

PERBANDINGAN REGRESI BINOMIAL NEGATIF DAN REGRESI CONWAY- MAXWELL-POISSON DALAM MENGATASI OVERDISPERSI PADA REGRESI POISSON

Lusi Eka Afri

lusiekaafri13@gmail.com

Program Studi Pendidikan Matematika

FKIP Universitas Pasir Pengaraian

2017

Abstrak

Regresi Binomial Negatif dan regresi Conway-Maxwell-Poisson merupakan solusi untuk mengatasi overdispersi pada regresi Poisson. Kedua model tersebut merupakan perluasan dari model regresi Poisson. Menurut Hinde dan Demetrio (2007), terdapat beberapa kemungkinan terjadi overdispersi pada regresi Poisson yaitu keragaman hasil pengamatan keragaman individu sebagai komponen yang tidak dijelaskan oleh model, korelasi antar respon individu, terjadinya pengelompokan dalam populasi dan peubah teramati yang dihilangkan. Akibatnya dapat menyebabkan pendugaan galat baku yang terlalu rendah dan akan menghasilkan pendugaan parameter yang bias ke bawah (*underestimate*). Penelitian ini bertujuan untuk membandingkan model Regresi Binomial Negatif dan model regresi Conway-Maxwell-Poisson (COM-Poisson) dalam mengatasi overdispersi pada data distribusi Poisson berdasarkan statistik uji devians. Data yang digunakan dalam penelitian ini terdiri dari dua sumber data yaitu data simulasi dan data kasus terapan. Data simulasi yang digunakan diperoleh dengan membangkitkan data berdistribusi Poisson yang mengandung overdispersi dengan menggunakan bahasa pemrograman R berdasarkan karakteristik data berupa μ , peluang munculnya nilai nol (p) serta ukuran sampel (n). Data dibangkitkan berguna untuk mendapatkan estimasi koefisien parameter pada regresi binomial negatif dan COM-Poisson.

Kata Kunci: overdispersi, regresi binomial negatif, regresi Conway-Maxwell-Poisson

Abstract

Negative binomial regression and Conway-Maxwell-Poisson regression could be used to overcome over dispersion on Poisson regression. Both models are the extension of Poisson regression model. According to Hinde and Demetrio (2007), there will be some over dispersion on Poisson regression: observed variance in individual variance cannot be described by a model, correlation among individual response, and the population group and the observed variables are eliminated. Consequently, this can lead to low standard error estimation and to downward bias parameter estimation (*underestimate*). This study aims to compare the Negative Binomial Regression model and Conway-Maxwell-Poisson (COM-Poisson) regression model to overcome over dispersion of Poisson distribution data based on deviance test statistics. The data used in this study are simulation data and applied case

data. The simulation data were obtained by generating the Poisson distribution data containing over dispersion using the R programming language based on data characteristic such as μ , the probability (p) of zero value and the sample size (n). The generated data is used to get the estimated parameter coefficient of the negative binomial regression and COM-Poisson.

Keywords: overdispersion, negative binomial regression and Conway-Maxwell-Poisson regression

I. Pendahuluan

Analisis regresi merupakan analisis statistika yang bertujuan untuk memodelkan hubungan antara variabel respon Y dengan satu atau lebih variabel prediktor X . Asumsi variabel responnya adalah data kontinu yang mengikuti distribusi normal. Namun dalam aplikasinya, banyak ditemukan penelitian yang menggunakan variabel respon yang berupa data cacah (*count data*). Analisis regresi yang menyatakan pola hubungan antara variabel respon berupa data cacah dengan variabel prediktor adalah regresi Poisson (Cameron dan Trivedi: 1998).

Analisis regresi Poisson memiliki asumsi nilai tengah yang sama dengan ragamnya yang dikenal dengan istilah *equidispersi*. Akan tetapi, pada beberapa penelitian ditemukan kondisi ragam lebih besar daripada nilai tengahnya atau disebut gejala overdispersi (McCullagh & Nelder 1989). Menurut Hinde dan Demetrio (2007), terdapat beberapa kemungkinan tidak dipenuhinya asumsi *equidispersi* tersebut yaitu keragaman hasil pengamatan keragaman individu sebagai komponen yang tidak dijelaskan oleh model, korelasi antar respon individu, terjadinya pengelompokan dalam populasi dan peubah teramati yang dihilangkan. Akibatnya signifikansi dari pengaruh peubah prediktor menjadi berbias ke atas (*overestimate*). Menurut Famoye et al (2004), regresi Poisson tidak sesuai untuk memodelkan data overdispersi.

