MATHunesa
Jurnal Ilmiah Matematika *Volume* 12 **No** 02
e-ISSN : 2716-506X | p-ISSN : 2301-9115 *Tahun* 2024

PEMODELAN JUMLAH KEMATIAN AKIBAT FILARIASIS DI JAWA TIMUR DENGAN MENGGUNAKAN REGRESI HURDLE NEGATIVE BINOMIAL

Mutia Eva Mustafidah
Jurusan Matematika, FMIPA, Universitas Negeri Surabaya
e-mail : mutia.20010@mhs.unesa.ac.id
A'yunin Sofro
Jurusan Matematika, FMIPA, Universitas Negeri Surabaya
e-mail : ayuninsofro@unesa.ac.id

Abstrak

Generalized Linear Model (GLM) merupakan perluasan dari model regresi umum untuk variabel respon yang mengikuti distribusi keluarga eksponensial, termasuk distribusi normal, binomial, poisson, negative binomial, eksponensial, dan gamma. Jika variabel respon bersifat diskrit dan mengikuti distribusi Poisson, maka model regresi Poisson dapat digunakan untuk pembentukan model. Namun, dalam penerapannya, sering kali terjadi overdispersi, di mana varians lebih besar dari rata-rata. Overdispersi dalam regresi Poisson dapat terjadi karena banyaknya pengamatan yang memiliki nilai nol dalam variabel respon (excess zeros). Data yang mengalami overdispersi dan excess zeros dapat diselesaikan menggunakan regresi *Hurdle Negative Binomial* (HNB). Permasalahan yang terkait metode HNB dapat ditemukan dalam kasus kematian, di mana data yang digunakan berkaitan dengan jumlah kematian akibat filariasis di Jawa Timur pada tahun 2022. Hasil pengujian signifikansi parameter secara parsial dalam model regresi HNB menunjukkan bahwa persentase penduduk miskin, persentase sanitasi layak, persentase air bersih, jumlah sarana kesehatan, persentase indeks pembangunan manusia, persentase keluhan kesehatan berpengaruh secara signifikan terhadap jumlah kematian akibat filariasis.
Kata Kunci: Overdispersi, *Excess Zeros, Hurdle Negative Binomial*, filariasis.

Abstract

Generalized Linear Model (GLM) is an extension of the general regression model for response variables that follow exponential family distributions, including normal, binomial, Poisson, negative binomial, exponential, and gamma distributions. If the response variable is discrete and follows a Poisson distribution, then the Poisson regression model can be used to build the model. However, in its application, overdispersion often occurs, where the variance is greater than the average. Overdispersion in Poisson regression can occur because there are many observations that have zero values in the response variable (excess zeros). The data experiencing overdispersion and excess zeros can be resolved using Hurdle Negative Binomial (HNB) regression. Problems related to the HNB method can be found in cases of death, where the data used relates to the number of deaths due to filariasis in East Java in 2022. The results of partial parameter significance testing in the HNB regression model show that the percentage of poor people, the percentage of adequate sanitation, the percentage clean water, number of health facilities, percentage of human development index, percentage of health complaints have a significant effect on the number of deaths due to filariasis.
Keywords: *Overdispersi, Excess Zeros, Hurdle Negative Binomial, filariasis.*

PENDAHULUAN

Generalized Linear Model (GLM) merupakan perluasan model regresi umum untuk variabel respon mengikuti sebaran keluarga eksponensial meliputi distribusi normal, binomial, Poisson, binomial negatif, eksponensial, dan gamma (Myers dkk., 2010). Jika variabel respon yang digunakan merupakan data diskrit dan berdistribusi Poisson, maka dapat digunakan model regresi Poisson untuk pembentukan model (Cameron & Trivedi, 2013).

Asumsi penting dalam analisis regresi Poisson adalah bahwa varians sama dengan rata-rata, yang biasa disebut equidispersi. Namun, asumsi ini sering tidak terpenuhi dalam beberapa kasus, menyebabkan overdispersi, di mana varians suatu variabel lebih besar dari rata-ratanya. Menurut Hilbe (2011), jika overdispersi terjadi, regresi Poisson tidak lagi tepat karena akan mengevaluasi standar kesalahan yang terlalu rendah, mengakibatkan kesalahan dalam signifikansi parameter regresi individual. Salah satu alternatif untuk mengatasi overdispersi dalam regresi Poisson adalah dengan menggunakan regresi

Negative Binomial (Saputro et al., 2021). Namun, dalam beberapa kasus, data hitung tidak hanya mengalami overdispersi tetapi juga kelebihan nol. Kelebihan nol terjadi ketika jumlah nilai nol dalam variabel respons lebih besar dari jumlah nilai diskrit lainnya, menyebabkan kesalahan dalam menarik kesimpulan. Dalam kondisi tersebut, regresi Negative Binomial tidak dapat digunakan, memerlukan model regresi lain (Bilgic et al., 2007).

Regresi yang dapat digunakan untuk mengatasi overdispersi dengan excess zeros adalah model regresi *Hurdle Negative Binomial* (HNB). Model regresi *Hurdle Negative Binomial* (HNB) merupakan model regresi yang dibentuk dari distribusi campuran Poisson-Gamma. Model *Hurdle Negative Binomial* mempunyai dua kondisi yaitu kondisi *Zero Hurdle State* dan *Truncated Negative Binomial State*. Kondisi pertama yaitu *Zero Hurdle State* dengan probabilitas (π_i) dan menghasilkan observasi bernilai nol dengan hambatan (*hurdle*). Sedangkan kondisi kedua adalah *Truncated Negative Binomial State* dengan probabilitas $(1-\pi_i)$ yang menghasilkan observasi bernilai non-negative setelah melewati hambatan (*truncated*) dan berdistribusi *Negative Binomial* dengan mean μ_i (Ma dkk., 2016).

Penelitian ini menerapkan model regresi *Hurdle Negative Binomial* (HNB) pada kasus jumlah kematian akibat filariasis di Jawa Timur tahun 2022. Data tersebut merupakan count data yang mengalami overdispersi dan *excess zeros* sehingga dapat diselesaikan menggunakan regresi HNB. Jumlah kematian akibat filariasis di Jawa Timur pada tahun 2022. Adapun penelitian-penelitian yang telah dilakukan untuk mengatasi overdispersi dan kelebihan nol diantaranya Oktari dkk (2016) menggunakan regresi negative binomial untuk mengatasi kasus overdispersi pada data jumlah bayi di Kota Padang, Yulian (2018) memodelkan regresi *Zero Inflated Negative Binomial* untuk mengatasi kasus overdispersi akibat kelebihan nol, dan Hadi (2019) menggunakan regresi *Hurdle Poisson* untuk mengatasi kasus overdispersi pada data demam berdarah di Kota Makassar. Namun, menurut Zhang dkk. (2018) model regresi *Hurdle Negative Binomial* lebih tepat digunakan untuk mengatasi kasus overdispersi akibat kelebihan nol jika dibandingkan dengan model regresi Poisson, regresi *Negative Binomial*, dan regresi *zero inflated*. Keunggulan dari model hurdle adalah kedua model didalamnya dapat dilakukan penaksiran parameter secara terpisah atau dengan kata lain dimaksimumkan secara terpisah sehingga diharapkan dapat lebih mudah dalam penginterpretasiannya (Cantoni dan Zedini, 2010). . Tujuan dari penelitian ini adalah untuk mendapatkan model regresi HNB dan mengidentifikasi faktor-faktor signifikan terhadap data.

