

BAB I PENDAHULUAN

1.1 Latar Belakang

Fokus utama dalam regresi berganda terletak pada analisis hubungan antara satu variabel respon dengan beberapa variabel prediktor (Sihombing, 2022). Metode ini memberikan gambaran kuantitatif mengenai seberapa besar pengaruh variabel prediktor terhadap variabel respon. Penerapan regresi berganda telah memberikan kontribusi besar dalam berbagai bidang seperti ekonomi, sosial, dan ilmu alam. Namun, dalam beberapa aplikasinya, peubah respon yang akan dianalisis dapat berupa data diskrit. Salah satu contoh di mana peubah responnya diskrit adalah banyaknya kejadian yang jarang terjadi (*rare event*). Model regresi yang dapat digunakan untuk menjelaskan hubungan antara peubah respon yang berupa data diskrit dengan peubah prediktor berupa data diskrit, kontinu, kategorik, atau campuran adalah model regresi Poisson.

Syarat utama penggunaan model regresi Poisson adalah kesamaan antara nilai rata-rata dan variansi dari variabel respon yang disebut juga dengan equidispersi (Pelangi dkk., 2020). Namun, kondisi ini jarang terpenuhi. Secara umum, data yang dimiliki menunjukkan variansi yang lebih besar dari rata-rata, kondisi yang disebut dengan overdispersi. Kondisi overdispersi mengakibatkan ketidaksesuaian model Poisson dengan data sehingga estimasi parameter menjadi tidak reliabel. Sebagai solusi, model Binomial Negatif (*Negative Binomial Regression/NBR*) digunakan sebagai generalisasi dari model Poisson. Model ini mampu menangani masalah overdispersi dan menghasilkan estimasi parameter yang lebih stabil serta akurat.

Model NBR efektif dalam menangani permasalahan overdispersi. Namun, model ini tetap memiliki keterbatasan apabila terjadi multikolinearitas, yaitu kondisi ketika terdapat hubungan yang sangat kuat antar variabel prediktor. Multikolinearitas dapat mengganggu interpretasi koefisien regresi dan menyebabkan ketidakstabilan estimasi parameter (Hendrianto dkk., 2023). Salah satu metode yang umum digunakan untuk mengatasi masalah multikolinearitas adalah *Principal Component Analysis* (PCA). PCA berfungsi untuk mereduksi dimensi data dan menghilangkan korelasi antar variabel prediktor dengan mengubahnya menjadi sejumlah komponen utama yang saling bebas (*orthogonal*), namun tetap mempertahankan sebagian besar informasi dalam data (Maulana dkk., 2024).

Metode NBR-PCA dikembangkan sebagai penyempurnaan dari pendekatan sebelumnya, yaitu Poisson-PCA. Model ini menggabungkan regresi Poisson dengan analisis komponen utama untuk menangani multikolinearitas pada data *count*. Poisson-PCA bekerja dengan mengaplikasikan PCA terhadap variabel prediktor sebelum dilakukan pemodelan regresi Poisson sehingga model yang dihasilkan lebih stabil dan *interpretable* meskipun terjadi korelasi antar prediktor. Salah satu penelitian yang



e Poisson PCA adalah Kenney et al., (2021) mengusulkan metode menangani bias pada data *count* yang dipengaruhi residual bertipe dalam konteks analisis mikrobioma. Metode ini menggunakan ametrik yang memperbaiki estimasi variansi dari rata-rata Poisson un sesudah transformasi *non-linear* seperti logaritmik. Namun, on mengasumsikan dispersi data yang setara antara rata-rata dan urang optimal ketika data mengalami overdispersi, sehingga dalam

kasus tertentu dibutuhkan alternatif seperti NBR-PCA.

Pendekatan NBR-PCA sangat relevan digunakan dalam menganalisis data kasus pneumonia pada balita. Pneumonia merupakan salah satu penyakit infeksi saluran pernapasan bawah yang paling mematikan bagi balita, terutama di negara berkembang. Penyakit ini dipengaruhi oleh berbagai faktor sosial-ekonomi dan kesehatan. Oleh karena itu, pemodelan menggunakan NBR dapat menangani data dengan sifat *count* yang menunjukkan overdispersi, sedangkan PCA dapat membantu mengatasi permasalahan multikolinearitas yang umum dijumpai pada data dengan banyak variabel prediktor. Berdasarkan data UNICEF, secara global terdapat lebih dari 1.400 kasus pneumonia per 100.000 anak, atau setara dengan 1 kasus untuk setiap 71 anak per tahun. Di Indonesia sendiri, pneumonia merupakan penyebab utama kematian balita. Setiap tahun diperkirakan lebih dari 2 juta balita meninggal akibat pneumonia (Kementerian Kesehatan Republik Indonesia, 2023). Provinsi Jawa Timur menempati peringkat kedua sebagai provinsi dengan jumlah kasus pneumonia terbanyak di Indonesia. Tingginya angka kejadian pneumonia pada balita ini berkaitan dengan beberapa faktor risiko seperti status gizi, pemberian ASI, rendahnya cakupan imunisasi (seperti DPT-Hb dan imunisasi dasar lengkap), kualitas udara yang buruk, serta keterbatasan akses terhadap sanitasi yang layak (Afriani & Oktavia, 2021; Faisal dkk., 2024).

