

PENANGANAN OVERDISPERSI PADA MODEL REGRESI POISSON MENGUNAKAN MODEL REGRESI BINOMIAL NEGATIF

Rio Tongaril Simarmata¹, Dwi Ispriyanti²

¹Alumni Program Studi Statistika Undip

²Staf Pengajar Program Studi Statistika Undip

Abstract

Poisson regression is the most popular tool for modeling the relationship between a discrete data in the response variable and a set of predictors with continue, discrete, categoric or mix data. Response variable with discrete data, however, may overdispersed or underdispersed, not conductive to Poisson regression which assumed that the mean value equals to variance (equidispersed). One of the model that be used to overdispersed the discrete data is a regression model based on mixture distribution namely Poisson-gamma mixture which result negative binomial distribution. This regression model usually known as binomial negative regression. Using Generalized Linier Model (GLM) approach, the given model, parameter estimate, diagnostics, and interpretation of negative binomial regression can be determined.

Keyword: Negative Binomial Distribution, Dispersion, Generalized Linier Model

1. Pendahuluan

Analisis regresi adalah suatu metode yang digunakan untuk menganalisis hubungan antara variabel respon dengan beberapa variabel prediktor. Pada umumnya analisis regresi digunakan untuk menganalisis data variabel respon yang berupa data kontinu. Namun dalam beberapa aplikasinya, data variabel respon yang akan dianalisis dapat berupa data diskrit. Salah satu model regresi yang dapat digunakan untuk menganalisis hubungan antara variabel respon Y yang berupa data diskrit dengan variabel prediktor X berupa data diskrit, kontinu, kategorik atau campuran adalah model regresi Poisson.

Dalam model regresi Poisson terdapat beberapa asumsi yang harus dipenuhi, salah satunya adalah nilai variansi dari variabel respon Y yang diberikan oleh $X = x$ harus sama dengan nilai meannya yaitu: $Var(Y | x) = E(Y | x) = \mu$. Namun dalam analisis data diskrit dengan menggunakan model regresi Poisson terkadang terjadi pelanggaran asumsi tersebut, dimana nilai variansinya lebih besar dari nilai mean yang disebut overdispersi atau varian lebih kecil dari nilai mean yang disebut underdispersi. Dalam model regresi linier klasik pelanggaran tersebut dinamakan pelanggaran asumsi homokedastisitas^[3]. Jika terjadi kasus overdispersi, penanganan model yang dapat digunakan adalah model regresi binomial negatif. Model ini dibentuk dari distribusi *mixture* Poisson-gamma dimana:

$$y | \mu \sim Poisson(\mu)$$

$$\mu \sim Gamma(\alpha, \beta)$$

Jika suatu distribusi Poisson (μ) dimana μ merupakan nilai variabel random yang berdistribusi gamma, maka akan dihasilkan distribusi *mixture* yang dinamakan distribusi binomial negatif. Model regresi binomial negatif mengasumsikan

$$E(Y | x) = \mu \quad \text{dan} \quad Var(Y | x) = \mu + k\mu^2$$

dengan k adalah parameter dispersi = $1/\alpha$ dimana α = parameter bentuk dari distribusi gamma.

Model ini dapat mengatasi masalah overdispersi karena tidak mengharuskan nilai mean yang sama dengan nilai variansi seperti pada model regresi Poisson.

2. Regresi Poisson

a. Model

Regresi Poisson merupakan penerapan dari Generalisasi Model Linier (GML) yang menggambarkan hubungan antara variabel respon Y data diskrit berdistribusi Poisson dengan variabel prediktor X . Jika Y merupakan data diskrit yang berdistribusi Poisson dengan parameter $\mu > 0$ maka fungsi massa peluangnya adalah^[2]:

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

dengan asumsi :

$$E(y) = \mu \quad \text{dan} \quad \text{Var}(y) = \mu$$

Persamaan (1) dapat ditulis dalam bentuk

$$\begin{aligned} f(y; \mu) &= \exp[y_i \ln(\mu_i) - \mu_i - \ln(y_i!)] \\ &= \exp[y_i \theta_i - b(\theta_i) - \ln(y_i!)] \end{aligned}$$

dimana $\theta_i = \ln(\mu_i)$. Persamaan ini merupakan suatu bentuk persamaan fungsi distribusi keluarga eksponensial.