KAJIAN TEORI

PENGUJIAN MULTIKOLINIERITAS

Multikolinieritas merupakan keadaan di mana variabel prediktor memiliki korelasi yang tinggi dengan variabel predictor yang lain. Pengaruh yang kuat dari multikolinieritas adalah mengurangi kekuatan dari prediksi yang dihasilkan oleh variabel predictor. Pendeteksian kasus multikolinieritas dilakukan menggunakan kriteria nilai VIF dengan rumus sebagai berikut (Hocking, 2013):

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

Dengan :
$j = 1,2,\dots,p$
$VIF =$ Angka Variance Inflation Factor (VIF)
$R_j^2 =$ koefisien determinasi antara satu variabel prediktor dengan variabel prediktor lainnya.

Apabila nilai VIF lebih dari 10 maka terjadi kasus multikolinieritas (Hocking, 2013).

KELUARGA EKSPONENSIAL

Sebuah konsep penting yang mendasari Generalized Linear Model adalah keluarga eksponensial. Variabel respon berdistribusi keluarga eksponensial apabila mengikuti fungsi probabilitas berikut:

$$P(y_i) = e^{\left(\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} - c(y_i;\phi)\right)} \qquad (2)$$

Dimana $a(\phi); b(\theta_i); c(y_i;\phi)$ adalah fungsi spesifik, θ_i adalah parameter natural, sedangkan ϕ adalah parameter disperse dan fungsi $a(\phi)$ umumnya berupa $a(\phi) = \phi.\omega$ dimana ω adalah konstanta yang diketahui (myeres, 2012).

DISTRIBUSI POISSON

Distribusi Poisson adalah salah satu distribusi pada Generalized Linear Model. Fungsi probabilitas distribusi Poisson ditunjukkan pada persamaan berikut (Agresti, 2015):

$$P(y_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \qquad (3)$$

Pada distribusi poisson nilai mean dan varians adalah $E[y_i] = \mu_i$ dan $Var[y_i] = \mu_i$

DISTRIBUSI GAMMA

Distribusi Gamma merupakan salah satu distribusi pada Generalized Linear Model. Fungsi probabilitas distribusi Gamma ditunjukkan pada persamaan berikut (Agresti, 2015):

$$P(y_i) = \frac{1}{\Gamma(\phi)\kappa^{\phi}} y_i^{\phi-1} e^{\frac{-y_i}{\kappa}} \qquad (4)$$

Pada distribusi Gamma nilai mean dan varians adalah $E[y_i] = \phi\kappa$ dan $Var(y_i) = \phi\kappa^2$

DISTRIBUSI CAMPURAN *POISSON-GAMMA*

Distribusi campuran Poisson-Gamma merupakan salah satu pendekatan yang digunakan pada distribusi Negative Binomial yang digunakan untuk menangani kasus overdispersi pada distribusi Poisson. Fungsi probabilitas distribusi campuran Poisson-Gamma ditunjukkan pada persamaan berikut (Agresti, 2015):

$$P(y_i) = \frac{\Gamma\left(y_i + \frac{1}{\phi}\right)}{\Gamma\left(\frac{1}{\phi}\right) y_i!} \left(\frac{1}{1+\phi\mu_i}\right)^{\frac{1}{\phi}} \left(\frac{\phi\mu_i}{1+\phi\mu_i}\right)^{y_i} \qquad (5)$$

GENERALIZED LINEAR MODEL

Generalized Linear Model (GLM) adalah perluasan dari proses pemodelan yang melibatkan lebih dari satu variable prediktor dengan variable respon diasumsikan berdistribusi kelurga eksponensial (Myers dkk., 2012).

Struktur Generalized Linear Model terdiri dari 3 komponen adalah sebagai berikut (Agresti, 2015):

1. Komponen acak yaitu komponen yang ditunjukkan dengan variabel respon yang bersifat independen dan mengikuti distribusi keluarga eksponensial.
2. Komponen sistematik merupakan kombinasi linier variabel prediktor x_i dengan koefisien parameter regresi $\boldsymbol{\beta}$ yang menghasilkan prediktor linier η_i yang dapat dinyatakan dengan persamaan berikut: $\eta_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} = \boldsymbol{x_i^T \beta}$
3. Fungsi penghubung yaitu komponen menghubungkan antara komponen acak dan komponen sistematik. Model yang menghubungkan μ_i dengan prediktor linier η_i dapat dinyatakan dengan persamaan berikut:

$$g(\mu_i) = \eta_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} = \boldsymbol{x_i^T \beta} \qquad (6)$$

REGRESI POISSON

Regresi Poisson digunakan untuk menganalisis count data. Regresi Poisson merupakan model regresi nonlinear yang sering digunakan untuk mengatasi count data dimana variabel respon mengikuti distribusi Poisson (Agresti, 2015). Model regresi Poisson menggunakan ln dapat dinyatakan sebagaimana persamaan berikut:

$$\ln(\mu_i) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \qquad (7)$$

Dimana $i = 1,2,\dots,n$ dan $j = 1,2,\dots,p$, $p =$ jumlah variabel predictor, $n =$ jumlah pengamata, $\beta =$ parameter model regresi yang diestimasi, $x_{ij} =$ variabel prediktor ke-j pada pengamatan ke-i.

Persamaan (7) juga dapat dinyatakan dalam bentuk matriks sehingga diperoleh persamaan berikut:

$$\ln(\mu_i) = \boldsymbol{x_i^T \beta}$$

Dimana $i = 1, 2, \dots, n$, $j = 1, 2, \dots, p$, n adalah jumlah pengamatan dan p adalah jumlah variabel prediktor, $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$

OVERDISPERSI PADA REGRESI POISSON

Suatu permasalahan yang terjadi akibat nilai variansi lebih besar daripada mean pada variabel respon ($Var[y_i] > E[y_i]$) disebut overdispersi (Hilbe, 2011). Rumus yang digunakan untuk mengitung overdispersi sebagai berikut:

$$\phi = \frac{X^2}{\text{derajat bebas}} \qquad (8)$$

Dengan X^2 adalah pearson chi square. Jika nilai dispersi $\phi > 1$ maka suatu data mengalami overdispersi.

REGRESI *NEGATIVE BINOMIAL*

Regresi *Negative Binomial* merupakan regresi yang berdistribusi Negative Binomial pada keluarga eksponensial. Pendekatan yang digunakan pada distribusi Negative Binomial adalah distribusi campuran Poisson-Gamma untuk mengatasi kasus overdispersi pada regresi Poisson. Model regresi Negative Binomial dengan fungsi penghubung ln dinyatakan sebagai berikut:

$$\ln(\mu_i) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \qquad (9)$$

Dimana $i = 1,2,\dots,n$ dan $j = 1,2,\dots,p$, $p =$ jumlah variabel predictor, $n =$ jumlah pengamata, $\beta =$ parameter model regresi yang diestimasi, $x_{ij} =$ variabel prediktor ke-j pada pengamatan ke-i.

Persamaan (7) juga dapat dinyatakan dalam bentuk matriks sehingga diperoleh persamaan berikut:

$$\ln(\mu_i) = \boldsymbol{x_i^T \beta}$$

Dimana $i = 1, 2, \dots, n$, $j = 1, 2, \dots, p$, n adalah jumlah pengamatan dan p adalah jumlah variabel prediktor, $\boldsymbol{x_i} = (\boldsymbol{x_{i1}}, \boldsymbol{x_{i2}}, \dots, \boldsymbol{x_{ip}})^T$ $\boldsymbol{\beta} = (\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \dots, \boldsymbol{\beta_p})^T$.

EXCESS ZEROS PADA VARIABEL RESPON

Excess zeros merupakan kondisi variable respon yang memiliki nilai no; berlebihan atau memiliki proporsi nilai nol lebih besar jika dibandingkan dengan nilai data diskrit lain pada variabel respon (Famoye dan Singh, 2006). Jika terjadi excess zeros pada variabel respon maka regresi Negative Binomial tidak dapat digunakan sehingga diperlukan metode lain untuk mengatasi permasalahan tersebut.