Beberapa penelitian terdahulu telah menggunakan Regresi Binomial Negatif dalam mengatasi masalah overdispersi pada data. Sebagai salah satu penerapan nyata, Winata, (2022) mengidentifikasi faktor-faktor yang memengaruhi Angka Kematian Ibu (AKI) di Kota Bandung menggunakan data *count*. Pada tahap awal, digunakan model regresi Poisson, tetapi hasil pengujian mengungkap adanya overdispersi sehingga model tersebut kurang sesuai. Peneliti kemudian menerapkan model regresi binomial negatif sebagai alternatif. Sebagai pendekatan lain dalam analisis data, (Khikmah, 2021) menerapkan *Principal Component Analysis* pada data cuaca BMKG periode April-September 2020 untuk mereduksi dimensi dan menentukan tiga faktor dominan sehingga ditemukan bahwa suhu, kelembapan, serta lama penyinaran matahari yang menjelaskan 87% variansi penyebaran COVID-19 di Surabaya. Dengan pendekatan yang serupa, studi ini bertujuan untuk mengeksplorasi dan mengidentifikasi faktor-faktor yang mempengaruhi jumlah kasus pneumonia pada balita melalui metode *Negative Binomial Regression* yang dikombinasikan dengan *Principal Component Analysis* (NBR-PCA).

Berdasarkan uraian sebelumnya, peneliti tertarik untuk mengkaji hal tersebut dengan judul “**Pemodelan *Negative Binomial Regression* Pada Komponen Utama untuk Kasus Multikolinearitas (Studi Kasus: Pneumonia pada Balita di Provinsi Jawa Timur 2023)**”.



Optimized using
trial version
www.balesio.com

ialah

salahan yang telah diuraikan sebelumnya, Batasan masalah pada kan sebagai berikut:

menggunakan metode *Maximum Likelihood Estimation* (MLE) dan ison untuk menentukan parameter dalam model *Negative Binomial* (NBR).

berfokus pada analisis jumlah kasus pneumonia pada balita dan

faktor-faktor yang mempengaruhinya di Provinsi Jawa Timur berdasarkan data tahun 2023.

3. Kriteria pemilihan komponen utama dalam PCA dibatasi pada *eigenvalue* lebih dari 1 dan *explained variance* minimal 80%.

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Memperoleh model *Negative Binomial Regression Principal Component Analysis* untuk data kasus pneumonia pada Balita di Provinsi Jawa Timur yang mengalami multikolinearitas dan overdispersi.
2. Memperoleh faktor-faktor yang berpengaruh terhadap jumlah kasus pneumonia pada Balita di Jawa Timur.

1.4 Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Memberikan kontribusi dalam pengembangan metode analisis regresi yang mengintegrasikan NBR dan PCA untuk data overdispersi dengan multikolinearitas.
2. Memberikan informasi yang lebih akurat mengenai faktor-faktor risiko pneumonia pada balita di Provinsi Jawa Timur.
3. Membantu pemerintah dan pemangku kebijakan dalam merancang intervensi Kesehatan yang lebih tepat sasaran berdasarkan faktor-faktor yang berpengaruh signifikan terhadap kasus pneumonia pada balita.
4. Menambah referensi bagi peneliti lain dalam kebijakan dalam bidang Kesehatan masyarakat, statistika terapan, dan epidemiologi yang memanfaatkan pendekatan serupa.

1.5 Landasan Teori

1.5.1 Analisis Regresi Linear Berganda

Regresi linear (*linear regression*) merupakan metode statistik yang digunakan untuk memodelkan hubungan linear antara 1 variabel dan 1 atau lebih variabel. Jika hanya digunakan 1 variabel independen dalam model, maka teknik ini disebut sebagai regresi linear sederhana (*simple linear regression*), sedangkan jika yang digunakan adalah beberapa variabel independen, teknik disebut regresi linear ganda (*multiple linear regression*) (Mufidah & Basuki, 2023). Adapun bentuk umum yang digunakan dalam regresi linear berganda ditunjukkan pada Persamaan (1) (Wijayanti & Prasetyo, 2024).

$$y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n + e \quad (1)$$

Keterangan:



pon
dikator

koefisien estimasi

1.5.2 Regresi Poisson

Distribusi Poisson adalah distribusi probabilitas yang digunakan untuk memodelkan kejadian langka yang terjadi dalam suatu interval waktu tertentu, dengan asumsi bahwa kejadian-kejadian tersebut terjadi secara independen dan rata-rata kejadiannya konstan dalam setiap interval waktu tertentu (Nurhidaya dkk., 2023). Distribusi Poisson dapat dinyatakan secara matematis dengan notasi $y \sim \text{Poisson}(\mu)$, di mana y merupakan variabel acak yang mengikuti distribusi Poisson dengan parameter μ di mana fungsi distribusi yang dinyatakan pada Persamaan (2).

$$f(y; \mu) = \frac{e^{-\mu} \mu^y}{y!}; y = 0, 1, 2, \dots \quad (2)$$

dengan μ merupakan rata-rata variabel random y yang berdistribusi Poisson dengan nilai rata-rata dan variansi dari y mempunyai nilai lebih dari nol. Dalam analisis regresi, hubungan antara respon dengan variabel-variabel prediktornya dinyatakan dalam suatu model yang disebut dengan model regresi. Hubungan antara variabel prediktor dan variabel respon dalam model tersebut didefinisikan oleh parameter-parameter regresi yang dinotasikan dengan β . Berdasarkan kajian Myers (1990), regresi Poisson merupakan salah satu teknik analisis yang efektif untuk memodelkan hubungan antara variabel respon yang berdistribusi Poisson dengan satu atau lebih variabel prediktor. Model regresi Poisson dirumuskan pada Persamaan (3).

