

PENERAPAN REGRESI *ZERO-INFLATED NEGATIVE BINOMIAL* (ZINB) UNTUK PENDUGAAN KEMATIAN ANAK BALITA

**NI MADE SEKARMINI¹, I KOMANG GDE SUKARSA²,
I GUSTI AYU MADE SRINADI³**

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

Abstract

One method of regression analysis used to analyze the count data is Poisson regression. Poisson regression requires that the mean value equal to the value of variance (equidispersion). However, sometimes the data is going overdispersion the state variance values greater than the mean value. One of the causes overdispersion is the excessive number of zero values on the response variable (excess zeros). One method of analysis that can be used on data that had overdispersion due to excess zeros is regression Zero-Inflated Negative Binomial (ZINB). The data that can be analyzed using the ZINB regression is the early childhood mortality in the province of Bali because much of the data is zero. The analysis showed that the data had overdispersion on Poisson regression, so the ZINB regression analysis was used. From the results of the ZINB regression can overcome overdispersion so it was better than the Poisson Regression Model.

Keywords: *Overdispersion, ZINB Regression, Poisson Regression, Early Childhood Mortality*

1. Pendahuluan

Analisis regresi linear adalah salah satu metode statistika yang dapat menjelaskan hubungan sebab akibat antara satu variabel respon (Y) dengan satu atau lebih variabel prediktor (X_i). Pada umumnya analisis regresi linear digunakan untuk menganalisa variabel respon yang berupa data kontinu dan berdistribusi normal. Namun dalam beberapa aplikasinya, variabel respon yang akan dianalisis dapat berupa data diskrit [2].

Salah satu model yang dapat digunakan untuk menganalisis variabel respon diskrit Y yang berdistribusi Poisson dengan variabel bebas berupa data kategorik, diskrit, kontinu ataupun campuran disebut dengan model regresi Poisson. Pada regresi Poisson harus memenuhi asumsi variannya yaitu $Var(Y_i) = E(Y_i)$, untuk setiap observasi i ($i = 0, 1, \dots, N$), dilain pihak untuk data yang bertipe diskrit terkadang terjadi overdispersi yaitu nilai varian lebih besar dari nilai mean pada data respon. Salah satu penyebab terjadinya overdispersi yaitu banyaknya nilai nol

¹ Mahasiswa Jurusan Matematika FMIPA Universitas Udayana

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

yang berlebih pada variabel respon (*excess zeros*). Penanganan model yang dapat digunakan untuk mengatasi masalah tersebut antara lain adalah model regresi *Zero-Inflated Negative Binomial* (ZINB). Regresi *Zero Inflated Negative Binomial* (ZINB) adalah model yang dibentuk dari distribusi campuran poisson gamma. Pada analisis mortalitas (kematian) sering dijumpai banyak data yang bernilai nol. Salah satu data mortalitas (kematian) yang cocok dianalisis dengan menggunakan model Regresi *Zero Inflated Negative Binomial* (ZINB) adalah data kematian anak balita. Fakta menunjukkan bahwa jarang terjadi peristiwa kematian anak balita di suatu daerah. Dalam penelitian ini, permasalahan yang dibahas adalah penggunaan model Regresi *Zero Inflated Negative Binomial* (ZINB) untuk pendugaan kematian anak balita di Provinsi Bali dan faktor-faktor yang memengaruhi kematian anak balita.

Model regresi Poisson adalah model regresi nonlinear yang berasal dari distribusi Poisson yang merupakan penerapan dari *Generalized Linear Model* (GLM). Model ini menggambarkan hubungan antara variabel respon dan variabel prediktor, dengan variabel respon dalam bentuk diskrit [1].

Model regresi Poisson diberikan sebagai berikut.

$$Y_i \sim \text{Poisson}(\mu_i) \\ \mu_i = \exp(X_i^T \beta)$$

Regresi *Zero-Inflated Negative Binomial* (ZINB) merupakan model yang dibentuk dari distribusi poisson dan distribusi gamma. Model regresi ini memiliki dua keadaan yaitu Keadaan pertama disebut *zero state* terjadi dengan probabilitas p_i dan menghasilkan hanya observasi bernilai nol, sementara keadaan kedua disebut *Negative Binomial state* terjadi dengan probabilitas $(1 - p_i)$ dan berdistribusi Binomial Negatif dengan mean μ , dengan $0 \leq p_i \leq 1$ [1].