Pendekatan klasik yang dapat digunakan untuk memodelkan kasus overdispersi pada model regresi Poisson adalah dengan memuat parameter tambahan yang memiliki distribusi Gamma di dalam nilai tengah sebaran Poisson

untuk mengakomodasi kelebihan ragam dari pengamatan (McCullagh & Nelder 1989). Peubah ini memiliki sebaran gamma dengan asumsi nilai tengah 1 dan ragam ϕ dalam nilai rata-rata sebaran Poisson. Dari pendekatan ini diperoleh distribusi campuran Poisson Gamma yang dikenal dengan distribusi Binomial Negatif.

Lord (2006) telah memodelkan kasus kecelakaan sepeda motor menggunakan model Poisson-Gamma dengan efek nilai rata-rata sampel rendah dan ukuran sampel kecil pada pendugaan parameter dispersi tetap. Hal itu menunjukkan bahwa penting pendekatan model Poisson-Gamma mengatasi kasus overdispersi pada model Poisson dengan nilai rata-rata sampel rendah dan ukuran sampel kecil.

Sellers dan Shmueli (2010) juga memberikan alternatif untuk kasus overdispersi pada model Poisson berupa perluasan model regresi Poisson yaitu model regresi Conway-Maxwell-Poisson (COM-Poisson). Model regresi COM-Poisson memiliki dua parameter yaitu parameter regresi (β) dan dispersi (ν). Kelebihan distribusi ini adalah memiliki fleksibilitas dalam memodelkan berbagai kasus data overdispersi maupun underdispersi dan memiliki sifat yang membuat metodologis menarik dan berguna dalam prakteknya.

Pada penelitian ini akan dilakukan tahapan estimasi parameter dan statistik uji devians pada model Regresi Binomial Negatif dan model regresi Conway-Maxwell-Poisson (COM-Poisson) serta membandingkan kedua model tersebut dalam mengatasi overdispersi pada data berdistribusi Poisson berdasarkan statistik uji devians.

II. Metode Penelitian

Penelitian ini mengkaji tahapan estimasi parameter untuk dua model regresi yang merupakan alternatif model untuk mengatasi overdispersi pada data cacah yang memiliki sebaran Poisson yaitu Model regresi binomial negatif dan model regresi COM-Poisson. Metode pendugaan parameter yang digunakan adalah metode pendugaan kemungkinan maksimum (*Maximum Likelihood*). Metode ini memaksimalkan fungsi kemungkinan dari fungsi kemungkinan binomial negatif dan COM-Poisson dengan cara mendiferensialkan masing-masing berturut-turut sebagai berikut :

$$L(\boldsymbol{\beta}, \phi | y, \mathbf{x}) = \prod_{i=1}^n \left\{ \frac{\Gamma(y_i + \phi^{-1})}{\Gamma(\phi^{-1})y_i!} \left(\frac{\phi\mu_i}{1 + \phi\mu_i} \right)^{y_i} \left(\frac{1}{1 + \phi\mu_i} \right)^{\phi^{-1}} \right\}$$

$$L(\boldsymbol{\beta}, \phi | y, \mathbf{x}) = \prod_{i=1}^n \left\{ \frac{[\exp(x_i^T \boldsymbol{\beta})]^y [\exp(x_i^T \boldsymbol{\beta})]^{\frac{v-1}{2v}} (2\pi)^{\frac{v-1}{2}} \sqrt{v}}{(y!)^v (\exp(v[\exp(x_i^T \boldsymbol{\beta})]^{\frac{1}{v}}))} \right\}$$

Data yang digunakan dalam penelitian ini terdiri dari dua sumber data yaitu data simulasi dan data kasus terapan. Data simulasi yang digunakan pada penelitian ini diperoleh dengan membangkitkan data berdistribusi Poisson yang mengandung overdispersi dengan menggunakan bahasa pemrograman R.

Data kasus terapan yang digunakan pada penelitian ini adalah data sekunder yang diperoleh Rumah Sakit Umum Kabupaten Rokan Hulu yaitu data banyaknya komplikasi penyakit *Diabetes Mellitus* sebagai variabel respon. Sedangkan variabel penjelas adalah X_1 sebagai usia, X_2 sebagai obesitas, X_3 sebagai jenis kelamin, X_4 sebagai riwayat *Diabetes Mellitus*, X_5 sebagai gula darah dan X_6 sebagai tensi. Langkah-langkah analisis yang dilakukan sebagai berikut :

1. Eksplorasi variabel respon Y
2. Menganalisis data terapan menggunakan metode Regresi Poisson yaitu untuk memeriksa terjadinya overdispersi pada regresi Poisson.