$$y_i = \mu_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (3)$$

dengan

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

Keterangan:

- y_i : jumlah kejadian ke- i , $i = 1, 2, \dots, n$
 μ_i : rata-rata jumlah kejadian yang terjadi dalam selang waktu ke- i
 ε_i : residual ke- i

1.5.3 Overdispersi

Overdispersi merupakan kondisi di mana variansi data jauh lebih besar dibandingkan dengan rata-ratanya. Penggunaan model regresi Poisson pada data yang mengalami overdispersi dapat menghasilkan estimasi parameter yang bias dan tidak efisien. Hal ini disebabkan oleh ketidakmampuan model regresi dalam mengakomodasi parameter dispersi yang melekat pada data (Hida dkk., 2022). Parameter dispersi (ϕ) dapat diestimasi menggunakan nilai *deviance* dibagi dengan derajat bebasnya. Secara matematis, hal ini dinyatakan dalam Persamaan (4).

$$\phi = \frac{\text{nilai deviance}}{df} \quad (4)$$

dengan nilai *deviance* didefinisikan sebagai berikut:



$$G^2 = 2 \sum_{i=1}^n y_i \ln \left(\frac{y_i}{\lambda_i} \right)$$

= 1, maka asumsi equidispersi terpenuhi pada data. Namun, jika $\phi > 1$, maka terjadi overdispersi pada data, dan jika nilai $\phi < 1$, maka underdispersi pada data. Overdispersi dapat ditulis sebagai berikut:

$$\text{Var}(y) > E(y)$$

Untuk menangani masalah *over dispersi* dapat menggunakan

pendekatan dengan sebaran binomial negatif.

1.5.4 Distribusi Binomial Negatif

Distribusi binomial negatif digunakan sebagai pendekatan terhadap suatu percobaan yang ingin melihat besarnya peluang terjadinya r keberhasilan setelah sebelumnya muncul sejumlah kejadian yang gagal (Wahyudi & Palit, 2022). Suatu variabel acak Y didefinisikan mempunyai distribusi binomial negatif jika fungsi peluangnya diberikan dalam Persamaan (5).

$$\begin{aligned} P(y) &= \binom{y+r-1}{y} p^r (1-p)^y \\ &= \frac{(y+r-1)!}{(r-1)!y!} p^r (1-p)^y, \quad y = 0, 1, 2, \dots \end{aligned} \quad (5)$$

Keterangan:

r : banyaknya kejadian berhasil

y : banyaknya kejadian yang gagal sebelum terjadinya kejadian berhasil ke-

p : peluang berhasil

$(1-p)$: peluang gagal

1.5.5 Negative Binomial Regression

Regresi binomial negatif merupakan suatu model regresi yang digunakan untuk menganalisis hubungan antara variabel respon dengan satu atau lebih variabel prediktor, terutama pada kondisi di mana data menunjukkan variansi yang lebih besar dari yang diharapkan (Widyaningsih dkk., 2021). Fungsi peluang binomial negatif dapat diturunkan dari model Poisson dengan Gamma yang mempunyai rata-rata sebesar 1. Persamaan (6) menunjukkan fungsi peluang distribusi Poisson.

$$f(y, \mu, x) = \frac{e^{-\mu_i x_i} (\mu_i x_i)^{y_i}}{y_i!} \quad (6)$$

dan Persamaan (7) merupakan fungsi peluang distribusi Gamma

$$g(x_i) = g(\beta, \alpha, x_i) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x_i^{\alpha-1} e^{-\frac{x_i}{\beta}} \quad (7)$$

Berdasarkan fungsi peluang distribusi Poisson dan distribusi Gamma tersebut diperoleh fungsi Binomial Negatif dengan cara sebagai berikut dalam Persamaan (8).

$$f(y, \mu, x) = \int_0^\infty \frac{e^{-\mu_i x_i} (\mu_i x_i)^{y_i}}{y_i!} g(x_i) dx_i \quad (8)$$

$g(x_i)$ merupakan fungsi Gamma dengan $\mu = 1$. Untuk membentuk suatu model regresi pada distribusi *mixture* Poisson-Gamma dinyatakan dalam bentuk $\mu = \alpha\beta$ dan $\theta = \frac{1}{\alpha}$ maka $\beta = \theta$. Dari penyelesaian integral pada Persamaan (8) diperoleh Persamaan (9).



$$= \frac{\Gamma(y_i + \frac{1}{\theta})}{y_i! \Gamma(\frac{1}{\theta})} \left(\frac{1}{1 + \theta\mu_i} \right)^{\frac{1}{\theta}} \left(\frac{\theta\mu_i}{1 + \theta\mu_i} \right)^{y_i}, \quad y_i = 0, 1, 2, \dots \quad (9)$$

dan lanjut akan diperoleh:

$$\left(y_i + \frac{1}{\theta} - 1 \right) \left(\frac{1}{1 + \theta\mu_i} \right)^{\frac{1}{\theta}} \left(\frac{\theta\mu_i}{1 + \theta\mu_i} \right)^{y_i}, \quad y_i = 0, 1, 2, \dots \quad (10)$$