REGRESI *HURDLE*

Salah satu pendekatan yang dapat digunakan untuk mengatasi overdispersi adalah model Hurdle. Salah satu penyebab terjadi overdispersi adalah banyaknya nilai nol pada variabel respon (excess zero). Model hurdle mampu mengatasi kasus excess zeros dengan membagi dua model ke dalam dua bagian yaitu:

1. Model Zero Hurdle State yaitu model untuk data yang bernilai nol dengan hambatan (hurdle).
2. Model Truncated Negative Binomial State yaitu model untuk data yang bernilai positif (positive counts) setelah melewati hambatan (truncated).

Fungsi probabilitas dari model regresi hurdle adalah sebagai berikut :

$$f(y_i) = \begin{cases} f_1(0), & y_i = 0 \\ \left(1 - f_1(0)\right)\frac{f_2(y_i)}{1-f_2(0)} & , y_i > 0 \end{cases} \tag{10}$$

REGRESI *HURDLE NEGATIVE BINOMIAL*

Regresi *Hurdle Negative Binomial* adalah model regresi yang dibentuk dari distribusi campuran Poisson-Gamma. Distribusi tersebut merupakan salah satu pendekatan pada distribusi *Negative Binomial* yang digunakan untuk mengatasi kasus overdispersi pada regresi Poisson. Regresi tersebut terdiri dari dua kondisi yaitu kondisi *Zero Hurdle State* dan *Truncated Negative Binomial State*. Kondisi pertama yaitu *Zero Hurdle State* dengan probabilitas (π_i) dan menghasilkan observasi bernilai nol dengan hambatan (*hurdle*). Sedangkan keadaan kedua adalah *Truncated Negative Binomial State* dengan probabilitas $(1 - \pi_i)$ yang menghasilkan observasi bernilai non-negative setelah melewati hambatan (*truncated*) dan berdistribusi *Negative Binomial* dengan mean μ_i . Fungsi probabilitas model regresi *Hurdle Negative Binomial* sebagai berikut (Greene, 2008):

$$P(y_i) = \begin{cases} \pi_i, & y_i = 0 \\ (1-\pi_i)\frac{\Gamma\left(y_i+\frac{1}{\phi}\right)}{\Gamma\left(\frac{1}{\phi}\right)y_i!}\left(\frac{\phi\mu_i}{1+\phi\mu_i}\right)^{y_i}\frac{(1+\phi\mu_i)^{-\phi^{-1}}}{1-(1+\phi\mu_i)^{-\phi^{-1}}}, & y_i > 0 \end{cases} \tag{11}$$

Dimana $i = 1,2,\dots,n$, $0 \le \pi_i \le 1$, $\mu_i \ge 0$ dan ϕ adalah parameter dispersi dengan $\phi > 0$, $\Gamma(.)$ adalah fungsi Gamma.

Model regresi *Hurdle Negative Binomial* terdiri dari dua model yang dinyatakan pada persamaan berikut :

1. Model *Zero Hurdle State*

   $$\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \gamma_0 + \sum_{j=1}^{p}\gamma_j x_{ij} \tag{12}$$

   Dimana $i = 1,2,\dots,n$ dan $j = 1,2,\dots,p$, $p =$ jumlah variabel predikto, $n =$ jumlah pengamat, γ =parameter model regresi yang diestimas, x_{ij} =variabel prediktor ke-j pada pengamatan ke-i.

   Persamaan (12) juga dapat dinyatakan dalam bentuk matriks sehingga diperoleh persamaan berikut:

   $$\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \boldsymbol{x_i^T\gamma} \tag{13}$$

   Dimana $i = 1,2,\dots,n$, $j = 1,2,\dots,p$, n adalah jumlah pengamatan dan p adalah jumlah variabel prediktor,

   $\boldsymbol{x_i} = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2},\dots,\boldsymbol{x}_{ip})^T \boldsymbol{\gamma} = (\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\dots,\boldsymbol{\gamma}_p)^T$.

2. Model *Truncated Negative Binomial State*

   $$\ln(\mu_i) = \beta_0 + \sum_{j=1}^{p}\beta_j x_{ij} \tag{14}$$

   Dimana $i = 1,2,\dots,n$ dan $j = 1,2,\dots,p$, $p =$ jumlah variabel predictor, $n =$ jumlah pengamata, $\beta =$ parameter model regresi yang diestimasi, x_{ij} = variabel prediktor ke-j pada pengamatan ke-i.

   Persamaan (14) juga dapat dinyatakan dalam bentuk matriks sehingga diperoleh persamaan berikut:

   $$\ln(\mu_i) = \boldsymbol{x_i^T\beta} \tag{15}$$

   Dimana $i = 1,2,\dots,n$, $j = 1,2,\dots,p$, n adalah jumlah pengamatan dan p adalah jumlah variabel prediktor,

   $\boldsymbol{x_i} = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2},\dots,\boldsymbol{x}_{ip})^T \boldsymbol{\beta} = (\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\dots,\boldsymbol{\beta}_p)^T$.

ESTIMASI PARAMETER REGRESI *HURDLE NEGATIVE BINOMIAL*

Metode *maximum likelihood estimation* (MLE) dapat digunakan untuk mengestimasi parameter model regresi *Hurdle Negative Binomial*.

Berdasarkan persamaan (13) diperoleh hasil :

$$\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \boldsymbol{x_i^T\gamma}$$

$$\pi_i = \frac{e^{x_i^T\gamma}}{1+e^{x_i^T\gamma}} \tag{16}$$

$$(1-\pi_i) = \frac{1+e^{x_i^T\gamma}}{1+e^{x_i^T\gamma}} - \frac{e^{x_i^T\gamma}}{1+e^{x_i^T\gamma}} = \frac{1}{1+e^{x_i^T\gamma}}$$

(17)

sedangkan berdasarkan (15) dihasilkan :

$$\ln(\mu_i) = \boldsymbol{x_i^T\beta}$$

$$\mu_i = e^{x_i^T\beta} \quad (18)$$

Kemudian persamaan (16), (17) dan (18) disubstitusikan ke persamaan (11) sehingga diperoleh fungsi probabilitas dari model regresi ZINB sebagai berikut:

$$P(y_i) = \begin{cases} \frac{e^{x_i^T\gamma}}{1+e^{x_i^T\gamma}}, & y_i = 0 \\ \frac{1}{1+e^{x_i^T\gamma}} \frac{\Gamma\left(y_i+\frac{1}{\phi}\right)}{\Gamma\left(\frac{1}{\phi}\right)y_i!}\left(\frac{\phi e^{x_i^T\beta}}{1+\phi e^{x_i^T\beta}}\right)^{y_i} \frac{\left(1+\phi e^{x_i^T\beta}\right)^{-\phi^{-1}}}{1-\left(1+\phi e^{x_i^T\beta}\right)^{-\phi^{-1}}}, & y_i > 0 \end{cases}$$

(19)

Dimana $i = 1, 2, \dots, n, j = 1, 2, \dots, p$, n adalah jumlah pengamatan, p adalah jumlah variabel prediktor $\boldsymbol{x_i} = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \dots, \boldsymbol{x}_{ip})^T$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_p)^T$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \dots, \boldsymbol{\gamma}_p)^T$, dan ϕ *adalah* parameter dispersi