1. Model data diskrit untuk μ_i

$$\ln(\mu_i) = x_i^T \beta, \quad \mu_i \geq 0, i = 1, \dots, n$$

2. Model *Zero-Inflation* untuk p_i

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = z_i^T \gamma, \quad 0 \leq p_i \leq 1, \quad i = 1, \dots, n$$

Jika nilai awal tidak bagus, maka fungsi log-likelihood dari fungsi probabilitas ZINB tidak linier, sehingga fungsi likelihood ini tidak dapat diselesaikan dengan metode numerik biasa. Oleh karena itu, digunakan algoritma EM (*Expectation Maximization*) [3].

2. Metode Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Dinas Kesehatan Provinsi Bali yaitu data kematian anak balita tahun 2011 dari seluruh kabupaten/kota di Provinsi Bali. Variabel respon dalam

penelitian ini yaitu jumlah kematian anak balita (Y) dan variabel prediktor dalam penelitian ini adalah persentase Pneumonia(X_1), persentase anak balita mendapat pelayanan kesehatan (X_2), persentase anak balita mendapat vitamin A sebanyak 2x (X_3), balita gizi buruk (X_4), jumlah posyandu aktif (X_5). Teknik analisis datanya yaitu pertama melakukan pengumpulan data lalu melakukan analisis regresi poisson. Dari model yang telah diperoleh dilakukan pemeriksaan terjadinya overdispersi. Jika terjadi overdispersi dilanjutkan dengan melakukan analisis regresi ZINB. Untuk menentukan model terbaik ZINB dilihat dari nilai AIC terkecil kemudian melakukan uji kesesuaian model. Selanjutnya melakukan pengujian signifikansi parameter secara individu. Dari model yang telah diperoleh dilakukan interpretasi model. Kemudian kembali dilakukan pengujian overdispersi. Setelah itu dilakukan perbandingan antara regresi Poisson dan regresi ZINB.

3. Hasil dan Pembahasan

3.1 Deskripsi Data Penelitian

Di Provinsi Bali dengan jumlah kecamatan sebanyak 57 kecamatan terjadi jumlah total kematian anak balita sebanyak 41 orang dengan jumlah terbanyak di Kabupaten Gianyar. Jumlah data yang bernilai nol (tidak ada anak balita yang meninggal) sebanyak 36 data, bernilai 1 sebanyak 12 data, bernilai 2 sebanyak 3 data, bernilai 3 sebanyak 4 data, bernilai 4 sebanyak 1 data dan bernilai 7 sebanyak 1 data.

3.2 Analisis Regresi Poisson

Dari data kematian anak balita di Provinsi Bali diperoleh nilai-nilai Koefisien (β) dan nilai uji Wald seperti tampak pada Tabel 3.1.

Tabel 3. 1 Koefisien Regresi Model Regresi Poisson

Variabel Prediktor	Koefisien (β)	Uji Wald	<i>P-Value</i>
Intercept	1,3232	0,78	0,377
X_1	0,0027	0,06	0,8062
X_2	0,0146	2,21	0,1372
X_3	-0,0394	5,85	0,0155*
X_4	0,3174	0,76	0,3830
X_5	0,0085	4,47	0,0344*

Sumber :Data diolah tahun 2013; *. Signifikan pada $\alpha = 0.05$

Berdasarkan Tabel 3.1 diperoleh model Regresi Poisson yaitu:

$$\mu = \exp(1,9769 - 0,0294X_3 + 0,0074X_5)$$

Sebagai langkah selanjutnya dilakukan uji kesesuaian Regresi Poisson.

