

PENERAPAN REGRESI BINOMIAL NEGATIF UNTUK MENGATASI OVERDISPERSI PADA REGRESI POISSON

PUTU SUSAN PRADAWATI¹, KOMANG GDE SUKARSA²,
I GUSTI AYU MADE SRINADI³

^{1,2,3}Jurusan Matematika FMIPA Universitas Udayana, Bukit Jimbaran-Bali,
e-mail: ¹susanpradawati@gmail.com, ²sukarsakomang@yahoo.com,
³igustiayumadesrinadi@yahoo.co.id

Abstract

Poisson regression was used to analyze the count data which Poisson distributed. Poisson regression analysis requires state equidispersion, in which the mean value of the response variable is equal to the value of the variance. However, there are deviations in which the value of the response variable variance is greater than the mean. This is called overdispersion. If overdispersion happens and Poisson Regression analysis is being used, then underestimated standard errors will be obtained. Negative Binomial Regression can handle overdispersion because it contains a dispersion parameter. From the simulation data which experienced overdispersion in the Poisson Regression model it was found that the Negative Binomial Regression was better than the Poisson Regression model.

Keywords: *Poisson Regression, Overdispersion, Negative Binomial Regression, best model.*

1. Pendahuluan

Analisis Regresi Poisson adalah suatu model yang digunakan untuk menganalisis hubungan antara variabel respon yang berdistribusi Poisson dengan variabel-variabel bebasnya. Pada model Regresi Poisson disyaratkan keadaan yang equidispersi yaitu nilai mean dan varians dari variabel respon sama. Kadangkala terjadi kasus nilai varians lebih besar dari meannya yang disebut overdispersi. Sedangkan, jika nilai varians lebih kecil dari nilai meannya disebut underdispersi (McCullagh & Nelder [2]).

Penerapan Regresi Poisson pada data yang mengandung overdispersi akan menghasilkan galat baku yang lebih kecil dari nilai sesungguhnya (*underestimate*) (Pamungkas [4]). Salah satu metode yang dapat digunakan untuk mengatasi overdispersi pada Regresi Poisson adalah Regresi Binomial Negatif, karena distribusi Binomial Negatif tidak mengharuskan nilai varians sama dengan meannya. Pada penelitian ini penulis akan menentukan model Regresi Poisson dan model Regresi Binomial Negatif pada data yang mengalami overdispersi serta

¹ Mahasiswa Jurusan Matematika FMIPA Universitas Udayana

^{2,3} Staf Pengajar Jurusan Matematika FMIPA Universitas Udayana

membandingkan kedua model tersebut yang mengetahui model yang lebih baik berdasarkan kriteria nilai AIC, Devians, *Pearson Chi-square*, *Log likelihood*, dan *Likelihood Ratio*.

Regresi Poisson menyaratkan keadaan nilai mean dan varians dari variabel respon sama (Myers *et al.* [3]), pada kenyataannya sering terjadi nilai varians yang lebih besar dari meannya yang disebut overdispersi. Overdispersi dapat dideteksi dari taksiran dispersi yaitu nilai Devians/db dan *Pearson Chi-square*/db yang lebih besar dari 1. Pengujian taksiran dispersi (α) secara statistik dapat dilakukan dengan uji nilai Devians, dengan hipotesis : $H_0 : \alpha = 1$ $H_1 : \alpha > 1$. Statistik uji nilai Deviansnya: $G = 2 \sum_{i=1}^n [y_i \ln \left(\frac{y_i}{\hat{y}_i} \right) - (y_i - \hat{y}_i)]$. Penolakan terhadap H_0 jika $G > \chi_{(n-k-1), \alpha}^2$. Hal ini berarti terjadi overdispersi pada Regresi Poisson. Metode Regresi Binomial Negatif dapat mengatasi overdispersi pada Regresi Poisson karena memiliki parameter dispersi (κ). Pemilihan model terbaik dapat dilihat dari beberapa kriteria, yaitu nilai AIC dan BSC yang lebih kecil, nilai Devians dan *Pearson Chi-square* yang lebih mendekati satu, nilai *Log Likelihood* yang lebih besar, dan pengujian secara statistik dengan *Likelihood Ratio*. Perumusan hipotesis pada pengujian dengan *Likelihood Ratio* sebagai berikut: H_0 : model Regresi Poisson = model Regresi Binomial Negatif, H_1 : model Regresi Binomial Negatif lebih baik dari model Regresi Poisson, uji statistik *Likelihood Ratio* $T = 2(l_1 - l_0)$. Dengan l_1 dan l_0 adalah log likelihood masing-masing model pada hipotesis. Penolakan H_0 pada taraf signifikansi α jika $T > \chi_{(1-2\alpha, 1)}^2$ (Ismail & Jemain [1]). Perbandingan antara Regresi Poisson dan Regresi Binomial Negatif disajikan dalam Tabel 1 berikut.

