

## PENERAPAN REGRESI *GENERALIZED POISSON* UNTUK MENGATASI FENOMENA OVERDISPERSI PADA KASUS REGRESI POISSON

I PUTU YUDANTA EKA PUTRA<sup>1</sup>, I PUTU EKA NILA KENCANA<sup>2</sup>,  
I GUSTI AYU MADE SRINADI<sup>3</sup>

<sup>1,2,3</sup>Jurusan Matematika FMIPA Universitas Udayana, Bukit Jimbaran-Bali  
e-mail: <sup>1</sup>yudantaekaputra@yahoo.com, <sup>2</sup>i.putu.enk@gmail.com,  
<sup>3</sup>srinadiigustiyumade@yahoo.co.id

### *Abstract*

*The Poisson regression is generally used to analyze the response variable that is a discrete data. Poisson regression has assumption which must be met, that is condition equidispersion. But in fact this assumption is often violated, that is the value of the variance is greater or less than the mean value. The condition when value of the variance is greater than the mean value is called overdispersion. One method that can be used for overdispersion data is Generalized Poisson regression. In this research, it was found that the Generalized Poisson regression method was better than Poisson regression method.*

**Keywords:** *Poisson Regression, Overdispersion, Generalized Poisson regression.*

### 1. Pendahuluan

Analisis regresi Poisson merupakan metode regresi yang digunakan untuk menganalisis data yang variabel responnya berupa data diskret. Pada regresi Poisson terdapat asumsi yang harus dipenuhi, yaitu nilai varians dan rata-rata dari variabel respon tersebut sama atau equidispersi (Myers *et al.* [3]). Namun dalam kenyataan di lapangan sering terjadi pelanggaran asumsi tersebut, yaitu nilai variansnya lebih besar dari nilai rata-rata yang dinamakan overdispersi atau nilai variansnya lebih kecil dari nilai rata-rata yang dinamakan underdispersi (Wang & Famoye [5]). Jika terjadi fenomena overdispersi pada data, maka regresi Poisson kurang akurat digunakan untuk analisis, karena berdampak pada nilai *standard error* menjadi *under estimate* (lebih kecil dari nilai sesungguhnya), sehingga kesimpulan yang diperoleh menjadi tidak valid (McCullagh & Nelder [2]). Untuk mengatasi masalah overdispersi tersebut, salah satu metode yang dapat digunakan adalah analisis regresi *Generalized Poisson* yang merupakan perluasan dari regresi Poisson.

Tujuan penelitian ini adalah untuk mengetahui model regresi Poisson dan model regresi *Generalized Poisson* untuk data variabel respon yang mengalami

<sup>1</sup> Mahasiswa Jurusan Matematika FMIPA Universitas Udayana

<sup>2,3</sup> Staf Pengajar Jurusan Matematika FMIPA Universitas Udayana

overdispersi, dan mengetahui metode analisis regresi yang lebih baik digunakan untuk data variabel respon mengalami overdispersi apabila dilihat dari nilai *Pearson chi-square*, AIC (*Akaike's Information Criterion*), BIC (*Bayesian Information Criterion*), *Log likelihood* dan *Likelihood ratio*.

Kondisi overdispersi dapat dilihat dari nilai taksiran dispersi yaitu nilai *Pearson chi-square/df* dan *Deviance/df* nilai keduanya lebih besar dari 1, kemudian dilanjutkan dengan melakukan uji rasio dispersi ( $\alpha$ ) (Setyawan [4]) untuk mempertegas nilai taksiran dispersi. Metode yang dapat digunakan menganalisis data yang mengalami overdispersi salah satunya adalah regresi *Generalized Poisson*, regresi *Generalized Poisson* merupakan perluasan dari regresi Poisson yang dapat mengatasi keadaan overdispersi/underdispersi (Wang & Famoye [5]). Hubungan nilai rata-rata dan varians dalam regresi *Generalized Poisson* dapat dikondisikan sebagai berikut, 1) Jika nilai varians sama dengan nilai rata-rata  $E(Y_i|x_i) = Var(Y_i|x_i)$ , maka nilai parameter dispersi  $k = 0$ , sehingga fungsi densitas peluang *Generalized Poisson*, akan diturunkan ke regresi Poisson. 2) Jika nilai varians lebih besar dari nilai rata-rata  $E(Y_i|x_i) < Var(Y_i|x_i)$ , maka nilai parameter disperse  $k > 0$ , sehingga dapat dikatakan pada data terjadi overdispersi. 3) Jika nilai varians lebih kecil dari pada nilai rata-rata  $E(Y_i|x_i) > Var(Y_i|x_i)$ , maka nilai parameter dispersi  $k < 0$ , sehingga pada data terjadi underdispersi.