Dengan menggunakan fungsi link diperoleh model regresi Poisson dalam bentuk:

$$\ln \mu = \mathbf{X}\boldsymbol{\beta}$$

$$\mu_i = \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})$$

dengan μ_i nilai ekspektasi Y_i berdistribusi Poisson

Penaksiran koefisien parameter regresi Poisson menggunakan metode maksimum likelihood yaitu dengan melakukan turunan parsial fungsi ln-likelihood terhadap parameter yang akan diestimasi dan diiterasikan dengan menggunakan metode iterasi Newton-Rhapon. Fungsi ln-likelihood untuk regresi Poisson adalah:

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n [y_i \ln(\mu_i) - \mu_i - \ln(y_i!)]$$

b. Overdispersi

Regresi Poisson dikatakan mengandung overdispersi apabila nilai variansinya lebih besar dari nilai meannya. Overdispersi memiliki dampak yang sama dengan pelanggaran asumsi homokedastisitas dalam model regresi linier^[3]. Jika pada data diskrit terjadi overdispersi namun tetap digunakan model regresi Poisson maka estimasi parameter koefisien regresinya tetap konsisten tetapi tidak efisien karena berdampak pada nilai standar error. Untuk mendeteksi terjadinya masalah overdispersi dalam model regresi Poisson, dapat dilakukan dengan menguji hubungan antara varian dan mean dalam bentuk persamaan : $v(\mu_i) = \phi \mu_i$ ^[1], nilai ϕ adalah constant (parameter gangguan) dan dapat diestimasi dengan menggunakan pendekatan nilai bagi pearson chi-square dengan derajat bebasnya^[1]. Jika nilainya lebih besar dari satu maka dapat disimpulkan terdapat overdispersi pada model regresi Poisson.

3. Regresi Binomial Negatif

a. Generalisasi Model Linier Binomial Negatif

Regresi Binomial Negatif merupakan salah satu model regresi terapan dari Generalisasi Model Linier (GML). Sebagai penerapan dari Generalisasi Model Linier (GML) maka distribusi binomial negatif memiliki ketiga komponen yang akan dijelaskan sebagai berikut^[6]:

i. Komponen Random

Pada regresi binomial negatif variabel respon Y_i diasumsikan berdistribusi binomial negatif yang dihasilkan dari distribusi *mixture* Poisson-gamma.

Misalkan :

$$y | \mu \sim \text{Poisson}(\mu)$$

$$\mu \sim \text{Gamma}(\alpha, \beta)$$

Fungsi massa peluang Poisson-gamma *mixture* dapat diperoleh dengan cara:

$$\begin{aligned} P(y | \alpha, \beta) &= \int_0^{\infty} \text{Poisson}(y | \mu) \cdot \text{Gamma}(\mu | \alpha, \beta) d\mu \\ &= \int_0^{\infty} \frac{e^{-\mu} \mu^y}{y!} \cdot \frac{1}{\Gamma(\alpha) \beta^\alpha} \mu^{\alpha-1} \exp(-\mu / \beta) d\mu \\ &= \int_0^{\infty} \frac{e^{-\mu} \mu^y \mu^{\alpha-1} e^{(-\mu / \beta)}}{y! \Gamma(\alpha) \beta^\alpha} d\mu \\ &= \frac{\beta^{-\alpha}}{y! \Gamma(\alpha)} \int_0^{\infty} e^{-\mu(1+1/\beta)} \mu^{y+\alpha-1} d\mu \end{aligned}$$

Misalkan $v = \left(1 + \frac{1}{\beta}\right) \mu$ maka $dv = \left(1 + \frac{1}{\beta}\right) d\mu$, dan untuk $\mu = 0 \rightarrow v = 0$,

$$\mu = \infty \rightarrow v = \infty$$

$$\begin{aligned} P(y | \alpha, \beta) &= \frac{\beta^{-\alpha}}{y! \Gamma(\alpha)} \int_0^{\infty} e^{-v} \left(\frac{\beta v}{1 + \beta}\right)^{y+\alpha-1} \frac{\beta}{1 + \beta} dv \\ &= \frac{\beta^{-\alpha}}{y! \Gamma(\alpha)} \left(\frac{\beta}{1 + \beta}\right)^{y+\alpha} \int_0^{\infty} e^{-v} v^{y+\alpha-1} dv \\ &= \frac{\beta^{-\alpha}}{y! \Gamma(\alpha)} \left(\frac{\beta}{1 + \beta}\right)^\alpha \left(\frac{\beta}{1 + \beta}\right)^y \Gamma(y + \alpha) \end{aligned}$$