Tabel 1. Perbandingan antara Regresi Poisson dan Regresi Binomial Negatif

Kriteria	Regresi Poisson	Regresi Binomial Negatif
Peubah respon	$Y_i \overset{iid}{\sim} Poi(\mu_i)$	$Y_i \overset{iid}{\sim} BN(\mu_i, \kappa)$
Fungsi distribusi peluang	$p(y; \mu) = \frac{e^{-\mu} \mu^y}{y!}$ (Myers <i>et al.</i> [3])	$f(y, \mu, \kappa) = \frac{\Gamma(y + \kappa^{-1})}{y! \Gamma(\kappa^{-1})} \left(\frac{\kappa \mu}{1 + \kappa \mu} \right)^y \cdot \left(\frac{1}{1 + \kappa \mu} \right)^{1/\kappa}$
Rata-rata dan Ragam	$E(Y_i) = Var(Y_i) = \mu_i$	$E(Y_i) = \mu_i$ $Var(Y_i) = \mu_i + \kappa \mu_i^2$
Penaksir parameter Fungsi Log Likelihood	<i>Maximum Likelihood Estimator</i> $\sum_{i=1}^n y_i \ln t_i \mu(x_i; \hat{\beta}) - \sum_{i=1}^n t_i \mu(x_i; \hat{\beta}) - \sum_{i=1}^n \ln(y_i)!$	<i>Maximum Likelihood Estimator</i> $\log \Gamma(\kappa^{-1} + y_i) - \log \Gamma(\kappa^{-1}) + y_i \log \left(\frac{\kappa \mu_i}{1 + \kappa \mu_i} \right) + \kappa^{-1} \log \left(\frac{1}{1 + \kappa \mu_i} \right)$
Model Regresi	$\mu_i = x_i' \beta + \varepsilon_i$	$\mu_i = x_i' \beta + \varepsilon_i$

2. Metode Penelitian

Data yang digunakan dalam penelitian ini adalah data simulasi yang diperoleh dengan membangkitkan data berdistribusi Binomial Negatif yang mengalami overdispersi pada Regresi Poisson dengan menggunakan *software SAS 9.1.3*. Tahapan penelitian diawali dengan melakukan pembangkitan data variabel bebas X_1 dan variabel bebas X_2 berdistribusi Gamma, variabel respon (Y) berdistribusi Binomial Negatif, dengan amatan masing-masing sebanyak 200, pada *software SAS 9.1.3*. Langkah selanjutnya menganalisis data dengan menggunakan metode Regresi Poisson. Kemudian memeriksa terjadinya overdispersi pada Regresi Poisson. Jika terjadi overdispersi pada Regresi Poisson, akan dilanjutkan menganalisis data dengan menggunakan metode Regresi Binomial Negatif. Langkah terakhir mengidentifikasi kebaikan model Regresi Poisson dan Regresi Binomial Negatif dengan melihat nilai AIC, Devians, *Pearson Chi-square*, *Log likelihood*, dan *Likelihood Ratio* dalam permasalahan overdispersi.

3. Hasil dan Pembahasan

Deskripsi data variabel bebas X_1 dan X_2 serta variabel respon (Y) yang digunakan pada penelitian ini dapat dilihat pada Tabel 2.

Tabel 2. Deskripsi data variabel respon serta variabel-variabel bebasnya

Peubah	N	Rata-rata (\bar{X})	Standar Deviasi	Minimum	Maksimum	Ragam/rata-rata
X_1	200	0.9108	0.8629	0.0000337	4.3748	0.8175
X_2	200	1.1154	0.9840	0.00408	5.3650	0.004343
Y	200	1.950	1.8590	0.00000	8.0000	1.7722

Hasil analisis Regresi Poisson untuk data tersebut diperoleh model sebagai berikut $\mu = \exp(-0.2988 + 0.4270 * X_1 + 0.3891 * X_2)$. Selanjutnya, pengecekan overdispersi pada Regresi Poisson dapat dilihat dari taksiran dispersi yaitu nilai Devians/DB dan *Pearson Chi-square* /DB pada Tabel 3 berikut.