Bentuk fungsi likelihood dari model regresi *Hurdle Negative Binomial* (HNB) dapat dituliskan sebagaimana persamaan berikut:

$$L(\boldsymbol{\beta},\boldsymbol{\gamma}) = \begin{cases} \Pi_{i=1}^{n} \frac{e^{x_i^T\gamma}}{1+e^{x_i^T\gamma}}, & y_i = 0 \\ \Pi_{i=1}^{n} \frac{1}{1+e^{x_i^T\gamma}} \frac{\Gamma\left(y_i+\frac{1}{\phi}\right)}{\Gamma\left(\frac{1}{\phi}\right)y_i!}\left(\frac{\phi e^{x_i^T\beta}}{1+\phi e^{x_i^T\beta}}\right)^{y_i} \frac{\left(1+\phi e^{x_i^T\beta}\right)^{-\phi^{-1}}}{1-\left(1+\phi e^{x_i^T\beta}\right)^{-\phi^{-1}}}, & y_i > 0 \end{cases} \quad (20)$$

Langkah selanjutnya setelah mendapatkan fungsi Likelihood dari model regresi HNB adalah membentuk fungsi ln Likelihood berdasarkan persamaan (20). Bentuk fungsi ln Likelihood dapat dinyatakan dalam persamaan berikut :

$$\ln L(\boldsymbol{\beta},\boldsymbol{\gamma}) = \begin{cases} \sum_{i=1}^{n} \ln e^{x_i^T\gamma} - \sum_{i=1}^{n} \ln(1+e^{x_i^T\gamma}), y_i = 0 \\ -\sum_{i=1}^{n} \ln(1+e^{x_i^T\gamma}) + y_i \sum_{i=1}^{n}\left[\ln\phi + \ln e^{x_i^T\beta}\right] - \left(\frac{1}{\phi}+y_i\right)\sum_{i=1}^{n}\left[\ln\left(1+\phi e^{x_i^T\beta}\right)\right] \\ + \sum_{i=1}^{n}\ln\Gamma\left(y_i+\frac{1}{\phi}\right) - \sum_{i=1}^{n}\ln(y_i!) - \sum_{i=1}^{n}\ln\Gamma\left(\frac{1}{\phi}\right) - \sum_{i=1}^{n}\left[\ln\left(1-\left(1+\phi e^{x_i^T\beta}\right)^{-\phi^{-1}}\right)\right], y_i > 0 \end{cases} \quad (21)$$

Fungsi ln likelihood pada persamaan (21) memiliki 2 kondisi yang digabungkan yaitu saat $y_i = 0$ yang mewakili kelompok *zero inflation state* dan $y_i > 0$ yang mewakili kelompok *negative binomial state*. Sehingga membuat sulit perhitungan dan tidak dapat diketahui mana nilai nol yang berasal dari *zero inflation state* dan *negative binomial state*. Untuk menggambarkan kondisi variabel y_i secara terperinci maka akan didefinisikan Kembali y_i dengan suatu vari-abel laten z_i untuk mengetahui nilai nol yang berasal *zero inflation state* dan *negative binomial state* yang dapat didefinisikan sebagai berikut (Hall, 2000):

$$z_i = \begin{cases} 1, \; untuk \; y_i = 0 \\ 0, \; untuk \; y_i > 0 \end{cases} \quad (22)$$

Kemudian menentukan fungsi probabilitas dari variabel laten z_i sebagai berikut:

$$P(z_i) = \begin{cases} \pi_i, \; untuk \; z_i = \mathbf{1} \\ 1-\pi_i, \; untuk \; z_i = 0 \end{cases} \quad (23)$$

Berdasarkan persamaan (23) dengan kondisi *Zero Inflation State* ($P(z_i = 1) = \pi_i$) dan kondisi *Negative Binomial State* ($P(z_i = 0) = 1 - \pi_i$). Kemudian dibentuk fungsi probabilitas gabungan bersama antara y_i dan z_i yaitu sebagaimana persamaan berikut:

$$f(\boldsymbol{\beta},\boldsymbol{\gamma}|\boldsymbol{y},\boldsymbol{z}) = f(\boldsymbol{z})f(\boldsymbol{y}|\boldsymbol{z})$$
$$= f(\boldsymbol{z}|1,\pi_i)f((\boldsymbol{y}|\boldsymbol{z},\mu_i)$$
$$= (\pi_i)^{z_i}(1-\pi_i)^{(1-z_i)}\left[\frac{\Gamma\left(y_i+\frac{1}{\phi}\right)}{\Gamma\left(\frac{1}{\phi}\right)y_i!}\left(\frac{\phi e^{x_i^T\beta}}{1+\phi e^{x_i^T\beta}}\right)^{y_i} \frac{\left(1+\phi e^{x_i^T\beta}\right)^{-\phi^{-1}}}{1-\left(1+\phi e^{x_i^T\beta}\right)^{-\phi^{-1}}}\right]^{(1-z_i)} \quad (24)$$

Kemudian persamaan (16), (17) dan (18) disubstitusikan pada persamaan (24) sehingga diperoleh:

$$f(\boldsymbol{\beta},\boldsymbol{\gamma}|\boldsymbol{y},\boldsymbol{z}) =$$
$$\left(\frac{e^{x_i^T\gamma}}{1+e^{x_i^T\gamma}}\right)^{z_i}\left(\frac{1}{1+e^{x_i^T\gamma}}\right)^{1-z_i}\left(\frac{\Gamma\left(y_i+\frac{1}{\phi}\right)}{\Gamma\left(\frac{1}{\phi}\right)y_i!}\left(\frac{\phi e^{x_i^T\beta}}{1+\phi e^{x_i^T\beta}}\right)^{y_i} \frac{\left(1+\phi e^{x_i^T\beta}\right)^{-\phi^{-1}}}{1-\left(1+\phi e^{x_i^T\beta}\right)^{-\phi^{-1}}}\right)^{(1-z_i)}$$
$$= \left(e^{x_i^T\gamma}\right)^{z_i}\left(\frac{1}{1+e^{x_i^T\gamma}}\right)\left(\frac{\Gamma\left(y_i+\frac{1}{\phi}\right)}{\Gamma\left(\frac{1}{\phi}\right)y_i!}\left(\frac{\phi e^{x_i^T\beta}}{1+\phi e^{x_i^T\beta}}\right)^{y_i} \frac{\left(1+\phi e^{x_i^T\beta}\right)^{-\phi^{-1}}}{1-\left(1+\phi e^{x_i^T\beta}\right)^{-\phi^{-1}}}\right)^{(1-z_i)} \quad (25)$$

Fungsi likelihood baru dinyatakan berikut:

$$L(\boldsymbol{\beta},\boldsymbol{\gamma}|\boldsymbol{y},\boldsymbol{z}) = \Pi_{i=1}^{n}\left(e^{x_i^T\gamma}\right)^{z_i}\left(\frac{1}{1+e^{x_i^T\gamma}}\right)\left(\frac{\Gamma\left(y_i+\frac{1}{\phi}\right)}{\Gamma\left(\frac{1}{\phi}\right)y_i!}\left(\frac{\phi e^{x_i^T\beta}}{1+\phi e^{x_i^T\beta}}\right)^{y_i} \frac{\left(1+\phi e^{x_i^T\beta}\right)^{-\phi^{-1}}}{1-\left(1+\phi e^{x_i^T\beta}\right)^{-\phi^{-1}}}\right)^{(1-z_i)} \quad (26)$$