3. Jika terjadi overdispersi dilanjutkan menganalisis data terapan menggunakan metode regresi binomial negatif dan metode regresi COM-Poisson menggunakan metode kemungkinan maksimum melalui iterasi Newton Raphson.
4. Membandingkan metode regresi binomial negatif dan metode regresi COM-Poisson dengan statistik uji devians.

Uji signifikansi parameternya dengan menggunakan uji Wald. Hipotesis untuk parameter koefisien β_k (Fleiss *et al.* 2003) adalah:

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

Dengan statistik uji Wald :

$$G_{\beta} = \left\{ \frac{\hat{\beta}_k}{\widehat{se}(\hat{\beta}_k)} \right\}^2$$

statistik G_{β} akan mengikuti sebaran χ^2 dengan derajat bebas 1.

Kriteria keputusan yang diambil yaitu menolak H_0 , jika $G_{\beta} \geq \chi_{(\alpha,1)}^2$. Simpangan baku diperoleh menggunakan matriks simetris informasi Fisher $\mathbf{I}(\boldsymbol{\beta}^*)$ (McCulloch dan Searle 2001), dengan rumus sebagai berikut :

$$\mathbf{I}(\boldsymbol{\beta}) = - \begin{pmatrix} \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_0^2} & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} & \dots & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_k} \\ \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_1^2} & \dots & \dots & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_k} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_k^2} \end{pmatrix}$$

ragam dari $\hat{\beta} \approx [\mathbf{I}(\boldsymbol{\beta})]^{-1}$, sehingga simpangan baku $= \sqrt{[\mathbf{I}(\boldsymbol{\beta})]^{-1}}$.

Penilaian terhadap model regresi dapat dilihat dari devian (*deviance*) sebagai berikut (Gill 2001) :

$$D = -2 \ln \left[\frac{L(y | \boldsymbol{\beta}(\widehat{u}_i, v_i))}{L(y | \hat{\mu})} \right]$$

III. Hasil dan Pembahasan

Estimasi Parameter Model Regresi Binomial

Negatif

Pendugaan parameter pada model regresi binomial negatif menggunakan metode kemungkinan maksimum (*Maximum Likelihood Estimation*). Langkah awal dalam pendugaan parameter adalah membentuk fungsi kemungkinan (*likelihood function*) sebagai berikut:

$$L(\boldsymbol{\beta}, \phi | y, \mathbf{x}) = \prod_{i=1}^n \left\{ \frac{\Gamma(y_i + \phi^{-1})}{\Gamma(\phi^{-1})\Gamma(y_i)} \left(\frac{\phi\mu_i}{1 + \phi\mu_i} \right)^{y_i} \left(\frac{1}{1 + \phi\mu_i} \right)^{\phi^{-1}} \right\}$$

Estimasi parameter $\boldsymbol{\beta}^*$ diperoleh dengan mendiferensialkan logaritma natural fungsi kemungkinan dinyatakan secara matematis sebagai berikut:

$$\frac{\partial \ln L(\boldsymbol{\beta}^* | y_j, \mathbf{x}_j)}{\partial \boldsymbol{\beta}^*} = \frac{\partial \left(\sum_{j=1}^n \ln \left(\frac{\Gamma(y_j + \phi^{-1})}{\Gamma(\phi^{-1})\Gamma(y_j + 1)} \right) + y_j \ln(\phi\mu_j) - (y_j + \phi^{-1}) \ln(1 + \phi\mu_j) \right)}{\partial \boldsymbol{\beta}^*} = 0$$

Hasilnya berupa fungsi nonlinier yang berbentuk implisit maka penyelesaiannya dilakukan secara iterasi numerik Newton-Raphson. Secara dinyatakan sebagai berikut:

$$\boldsymbol{\beta}^*_{(m+1)} = \boldsymbol{\beta}^*_{(m)} - \mathbf{H}_{(m)}^{-1}(\boldsymbol{\beta}^*_{(m)}) \mathbf{g}_{(m)}(\boldsymbol{\beta}^*_{(m)})$$

dengan

$$\mathbf{g}^T(\boldsymbol{\beta}^*_{(m)}) = \left(\frac{\partial \ln L(\boldsymbol{\beta}^*)}{\partial \phi}, \frac{\partial \ln L(\boldsymbol{\beta}^*)}{\partial \beta_0}, \frac{\partial \ln L(\boldsymbol{\beta}^*)}{\partial \beta_1}, \dots, \frac{\partial \ln L(\boldsymbol{\beta}^*)}{\partial \beta_k} \right)$$

