunjukkan pada Persamaan 2.20 berikut.

$$L(\beta, \theta) = \prod_{l=1}^{N}\frac{\Gamma(y+1/\theta)}{\Gamma(1/\theta)y!}\left(\frac{1}{1+\theta\mu_i}\right)^{1/\theta}\left(\frac{\theta}{1+\theta\mu_i}\right)^{y} \quad (2.20)$$

Keterangan :

y = Variabel respon

μ = Ekspektasi prediktor linier

θ = Parameter dispersi

2.6.2 Penguji Parameter Model Regresi Binomial Negatif

Model Regresi Binomial Negatif (Negative Binomial Regression) digunakan untuk menganalisis data hitung (count data) yang menunjukkan overdispersi, yaitu ketika varians lebih besar dari nilai rata-rata (var(Y) > E(Y)) suatu kondisi yang tidak bisa ditangani oleh model regresi Poisson (Agresti, 2015). Pengujian signifikansi parameter regresi binomial negatif secara serentak dilakukan dengan hipotesis sebagai berikut.

Hipotesis :

$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$

H_1 : minimal ada satu $\beta_k \neq 0$; k = 1, 2, ..., p

Statistik uji :

$$G^2 = -2ln\Lambda - 2ln\left(\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right) \quad (2.21)$$

Keterangan :
G^2 = Statistik uji *likelihood rasio*
Λ = Rasio *likelihood* antara dua model
$L(\hat{\omega})$ = Nilai *likelihood* dari model yang disederhanakan
$L(\hat{\Omega})$ = Nilai *likelihood* dari model penuh
Daerah kritis : Tolak H_0 jika nilai G^2 lebih dari $\chi^2_{(\alpha,p)}$.
Pengujian signifikansi parameter regresi binomial negatif secara parsial dilakukan dengan hipotesis sebagai berikut
Hipotesis :
$H_0 : \beta_k = 0$; k = 1, 2, ..., p
$H_1 : \beta_k \neq 0$; k = 1, 2, ..., p
Statistik uji :

$$t = \frac{\hat{\beta}_k}{s\hat{e}(\hat{\beta}_k)} \quad (2.22)$$

Keterangan :
t = Statistik uji parameter parsial
$s\hat{e}(\hat{\beta}_k)$ = Statistik uji parameter $\hat{\beta}_k$
Daerah kritis : Tolak H_0 jika nilai $|t|$ lebih dari $t_{(n-k,\alpha/2)}$

2.7 Pneumonia

Pneumonia adalah infeksi atau peradangan akut pada jaringan paru-paru (alveoli) yang disebabkan oleh mikroorganisme seperti bakteri, virus, jamur, atau parasit. Pada penderita pneumonia, alveoli (kantung udara kecil di paru-paru) yang seharusnya berisi udara menjadi terisi oleh cairan atau nanah, yang menyebabkan gangguan pertukaran oksigen (Murray, 2016).

2.8 BBRL (Berat Badan Rendah Lahir)

BBRL (Berat Badan Rendah Lahir) adalah kondisi di mana bayi lahir dengan berat badan kurang dari 2.500 gram (2,5 kg) tanpa memandang usia kehamilan. Ini merupakan indikator penting dalam menilai kesehatan dan status gizi ibu selama kehamilan, serta berhubungan erat dengan morbiditas (kesakitan) dan mortalitas (kematian) neonatal dan bayi (Saifuddin, 2020).

2.9 Batuk

Batuk adalah respons refleks tubuh yang bertujuan untuk membersihkan saluran pernapasan dari zat asing, lendir, iritan, atau mikroorganisme seperti virus dan bakteri. Batuk dapat terjadi secara sadar (sengaja) maupun refleksif (tak sadar) sebagai mekanisme pertahanan alami tubuh terhadap gangguan pada saluran napas (Clinic, 2022).