Pada penelitian ini dalam menentukan metode analisis terbaik antara regresi Poisson dengan regresi *Generalized Poisson* dapat dilihat dari nilai *Pearson chi-square/df*, AIC, BIC, *Log likelihood*, dan *Likelihood ratio*, dengan kriteria nilai *Pearson Chi-square* yang mendekati *df* yang lebih baik, nilai AIC dan BIC yang mempunyai nilai lebih kecil menunjukkan model yang lebih baik, dan nilai *Log likelihood* yang mempunyai nilai lebih besar menunjukkan metode yang lebih baik Perbedaan antara metode regresi Poisson dan regresi *Generalized Poisson* secara teori disajikan pada Tabel 1.

Tabel 1. Perbedaan Regresi Poisson dengan Regresi *Generalized Poisson*

Kriteria	Poisson	<i>Generalized Poisson</i>
Fungsi densitas	$f(y_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}$	$P(Y_i = y_i) = \left(\frac{\mu_i}{1 + k\mu_i}\right)^{y_i} \frac{(1 + ky_i)^{y_i-1}}{y_i!} \exp\left(-\frac{\mu_i(1 + ky_i)}{1 + k\mu_i}\right)$
Rata-rata dan ragam	$E(y) = Var(y) = \mu$	$E(y) = \mu$ $Var(y) = \mu_i(1 + k\mu_i)^2$
Penduga parameter	$\frac{\text{Maximum likelihood}}{\prod_{i=1}^n \mu_i^{y_i} \exp(-\sum_{i=1}^n \mu_i)} \prod_{i=1}^n y_i!$	$\frac{\text{Maximum Likelihood}}{\prod_{i=1}^n \left\{ \left( \frac{\exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ji})}{1 + k \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ji})} \right)^{y_i} \frac{(1 + ky_i)^{y_i-1}}{y_i!} \right\} \exp\left(-\frac{\exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ji})(1 + ky_i)}{1 + k \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ji})}\right)}$
Uji signifikansi	$W = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2$ $H_0: \beta_j=0$ $H_1: \beta_j \neq 0$	$W = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2$ $H_0: \beta_j=0$ $H_1: \beta_j \neq 0$
Parameter disperse	-	K
Model	$y_i = \exp(x_i^T \hat{\beta})$	$y_i = \exp(x_i^T \hat{\beta})$

## 2. Metode Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Dinas Pendidikan Pemuda dan Olahraga Provinsi Bali, berupa variabel respon pada penelitian ini adalah jumlah siswa SMA/SMK yang tidak lulus UN tahun 2012 tiap kabupaten di Provinsi Bali. Variabel bebas pada penelitian ini adalah jumlah sekolah SMA/SMK negeri tiap kabupaten di Provinsi Bali, jumlah ruang kelas SMA/SMK rusak tiap kabupaten di Provinsi Bali, jumlah guru S2 untuk SMA/SMK tiap kabupaten di Provinsi Bali, jumlah siswa SMA/SMK yang mengikuti UN tiap kabupaten di Provinsi Bali.

Adapun langkah-langkah analisis pada penelitian ini ialah mendeskripsikan data siswa SMA/SMK yang tidak lulus UN 2012 di Provinsi Bali, selanjutnya mencari model regresi Poisson dari data dan memeriksa terjadinya overdispersi, kemudian melakukan analisis data menggunakan analisis regresi *Generalized Poisson*, setelah itu mencari metode analisis yang lebih baik antara analisis regresi Poisson dengan regresi *Generalized Poisson* apabila dilihat dari kriteria *Pearson Chi-square*, AIC, BIC, *Likelihood ratio* (Ismail & Jemain [1]), dan sebagai langkah terakhir adalah menginterpretasi hasil perbandingan dari model regresi Poisson dan *Generalized Poisson*.

### 3. Hasil dan Pembahasan

Deskripsi data siswa SMA/SMK tidak lulus UN 2012 pada penelitian ini disajikan pada Tabel 2.