$$\mathbf{g}(\boldsymbol{\beta}^*_{(m)})_{(k+2) \times 1} = \begin{pmatrix} \sum_{j=1}^{38} \left\{ \frac{\ln(1 + \phi\mu_j) + \Psi(\phi^{-1}) - \Psi(y_j + \phi^{-1})}{\phi^2} + \frac{y_j - \mu_j}{\phi(1 + \phi\mu_j)} \right\} \\ \sum_{j=1}^{38} \left[x_{j,0} \left(\frac{y_j - \mu_j}{1 + \phi\mu_j} \right) \right] \\ \sum_{j=1}^{38} \left[x_{j,1} \left(\frac{y_j - \mu_j}{1 + \phi\mu_j} \right) \right] \\ \vdots \\ \sum_{j=1}^{38} \left[x_{j,k} \left(\frac{y_j - \mu_j}{1 + \phi\mu_j} \right) \right] \end{pmatrix}$$

$$\mathbf{H}(\boldsymbol{\beta}^*_{(m)})_{(k+1) \times (k+1)} = \begin{pmatrix} \frac{\partial^2 \ln L(\boldsymbol{\beta}^*)}{\partial \phi^2} & \frac{\partial^2 \ln L(\boldsymbol{\beta}^*)}{\partial \phi \partial \beta_0} & \dots & \frac{\partial^2 \ln L(\boldsymbol{\beta}^*)}{\partial \phi \partial \beta_k} \\ & \frac{\partial^2 \ln L(\boldsymbol{\beta}^*)}{\partial \beta_0^2} & \dots & \frac{\partial^2 \ln L(\boldsymbol{\beta}^*)}{\partial \beta_0 \partial \beta_k} \\ & & \ddots & \vdots \\ & & & \frac{\partial^2 \ln L(\boldsymbol{\beta}^*)}{\partial \beta_k^2} \end{pmatrix}$$

Statistik uji deviance model regresi binomial negatif dinyatakan secara matematis sebagai berikut:

$$D = -2 \ln \left[\frac{L(y | \boldsymbol{\beta}^*(\hat{u}_i, \hat{v}_i))}{L(y | \hat{\mu})} \right] = 2 \sum_{i=1}^n \left[y_i \ln \left(\frac{y_i}{\hat{\mu}_i(u_i, v_i)} \right) + (1 + y_i) \ln \left(\frac{1 + \hat{\mu}_i(u_i, v_i)}{1 + y_i} \right) \right]$$

Estimasi Parameter Model Regresi COM-Poisson

Pendugaan parameter model regresi COM-Poisson menggunakan pendekatan kemungkinan maksimum (*maximum likelihood*). Fungsi kemungkinan dari fungsi peluang COM-Poisson sebagai berikut:

$$L(\boldsymbol{\beta}, \phi | y, \mathbf{x}) = \prod_{i=1}^n \left\{ \frac{[\exp(x_i^T \boldsymbol{\beta})]^y [\exp(x_i^T \boldsymbol{\beta})]^{\frac{v-1}{2v}} (2\pi)^{\frac{v-1}{2}} \sqrt{v}}{(y!)^v (\exp(v[\exp(x_i^T \boldsymbol{\beta})]^{\frac{1}{v}}))} \right\}$$

Diferensial logaritma natural fungsi kemungkinan merupakan suatu fungsi nonlinier yang berbentuk implisit. Maka solusi yang digunakan untuk menyelesaikan persamaan tersebut adalah menggunakan metode numerik yaitu metode numerik Newton Raphson. Persamaan Newton Raphson secara umum dapat dituliskan sebagai berikut:

$$\begin{aligned} & \begin{pmatrix} \beta_{m+1} \\ v_{m+1} \end{pmatrix} \\ &= \begin{pmatrix} \beta_m \\ v_m \end{pmatrix} \\ &+ \begin{pmatrix} \frac{\partial^2 \ln L(\beta)}{\partial \beta^2} & \frac{\partial^2 \ln L(\beta)}{\partial \beta \partial v} \\ \frac{\partial^2 \ln L(\beta)}{\partial \beta \partial v} & \frac{\partial^2 \ln L(\beta)}{\partial v^2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial \ln L(\beta)}{\partial \beta} \\ \frac{\partial \ln L(\beta)}{\partial v} \end{pmatrix} \end{aligned}$$

Dengan diferensial pertama logaritma natural fungsi kemungkinan

$$\frac{\partial \ln L(\beta)}{\partial \beta_j} = \sum_{i=1}^n y_i x_{ij} - x_{ij} [\exp(x_i^T \beta)]^{\frac{1}{v}} + \frac{1}{2} \left(1 - \frac{1}{v}\right) x_{ij}$$

$$\begin{aligned} \frac{\partial \ln L(\beta)}{\partial v} &= \frac{n}{2v} + \frac{n}{2} \ln(2\pi) \\ &+ \sum_{i=1}^n \frac{(x_i^T \beta) [\exp(x_i^T \beta)]^{\frac{1}{v}}}{v} + \frac{x_i^T \beta}{2v} \\ &- [\exp(x_i^T \beta)]^{\frac{1}{v}} - \ln(y!) \end{aligned}$$