Bentuk *mixture* Poisson-Gamma pada Persamaan (10) menyerupai distribusi Binomial Negatif $f(y; p, r) = \binom{y+r-1}{r-1} p^r (1-p)^y$ apabila $r = \frac{1}{\theta}$ dan $p_i = \left(\frac{1}{1+\theta\mu_i}\right)$

Link function yang digunakan dalam regresi Binomial Negatif adalah $\ln$ sehingga model regresi Binomial Negatif dapat dinyatakan sebagai berikut:

$$\ln = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} ; i = 1, 2, \dots, n \quad (11)$$

1.5.6 Estimasi Parameter *Negative Binomial Regression*

Estimasi Parameter dalam *Negative Binomial Regression* menggunakan metode *maximum likelihood* dengan prosedur iterasi *Newton Raphson*. Parameter yang akan diduga adalah parameter β dan parameter dispersi (θ). Parameter θ diestimasi untuk menilai pemenuhan asumsi equidispersi dalam model. Metode ini membutuhkan turunan pertama dan kedua dari fungsi *likelihood*. y_i mempunyai fungsi massa probabilitas sebaran binomial negatif, yaitu Persamaan (12).

$$f(y; \mu, \theta) = \frac{\Gamma(y_i + \frac{1}{\theta})}{\Gamma(\frac{1}{\theta}) \Gamma(y_i + 1)} \left(\frac{1}{1 + \theta\mu_i}\right)^{\frac{1}{\theta}} \left(\frac{\theta\mu_i}{1 + \theta\mu_i}\right)^{y_i}, y_i = 0, 1, 2, \dots \quad (12)$$

θ adalah parameter dispersi yang didefinisikan sebagai $1/\alpha$ dengan α merupakan parameter bentuk dari sebaran gamma. Menurut (Pagui dkk., 2022), karena fungsi saling bebas maka fungsi *likelihood* dapat dituliskan sebagai Persamaan (13).

$$\begin{aligned} L(\beta, \theta) &= \prod_{i=1}^n \frac{\Gamma(y_i + \frac{1}{\theta})}{\Gamma(\frac{1}{\theta}) \Gamma(y_i + 1)} \left[\frac{1}{1 + \theta\mu_i}\right]^{\frac{1}{\theta}} \left[\frac{\theta\mu_i}{1 + \theta\mu_i}\right]^{y_i} \\ &= \prod_{i=1}^n \left[\left(\prod_{r=0}^{y_i-1} (r + \theta^{-1}) \right) \frac{1}{y_i!} \left[\frac{1}{1 + \theta\mu_i}\right]^{\frac{1}{\theta}} \left[\frac{\theta\mu_i}{1 + \theta\mu_i}\right]^{y_i} \right] \end{aligned} \quad (13)$$

dengan $\frac{\Gamma(y+\frac{1}{\theta})}{\Gamma(\frac{1}{\theta})} = \prod_{r=0}^{y-1} (r + \theta^{-1})$ yang dapat dibuktikan dengan menggunakan sifat rekurens dari fungsi gamma pada Persamaan (14).

Deret rekurens untuk $\Gamma(a + n)$:

$$\Gamma(a + n) = (a + n - 1)(a + n - 2) \dots (a) \Gamma(a)$$

$$\frac{\Gamma(a + n)}{\Gamma(a)} = (a)(a + 1)(a + 2) \dots (a + n - 1) \quad (14)$$

$$\frac{\Gamma(a + n)}{\Gamma(a)} = \prod_{r=0}^{y-1} (r + a)$$

dengan $a = \frac{1}{\theta}$ dan $n = y_i$, sehingga:

$$\frac{\Gamma(y_i + \frac{1}{\theta})}{\Gamma(\frac{1}{\theta})} = \prod_{r=0}^{y_i-1} \left(r + \frac{1}{\theta}\right)$$

↳ bilangan bulat *non-negative*, maka $\Gamma(y_i + 1) = y_i!$.

(15)



$$l(\beta, \theta) = \sum_{i=1}^n \left(\prod_{r=0}^{y_i-1} \ln(r + \theta^{-1}) - \ln(y_i!) + \frac{1}{\theta} \ln \left[\frac{1}{1 + \theta \mu_i} \right] + y_i \ln \left[\frac{\theta \mu_i}{1 + \theta \mu_i} \right] \right)$$

Turunan pertama fungsi *log-likelihood* terhadap koefisien β dinyatakan dalam Persamaan (16).

$$\begin{aligned} \frac{\partial L(\beta, \theta)}{\partial \beta_0} &= \sum_{i=1}^n \left[y_i - \left(y_i + \frac{1}{\theta} \right) \left(\frac{\theta \mu_i}{1 + \theta \mu_i} \right) \right] = \sum_{i=1}^n \left(\frac{y_i \mu_i}{1 + \theta \mu_i} \right) = 0 \\ &\vdots \\ \frac{\partial L(\beta, \theta)}{\partial \beta_p} &= \sum_{i=1}^n \left[y_i x_{ip} - \left(y_i + \frac{1}{\theta} \right) \left(\frac{\theta \mu_i x_{ip}}{1 + \theta \mu_i} \right) \right] \\ &= \sum_{i=1}^n \left(\frac{(y_i - \mu_i) x_{ip}}{1 + \theta \mu_i} \right) = 0 \end{aligned} \quad (16)$$

Bentuk persamaan matriks dari turunan pertama fungsi *likelihood* terhadap parameter β yaitu: $q = X^T W z$,