2.10 COVID-19

COVID-19 (Coronavirus Disease 2019) adalah penyakit infeksi saluran pernapasan yang disebabkan oleh SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) sejenis virus corona yang pertama kali teridentifikasi di Wuhan, Tiongkok pada akhir tahun 2019. Penyakit ini menyebar melalui droplet (percikan air liur) dari orang yang terinfeksi, terutama saat batuk, bersin, berbicara, atau melalui kontak dengan permukaan yang terkontaminasi (Aditia, 2021).

3. RESULTS AND DISCUSSION

Metode yang digunakan adalah regresi *poisson* yang terdiri dari estimasi parameter, uji serentak dan uji parsial. Jika data terdeteksi *overdispersion*, maka dilanjutkan penanganan dengan regresi binomial negatif dan diakhiri dengan analisis pemilihan model terbaik.

3.1. Karakteristik Data

Adapun karakteristik data penelitian disajikan pada tabel berikut:

Tabel 2. Statistika Deskriptif

Variabel	Rata-rata	Varians	Minimum	Maximum
Y	2424	5955193	113	11692
X_1	5,189	2,404	5,150	8,500
X_2	22120	36709	16684	86932
X_3	10528	10461	8186	67078

3.2. Analisis Regresi Poisson

3.2.1. Estimasi parameter

Hasil estimasi parameter disajikan pada tabel berikut:

Tabel 3. Estimasi Parameter

Variabel	Koefisien
Intersep	7,2140
X_1	-0,04609
X_2	0,000027
X_3	0,000
	AIC = 17964
	RMSE= 1001,165

Berdasarkan Tabel 4.4 menujukkan model regresi *Poisson* yang diperoleh sebagai berikut.

$$\hat{\mu} = exp(7{,}2140 - 0{,}04609X_1 + 0{,}000027X_2 + 0{,}000X_3)$$

3.2.2. Uji Serentak dan Uji Parsial

Statistik uji serentak regresi Poisson menunjukkan bahwa minimal terdapat satu variabel prediktor yang berpengaruh signifikan terhadap jumlah kasus pneumonia pada anak di Jawa Timur Tahun 2022. Adapun hasil statistik uji parsial disajikan pada tabel berikut:

Tabel 4. Uji Parsial Regresi Poisson

Parameter	$\lvert Z_{hitung}\rvert$	$z_{(0.025)}$	P-*value*	Keputusan
β_0	460,816	1,96	0,000	Tolak H_0
β_1	18,719		0,000	Tolak H_0
β_2	119,914		0,000	Tolak H_0
β_3	1,242		0,214	Gagal tolak H_0

3.2.3. *Overdispersion*

Hasil pemeriksaan *overdispersion* disajikan pada tabel berikut.

Tabel 5. Pemeriksaan *Overdispersion*

Deviance	*df*	*Deviance/df*
17.605	34	517,794

Tabel 5 menunjukkan bahwa rasio nilai *deviance* terhadap derajat bebas adalah sebesar 517,794 yang lebih besar dari 1. Hal tersebut mengindikasikan adanya *overdispersion.*

3.3. Analisis Regresi Binomial Negatif Dengan Seluruh Variabel Prediktor X_1, X_2, dan X_3

Pemodelan regresi binomial negatif dilakukan dengan menggunakan nilai inisial θ sebesar 2,686.

3.3.1. Estimasi Parameter

Model yang dihasilkan sebagai berikut.

$$\hat{\mu} = exp(6{,}888 - 0{,}0102X_1 + 0{,}00003\, X_2 + 0{,}000004X_3)$$

3.3.2. Uji Serentak dan Uji Parsial

Statistik uji serentak regresi binomial negatif menunjukkan bahwa minimal terdapat satu variabel prediktor yang berpengaruh signifikan terhadap jumlah kasus pneumonia pada anak di Jawa Timur Tahun 2022. Adapun hasil statistik uji parsial disajikan pada tabel berikut:

Tabel 6. Uji Parsial Binomial Negatif

Parameter	$\lvert Z_{hitung}\rvert$	$Z_{(0.025)}$	P-*value*	Keputusan
β_0	17,002	1,96	0,000	Tolak H_0
β_1	0,159		0,874	Gagal Tolak H_0
β_2	4,289		0,000	Tolak H_0
β_3	0,385		0,703	Gagal tolak H_0

3.4. Analisis Regresi Binomial Negatif dengan Menggunakan Variabel Prediktor yang Signifikan

Pemodelan regresi binomial negatif selanjutnya dilakukan dengan menggunakan nilai inisial θ sebesar 2,686.