Sehingga fungsi ln likelihood dapat ditulis sebagai berikut (Garay dkk., 2011):

$$\ln L(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{z}) = \sum_{i=1}^{n}(z_i)(\boldsymbol{x}_i^T\boldsymbol{\gamma}) - \sum_{i=1}^{n}\ln\left(1+e^{x_i^T\gamma}\right) + \left(\sum_{i=1}^{n}(1-z_i)\right)\sum_{i=1}^{n}\Gamma\left(y_i+\frac{1}{\phi}\right)$$
$$-\sum_{i=1}^{n}\Gamma\left(\frac{1}{\phi}\right) - \sum_{i=1}^{n}\ln(y_i!) - \left(\frac{1}{\phi}+y_i\right)\left[\sum_{i=1}^{n}\ln\left(1+\phi e^{x_i^T\beta}\right)\right]$$
$$+y_i\sum_{i=1}^{n}\left(\ln\phi + \ln e^{x_i^T\beta}\right) - \left[\sum_{i=1}^{n}\ln\left(1-\left(1+\phi e^{x_i^T\beta}\right)^{-\phi^{-1}}\right)\right] \quad (27)$$

Proses estimasi parameter $\boldsymbol{\beta},\boldsymbol{\gamma}$ dilakukan secara terpisah, sehingga fungsi likelihood lengkap dapat dituliskan kembali sebagai mana persamaan (28) dan (29).

$$\ln L(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{z}) = \sum_{i=1}^{n}(1-z_i)\sum_{i=1}^{n}\Gamma\left(y_i+\frac{1}{\phi}\right) - \sum_{i=1}^{n}\ln(y_i!) - \sum_{i=1}^{n}\Gamma\left(\frac{1}{\phi}\right) - \left(\frac{1}{\phi}+y_i\right)$$

$$\left[\sum_{i=1}^{n}\ln\left(1+\phi e^{x_i^T\beta}\right)\right]+y_i\sum_{i=1}^{n}\left(\ln\phi+\ln e^{x_i^T\beta}\right)$$
$$-\left[\sum_{i=1}^{n}\ln\left(1-\left(1+\phi e^{x_i^T\beta}\right)^{-\phi^{-1}}\right)\right] \quad (28)$$

Dan

$$\ln L\,(\boldsymbol{\gamma}|\boldsymbol{y},\boldsymbol{z})=\sum_{i=1}^{n}(z_i)(\boldsymbol{x}_i^T\boldsymbol{\gamma})-\sum_{i=1}^{n}\ln\left(1+e^{x_i^T\gamma}\right) \quad (29)$$

Untuk mendapatkan estimasi parameter $\boldsymbol{\beta}$ fungsi ln likelihood pada persamaan (28) diturunkan secara parsial terhadap parameter $\boldsymbol{\beta}$.

Turunan pertama fungsi ln likelihood pada persamaan (28) terhadap $\boldsymbol{\beta}$ dapat dinyatakan sebagaimana persamaan berikut:

$$\frac{\partial\ln L\,(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{z})}{\partial(\boldsymbol{\beta})}=\sum_{i=1}^{n}(1-z_i)y_i\sum_{i=1}^{n}\boldsymbol{x}_i^T-(y_i+\phi^{-1})\sum_{i=1}^{n}\frac{\phi e^{x_i^T\beta}\boldsymbol{x}_i^T}{1+\phi e^{x_i^T\beta}}-$$
$$\sum_{i=1}^{n}\frac{1}{\phi}\frac{\left(1+\phi e^{x_i^T\beta}\right)^{-\frac{(\phi+1)}{\phi}}\left(\phi\boldsymbol{x}_i^T e^{x_i^T\beta}\right)}{1-\left(1+\phi e^{x_i^T\beta}\right)^{-\phi^{-1}}} \quad (30)$$

Kemudian turunan kedua fungsi ln likelihood pada persamaan (28) terhadap $\boldsymbol{\beta}$ dapat dinyatakan sebagaimana persamaan berikut:

$$\frac{\partial^2\ln L\,(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{z})}{\partial^2(\boldsymbol{\beta})}=-\sum_{i=1}^{n}(1-z_i)\left((y_i+\phi^{-1})\sum_{i=1}^{n}\frac{\phi e^{2x_i^T\beta}(\boldsymbol{X})^2}{\left(1+\phi e^{x_i^T\beta}\right)^2}\right)-$$
$$\sum_{i=1}^{n}\left(\frac{\phi(x_i^T)^2e^{2x_i^T\beta}-(\phi+1)(1+\phi e^{x_i^T\beta})^{-\frac{(2\phi+1)}{\phi}}+(1+\phi e^{x_i^T\beta})^{-\frac{(\phi+1)}{\phi}}-\phi(1+\phi e^{x_i^T\beta})^{-\frac{2(\phi+1)}{\phi}}}{\phi^2\left(1-(1+\phi e^{x_i^T\beta})^{-\phi^{-1}}\right)^2}\right) \quad (31)$$

Estimasi untuk $\boldsymbol{\beta}$ tidak dapat secara langsung diperoleh. Karena hasil persamaan yang diperoleh dari turunan pertama fungsi ln likelihood terhadap masing-masing parameter tidak closed form. Untuk mengatasi kasus tersebut dilakukan dengan menggunakan metode iterasi numerik untuk memudahkan perhitungan yaitu metode iterasi Newton Raphson untuk memperoleh estimasi parameternya (Hilbe, 2011). Langkah-langkah metode iterasi numerik Newton Raphson adalah sebagai berikut:

a. Menentukan taksiran awal parameter $\hat{\beta}^{(0)}$ yang diperoleh dengan metode OLS, yaitu:
$$\widehat{\boldsymbol{\beta}}^{(\mathbf{0})}=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$
b. Membentuk vektor gradien $\mathbf{u}$ yaitu vektor yang terdiri dari turunan parsial pertama.
c. Membetuk matriks Hessian $\mathbf{H}$ yaitu matriks yang berisi turunan parsial kedua.
d. Memasukkan nilai $\widehat{\boldsymbol{\beta}}^{(\mathbf{0})}$ ke dalam elemen-elemen vector $\mathbf{u}$ dan matriks $\mathbf{H}$ sehingga diperoleh vektor $\boldsymbol{u}(\widehat{\boldsymbol{\beta}}^{(\mathbf{0})})$dan $\boldsymbol{H}(\widehat{\boldsymbol{\beta}}^{(\mathbf{0})})$
e. Melakukan iterasi mulai dari $r=0$ sebagaimana persamaan berikut :
$$\widehat{\boldsymbol{\beta}^{(r+1)}}=\widehat{\boldsymbol{\beta}^{(r)}}-\boldsymbol{H}^{-1}\left(\widehat{\boldsymbol{\beta}^{(r)}}\right)\boldsymbol{u}\left(\widehat{\boldsymbol{\beta}^{(r)}}\right)$$
f. Jika parameter yang didapatkan belum konvergen, maka dilakukan kembali langkah pada bagian ke- e hingga iterasi ke dari $r=r+1$. Proses Iterasi akan berhenti jika telah diperoleh estimasi parameter yang konvergen dngan memenuhi. Proses Iterasi akan berhenti jika telah diperoleh estimasi parameter yang konvergen dengan memenuhi $\left(\left|\widehat{\boldsymbol{\beta}^{(r+1)}}-\widehat{\boldsymbol{\beta}^{(r)}}\right|\right)\leq\epsilon$ dimana ϵ dalah nilai yang sangat kecil dan telah diterapkan sebelumnya, misal 10^{-4}

Sama halnya dengan tahap maksimalisasi parameter $\boldsymbol{\beta}$, untuk tahap maksimalisasi parameter $\boldsymbol{\gamma}$ juga digunakan metode iterasi numerik yaitu metode Newton Raphson. Untuk mendapatkan estimasi parameter $\boldsymbol{\gamma}$, fungsi ln likelihood persamaan (29) diturunkan secara parsial terhadap parameter $\boldsymbol{\gamma}$. Turunan pertamanya sebagai berikut:

$$\frac{\partial\ln L\,(\boldsymbol{\gamma}|\boldsymbol{y},\boldsymbol{z})}{\partial(\boldsymbol{\gamma})}=\sum_{i=1}^{n}(z_i)\left(\boldsymbol{x}_i^T-\sum_{i=1}^{n}\frac{e^{x_i^T\gamma}}{1+e^{x_i^T\gamma}}\boldsymbol{x}_i^T\right) \quad (32)$$