Tabel 2. Deskripsi Data Penelitian

	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	Y
Mean	10,89	125,90	34,11	4983,00	4,11
Simpangan Baku	4,57	291,90	13,75	2512,00	5,42
Ragam/Mean	1,91	676,77	5,54	1.266,33	7,14
Nilai Minimum	7	12	17	2171	0
Nilai Maksimum	22	903	58	10176	17

Model *full* regresi Poisson dari data yang dianalisis adalah sebagai berikut:

$$\mu = \exp(-1,2261 + 0,1619X_1 + 0X_2 - 0,0299X_3 + 0,0002X_4)$$

Selanjutnya dilakukan pemeriksaan overdispersi dengan menghitung nilai taksiran dispersi seperti disajikan pada Tabel 3 berikut:

Tabel 3. Taksiran Dispersi

kriteria	df	Nilai	nilai/df
<i>Deviance</i>	4	21,8662	5,4666
<i>Pearson Chi-square</i>	4	19,3321	4,8330

Namun untuk lebih jelasnya akan dilakukan uji lanjutan menggunakan uji statistika rasio dispersi ( $a$ ) dengan hipotesis uji sebagai berikut

$$H_0: a = 1 \text{ (pada data tidak terjadi overdispersi)}$$

$$H_1: a > 1 \text{ (pada data terjadi overdispersi)}$$

dengan statistik uji nilai *deviance* yaitu  $21,8662 > \chi_{0,05,4}^2 = 9,48773$ , yang berarti  $H_0$  ditolak. Hal ini menunjukkan bahwa terjadi overdispersi pada regresi Poisson. Oleh karena itu, selanjutnya akan dianalisis dengan menggunakan metode regresi *Generalized Poisson*.

Model *full* regresi *Generalized Poisson* dari data yang dianalisis adalah:

$$\mu = \exp(-1,4116 + 0,3425X_1 + 0,000592X_2 - 0,1548X_3 + 0,000573X_4)$$

Untuk menentukan metode yang terbaik berikut disajikan Tabel Kriteria Model

Tabel 4. Kriteria Model

Kriteria	Poisson	<i>Generalized Poisson</i>
<i>Pearson Chi-square</i>	19,3321	4,64
AIC	53,8532	52,30
BIC	54,8393	53,48
<i>Log Likelihood</i>	-21,9266	-20,15

Berdasarkan Tabel 4, dapat diketahui metode regresi *Generalized Poisson* lebih baik dari pada regresi Poisson. Namun untuk lebih mpertegas berikut disajikan Tabel nilai *Standard Error*

Tabel 5. Nilai *Standard Error*

Variabel bebas	Standar error regresi Poisson	Standar error regresi <i>Generalized Poisson</i>
Intercept	0,6900	1,338500
X1	0,0822	0,292200
X2	0,0009	0,001571
X3	0,0509	0,180500
X4	0,0002	0,000587

Berdasarkan nilai *standard error* yang disajikan pada tabel 5 terlihat nilai *standard error* pada regresi *Generalized Poisson* mengalami peningkatan atau *under estimate* yang terjadi pada regresi Poisson telah diatasi sehingga metode *Generalized Poisson* lebih baik.

#### 4. Kesimpulan

Berdasarkan data ketidaklulusan siswa SMA/SMK dalam mengikuti UN 2012 yang diperoleh dari Dinas Pendidikan Pemuda dan Olahraga Provinsi Bali yang telah dianalisis menggunakan *software SAS 9.2*, model regresi Poisson yang diperoleh untuk data tersebut adalah

$$\mu = \exp(-1,2261 + 0,1619X_1 + 0X_2 - 0,0299X_3 + 0,0002X_4)$$

Sedangkan model regresi *Generalized Poisson* yang diperoleh adalah

$$\mu = \exp(-1,4116 + 0,3425X_1 + 0,000592X_2 - 0,1548X_3 + 0,000573X_4)$$

Dilihat dari kriteria model terbaik yaitu nilai *Pearson Chi-square*, AIC, BIC, *Log Likelihood*, dan *Likelihood Ratio* maka dapat disimpulkan metode regresi *Generalized Poisson* lebih baik untuk menganalisis.

#### Daftar Pustaka

- [1] Ismail, Noriszura & Abdul Aziz Jemain. 2005. "Generalized Poisson Regression: An Alternative For Risk Classification". *Jurnal Teknologi*, 43, 1, 39-50. <http://www.jurnalteknologi.utm>. diakses tanggal 21 Oktober 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 Applications in Engineering and the Sciences*. Canada : A John Wiley & Sons, Inc., Publication.
- [4] Setyawan, Aji. 2012. "Perbandingan antara Regresi Poisson, Binomial Negatif dan Zero-Inflated Poisson pada Data Overdispersi. [Skripsi]. <http://www.repository.ipb.ac.id> pada tanggal 21 Maret 2013.
- [5] Wang, W. & Famoye, F. 1997. Modeling household fertility decision with generalized Poisson regression. *Journal of Population Economics*, 10, 3, 273-283. <http://www.jstor.org> diakses pada tanggal 27 Januari 2013.