3.4.1. Estimasi Parameter

Model yang dihasilkan sebagai berikut.

$$\hat{\mu} = exp(6{,}888 + 0{,}00003\, X_2)$$

3.4.2. Uji Serentak dan Uji Parsial

Statistik uji serentak menunjukkan bahwa jumlah balita batuk atau kesukaran bernapas berpengaruh signifikan terhadap jumlah kasus pneumonia pada anak di Jawa Timur Tahun 2022. Adapun hasil statistik uji parsial disajikan pada tabel berikut:

Tabel 7. Uji Parsial Binomial Negatif

Parameter	$\lvert Z_{hitung} \rvert$	$Z_{(0.025)}$	P-*value*	Keputusan
β_0	51,045	1,96	0,000	Tolak H_0
β_2	7,124		0,000	Tolak H_0

3.5. Analisis *Generalized Poisson Regression* dengan Seluruh Variabel Prediktor X_1, X_2, dan X_3

3.5.1. Estimasi Parameter

Model yang dihasilkan sebagai berikut.

$$\hat{\mu} = exp(4{,}030 - 0.0736X_1 + 2{,}8x10^{-5}\,X_2 + 5{,}4x10^{-7}X_3)$$

3.5.2. Uji Serentak dan Uji Parsial

Statistik uji serentak menunjukkan bahwa minimal terdapat satu faktor yang berpengaruh signifikan terhadap jumlah kasus pneumonia pada anak di Jawa Timur Tahun 2022. Adapun hasil statistik uji parsial disajikan pada tabel berikut:

Tabel 8. Uji Parsial Model GPR

Parameter	$\lvert Z_{hitung} \rvert$	$Z_{(0.025)}$	P-*value*	Keputusan
β_0	11,912	1,96	0,000	Tolak H_0
β_1	1,408		0,159	Gagal Tolak H_0
β_2	5,530		0,000	Tolak H_0
β_3	0,079		0,937	Gagal tolak H_0

3.6. Analisis *Generalized Poisson Regression* dengan Menggunakan Variabel Prediktor yang Signifikan

3.6.1. Estimasi Parameter

Model yang dihasilkan sebagai berikut.

$$\hat{\mu} = exp(3{,}58 + 3{,}05x10^{-5}\,X_2)$$

3.6.2. Uji Parsial

Adapun hasil statistik uji parsial disajikan pada tabel berikut:

Tabel 9. Uji Parsial Model GPR

Parameter	$\lvert Z_{hitung} \rvert$	$Z_{(0.025)}$	P-*value*	Keputusan
β_0	51,045	1,96	0,000	Tolak H_0
β_2	7,124		0,000	Tolak H_0

3.6. Pemilihan Model Terbaik Menggunakan Kriteria AIC dan RMSE

Perbandingan model dapat dilihat pada tabel berikut.

Tabel 10. Pemilihan Model Terbaik

Model Regresi	AIC	RMSE
Poisson	17.963,59	978,600
Binomial Negatif dengan seluruh variabel prediktor (X_1, X_2, X_3)	638,9	1001,165
Binomial Negatif dengan variabel prediktor yang signifikan (X_2)	635,01	1060,727
Generalized Poisson Regression dengan seluruh variabel prediktor (X_1, X_2, X_3)	638,197	1497.656
Generalized Poisson Regression dengan variabel prediktor yang signifikan (X_2)	635,537	1231,675

Model yang optimal ditentukan berdasarkan nilai AIC terkecil. Berdasarkan nilai AIC yang tercantum pada Tabel 10, dapat disimpulkan bahwa model regresi binomial negatif yang hanya menggunakan variabel prediktor signifikan memberikan performa terbaik dalam memodelkan jumlah kasus pneumonia di Jawa Timur tahun 2022.