Kemudian turunan kedua fungsi ln likelihood pada persamaan (29) terhadap $(\boldsymbol{\gamma})$ dinyatakan sebagaimana persamaan berikut:

$$\frac{\partial^2\ln L\,(\boldsymbol{\gamma}|\boldsymbol{y},\boldsymbol{z})}{\partial^2(\boldsymbol{\gamma})}=\sum_{i=1}^{n}\left(\frac{e^{x_i^T\gamma}(\boldsymbol{x}_i^T)^2}{\left(1+e^{x_i^T\gamma}\right)^2}\right) \quad (33)$$

Estimasi untuk $\boldsymbol{\gamma}$ tidak dapat secara langsung diperoleh. Karena hasil persamaan yang diperoleh dari turunan pertama fungsi ln likelihood terhadap masing-masing parameter tidak closed form. Untuk mengatasi kasus tersebut dilakukan dengan menggunakan metode iterasi numerik untuk memudahkan perhitungan yaitu metode iterasi Newton Raphson untuk memperoleh estimasi parameternya (Hilbe, 2011). Langkah-langkah metode iterasi numerik Newton Raphson adalah sebagai berikut:

a. Menentukan taksiran awal parameter $\widehat{\boldsymbol{\gamma}}^{(\mathbf{0})}$ yang diperoleh dengan metode OLS, yaitu:

$$\widehat{\boldsymbol{\gamma}}^{(\mathbf{0})} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}$$

b. Membentuk vektor gradien $\mathbf{u}$ yaitu vektor yang terdiri dari turunan parsial pertama.

c. Membetuk matriks Hessian $\mathbf{H}$ yaitu matriks yang berisi turunan parsial kedua.

d. Memasukkan nilai $\widehat{\boldsymbol{\gamma}}^{(\mathbf{0})}$ ke dalam elemen-elemen vector $\mathbf{u}$ dan matriks $\mathbf{H}$ sehingga diperoleh vektor $\boldsymbol{u}(\widehat{\boldsymbol{\gamma}}^{(\mathbf{0})})$dan $\boldsymbol{H}(\widehat{\boldsymbol{\gamma}}^{(\mathbf{0})})$

e. Melakukan iterasi mulai dari $r = 0$ sebagaimana persamaan berikut :

$$\widehat{\boldsymbol{\gamma}^{(r+1)}} = \widehat{\boldsymbol{\gamma}^{(r)}} - \boldsymbol{H}^{-1}(\widehat{\boldsymbol{\gamma}^{(r)}})\boldsymbol{u}(\widehat{\boldsymbol{\gamma}^{(r)}})$$

f. Jika parameter yang didapatkan belum konvergen, maka dilakukan kembali langkah pada bagian ke- e hingga iterasi ke dari$r = r + 1$. Proses Iterasi akan berhenti jika telah diperoleh estimasi parameter yang konvergen dngan memenuhi. Proses Iterasi akan berhenti jika telah diperoleh estimasi parameter yang konvergen dengan memenuhi $\left(\left|\widehat{\boldsymbol{\gamma}^{(r+1)}} - \widehat{\boldsymbol{\gamma}^{(r)}}\right|\right) \leq \epsilon$ dimana ϵ dalah nilai yang sangat kecil dan telah diterapkan sebelumnya, misal 10^{-4}

UJI SIGNIFIKANSI SIMULTAN REGRESI HNB

Pengujian Simultan Model Regresi Hurdle Negative Binomial yang beranggotakan seluruh variabel prediktor menggunakan Likelihood Ratio Test untuk menentukan apakah variabel prediktor secara bersama-sama berpengaruh signifikan terhadap variable respon disebut uji signifikansi parameter (Agresti, 2015). Hipotesis yang digunakan adalah:

$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = \gamma_1 = \gamma_2 = \cdots = \gamma_p = 0$

$H_1: paling\ sedikit\ ada\ satu\ \beta_j \neq 0\ atau\ \gamma_j \neq 0$

Uji yang digunakan adalah uji G dengan rumus sebagai berikut:

$$G = -2\ln\left(\frac{L(1)}{L(2)}\right) = -2[\ln L\,(1) - \ln L\,(2)] \qquad (34)$$

Kriteria pengujian Tolak H_0 jika nilai $G > X^2_{\alpha,p}$ atau nilai p−value < α yang berarti paling sedikit ada satu variabel prediktor yang berpengaruh signifikan terhadap variabel respon.

UJI SIGNIFIKANSI PARSIAL REGRESI HNB

Pengujian signifikansi parameter secara parsial pada regresi *Hurdle Negative Binomial* adalah uji parameter pada model secara individual yang terdiri dari masing-masing variabel bebas atau digunakan untuk mengetahui pengaruh variabel predictor terhadap variabel respon dengan menggunakan uji Wald (Hosmer dan Lemeshow, 2000). Hipotesis yang digunakan adalah:

a. $H_0: \beta_j = 0$ (Varibel prediktor tidak berpengaruh secara signifikan)
$H_1: \beta_j \neq 0, j = 1,2, \dots, p$ (Varibel prediktor berpengaruh secara signifikan)
Uji yang digunakan adalah uji Wald dengan rumus sebagai berikut:

$$W_j = \left(\frac{\beta_j}{SE(\beta_j)}\right)^2 \qquad (35)$$

b. $H_0: \gamma_j = 0$ (Varibel prediktor tidak berpengaruh secara signifikan)
$H_1: \gamma_j \neq 0, j = 1,2, \dots, p$ (Varibel prediktor berpengaruh secara signifikan)
Uji yang digunakan adalah uji Wald dengan rumus sebagai berikut:

$$W_j = \left(\frac{\gamma_j}{SE(\gamma_j)}\right)^2 \qquad (36)$$

Kriteria pengujian tolak H_0 jika $|W_j| > Z_{\frac{\alpha}{2}}$ atau p−value < α, yang berarti suatu variabel prediktor berpengaruh secara signifikan terhadap variabel respon.

METODE

DATA PENELITIAN

Jenis penelitian yang digunakan merupakan penelitian kuantitatif. Data penelitian diperoleh dari Profil Kesehatan Provinsi Jawa Timur yang dipublikasikan oleh Dinas Kesehatan Jawa Timur dan Jawa Timur Dalam Angka yang dipublikasikan oleh Badan Pusat Statistik Jawa Timur. Jumlah pengamatan adalah seluruh Kabupaten/Kota di Provinsi Jawa Timur yang berjumlah 38 Kabupaten/Kota. Dalam penelitian ini variabel respon (Y) yaitu Jumlah kematian akibat filariasis. Sedangkan variable prediktor (x) adalah persentase penduduk miskin, persentase sanitasi layak, persentase air bersih, jumlah sarana kesehatan, persentase indeks pembangunan manusia, persentase keluhan kesehatan.