$$\begin{aligned} P(Y | \alpha, \beta) &= \frac{\Gamma(y + \alpha)}{y! \Gamma(\alpha)} \left(\frac{1}{1 + \beta}\right)^\alpha \left(\frac{\beta}{1 + \beta}\right)^y, \text{ dengan } y = 0, 1, 2, \dots \\ &= \frac{\Gamma(y + \alpha)}{y! \Gamma(\alpha)} \left(\frac{1}{1 + \beta}\right)^\alpha \left(1 - \frac{1}{1 + \beta}\right)^y \end{aligned}$$

$P(Y | \alpha, \beta)$ merupakan fungsi massa peluang binomial negatif yang dihasilkan dari distribusi *mixture* Poisson-gamma. Nilai mean dan variansi Poisson-gamma *mixture* adalah : $E[Y] = \alpha\beta$ dan $V[Y] = \alpha\beta + \alpha\beta^2$

Untuk membentuk suatu model regresi pada distribusi binomial negatif, maka nilai parameter dari distribusi Poisson gamma *mixture* dinyatakan dalam bentuk $\mu = \alpha\beta$ dan $k = 1/\alpha$ sehingga diperoleh mean dan variansi dalam bentuk

$$E[Y] = \mu \text{ dan } V[Y] = \mu + k\mu^2$$

Kemudian fungsi massa peluang binomial negatif menjadi:

$$f(y; \mu, k) = \frac{\Gamma(y + 1/k)}{\Gamma(1/k)y!} \left(\frac{1}{1 + k\mu}\right)^{1/k} \left(\frac{k\mu}{1 + k\mu}\right)^y \quad y = 0, 1, 2, \dots \quad (2)$$

Saat $k \rightarrow 0$ maka distribusi binomial negatif memiliki varian $V[Y] \rightarrow \mu$. Distribusi binomial negatif akan mendekati suatu distribusi Poisson yang mengasumsikan mean dan variansi sama yaitu $E[Y] = V[Y] = \mu$. Fungsi distribusi keluarga eksponensial dari distribusi binomial negatif adalah

$$f(y; \mu, k) = \exp \left\{ y \ln \left(\frac{k\mu}{1 + k\mu} \right) + \frac{1}{k} \ln \left(\frac{1}{1 + k\mu} \right) + \ln \left(\frac{\Gamma(y + 1/k)}{\Gamma(1/k)y!} \right) \right\} \quad (3)$$

ii. Komponen Sistematis

Kontribusi variabel prediktor dalam model regresi binomial negatif dinyatakan dalam bentuk kombinasi linier antara parameter (η) dengan parameter regresi yang akan diestimasi yaitu :

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

Atau dalam matriks dituliskan dalam bentuk

$$\eta = \mathbf{X}\beta$$

dengan η adalah vektor ($n \times 1$) dari observasi, \mathbf{X} adalah matriks ($n \times c$) dari variabel bebas, β adalah matriks ($c \times 1$) dari koefisien regresi, dengan $c = p + 1$

iii. Fungsi Link

Nilai ekspektasi dari variabel respon Y adalah diskrit dan bernilai positif. Maka untuk mentransformasikan nilai η_i (bilangan riil) ke rentang yang sesuai dengan rentang pada respon y diperlukan suatu fungsi link $g(\cdot)$ yaitu:

$$g(\mu) = \ln \mu_i = \mathbf{X}\beta$$

b. Estimasi Parameter dan Uji Kesesuaian Model Regresi Binomial Negatif

Estimasi parameter dari regresi binomial negatif digunakan metode maksimum likelihood dengan prosedur iterasi Fisher Scoring dan Newton Rhapsion^[5]. Metode ini membutuhkan turunan pertama dan kedua dari fungsi likelihood. Y_i mempunyai fungsi massa probabilitas distribusi binomial negatif seperti pada persamaan (2) yaitu:

$$f(y_i | \mu_i, k) = \frac{\Gamma(y_i + 1/k)}{\Gamma(1/k)\Gamma(y_i + 1)} \left(\frac{1}{1 + k\mu_i}\right)^{1/k} \left(\frac{k\mu_i}{1 + k\mu_i}\right)^{y_i}, \text{ dengan } y = 0, 1, 2, \dots \quad (4)$$