4. KESIMPULAN

Berdasarkan hasil penelitian tersebut, maka dapat disimpulkan dalam penelitian sebagai berikut:

- Kasus pneumonia pada anak di Jawa Timur Tahun 2022 masih tinggi terutama di Kota Surabaya. Sementara persentase Berat Bayi Lahir Rendah BBRL) masih tinggi. Jumlah balita batuk atau kesukaran bernapas diperoleh rata-rata kasus tinggi dengan varians tinggi berarti tidak semua wilayah memiliki kasus yang sama besar. Lalu persentase penderita COVID-19 memiliki rata-rata dan varians yang hampir sama besar artinya, secara umum wilayah cenderung memiliki jumlah kasus yang tidak terlalu jauh berbeda satu sama lain, meskipun tetap ada varians.
- Model terbaik yang diperoleh adalah modeldengan metode Regresi Binomial Negatif. Variabel yang mempengaruhi pneumonia pada anak secara signifikan adalah jumlah balita batuk atau kesukaran bernapas di Jawa Timur Tahun 2022.

ACKNOWLEDGMENTS
Penulis menyatakan bahwa tidak terdapat individu atau pihak yang memberikan kontribusi secara personal dalam penelitian ini.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Muhammad Zulfadhli	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT
Penulis menyatakan tidak ada konflik kepentingan.

DATA AVAILABILITY
Ketersediaan data tidak berlaku untuk makalah ini karena tidak ada data baru yang dibuat atau dianalisis dalam studi ini.

REFERENCES
[1] A. M. Achmad, M. Zulfadhli, dan N. Nurfaudzan, "Binary logistic regression on the status of unemployment rate in West Java Province in 2023," *Likelihood: Journal of Statistics and Its Application*, vol. 1, no. 1, pp. 30–34, 2025.
[2] A. Aditia, "Covid-19: Epidemiologi, virologi, penularan, gejala klinis, diagnosa, tatalaksana, faktor risiko, dan pencegahan," *Jurnal Penelitian Perawat Profesional*, 2021. [Online]. Available: https://jurnal.globalhealthsciencegroup.com/index.php/JPPP/article/view/574
[3] A. Agresti, *Foundations of Linear and Generalized Linear Models*. Hoboken, NJ, USA: Wiley, 2015.
[4] S. M. Asnawati, *Analisis Inovasi dan Produk*. 2022.
[5] Badan Pusat Statistik (BPS), "Persentase penduduk miskin," 2020.
[6] Badan Pusat Statistik (BPS), "Tingkat Pengangguran Terbuka (TPT)," 2020. [Online]. Available: https://sirusa.bps.go.id/sirusa/index.php/indikator/44
[7] S. S. Bogusławski et al., "The outcomes of COVID-19 pneumonia in children: Clinical, radiographic, and pulmonary function assessment," *Pediatric Pulmonology*, 2023.
[8] R. Cahyandari, "Pengujian overdispersi pada model regresi Poisson (Studi kasus: kecelakaan lalu lintas mobil penumpang di Provinsi Jawa Barat)," *Jurnal Statistika*, 2014.
[9] Mayo Clinic, "Cough," 2022. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/cough
[10] I. I. Daqiqil, *Teori, Studi Kasus, dan Implementasi Menggunakan Python*. 2021.
[11] M. I. Dirwan, N. W. Surtini, dan M. Zulfadhli, "Learning by gaming: Extramural English gaming effect on Indonesian senior high students' pragmatic competence," *Journal of Pragmatics Research*, vol. 7, no. 1, pp. 1–20, 2025.
[12] A. N. Hidayati dan B. W., "Pelayanan puskesmas berbasis manajemen terpadu balita sakit dengan kejadian pneumonia balita," *Jurnal Kesehatan Masyarakat*, 2022.
[13] V. Silvia, *Statistika Deskriptif*. 2021.
[14] L. D. Martias, "Statistika deskriptif sebagai kumpulan informasi," *FIHRIS: Jurnal Ilmu Perpustakaan dan Informasi*, pp. 42–57, 2021.
[15] R. D. Mason, *Teknik Statistika untuk Bisnis dan Ekonomi*. Jakarta, Indonesia, 1996.
[16] J. F. Murray, *Textbook of Respiratory Medicine*. Amsterdam, Netherlands: Elsevier, 2016.
[17] Musbahuddin, *Analisis Data Penelitian dengan Statistik*. 2013.