Diferensial kedua logaritma natural fungsi kemungkinan

$$\begin{aligned} \frac{\partial^2 \ln L(\beta)}{\partial \beta^2} &= \sum_{i=1}^n \frac{-x_{ij}^2 [\exp(x_i^T \beta)]^{\frac{1}{v}}}{v} \\ \frac{\partial^2 \ln L(\beta)}{\partial \beta_j \partial v} &= \sum_{i=1}^n \frac{-x_{ij}^2 [\exp(x_i^T \beta)]^{\frac{1}{v}}}{v^2} + \frac{x_{ij}}{2v^2} \\ \frac{\partial^2 \ln L(\beta)}{\partial v^2} &= \frac{-n}{2v^2} + \sum_{i=1}^n \left[- (x_i^T \beta)^2 \frac{[\exp(x_i^T \beta)]^{\frac{1}{v}}}{v^3} \left(\frac{x_i^T \beta}{v^3} \right) \right] \end{aligned}$$

Iterasi terus dilakukan hingga mencapai penduga parameter yang konvergen.

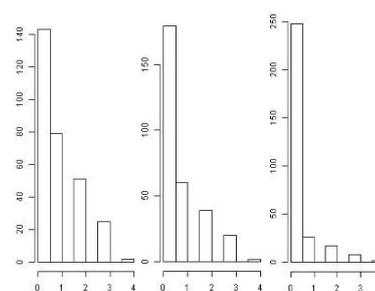
Statistik uji deviance model regresi COM-Poisson sebagai berikut :

$$\begin{aligned} d_i &= 2 \left[y_i \hat{v} \log \left(\left(y_i + \frac{\hat{v} - 1}{2\hat{v}} \right) / \left(\hat{\mu}_i + \frac{\hat{v} - 1}{2\hat{v}} \right) \right) \right. \\ &+ \log \left(Z \left(\left(\hat{\mu}_i + \frac{\hat{v} - 1}{2\hat{v}} \right)^{\hat{v}}, \hat{v} \right) \right. \\ &\left. \left. / Z \left(\left(y_i + \frac{\hat{v} - 1}{2\hat{v}} \right)^{\hat{v}}, \hat{v} \right) \right) \right] \end{aligned}$$

Kajian Simulasi

Karakteristik data simulasi dilakukan terhadap μ , n , dan p . Hasil yang dicobakan menunjukkan bahwa munculnya nilai p berpengaruh terhadap μ . Nilai μ yang dicobakan untuk 0.6, 1, dan 8.

Kondisi overdispersi pada variabel respon Y yang distribusi Poisson ditunjukkan dengan histogram pada Gambar 1 ketika variabel Y memiliki $\mu=1$ dengan $p=0.1, 0.3, \text{ dan } 0.7$. berdasarkan Gambar 1 terlihat bahwa kenaikan nilai p , maka akan terjadi perubahan nilai μ yang signifikan yaitu distribusi memiliki banyak nilai nol berlebih. Hal ini salah satu indikasi adanya overdispersi pada data cacah berdistribusi Poisson.



(a) p=0. 1 (b) p=0.3 (c) p=0.7

Gambar 1 Histogram peubah Y

Hasil uji khi-kuadrat dengan taraf signifikansi sebesar 5% dengan kombinasi nilai μ , n , p disajikan pada Tabel 2. Kondisi ini ditunjukkan dengan semakin besar nilai μ , maka persentase distribusi Poisson mencapai 0 % sedangkan persentase distribusi lainnya mencapai mendekati 80 %. Indikasi peluang nol berlebih yang menyebabkan terjadinya overdispersi pada

data cacah, sehingga data tidak lagi mengikuti suatu distribusi Poisson. Semakin besar μ dan p untuk setiap n yang dicobakan menunjukkan terjadinya overdispersi.