RANCANGAN PENELITIAN

Rancangan penelitian digunakan untuk mencapai tujuan dari penelitian yang dilakukan. Langkah pertama melakukan pemeriksaan multikolinieritas untuk mengetahui ada atau tidaknya multikolenieritas pada masing-masing variable prediktor, kemudian melakukan pengecekan overdispersi untuk mengetahui data tersebut mengalami overdispersi atau tidak. Salah satu permasalahan yang menyebabkan overdispersi adalah kondisi saat nilai nol pada variabel respon lebih banyak dibandingkan nilai lainnya (excess zeros) sehingga dilakukan pemeriksaan excess zeros. Jika data mengalami overdispersi dan excess zeros maka regresi poisson tidak dapat digunakan sehingga diperlukan regresi lain untuk mengatasi permasalahan tersebut. Regresi yang dapat digunakan adalah regresi *Hurdle Negative Binomial* (HNB). Berdasarkan tujuan dari penelitian ini maka dilakukan pemodelan menggunakan regresi HNB pada jumlah kematian akibat filariasis. Parameter pada model regresi HNB diperoleh dengan melakukan penaksiran parameter menggunakan *Maximum Likelihood Estimation* (MLE). Langkah selanjutnya melakukan pengujian signifikansi parameter model regresi HNB yang dilakukan secara simultan menggunkan uji G dan secara parsial menggunakan uji Wald.

HASIL DAN PEMBAHASAN

PEMERIKSAAN MULTIKOLINEARITAS

Hasil pengecekan multikolinieritas pada sepuluh variabel prediktor ditunjukkan pada tabel berikut:

Tabel 1. Pemeriksaan Multikolinearitas

Variable Prediktor	Hasil VIF
x_1	3,666947
x_2	2,568165
x_3	1,378986
x_4	1,059302
x_5	5,516152
x_6	1,419810

Berdasarkan tabel berikut dapat diketahui bahwa hasil perhitungan nilai VIF dari masing – masing variabel prediktor bernilai kurang dari 10 sehingga dapat disimpulkan bahwa tidak terjadi multikolinearitas pada setiap variabel prediktor.

PEMERIKSAAN OVERDISPERSI

Hasil pemeriksaan overdispersi ditunjukkan oleh tabel berikut:

Tabel 2. Pemeriksaan Overdispersi

Kriteria	Nilai	Db	Rasio
Deviance	88,24376	31	2,846573
Pearson Chi Square	99,1792	31	3,199329

Berdasarkan tabel tersebut diperoleh nilai deviance sebesar 88,24376 dengan derajat bebas 31 dan nilai Pearson Chi Square sebesar 99,1792 dengan derajat bebas 31. Sehingga rasio yang diperoleh dengan membagi nilai deviance dengan derajat bebas memperoleh hasil 2,846573, sedangkan nilai Pearson Chi Square yang dibagi dengan derajat bebas memperoleh hasil 3,199329. Karena hasil yang diperoleh lebih besar dari satu yaitu 2,846573 > 1 dan 3,199329 > 1, maka keputusannya tolak H_0 yang berarti data jumlah kematian akibat filariasis di Jawa Timur 2022 mengalami overdispersi. Sehingga tidak memenuhi asumsi pada regresi Poisson yaitu equidispersi.

PEMERIKSAAN EXCESS ZEROS

Hasil pemeriksaan excess zeros pada penelitian ini sebagai berikut:

Tabel 3. Pemeriksaan Excess Zeros

Variablel Respon	Frekuensi dari variable respon	Persentase
0	24	63,158
1	5	13,158
2	1	2,63
3	3	7,895
5	2	5,263
6	1	2,63
9	1	2,63
18	1	2,63

Berdasarkan tabel di atas dapat disimpulkan bahwa terjadi masalah excess zeros pada data jumlah kematian akibat filariasis di Jawa Timur 2022 karena persentase observasi bernilai nol lebih dari 50% yaitu 63,158 %.

PEMODELAN REGRESI HNB

ESTIMASI PARAMETER DENGAN MLE

Hasil estimasi parameter pada regresi Hurdle Negative Binomial menggunakan Maximum Likelihood Estimation (MLE) ditunjukkan pada tabel berikut:

Tabel 4. Hasil Estimasi Parameter

Parameter	Estimasi	Std. Error	Z-value	P-value
β_0	27,42753	14,02936	1,955	0,050582
β_1	0,55495	0,19771	2,807	0,005002
β_2	-0,18652	0,05678	3,285	0,001021
β_3	-0,05563	0,01422	-3,911	9,2e-05

β_4	-0,08274	0,01835	4,510	6,5e-06
β_5	-0,74279	0,25303	-2,936	0,003330
β_6	0,13657	0,03767	3,626	0,000288
γ_0	15,903476	14,003811	1,136	0,2561
γ_1	-0,105121	0,1825281	-0,576	0,5647
γ_2	0,0635626	0,0538645	1,180	0,2380
γ_3	-0,000181	0,056921	-0,007	0,9944
γ_4	0,0455961	0,0271498	1,679	0,0931
γ_5	-0,355480	0,2018724	-1,761	0,0783
γ_6	0,1030664	0,0509330	2,024	0,0430

Berdasarkan tabel di atas dapat dibentuk model regresi *Hurdle Negative Binomial* dalam fungsi berikut ini :

a. Model *Zero Hurdle State*

$$logit(\pi_i) = 15{,}903476 - 0{,}105121x_{i1} + 0{,}0635626\, x_{i2} - 0{,}000181x_{i3} + 0{,}0455961x_{i4} - 0{,}355480x_{i5} + 0{,}1030664x_{i6}$$

b. Model *Truncated Negative Binomial State*

$$\ln(\mu_i) = 27{,}42753 + 0{,}55495x_{i1} - 0{,}18652\, x_{i2} - 0{,}05563x_{i3} - 0{,}08274x_{i4} - 0{,}74279x_{i5} + 0{,}13657x_{i6}$$

UJI SIGNIFIKANSI PARAMETER SIMULTAN

Hasil uji G diperoleh nilai sebesar 32,719. Dengan kriteria pengujian yang digunakan tolak H_0 jika $G > X^2_{\alpha;p}$ dengan $p = 6$ dan taraf signifikansi $\alpha = 0{,}05$. Sehingga diperoleh $32{,}719 > 12{,}592$ ($G > X^2_{\alpha;p}$. Maka keputusannya adalah tolak H_0 yang berarti paling sedikit ada satu variabel prediktor terhadap variabel respon.

UJI SIGNIFIKANSI PARAMETER PARSIAL

Nilai uji Wald diperoleh dengan membagi nilai estimasi dengan standar eror atau dapat dilihat pada statistik uji Wald bahwa nilainya setara dengan nilai Z – value. Kriteria pengujiannya adalah tolak H_0 jika $|W_j| > Z_{\frac{\alpha}{2}}$ atau $p - value < \alpha$ dengan $\alpha = 0{,}05$ yang berarti suatu variabel prediktor berpengaruh secara signifikan terhadap variabel respon. Sehingga berdasarkan tabel (4) diperoleh hasil variable prediktor yang berpengaruh secara signifikan terhadap variabel respon adalah persentase penduduk miskin, persentase sanitasi layak, persentase air bersih, jumlah sarana kesehatan, persentase indeks pembangunan manusia, persentase keluhan kesehatan.