Karena fungsinya saling bebas, maka fungsi likelihood adalah:

$$l(\beta, k) = \prod_{i=1}^n \frac{\Gamma(y_i + 1/k)}{\Gamma(1/k)\Gamma(y_i + 1)} \left(\frac{1}{1 + k\mu_i}\right)^{1/k} \left(\frac{k\mu_i}{1 + k\mu_i}\right)^{y_i}$$

dengan $\frac{\Gamma(y + 1/k)}{\Gamma(1/k)} = \prod_{r=1}^{y-1} (r + k^{-1})$

$$l(\beta, k) = \prod_{i=1}^n \left(\prod_{r=1}^{y_i-1} (r + k^{-1}) \right) \frac{1}{(y_i!)} \left(\frac{1}{1 + k\mu_i}\right)^{1/k} \left(\frac{k\mu_i}{1 + k\mu_i}\right)^{y_i}$$

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

Turunan pertama dari fungsi likelihood terhadap koefisien regresi β adalah

$$\frac{\partial L(\beta, k)}{\partial \beta_0} = \sum_{i=1}^n \left[y_i - (y_i + k^{-1}) \left(\frac{k\mu_i}{1+k\mu_i} \right) \right] = \sum_{i=1}^n \left[\frac{y_i - \mu_i}{1+k\mu_i} \right] = 0$$

:

$$\frac{\partial L(\beta, k)}{\partial \beta_p} = \sum_{i=1}^n \left[y_i x_{ip} - (y_i + k^{-1}) \left(\frac{k\mu_i x_{ip}}{1+k\mu_i} \right) \right] \\ = \sum_{i=1}^n \left[\frac{(y_i - \mu_i) x_{ip}}{1+k\mu_i} \right] \\ = \sum_{i=1}^n \left[\frac{\mu_i}{1+k\mu_i} \frac{(y_i - \mu_i) x_{ip}}{\mu_i} \right] = 0$$

bentuk persamaan matriks dari turunan pertama fungsi likelihood terhadap parameter β yaitu : $\mathbf{q} = \mathbf{X}^T \mathbf{Wz}$, dengan \mathbf{X} adalah matriks ($n \times c$) dari variabel prediktor, \mathbf{W} adalah matriks weight diagonal ke- i dan \mathbf{z} adalah vektor matriks dengan baris ke- i , dengan masing-masing elemennya adalah:

$$w_i = \frac{\mu_i}{1+k\mu_i} \quad \text{dan} \quad z_i = \frac{(y_i - \mu_i)}{\mu_i}$$

Turunan pertama dari fungsi log-likelihood terhadap parameter dispersi k adalah :

$$f'(k) = \frac{\partial L(\beta, k)}{\partial k} = \sum_{i=1}^n \left[-k^{-2} \sum_{r=0}^{y_i-1} \frac{1}{(r+k^{-1})} + \frac{y_i}{k} + k^{-2} \ln(1+k\mu_i) - \frac{(k^{-1} + y_i)\mu_i}{(1+k\mu_i)} \right] = 0$$

$$f'(k) = \frac{\partial L(\beta, k)}{\partial k} = \sum_{i=1}^n \left[-k^{-2} \sum_{r=0}^{y_i-1} \frac{1}{(r+k^{-1})} + k^{-2} \ln(1+k\mu_i) + \left(\frac{y_i - \mu_i}{k(1+k\mu_i)} \right) \right] = 0$$

Turunan parsial kedua fungsi likelihood terhadap parameter koefisien regresi β adalah

$$\frac{\partial^2 L(\beta, k)}{\partial \beta_0^2} = - \sum_{i=1}^n \left[\frac{(1+ky_i)\mu_i}{(1+k\mu_i)^2} \right] \\ \frac{\partial^2 L(\beta, k)}{\partial \beta_0 \partial \beta_j} = \sum_{i=1}^n \left[\frac{-\mu_i x_{ij} (1+k\mu_i) - (y_i - \mu_i) (x_{ij} k\mu_i)}{(1+k\mu_i)^2} \right] \\ \frac{\partial^2 L(\beta, k)}{\partial \beta_0 \partial \beta_j} = - \sum_{i=1}^n \left[\frac{(1+ky_i) x_{ij} \mu_i}{(1+k\mu_i)^2} \right]$$