X : matriks ($n \times c$) dengan n merupakan banyaknya observasi dan c merupakan banyaknya peubah $k+1$

W : matriks *weight* diagonal ke- i dan berukuran $n \times n$

z : vektor matriks dengan baris ke- i berukuran $p \times 1$

Untuk mencari nilai W digunakan Persamaan (17).

$$W_i = \frac{\mu_i}{1 + \theta \mu_i} \quad (17)$$

dan

$$z_i = \frac{y_i - \mu_i}{\mu_i} \quad (18)$$

Turunan pertama dari fungsi *log-likelihood* terhadap parameter dispersi θ dinyatakan dalam Persamaan (19).

$$\begin{aligned} f'(\theta) &= \frac{\partial L(\beta, \theta)}{\partial \beta_0} = \sum_{i=1}^n \left[\frac{1}{\theta^2} \sum_{r=0}^{y_i-1} \frac{1}{r + \theta^{-1}} + \frac{y_i}{\theta} + \frac{1}{\theta^2} \ln(1 + \theta \mu_i) - \frac{(\theta^{-1} + y_i) \mu_i}{1 + \theta \mu_i} \right] = 0 \\ f'(\theta) &= \frac{\partial L(\beta, \theta)}{\partial \beta_0} \\ &= \sum_{i=1}^n \left[\frac{1}{\theta^2} \sum_{r=0}^{y_i-1} \frac{1}{r + \theta^{-1}} + \frac{1}{\theta^2} \ln(1 + \theta \mu_i) - \frac{y_i - \mu_i}{\theta(1 + \theta \mu_i)} \right] = 0 \end{aligned} \quad (19)$$

Turunan parsial kedua fungsi *log-likelihood* terhadap parameter koefisien regresi β yaitu:

$$\begin{aligned} \frac{\partial^2 L(\beta, \theta)}{\partial \beta_0^2} &= - \sum_{i=1}^n \left[\frac{(1 + \theta y_i) \mu_i}{(1 + \theta \mu_i)^2} \right] \\ \frac{\partial^2 L(\beta, \theta)}{\partial \beta_j \partial \beta_u} &= \sum_{i=1}^n \left[\frac{-\mu_i x_{ij} (1 + \theta \mu_i) - (y_i - \mu_i) (\theta \mu_i x_{ij})}{(1 + \theta \mu_i)^2} \right] \\ &\quad \frac{x_{ij} (1 + \theta y_i)}{(1 + \theta \mu_i)^2} \end{aligned}$$



arsial pertama dari $L(\beta, \theta)$ terhadap β_j , $j \leq p$ adalah $\frac{\partial L(\beta, \theta)}{\partial \beta_j} =$

aka turunan parsial kedua terhadap β_u , $u \leq p$ adalah

$$\frac{\partial^2 L(\boldsymbol{\beta}, \theta)}{\partial \beta_u \partial \beta_k} = \sum_{i=1}^n \left[\frac{-\mu_i x_{ik} x_{iu} (1 + \theta \mu_i) - (y_i - \mu_i) (\theta \mu_i x_{ik} x_{iu})}{(1 + \theta \mu_i)^2} \right]$$

$$\frac{\partial^2 L(\boldsymbol{\beta}, \theta)}{\partial \beta_u \partial \beta_k} = - \sum_{i=1}^n \left[\frac{\mu_i x_{ik} x_{iu} (1 + \theta y_i)}{(1 + \theta \mu_i)^2} \right]$$

Ekspektasi turunan kedua *log-likelihood* dapat dinyatakan pada Persamaan (20)

$$E \left(- \frac{\partial^2 L(\boldsymbol{\beta}, \theta)}{\partial \beta_u \partial \beta_k} \right) = - \sum_{i=1}^n \left[\frac{\mu_i x_{ik} x_{iu}}{(1 + \theta \mu_i)} \right] \quad (20)$$

Jika Persamaan (20) dinyatakan dalam matriks $\mathbf{I}$ yaitu matriks yang mengandung ekspektasi negatif dari turunan kedua *log-likelihood* maka $\mathbf{I} = \mathbf{X}^T \mathbf{W} \mathbf{X}$, dengan $\mathbf{X}$ adalah matriks variabel prediktor, $\mathbf{W}$ adalah matriks diagonal pembobot ke- i . Turunan kedua fungsi *log-likelihood* terhadap parameter dispersi θ dinyatakan dalam Persamaan (21).

$$f''(\theta) = \frac{\partial^2 L(\boldsymbol{\beta}, \theta)}{\partial \theta^2}$$

$$= \sum_{i=1}^n \left[\theta^{-3} \sum_{r=0}^{y_i-1} \frac{(2r + \theta^{-1})}{(r + \theta^{-1})^2} - 2\theta^{-3} \ln(1 + \theta \mu_i) + \frac{-\theta^2 \mu}{1 + \theta \mu_i} - \frac{(y_i - \mu_i)(1 + 2\theta \mu_i)}{(\theta + \theta^2 \mu_i)^2} \right] \quad (21)$$

Menurut (Agresti, 2002), menentukan penduga awal dari θ misal $\hat{\theta}_1 = 0$

1. Menentukan penduga maksimum *likelihood* dari parameter β menggunakan prosedur iterasi *Newton Raphson* dengan asumsi $\theta = \hat{\theta}_1$

$$\hat{\beta}_i = \hat{\beta}_{i+1} + (\mathbf{X}^T \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_i \mathbf{z}_i \quad (22)$$

Iterasi berakhir jika diperoleh $\hat{\beta}_{i+1} = \hat{\beta}_i$

2. Menggunakan $\hat{\beta}$ untuk menghasilkan penduga dari parameter θ dengan menggunakan prosedur iterasi *Newton Raphson*.
3. Jika $|\hat{\theta}_{i+1} - \hat{\theta}_i| < \varepsilon$ selesai. Bila tidak, gunakan parameter $\theta = \theta_{i+1}$ dan kembali ke langkah 2. Nilai ε merupakan nilai bilangan positif yang sangat kecil, misalnya $\varepsilon = 0,001$.