[18] B. W. Otok, M. Zulfadhli, R. D. Pangesti, M. I. Kurniawan, A. E. P. Haryanto, dan I. Kurniawan, "Comparison of binary probit regression and Fourier series nonparametric logistic regression in modeling diabetes status at Hajj General Hospital Surabaya," *Barekeng*, vol. 20, no. 1, pp. 255–270, 2026.
[19] World Health Organization (WHO), "Pneumonia in children," 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia
[20] H. Purbaya, "Faktor-faktor yang berhubungan dengan kejadian pneumonia pada balita," 2023.
[21] F. Rachmillah, "Prediksi jumlah kasus baru kusta dengan metode geographically weighted Poisson regression (GWPR)," *Jurnal Stikes Mandala*, 2015.
[22] R. T. Pamungkas, *Bahan Ajar Matematika Mememo (Mean, Median, Modus) untuk Siswa Kelas 6 SD*. CV Graf Literasi, 2020.
[23] S. Saida, M. Zulfadhli, dan M. Jurais, "Analisis faktor-faktor yang mempengaruhi *vaccine hesitancy* (keragu-raguan vaksin) pada mahasiswa di era pandemi COVID-19," *Preventif: Jurnal Kesehatan Masyarakat*, vol. 13, no. 1, pp. 144–154, 2022.
[24] A. Saifuddin, *Buku Acuan Nasional Pelayanan Kesehatan Maternal dan Neonatal*. Jakarta, Indonesia: Yayasan Bina Pustaka Sarwono Prawirohardjo, 2020.
[25] S. Santoso, *Seri Solusi Bisnis Berbasis TI*. 2007.
[26] Y. R. Maharani dan S. Shinta, "Faktor risiko frekuensi kunjungan balita dengan kasus batuk," *Jurnal Kesehatan Poltekkes Palembang*, 2020.
[27] D. Siagian, *Metode Statistik*. 2006.
[28] S. R. Wahyuningrum dan A. Muhlis, *Statistika Pendidikan Edisi Kedua (Dengan Statistika Al-Qur'an)*. 2019.
[29] M. H. Sriningsih, "Penanganan multikolinearitas dengan menggunakan analisis regresi komponen utama pada kasus impor beras di Provinsi Sulawesi Utara," *Jurnal Ilmiah Sains*, vol. 18, no. 1, pp. 18–24, 2018.
[30] Dinas Kesehatan Provinsi Jawa Timur, *Profil Kesehatan Provinsi Jawa Timur*. Jawa Timur, Indonesia, 2022.
[31] K. A. Yulianingsih, "Penerapan regresi Poisson untuk mengetahui faktor-faktor yang memengaruhi jumlah siswa SMA/SMK yang tidak lulus UN di Bali," *E-Jurnal Matematika*, vol. 1, no. 1, pp. 59–63, 2012.
[32] B. Yuniarto, *Analisis Regresi*. 2016.
[33] M. Zulfadhli, I. N. Budiantara, V. Ratnasari, A. S. Suriaslan, dan R. C. Dhewy, "A statistical inference framework for FSNBLR: Modeling underdeveloped regional status in Eastern Indonesia," *MethodsX*, Art. no. 103746, 2025.
[34] M. Zulfadhli, I. N. Budiantara, V. Ratnasari, dan A. S. Suriaslan, "Semiparametric regression estimator of Fourier series for categorical data," *IAENG International Journal of Applied Mathematics*, vol. 55, no. 7, pp. 2157–2164, 2025.