Misalkan turunan parsial pertama dari $L(\beta, k)$ terhadap β_j , $j \leq p$ adalah

$$\frac{\partial L(\beta, k)}{\partial \beta_p} = \sum_{i=1}^n \left[\frac{(y_i - \mu_i) x_{ij}}{1+k\mu_i} \right] = 0$$

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

$$\begin{aligned} \frac{\partial^2 L(\beta, k)}{\partial \beta_u \beta_j} &= \sum_{i=1}^n \left[\frac{-\mu_i x_{iu} x_{ij} (1 + k\mu_i) - (y_i - \mu_i)(x_{iu} x_{ij} k\mu_i)}{(1 + k\mu_i)^2} \right] \\ \frac{\partial^2 L(\beta, k)}{\partial \beta_u \beta_j} &= -\sum_{i=1}^n \left[\frac{x_{iu} x_{ij} \mu_i (1 + k\mu_i)}{(1 + k\mu_i)^2} \right] \end{aligned} \quad (5)$$

Ekspektasi dari turunan kedua log-likelihood adalah:

$$E\left(-\frac{\partial^2 L(\beta, k)}{\partial \beta_u \beta_j}\right) = \sum_{i=1}^n \left[\frac{x_{iu} x_{ij} \mu_i}{(1 + k\mu_i)} \right] \quad (6)$$

Jika persamaan (6) dinyatakan dalam matriks **I** (matriks informasi) yaitu matriks yang mengandung ekspektasi negatif dari turunan kedua log-likelihood maka:

$$\mathbf{I} = \mathbf{X}^T \mathbf{W} \mathbf{X}$$

dengan **X** adalah matriks dari variabel prediktor, **W** adalah matriks weight diagonal ke-i

dengan elemen: $w_i = \frac{\mu_i}{1 + k\mu_i}$

Turunan kedua fungsi likelihood terhadap parameter dispersi k adalah

$$\begin{aligned} f'(k) &= \frac{\partial L(\beta, k)}{\partial k} = \sum_{i=1}^n \left[-k^{-2} \sum_{r=0}^{y_i-1} \frac{1}{(r + k^{-1})} + k^{-2} \ln(1 + k\mu_i) + \left(\frac{y_i - \mu_i}{k(1 + k\mu_i)} \right) \right] \\ f''(k) &= \frac{\partial L(\beta, k)}{\partial k^2} \\ &= \sum_{i=1}^n \left[k^{-3} \sum_{r=0}^{y_i-1} \frac{2r + k^{-1}}{(r + k^{-1})^2} - 2k^{-3} \ln(1 + k\mu_i) + \frac{k^{-2} \mu_i}{1 + k\mu_i} - \frac{(y_i - \mu_i)(1 + 2k\mu_i)}{(k + \mu_i k^2)^2} \right] \end{aligned}$$

Estimasi parameter regresi binomial negatif dilakukan dengan langkah sebagai berikut:

Langkah 1 : Tentukan taksiran awal dari k, misal $\hat{k}_1 = 0$

Langkah 2 : Tentukan estimasi maksimum likelihood dari parameter β menggunakan

prosedur iterasi Fisher Scoring dengan asumsi $k = \hat{k}_1$

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

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

Langkah 3 : Gunakan $\hat{\beta}$ untuk menghasilkan estimasi dari parameter k dengan menggunakan prosedur iterasi Newthton-Rhapon satu dimensi .

$$\hat{k}_{i+1} = \hat{k}_i - \frac{f'(\hat{k}_i)}{f''(\hat{k}_i)}$$

Iterasi berakhir jika diperoleh $\hat{k}_{i+1} \cong \hat{k}_i$

Langkah 4 : Jika $|\hat{k}_{i+1} - \hat{k}_1| < \varepsilon$ selesai; bila tidak, gunakan parameter $k = \hat{k}_{i+1}$ dan kembali ke langkah 2. Nilai ε merupakan nilai bilangan positif yang sangat kecil, misalnya $\varepsilon = 0.001$.