1.5.7 Uji Simultan

Menurut (Hosmer & Lemeshow, 2000) menyatakan bahwa pengujian parameter simultan bertujuan untuk mengetahui sejauh mana koefisien regresi memberikan pengaruh signifikan secara keseluruhan terhadap variabel respon. Pengujian parameter model secara simultan dilakukan menggunakan statistik uji G, yaitu metode *likelihood ratio test* yang berfungsi untuk menilai kontribusi kolektif variabel prediktor dalam model. Statistik uji G dirumuskan dalam Persamaan (23).

$$G = -2 \ln |L_0 - L_1| \quad (23)$$

Keterangan:



tanpa peubah prediktor
dengan peubah prediktor
akan adalah:

$$\beta_k = 0$$

$$\text{atau } \beta_k \neq 0$$

$$k = 1, 2, 3, \dots, p$$

1.5.8 Uji Parsial

Uji Parsial merupakan pengujian yang digunakan untuk mengetahui pengaruh masing-masing variabel prediktor terhadap variabel respon secara individu (Hosmer & Lemeshow, 2000). Statistik uji yang digunakan untuk uji parsial adalah uji *Wald*. Hipotesis yang digunakan adalah sebagai berikut:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0, i = 1, 2, 3, \dots, n$$

Persamaan yang digunakan untuk menghitung statistik uji *Wald* dituliskan dalam Persamaan (24).

$$W_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (24)$$

dengan:

$$SE(\hat{\beta}) = \frac{s}{\sqrt{n}} \quad (25)$$

Kriteria hipotesis nol ditolak jika $|W| > Z_{\frac{\alpha}{2}}$ dengan α adalah tingkat taraf signifikansi yang digunakan.

1.5.9 Uji Kelayakan Model

Uji *Chi-Square* Pearson merupakan salah satu teknik yang dapat diterapkan untuk menilai kelayakan suatu model (Agresti, 2007). Residual ini memiliki distribusi mendekati simpangan baku dengan nilai tengah 0 dan variansi mendekati 1. Hipotesis yang digunakan dalam uji kelayakan model adalah:

$$H_0 : \text{model layak}$$

$$H_1 : \text{model tidak layak}$$

Persamaan yang digunakan untuk menghitung statistik uji residual Pearson dituliskan dalam Persamaan (26).

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i(1 - \hat{y}_i)} \quad (26)$$

Model yang dihasilkan dapat dikatakan layak apabila nilai $\chi^2_{pearson} < \chi^2_{\alpha(n-p)}$ sehingga dapat diputuskan bahwa H_0 diterima.

1.5.10 Pemilihan Model Terbaik

Metode yang digunakan dalam menentukan model terbaik adalah AIC (*Akaike Information Criterion*). Metode AIC didasarkan pada metode *Maximum Likelihood Estimation* (MLE). Perhitungan nilai AIC menggunakan Persamaan (27).

$$AIC = \frac{2k}{n} + \ln \sum_{i=1}^n \hat{u}_i^2 \quad (27)$$



meter yang diestimasi dalam model
jaminan

i

i terbaik adalah model regresi yang memiliki nilai AIC terkecil

1.5.11 Multikolinearitas

Salah satu kondisi yang harus dipenuhi dalam membangun model regresi dengan banyak variabel adalah tidak adanya multikolinearitas di antara variabel-variabel prediktor. Multikolinearitas adalah suatu kondisi di mana variabel-variabel bebas dalam model regresi saling berkaitan erat, sehingga sulit untuk memisahkan pengaruh masing-masing variabel terhadap variabel dependen (Ginting dkk., 2023). Model regresi yang baik memiliki variabel-variabel prediktor yang independen atau tidak berkorelasi. Pada pengujian asumsi ini diharapkan asumsi multikolinearitas tidak terpenuhi. Dalam model regresi, adanya korelasi antar variabel prediktor menyebabkan estimasi parameter regresi yang dihasilkan akan memiliki residual yang sangat besar. Pendeteksian kasus multikolinearitas menurut (Hocking, 1996) dapat dilihat melalui beberapa cara yaitu sebagai berikut:

1. Jika koefisien korelasi Pearson (r_{ij}) antar variabel prediktor lebih dari 0,95 maka terdapat multikolinearitas antara variabel tersebut.
2. Nilai VIF (*Variance Inflation Factor*) lebih besar dari 10 menunjukkan adanya multikolinearitas antar variabel prediktor. Perhitungan nilai VIF dinyatakan dalam Persamaan (28).

$$VIF_j = \frac{1}{1 - R_j^2} \quad (28)$$

Perhitungan koefisien determinasi dilakukan menggunakan rumus yang ditunjukkan pada Persamaan (29).

$$R_j^2 = 1 - \frac{SSE}{(n - 1)s_j^2} \quad (29)$$

Keterangan:

R_j^2 : koefisien determinasi antara x_j dengan variabel prediktor lainnya

SSE : *sum of squared error*

s_j^2 : variansi dari variabel prediktor j

n : jumlah pengamatan

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (30)$$

$$s_j^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 1} \quad (31)$$

Solusi untuk mengatasi adanya kasus multikolinearitas adalah dengan mengeliminasi variabel independen yang tidak signifikan secara statistik atau dengan menerapkan teknik *Principal Component Analysis* (PCA) untuk mereduksi dimensi dan mengonstruksi variabel-variabel komponen utama yang tidak berkorelasi.