Tabel 3. Taksiran Dispersi pada Regresi Poisson

Kriteria	DB	Nilai	Nilai/DB
Devians	197	265.7390	1.3489
<i>Pearson Chi-square</i>	197	227.7251	1.1560

Dari Tabel 3 terlihat bahwa nilai Devians/DB sebesar 1.3489 dan nilai *Pearson Chi-square*/DB sebesar 1.1560. Kedua nilai tersebut lebih besar dari 1 sehingga dapat disimpulkan pada data respon (Y) terjadi overdispersi. Pengujian taksiran dispersi (α) secara statistik dengan hipotesis: $H_0 : \alpha = 1$ $H_1 : \alpha > 1$ dengan nilai Devians sebesar 265.7390 dan $\chi_{0.05;197}^2 = 230.746$. Nilai Devians lebih besar dari $\chi_{197;0.05}^2$, sehingga dapat diputuskan menolak H_0 yang artinya

terjadi overdispersi pada Regresi Poisson. Hal ini berarti penggunaan model Regresi Poisson kurang tepat digunakan untuk menganalisis data tersebut, untuk itu dilakukan analisis dengan Regresi Binomial Negatif. Hasil analisis Regresi Binomial Negatif untuk data tersebut diperoleh model sebagai berikut:

$$\mu = \exp(-0.3796 + 0.4601 * X_1 + 0.4231 * X_2).$$

Penentuan model yang lebih baik antara model Regresi Poisson dan model Regresi Binomial Negatif pada data yang digunakan dapat dilihat dari hasil uji model terbaik Tabel 4 berikut

Tabel 4. Uji Model Terbaik

Kriteria	Regresi Poisson	Regresi Binomial Negatif
Devians/DB	1.3489	1.1503
<i>Pearson Chi-square</i> /DB	1.1560	0.9661
AIC	666.25	664.09
<i>Log-likelihood</i>	-65.8503	-63.7705

Tabel 4 memperlihatkan bahwa dari kriteria model terbaik berdasarkan nilai Devians/DB dan *Pearson Chi-square*/DB yang lebih mendekati 1, nilai AIC yang lebih kecil, dan nilai *Log Likelihood* yang lebih besar model Regresi Binomial Negatif lebih baik dibandingkan model Regresi Poisson.

Pengujian secara statistik dengan menggunakan nilai *Likelihood Ratio* dengan hipotesis sebagai berikut : $H_0 : \kappa = 0$ (model Regresi Poisson = model Regresi Binomial Negatif), $H_1 : \kappa > 0$ (model Regresi Binomial Negatif lebih baik dari model Regresi Poisson). Berdasarkan Tabel 3.3 diperoleh nilai *Likelihood Ratio* = $2 \times (-63.7705 - (-65.8503)) = 4.1596$.

Dengan menggunakan taraf nyata $\alpha = 5\%$, nilai tabel $\chi^2_{0.90;1} = 2.706$, maka nilai *Likelihood Ratio* > 2.706 , sehingga H_0 ditolak yang berarti model Regresi Binomial Negatif lebih baik dibandingkan model Regresi Poisson.

4. Kesimpulan

Berdasarkan data simulasi yang dibangkitkan dari *Software SAS 9.1.3* model Regresi Poisson untuk data tersebut adalah

$$\mu = \exp(-0.2988 + 0.4270 * X_1 + 0.3891 * X_2)$$

sedangkan model Regresi Binomial Negatif untuk data tersebut adalah

$$\mu = \exp(-0.3796 + 0.4601 * X_1 + 0.4231 * X_2).$$

Berdasarkan nilai Devians, *Pearson Chi-square*, AIC, *Log-likelihood*, dan *Likelihood Ratio* diperoleh bahwa model Regresi Binomial Negatif lebih baik dibandingkan model Regresi Poisson.

Berdasarkan kesimpulan di atas disarankan pembangkitan data yang mengalami overdispersi pada Regresi Poisson perlu dilakukan dengan

membangkitkan parameter overdispersinya (κ). Selain itu, perlu dilakukan penelitian lanjutan untuk mengetahui kriteria data yaitu tingkat overdispersi yang cocok digunakan metode Regresi Binomial Negatif untuk mengatasi overdispersi pada Regresi Poisson.

Daftar Pustaka

- [1] Ismail, N. & Jemain, A.A. 2007. "Handling Overdispersion with Binomial Negative and Generalized Poisson Regression Models". *Casualty Actuarial Society Forum Casualty Actuarial Society – Arlington, Virginia: Winter*, 103-158. <http://www.casact.org/pubs/forum/07wforum/07w109.pdf> . Diakses 1 April 2012
- [2] McCullagh, P. & Nelder, J.A. 1989. *Generalized Linier Models, second edition*. London: Chapman & Hall.
- [3] Myers, R.H., Montgomery, D.C., Vining, G.G., & Robinson, T.J. 2010. *Generalized Linier Models with Aplications in Engineering and the Sciences*. Canada : A John Wiley & Sons, Inc., Publication.
- [4] Pamungkas, Dimas Haryanto. 2003. Kajian Pengaruh Overdispersi dalam Regresi Poisson. [Skripsi]. Tersedia : <http://repository.ipb.ac.id.pdf>. Diakses 18 September 2012.