Untuk menguji kesesuaian model regresi binomial negatif digunakan uji deviansi dengan:

Hipotesis:

$$H_0 : \mu_i = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip})$$

(model regresi binomial negatif tepat digunakan sebagai model)

$$H_1 : \mu_i \neq \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip})$$

(model regresi binomial negatif tidak tepat digunakan sebagai model)

Statistik uji:

$$D = 2 \left[\sum y_i \ln \left(\frac{y_i}{\bar{y}} \right) - \sum (k^{-1} + y_i) \ln \left(\frac{1 + ky_i}{1 + k\bar{y}} \right) \right]$$

Kriteria Uji:

$$H_0 \text{ ditolak jika statistik uji } D > \chi_{\delta; n-p-1}^2$$

Uji signifikansi individu variabel prediktornya dengan menggunakan uji Wald dengan

Hipotesis:

$$H_0 : \beta_j = 0, \text{ dengan } j = 1, 2, \dots, p$$

$$H_1 : \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, p$$

Statistik uji:

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

Kriteria Uji:

$$H_0 \text{ ditolak jika statistik uji } W_j \geq \chi_{\delta; 1}^2$$

4. Contoh Terapan

Sebagai contoh penerapan model regresi binomial negatif digunakan data suatu kasus penelitian di daerah Connecticut, Massachusetts dan Vermont (Amerika Serikat) yang diambil dari (<http://www.sci.usq.edu.au/staff/dunn/Dataset/tech-GLM.html>) tentang pengaruh lokasi, ketinggian dari permukaan laut (m), dan posisi garis lintang (cm) terhadap jumlah semut yang ditemukan di dalam 44 blok penyampelan (observasi) dengan ukuran 64 m², 22 blok ditempatkan di daerah berlumpur dan 22 blok lainnya ditempatkan di hutan disekitar daerah berlumpur tersebut. Lokasi tersebut berada di sekitar 3⁰ dari garis lintang di New England. Untuk menganalisis model regresi binomial negatif maka terlebih dahulu dilakukan pengecekan asumsi bahwa variabel respon Y_i berdistribusi Poisson menggunakan uji *kolmogorov-smirnov* dengan bantuan program *SPSS 13 for Windows*

Tabel 1. One-Sample Kolmogorov-Smirnov Test

		Y
N		44
Poisson Parameter(a,b)	Mean	7.75
Most Extreme Differences	Absolute	0.158
	Positive	0.158
	Negative	-0.154
Kolmogorov-Smirnov Z		1.047
Asymp. Sig. (2-tailed)		0.223

Analisis Output untuk uji *kolmogorov-smirnov* adalah sebagai berikut:

Hipotesis:

H_0 : $F(X) = F_0(X)$ (Data berdistribusi Poisson)

H_1 : $F(X) \neq F_0(X)$ (Data tidak berdistribusi Poisson)

Tingkat signifikansi $\delta = 5\%$

Statistik Uji:

$$D = \max|F(X) - F_0(X)|$$

Tolak H_0 jika $D \geq D^*(\delta)$, dimana $D^*(\delta)$ merupakan nilai kritis yang diperoleh dari tabel “*Kolmogorov-Smirnov*”, atau menggunakan tolak H_0 jika $p\text{-value} \leq \delta$.

Berdasarkan hasil output pada Tabel 1 diperoleh nilai *asymp. sig (2-tailed)* = 0.223, jika dibandingkan dengan $\delta = 0.05$ maka nilai $p\text{-value} = 0.223 > \delta = 0.05$ yang berarti H_0 diterima. Sehingga dapat disimpulkan bahwa variabel respon (Y) adalah berdistribusi Poisson.

- **Overdispersi Model Regresi Poisson**

Tabel 2. Output Model Regresi Poisson

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	40	56.5205	1.4130
Scaled Deviance	40	56.5205	1.4130
Pearson Chi-Square	40	54.8447	1.3711
Scaled Pearson X2	40	54.8447	1.3711
Log Likelihood		385.8606	

Nilai *Pearson Chi-Square* model dibagi derajat bebasnya adalah 1.3711. Nilai tersebut lebih besar dari satu yang berarti dapat disimpulkan model mengalami overdispersi.

- **Uji Kesesuaian Model Regresi Binomial Negatif**

Uji kesesuaian model regresi binomial negative dilakukan dengan menggunakan prosedur pengujian nilai deviansi.