Principal Component Analysis

Analysis (PCA) merupakan sebuah teknik statistik multivariat yang bah bentuk sekelompok variabel asli menjadi kumpulan variabel j tidak berkorelasi yang dapat mewakili informasi dari kumpulan ng dkk., 2023). Sebelum melakukan PCA, pengujian diawali dengan s korelasi. Matriks ini bertujuan mengukur tingkat hubungan linear

antar variabel prediktor. Selanjutnya, nilai kedekatan ini digunakan untuk menentukan jumlah komponen utama yang optimal dalam analisis PCA dan untuk menginterpretasi makna dari masing-masing komponen utama. Berdasarkan hal tersebut, terdapat asumsi yang harus dipenuhi yaitu asumsi kelayakan sampel menggunakan uji *Kaiser-Meyer Olkin Measure of Sampling Adequacy* (KMO MSA).

1.5.12.1 Uji Kaiser-Meyer-Olkin

Uji *Kaiser-Meyer-Olkin* (KMO) bertujuan untuk mengevaluasi kecukupan sampel yang dianalisis dengan membandingkan besarnya koefisien korelasi dan koefisien korelasi parsial (Alamsyah & Husnan, 2023). Adapun hipotesis uji KMO adalah sebagai berikut:

- Hipotesis
 H_0 : ukuran data cukup untuk melakukan analisis PCA
 H_1 : ukuran data tidak cukup untuk melakukan analisis PCA
- Taraf Signifikansi
 $\alpha = 5\%$
- Statistik Uji

$$KMO = \frac{\sum_{i=1}^n \sum_{j=1}^n r_{x_i x_j}^2}{\sum_{i=1}^n \sum_{j=1}^n r_{x_i x_j}^2 + \sum_{i=1}^n \sum_{j=1}^n a_{x_i x_j}^2} \quad (32)$$

dengan:

$$r_{ij}^2 = \frac{\sum_{k=1}^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ik} - \bar{x}_i)^2 \sum_{k=1}^n (x_{jk} - \bar{x}_j)^2}} \quad (33)$$

Keterangan:

r_{ij}^2 : koefisien korelasi antara variabel i dan j

a_{ij}^2 : koefisien korelasi parsial antara variabel i dan j

x_{ik} : nilai variabel i pada observasi ke- k

x_{jk} : nilai variabel j pada observasi ke- k

$\bar{x}_i$: rata-rata variabel i

$\bar{x}_j$: rata-rata variabel ke- j

n : jumlah observasi

- Kriteria Kesimpulan

Tolak H_0 jika nilai $KMO < 0,5$ yang berarti ukuran data tidak cukup untuk melakukan analisis PCA dan sebaliknya terima H_0 jika nilai $KMO \geq 0,5$ yang berarti ukuran data cukup untuk melakukan analisis PCA.

1.5.12.2 Uji Measure of Sampling Adequacy

Selain uji KMO, terdapat uji asumsi yang dilakukan untuk menunjukkan ukuran kecukupan pada setiap variabel yaitu uji *Measure of Sampling Adequacy* (MSA). Nilai



0 hingga 1, dengan nilai 0,5 sebagai ambang batas. Variabel atas 0,5 dianggap memiliki tingkat kesesuaian yang memadai untuk edangkan variabel dengan nilai di bawah 0,5 sebaiknya dieliminasi an dkk., 2023). Adapun bentuk hipotesis uji MSA adalah sebagai

- H_1 : variabel belum memadai untuk dianalisis lebih lanjut
- b. Taraf Signifikansi
 $\alpha = 5\%$
- c. Statistik Uji

$$MSA = \frac{\sum_{i=1}^n r_{x_i x_j}^2}{\sum_{i=1}^n r_{x_i x_j}^2 + \sum_{i=1}^n a_{x_i x_j}^2} \quad (34)$$

- d. Kriteria Kesimpulan

Tolak H_0 jika nilai $MSA < 0,5$ yang berarti variabel belum memadai untuk dianalisis lebih lanjut dan sebaliknya terima H_0 jika nilai $MSA \geq 0,5$ yang berarti variabel belum memadai untuk dianalisis lebih lanjut.

1.5.12.3 Loading Factor

Loading factor (bobot faktor) pada PCA adalah nilai yang menunjukkan korelasi antara variabel asli dengan komponen utama. Nilai ini menggambarkan seberapa besar kontribusi suatu variabel terhadap pembentukan komponen utama tertentu. *Loading factor* dianggap sebagai koefisien dalam transformasi linier dari data asli ke ruang komponen utama. Semakin tinggi nilai absolut *loading factor* (mendekati +1 atau -1), semakin besar kontribusi variabel tersebut dalam menjelaskan variasi data pada komponen tersebut (Jolliffe, 2002).