Tabel 1. Persentase khi-kuadrat terhadap kombinasi λ, n, p

n	λ	$p=0.1$		$p=0.3$		$p=0.7$	
		Poiss on	els e	Poiss on	els e	Poiss on	els e
10	0		80.		82.		85.
	8	89.2	2	62.6	0	18.0	4
30	0		82.		86.		82.
	8	88.2	2	51.0	0	7.8	8
50	0		81.		79.		76.
	8	0.0	6	0.0	6	0.0	8
100	0		85.		86.		83.
	8	82.4	4	25.2	6	0.2	2
150	0		85.		82.		85.
	8	75.0	4	8.2	6	0.0	8
200	0		84.		83.		81.
	8	0.0	0	0.0	4	0.0	6
300	0		81.		85.		87.
	8	76.2	8	7.6	8	0.0	0
400	0		87.		86.		83.
	8	67.4	0	0.2	4	0.0	2
500	0		84.		82.		84.
	8	0.0	6	0.0	6	0.0	6

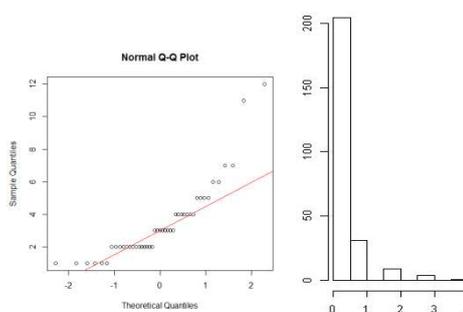
Kondisi overdispersi berpengaruh pada perubahan distribusi data. Dalam hal ini diestimasi bahwa distribusi yang cocok adalah distribusi binomial negatif dan distribusi COM-Poisson. Penelusuran dilakukan dengan uji Pearson khi-kuadrat terhadap derajat bebas. Hasilnya disajikan pada Tabel 2 bahwa rasio τ yang kurang dari satu menunjukkan bahwa overdispersi dapat diatas oleh kedua model tersebut. Akan tetapi jika dibandingkan kedua model regresi binomial negatif memiliki rasio τ lebih besar dibandingkan model regresi COM-Poisson.

Tabel 2. Rasio dispersi terhadap regresi Binomial Negatif dan COM-Poisson

n	λ	$p=0.1$		$p=0.3$		$p=0.7$	
		BN	CO M-P	BN	CO M-P	BN	CO M-P
10	0	0.5	0.50	0.6	0.62	0.8	0.69
	1	9		8		3	
	8	0.5	0.47	0.6	0.61	0.8	0.68
30	0	4		5		1	
	1	0.5	0.40	0.6	0.60	0.8	0.68
	8	0.3	0.40	0.6	0.60	0.8	0.68
50	0	8		5		2	
	1	0.5	0.43	0.6	0.64	0.8	0.68
	8	0.3	0.39	0.6	0.61	0.8	0.65
100	0	7		4		0	
	1	0.5	0.41	0.6	0.60	0.7	0.63
	8	0.3	0.35	0.6	0.59	0.8	0.60
150	0	7		3		6	
	1	0.5	0.43	0.6	0.62	0.8	0.63
	8	0.3	0.39	0.6	0.61	0.8	0.63
200	0	7		6		2	
	1	0.5	0.41	0.6	0.60	0.7	0.63
	8	0.3	0.35	0.6	0.59	0.8	0.60
300	0	2		3		1	
	1	0.5	0.41	0.6	0.60	0.7	0.63
	8	0.3	0.35	0.6	0.59	0.8	0.60
400	0	7		3		6	
	1	0.5	0.41	0.6	0.60	0.7	0.63
	8	0.3	0.35	0.6	0.59	0.8	0.60

Kajian Terapan

Data terapan yang digunakan adalah data banyaknya komplikasi penyakit *Diabetes Mellitu* dari suatu pasien di Rumah Sakit Umum di Kabupaten Rokan Hulu. Indikasi data banyaknya komplikasi penyakit ini mengikuti suatu distribusi Poisson ditunjukkan secara visual pada plot kuantil normal dan histogram pada Gambar 2. Plot kuantil menunjukkan distribusi data tidak mengikuti garis lurus dan histogram terlihat tidak simetris, sehingga plot data ini menunjukkan penyimpangan dari distribusi normal.



Gambar 2 (a) Plot kuantil-kuantil normal

(b) Histogram data

Hubungan antara banyaknya komplikasi penyakit *Diabetes Mellitus* dan faktor-faktor yang mempengaruhinya dapat diketahui menggunakan analisis regresi Poisson. Adapun faktor-faktor yang berkorelasi terdiri dari X_1 sebagai usia, X_2 sebagai obesitas, X_3 sebagai jenis kelamin, X_4 sebagai riwayat *Diabetes Mellitus*, X_5 sebagai gula darah dan X_6 sebagai tensi. Kriteria yang digunakan untuk mengetahui adanya kolinieritas antar variabel penjelas dengan menggunakan *Variance Inflation Factor* (VIF) pada Tabel 3.

Tabel 3. Nilai VIF Variabel Penjelas

Variabel	VIF
X_1	6.019
X_2	1.485
X_3	4.481
X_4	2.452
X_5	3.106
X_6	3.266

Myers (1990) bahwa antar variabel dikatakan saling bebas apabila nilai VIF kurang dari 10. Hasilnya menunjukkan bahwa tidak ada multikolinieritas antar variabel penjelas karena nilai statistik VIF kurang dari 10.