Tabel 3. Ouput Regresi Binomial Negatif

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	40	43.9689	1.0992
Scaled Deviance	40	43.9689	1.0992
Pearson Chi-Square	40	42.3412	1.0585
Scaled Pearson X2	40	42.3412	1.0585
Log Likelihood		386.6536	

Terlihat pada Tabel 3 nilai deviansi = 43.9689 lebih kecil dari nilai tabel chi-square $\chi^2_{0.5;40} = 55.78$). Sehingga H_0 diterima yang berarti model regresi binomial negatif dapat digunakan untuk menganalisis pengaruh lokasi, ketinggian, dan posisi garis lintang blok sampel yang dibentuk terhadap jumlah semut yang ditemukan didalam blok tersebut. Selanjutnya akan dilakukan uji individu koefisien regresi binomial negatif dengan menggunakan prosedur uji Wald.

Tabel 4. Output Regresi Binomial Negatif

Analysis Of Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	14.9449	2.8584	9.3426	20.5472	27.34	<.0001
hutan	1	0.3785	0.1263	0.1310	0.6260	8.98	0.0027
lintang	1	-0.3007	0.0673	-0.4325	-0.1689	19.99	<.0001
ketinggian	1	-0.0010	0.0004	-0.0018	-0.0002	6.42	0.0113
Dispersion	1	0.0377	0.0362	-0.0332	0.1086		

Berdasarkan uji individu koefisien regresi disimpulkan bahwa masing-masing variabel prediktor (lokasi, posisi garis lintang, dan ketinggian) memberikan pengaruh yang signifikan terhadap jumlah semut yang ditemukan dalam blok penyampelan (observasi). Sehingga diperoleh model regresi binomial negatif seperti berikut:

$$\mu_i = \exp(14.9449 + 0.3785X_{i1} - 0.3007X_{i2} - 0.001X_{i3})$$

5. Kesimpulan

Kesimpulan yang dapat diperoleh dari pembahasan diatas adalah sebagai berikut:

1. Distribusi yang digunakan dalam model regresi binomial negatif adalah distribusi Poisson-gamma mixture yaitu :

$$f(y; \mu, k) = \frac{\Gamma(y + 1/k)}{\Gamma(1/k)y!} \left(\frac{1}{1 + k\mu} \right)^{1/k} \left(\frac{k\mu}{1 + k\mu} \right)^y \quad y = 0, 1, 2, \dots$$

dengan: $E[Y] = \mu$ dan $V[Y] = \mu + k\mu^2$, k dinamakan parameter dispersi.

2. Regresi binomial negatif memenuhi ketiga komponen dari GML dengan asumsi variabel respon berdistribusi binomial negatif (Poisson-gamma mixture) dan fungsi link yang digunakan adalah log-link. Bentuk modelnya adalah

$$\mu_i = \exp\left(\beta_0 + \sum_{j=1}^p X_{ij}\beta_j\right)$$

3. Estimasi maksimum likelihood parameter regresi dan parameter dispersi binomial negatif tidak menghasilkan persamaan yang linier pada turunan parsial pertama sehingga hasil estimasinya diperoleh dengan menggunakan metode iterasi yaitu metode iterasi Fisher Scoring dan Newton-Rhapson.

DAFTAR PUSTAKA

1. Agresti, A., *Categorical Data Analysis*, Second edition, Jhon Wiley & Sons, Inc., New York, 2002.
2. Casella, G and Berger, R. L., *Statistical Inference*, Wadsworth, Inc., California, 1990.
3. Cameron, A.C and Tivedi, P.K., *Regression Analysis of Count Data*, Cambridge University Press., New York, 1998.
4. Cameron, A.C and Tivedi, P.K., *Essentials of Count Data Regression*. in Badi H. Baltagi ed., *A Companion to Theoretical Econometrics*, Blackwell, Oxford (U.K.), 1999: 331-348.
5. Dominique and Park, B., *Adjustment for The Maximum Likelihood Estimate of The Negative Binomial Dispersion Parameter*. Texas A&M University, 2007.
6. Greene, W., *Functional Forms For The Negative Binomial Model For Count Data*. Foundations and Trends in Econometrics. Working Paper, Department of Economics, Stern School of Business, New York University, 2008: 585-590.