INTERPRETASI HASIL

MODEL REGRESI HNB

Interpretasi model *Truncated Negative Binomial State* (ln) sebagai berikut:

a. Nilai koefisien persentase penduduk miskin (x_1) adalah 0,55495, sehingga dapat diinterpretasikan bahwa apabila terjadi peningkatan persentase penduduk miskin sebesar 1% dapat menyebabkan terjadinya peningkatan jumlah kasus kematian akibat filariasis sebesar $e^{0{,}55495} = 1{,}5549$ kali.
b. Nilai koefisien persentase sanitasi layak (x_2) adalah -0,18652, sehingga dapat diinterpretasikan bahwa apabila terjadi peningkatan persentase sanitasi layak sebesar 1% dapat menyebabkan terjadinya penurunan jumlah kasus kematian akibat filariasis sebesar $e^{-0{,}18652} = 0{,}8298$ kali.
c. Nilai koefisien persentase air bersih (x_3) adalah -0,05563, sehingga dapat diinterpretasikan bahwa apabila terjadi peningkatan persentase air bersih sebesar 1% dapat menyebabkan terjadinya penurunan jumlah kasus kematian akibat filariasis sebesar $e^{-0{,}05563} = 0{,}5733$ kali.
d. Nilai koefisien jumlah sarana kesehatan (x_4) adalah -0,08274, sehingga dapat diinterpretasikan bahwa apabila terjadi peningkatan jumlah sarana kesehatan sebesar 1% dapat menyebabkan terjadinya penurunan jumlah kasus kematian akibat filariasis sebesar $e^{-0{,}08274} = 09206$ kali.
e. Nilai koefisien persentase indeks pembangunan manusia (x_5) adalah -0,74279, sehingga dapat diinterpretasikan bahwa apabila terjadi peningkatan persentase indeks pembangunan manusia sebesar 1% dapat menyebabkan terjadinya penurunan jumlah kasus kematian akibat filariasis sebesar $e^{-0{,}74279} = 0{,}4758$ kali.
f. Nilai koefisien persentase keluhan kesehatan (x_6) adalah 0,13657, sehingga dapat diinterpretasikan bahwa apabila terjadi peningkatan persentase keluhan kesehatan sebesar 1% dapat menyebabkan terjadinya peningkatan jumlah kasus kematian akibat filariasis sebesar $e^{0{,}13657} = 1{,}1463$ kali.

Sedangkan interpretasi model *Zero Hurdle State* (logit) sebagai berikut:

a. Nilai koefisien persentase keluhan kesehatan (x_6) adalah 0,1030664, sehingga dapat diinterpretasikan bahwa apabila terjadi peningkatan persentase keluhan kesehatan sebesar 1% dapat menyebabkan terjadinya peningkatan jumlah kasus kematian akibat filariasis sebesar $e^{0,1030664} = 1,1086$ kali.

PENUTUP

SIMPULAN

Model Regresi *Hurdle Negative Binomial*

a. Model yang terbentuk terdiri dari dua model yaitu *Zero Hurdle State* dan *Truncated Negative Binomial State* sebagai berikut:

$$logit(\pi_i) = 15{,}903476 - 0{,}105121x_{i1} + 0{,}0635626\, x_{i2} - 0{,}000181x_{i3} + 0{,}0455961x_{i4} - 0{,}355480x_{i5} + 0{,}1030664x_{i6}$$

Dan

$$logit(\pi_i) = 15{,}903476 - 0{,}105121x_{i1} + 0{,}0635626\, x_{i2} - 0{,}000181x_{i3} + 0{,}0455961x_{i4} - 0{,}355480x_{i5} + 0{,}1030664x_{i6}$$

Berdasarkan hasil pengujian signifikansi parameter secara simultan dipeoleh hasil bahwa variabel prediktor berpengaruh secara simultan terhadap variabel respon dan uji signifikansi parameter secara parsial diperoleh hasil variable prediktor yang berpengaruh secara signifikan terhadap variabel respon adalah persentase penduduk miskin, persentase sanitasi layak, persentase air bersih, jumlah sarana kesehatan, persentase indeks pembangunan manusia, persentase keluhan kesehatan.

SARAN

Saran untuk penelitian selanjutnya dapat melakukan komparasi dengan metode lain karena penanganan masalah overdispersi dan excess zeros pada model regresi poisson tidak hanya terbatas pada regresi *Hurdle Negative Binomial*. Pada penelitian berikutnya diharapkan dapat menambahkan variabel prediktor yang lebih berkaitan dengan faktor kesehatan manusia dan lingkungan yang kemungkinan berpengaruh terhadap data kasus kematian akibat suatu penyakit sehingga dapat mengetahui faktor lainnya yang berpengaruh signifikan.

DAFTAR PUSTAKA

Agresti, A., 2015. Foundations of linear and generalized linear models. John Wiley & Sons.

BPS, 2023. Provinsi jawa timur dalam angka 2023. Accessed: 2 oktober 2023.URL https://jatim.bps.go.id/publication/2023/02/28/446036fbb58d36b009212dbc/provinsi-jawa-timur-dalam-angka-2023.html.

Cameron, A. C., dan Trivedi, P. K., 2013. Regression analysis of count data, vol. 53. Cambridge university press.

Dinkes, 2022. Profil kesehatan jawa timur 2022. Accessed: 2 oktober 2023. URL https://dinkes.jatimprov.go.id/userfile/dokumen/PROFIL%20KESEHATAN%20JATIM%202022.pdf.

Ehsan Saffari, S., Adnan, R., dan Greene, W., 2012. Hurdle negative binomial regression model with right censored count data. SORT: statistics and operations researchtransactions, 36(2), 0181–194.

Famoye, F., dan Singh, K. P., 2006. Zero-inflated generalized poisson regression model with an application to domestic violence data. Journal of Data Science, 4(1), 117–130.

Garay, A. M., Hashimoto, E. M., Ortega, E. M., dan Lachos, V. H., 2011. On estimation and influence diagnostics for zero-inflated negative binomial regression models. Computational Statistics & Data Analysis, 55(3), 1304–1318.

Greene, W., 2008. Functional forms for the negative binomial model for count data. Economics Letters, 99(3), 585–590.

Hall, D. B., 2000. Zero-inflated poisson and binomial regression with random effects: a case study. Biometrics, 56(4), 1030–1039.

Hilbe, J. M., 2011. Negative binomial regression. Cambridge University Press.

Hocking, R. R., 2013. Methods and applications of linear models: regression and the analysis of variance. John Wiley & Sons.

Hosmer, D., dan Lemeshow, S., 2000. Applied logistic regression, second eition.

Jumiati, J., Kalsum, U., dan Ilham, I., 2020. Analisis faktor risiko lingkungan terhadap kejadian filariasis di kabupaten tanjung jabung timur. *Jurnal Pembangunan Berkelanjutan*, 3(2), 13–19.

Kemenkes, 2022. Laporan kementrian kesehatan 2022. Accessed: 5 oktober 2023. URL https://p2pm.kemkes.go.id/storage/informasi-publik/content/GHwE3BiLbOrvZ ZPKY1Pm91BIRWqzE4metaTGFwa2luIFAyUE0gMjAyMi5wZGY=pdf.

Ma, L., Yan, X., Wei, C., dan Wang, J., 2016. Modeling the equivalent property damage only crash rate for road segments using the hurdle regression framework. Analytic methods in accident research, 11, 48–61.

Saengthong, P., Bodhisuwan, W., dan Thongteeraparp, A., 2015. The zero inflated negative binomial–crack distribution: some properties and parameter estimation. Songklanakarin J. Sci. Technol, 37(6), 701–711.

Saputro, M. I. A., dan Qudratullah, M. F., 2021. Estimation of zero-inflated negative binomial regression parameters using the maximum likelihood method (case study: Factors affecting infant mortality in wonogiri in 2015). Dalam Proceeding International Conference on Science and Engineering, vol. 4, (hal. 240–254).

Sinurat, S. R. Y., dan Pasaribu, E., 2022. Pemodelan kasus kronis filariasis di Indonesia tahun 2019 menggunakan geographically weighted negative binomial regression (gwnbr). *Indonesian Journal of Applied Statistics*, 5(1), 19–30.

Sularno, S., Nurjazuli, R. M., dan Raharjo, M., 2017. Faktor-faktor yang berhubungan dengan kejadian filariasis di kecamatan buaran kabupaten pekalongan. *J Kesehat Lingkung Indones*, 16(1), 22–8.