Tabel 4. Estimasi Model Regresi Poisson

Variabel	Estimasi	Std. Error	W_i	Keputusan
<i>Intercept</i>	-1.052	19,127	0.05	Terima H_0
Usia	-0.022	-0,003	8.46	Tolak H_0
Obesitas	1.247	0,898	1.38	Terima H_0
Jenis Kelamin	0.003	0,009	0.34	Terima H_0
Riwayat DM	0.003	0,016	0.18	Terima H_0
Gula Darah	-0.010	-0,003	3.16	Terima H_0
Tensi	-0.009	-0,003	3.06	Terima H_0

Hasil dari Pearson khi-kuadrat diperkuat dengan rasio τ sebesar 1.342. Hal ini menunjukkan bahwa model regresi Poisson yang mengalami overdispersi ini tidak cocok digunakan untuk memodelkan data banyaknya komplikasi penyakit *Diabetes Mellitus*.

Model regresi binomial negatif merupakan salah model yang dapat mengatasi overdispersi pada model regresi Poisson. Model ini merupakan campuran dari Poisson dan Gamma. Estimasi parameter model ditampilkan pada Tabel 5.

Tabel 5. Estimasi model regresi Binomial Negatif

Variabel	Estimasi	Std. Error	W_i	Keputusan
<i>Intercept</i>	-1.052	-19,127	0.055	Terima H_0
Usia	0.015	0.0018	8.468	Tolak H_0
Obesitas	0.189	0.0431	4.388	Tolak H_0
Jenis Kelamin	0.003	0,009	0.340	Terima H_0
Riwayat DM	0.041	0.0086	4.762	Tolak H_0
Gula Darah	0.056	0.0093	5.978	Tolak H_0
Tensi	-0.009	-0,003	3.066	Terima H_0

Pada tabel 4 terlihat bahwa variabel yang mempunyai pengaruh signifikan terhadap model adalah usia, obesitas, riwayat DM dan gula darah. Hasil dari Pearsonkhi-kuadrat diperkuat dengan rasio τ sebesar 0.898 yang menunjukkan bahwa rasio τ bernilai kurang dari 1. Hal ini berarti model regresi binomial negatif dapat mengatasi overdispersi pada data cacah yang mengikuti distribusi Poisson.

Model regresi Conway-Maxwell-Poisson (COM-Poisson) merupakan perluasan model regresi Poisson memiliki dua parameter yaitu yang memiliki parameter regresi (β) dan dispersi (ν). Kelebihan distribusi ini adalah memiliki fleksibilitas dalam memodelkan berbagai kasus data overdispersi maupun underdispersi dan memiliki sifat yang membuat metodologis menarik dan berguna dalam prakteknya.

Model ini merupakan campuran dari Poisson dan Gamma. Estimasi parameter model ditampilkan pada Tabel 6.

Tabel 6. Estimasi model regresi COM-Poisson

Variabel	Estimas i	Std. Erro r	W_i	Keputusa n
		-		
<i>Intercept</i>	-2.053	0.74	2.75	Terima H_0
		6	3	
		0.00	4.56	
Usia	0.027	6	8	Tolak H_0
		0.07	4.00	
Obesitas	0.283	1	8	Tolak H_0
Jenis		0,00	0.34	Terima
Kelamin	0.006	9	0	H_0
		0.04	4.76	
Riwayat DM	0.218	6	2	Tolak H_0
		0.07	5.97	
Gula Darah	0.474	9	8	Tolak H_0
		-		
		0,00	5.07	
Tensi	0.211	3	1	Tolak H_0
<i>v</i>	0.348			

Pada tabel 5 terlihat bahwa variabel yang mempunyai pengaruh signifikan terhadap model adalah usia, obesitas, riwayat DM, gula darah dan tensi. Hasil dari Pearsonkhi-kuadrat diperkuat dengan rasio τ sebesar 0.652 yang menunjukkan bahwa rasio τ bernilai kurang dari 1. Hal ini berarti model regresi binomial negatif COM-Poisson dapat mengatasi overdispersi pada data cacah yang mengikuti distribusi Poisson.

Perbandingan dari hasil analisis regresi Binomial negatif dan COM-Poisson pada kajian terapan menunjukkan bahwa kedua model ini dapat mengatasi overdispersi padadata cacah yang mengikuti distribusi Poisson. Hal ini ditunjukkan pada uji Pearson khi-kuadrat pada taraf nyata $\alpha = 0.05$ dengan rasio τ pada kedua model kurang dari 1. Berdasarkan uji Wald menunjukkan bahwa model binomial negatif dan COM-Poisson memberikan keputusan yang sama bahwa variabel usia, obesitas, riwayat DM, gula darah berpengaruh nyata terhadap variabel respon Y

yaitu banyak komplikasi penyakit *Diabetes Mellitus*.