1.5.13 Kasus Pneumonia pada Balita

Pneumonia yang terjadi pada anak-anak bertepatan dengan proses infeksi bronkial akut. Pneumonia merupakan pembunuh utama balita pada dunia yang paling banyak dibandingkan dengan infeksi, campak, dan penyakit AIDS, malaria dan TBC (Anjaswanti dkk., 2022). Faktor -faktor yang memengaruhi pneumonia pada balita meliputi berbagai aspek kesehatan dan lingkungan. Status gizi, misalnya, berperan penting karena anak dengan gizi buruk cenderung memiliki sistem imun yang lemah, sehingga lebih rentan terhadap infeksi pernapasan. ASI eksklusif juga menjadi faktor penting, karena memberikan zat antibodi alami yang dapat meningkatkan daya tahan tubuh bayi terhadap penyakit. Hal ini sejalan dengan penelitian yang dilakukan oleh Nurhayati pada tahun 2021 yang menyatakan bahwa status gizi dan ASI eksklusif memiliki hubungan dengan kasus pneumonia pada balita. Apabila rata-rata jumlah status gizi meningkat, maka diperkirakan dapat menurunkan kasus pneumonia pada balita sebesar 0,264. Selanjutnya, apabila rata-rata jumlah ASI eksklusif meningkat, maka diperkirakan dapat menurunkan kasus pneumonia pada balita sebesar 0,473 (Nurhayati dkk., 2021).

Cakupan imunisasi, baik imunisasi DPT-Hb maupun imunisasi dasar lengkap (Indarwati dkk., 2023), sangat berkontribusi dalam pencegahan pneumonia, karena imunisasi membantu tubuh melawan patogen tertentu yang dapat menyebabkan infeksi saluran pernapasan. Selain itu, sanitasi yang tidak layak meningkatkan risiko



dan virus penyebab pneumonia, terutama lingkungan padat sanitasi yang tidak memadai dapat memperbesar risiko kontak dengan virus tersebut di udara yang menempel di permukaan lingkungan (2023). Kualitas udara juga memainkan peran penting, karena debu atau asap, termasuk asap rokok, dapat merusak jaringan paru-paru dan meningkatkan risiko infeksi pernapasan seperti pneumonia. Semua faktor ini saling memengaruhi kasus pneumonia pada balita.

BAB II METODE PENELITIAN

2.1 Sumber Data dan Variabel

Penelitian ini menggunakan data sekunder yang diperoleh dari Dinas Kesehatan Jawa Timur (<https://dinkes.jatimprov.go.id>) dan Direktorat Jenderal Pengendalian Pencemaran Lingkungan tahun 2023 (<https://ppkl.menlhk.go.id>). Variabel yang digunakan adalah sebagai berikut:

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Satuan
Y	Jumlah kasus pneumonia pada balita di setiap kabupaten/kota Provinsi Jawa Timur	Jiwa
X_1	Balita gizi buruk	Jiwa
X_2	Pemberian ASI Eksklusif	%
X_3	Cakupan imunisasi DPT-Hb-Hib	Jiwa
X_4	Kualitas udara	%
X_5	Sanitasi layak	%
X_6	Imunisasi dasar lengkap	Jiwa

2.2 Metode Analisis

1. Mendeskripsikan karakteristik jumlah kasus pneumonia dan faktor-faktor yang mempengaruhinya di Provinsi Jawa Timur pada tahun 2023 menggunakan analisis deskriptif untuk masing-masing variabel.
2. Mendapatkan faktor-faktor yang mempengaruhi jumlah kasus pneumonia pada balita dengan metode NBR:
 - a. Pengujian multikolinearitas. Pengujian multikolinearitas dilakukan dengan memperhatikan nilai VIF (*Variance Inflation Factor*) dan koefisien korelasi Pearson (r_{ij}). Jika nilai VIF lebih besar dari 10 dan koefisien korelasi Pearson lebih besar dari 0,95 maka terdapat multikolinearitas antar variabel tersebut.
 - b. Mengatasi masalah multikolinearitas menggunakan PCA di antara variabel prediktor dengan mengevaluasi jumlah komponen utama yang signifikan berdasarkan kriteria *eigenvalue* (lebih dari 1) dan *explained variance* (minimal 80).
 - c. Mengestimasi parameter model regresi Poisson untuk mendapatkan model regresi Poisson.
 - d. Mendeteksi adanya kasus overdispersi dengan menghitung nilai rasio devians dengan derajat kebebasannya lebih dari 1 maka terjadi kasus overdispersi.



masi parameter model regresi Binomial Negatif menggunakan *Maximum Likelihood Estimation* (MLE) serta metode *Newton*

kan NBR untuk mendapatkan faktor-faktor yang diduga garuhi pneumonia dengan langkah-langkah sebagai berikut:

kukan iterasi *newton Raphson* untuk menghitung $\hat{\beta}$

- 2) Melakukan pengujian parameter model NBR secara serentak dengan hipotesis sebagai berikut:

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

$$H_1: \text{paling tidak ada satu } \beta_k \neq 0. \quad k = 1, 2, 3$$

- 3) Melakukan pengujian parameter model NBR secara parsial dengan hipotesis sebagai berikut:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

Kriteria penolakan yaitu tolak H_0 jika $G > \chi^2_{p;a}$

- 4) Melakukan interpretasi model NBR yang diperoleh.