Penentuan model terbaik dari kedua model yaitu model regresi binomial negatif dan COM-Poisson dilakukan berdasarkan statistik deviance terkecil pada Tabel 7.

Tabel 7. Statistik Deviance

Model	Deviance
Regersi Binomial Negatif	3.478
Regresi COM-Poisson	2.051

Kajian terapan berupa eksplorasi dan pengujian variabel respon Y, pengujian overdispersi serta evaluasi estimasi berdasarkan statistik deviance menunjukkan bahwa model regresi COM-Poisson lebih baik dibandingkan model regresi Binomial Negatif dalam mengatasi overdispersi pada data cacah yang mengikuti distri busi Poisson. Regresi COM-Poisson dapat menanggulangi faktor-faktor yang mempengaruhi banyaknya komplikasi penyakit *Diabetes Mellitus*.

IV. Kesimpulan

Estimasi model regresi binomial negatif dan model regresi COM-Poisson menggunakan metode kemungkinan maksimum (*Maximum Likelihood*) menghasilkan persamaan non linier yang diselesaikan dengan metode numerik Newton Raphson.

Kajian overdispersi terhadap data simulasi dari kombinasi λ , n , p yang dicobakan menunjukkan bahwa model regresi COM-Poisson memberikan hasil yang lebh baik dibandingkan dengan model regresi binomial negatif dalam mengatasi overdispersi pada data yang berdistribusi Poisson.

Kajian overdispersi terhadap terapan pada kasus faktor-faktor yang mempengaruhi banyak komplikasi penyakit *Diabetes Mellitus* dengan menggunakan regresi COM-Poisson lebih baik daripada regresi binomial negatif dalam mengatasi overdispersi pada variabel yang berdistribusi Poisson. Faktor yang mempengaruhi banyaknya komplikasi penyakit *Diabetes Mellitus* adalah usia, obesitas, riwayat DM, gula darah dan tensi.

Daftar Pustaka

- Agresti A. 2002. *Categorical Data Analysis Second Edition*. New York: John Wiley & Sons.
- Cameron A.C dan Trivedi P.K. 1998. *Regression analysis of count data*. Cambridge: Cambridge University Press.
- Draper NR and H Smith. 1981. *Applied Regression Analysis*. New York: John Wiley & Sons.
- Hardin JW, Hilbe JM. 2007. *Generalized Linier Models and Extensions*. Texas: Stata Press.
- Hilbe JM. 2008. *Negative Binomial Regression*. New York: Cambridge University Press.
- Hinde J, Dem'etrio CGB. 1998. Overdispersion: Models and Estimation. *Computational Statistics and Data Analisis* 27: 151-170.
- Jain M.K, Lyengar S R K, and Jain R K. 2004. *Numerical methods*. New Delhi: New Age.
- Lord D. 2006. Modeling motor vehicle crashes using poisson-gamma models: examining the effect of low sample mean value and small sample size on the estimation of the fixed dispersion parameter. *Accident Analysis & Prevention*, 38(4): 751-766
- Knowlton K, Solomon G. 2009. Mosquito-Borne Dengue Feer Threat Spreading in the Americas. New York: Natural Resources Defense Council Issue Paper.
- McCullagh P, Nelder JA. 1989. *Generalized Linear Models Second Edition*, London: Chapman and Hall.
- McCulloch CE, Searle SR. 2001. *Generalized Linear and Mixed Models*. Canada: John Wiley & Sons, Inc.
- Osgood D Wayne. 2000. Poisson-Based Regression Analysis of Aggregate Crime Rates. *Journal of Quantitative Criminology* 16: 21-43.
- Setyorini E. 2008. *Pemodelan Regresi Poisson Pada Maternal Mortality di Jawa Timur*. Surabaya: Tugas Akhir Jurusan Statistika FMIPA ITS.
- Sellers, K.F., Shmueli G. A 2010. Flexible Regressi on Model for Count Data. *Annals of Applied Statistics*, 2010, in press. (http://imstat.org/aoas/next_issue.html)
- Shmueli G, Tminka, J Borle and P Boatwright. 2005. A useful Distribution for Fitting Discrete Data: Revival of The Conway-Maxwell-Poisson Distribution. *Applied Statistics. Journal of Royal Statistical* 54(1):127-142

JURNAL GANTANG. Maret 2017; II(1): 79 – 87

p-ISSN. 2503-0671

e-ISSN. 2548-5547