Uji kesesuaian regresi poisson dilakukan dengan menggunakan prosedur pengujian nilai Deviansi dengan hipotesis sebagai berikut:

H_0 : Model Regresi Poisson cocok pada data

H_1 : Model Regresi Poisson tidak cocok pada data

Setelah dilakukan pengolahan data, diperoleh nilai devians dan *pearson chi-square* seperti tampak pada Tabel 3.2.

Tabel 3.2 Uji Devians dan Taksiran Dispersi pada Regresi Poisson

Kriteria	DB	Nilai	Nilai/DB
Devians	54	94,0535	1,7417
Pearson Chi-square	54	135,3078	2,5057

Sumber : Data diolah tahun 2013

Tabel 3.2 memperlihatkan bahwa nilai devians lebih besar dari nilai tabel $\chi_{0,05;54}^2 = 72,1358$. Oleh karena itu, H_0 ditolak yang berarti bahwa model regresi Poisson tidak dapat menjelaskan hubungan antara variabel respon (Y) dengan variabel-variabel bebas (X_3, X_5). Lebih lanjut lagi, dilakukan pemeriksaan terjadinya overdispersi.

Dari Tabel 3.2 terlihat bahwa nilai Devians/ DB dan nilai Pearson Chi-square/ DB lebih besar dari 1 sehingga dapat disimpulkan bahwa terjadi overdispersi pada data respon (Y). Oleh karena itu, harus dilakukan analisis Regresi *ZINB*.

3.3 Analisis Regresi *Zero-Inflated Negative Binomial (ZINB)*

Input data persentase anak balita mendapat vitamin A sebanyak 2x menghasilkan model awal *ZINB* untuk variabel prediktor X_3 yaitu:

1. Model data diskrit untuk μ_i yaitu:

$$\ln \mu_i = 0,7425 - 0,0069X_3$$

2. Model *zero-inflated* untuk p_i yaitu:

$$\text{logit}(p_i) = -8,8337 + 0,0854X_3$$

dengan nilai AIC sebesar 139,5338.

Sedangkan input data jumlah posyandu aktif menghasilkan model awal *ZINB* untuk variabel prediktor X_5 yaitu:

1. Model data diskrit untuk μ_i yaitu:

$$\ln \mu_i = 0,5827 - 0,0054X_5$$

2. Model *zero-inflated* untuk p_i yaitu:

$$\text{logit}(p_i) = 2,3973 - 0,0523X_5$$

dengan nilai AIC sebesar 134,3934.

Lebih lanjut lagi, data-data persentase anak balita mendapat vitamin A sebanyak 2x dan jumlah posyandu aktif menghasilkan model awal *ZINB* untuk variabel prediktor X_3 dan X_5 yaitu:

1. Model data diskrit untuk μ_i yaitu:

$$\ln \mu_i = 0,2724 + 0,0061X_3 - 0,0089X_5$$

2. Model *zero-inflated* untuk p_i yaitu:

$$\text{logit}(p_i) = -49,1936 + 0,6367X_3 - 0,2412X_5$$

dengan nilai AIC sebesar 123,1146.

Sebagai langkah selanjutnya, dilakukan pemilihan model Regresi *ZINB* terbaik. Model dengan AIC terkecil adalah model regresi *ZINB* dengan dua variabel yaitu X_3 dan X_5 . Hal ini berarti bahwa peluang respon Y bernilai nol dipengaruhi oleh persentase anak balita mendapat vitamin A 2x (X_3) dan jumlah posyandu aktif (X_5). Lebih lanjut lagi, dilakukan pengujian kesesuaian model Regresi *ZINB*. Adapun Hipotesisnya adalah:

$$H_0: \beta_0 = \beta_3 = \beta_5 = \gamma_0 = \gamma_3 = \gamma_5 = 0$$

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

$$\text{dengan } j = 1, 2, \dots, 5$$

Dari data diperoleh nilai uji G:

$$G = -2(-65,804 - (-54,557)) = 22,494$$

Kriteria uji dengan menggunakan $\alpha = 0,05$, dari tabel *chi-square* diperoleh $\chi_{0,05;4}^2 = 9,488$. Tampak bahwa $G_{hitung} > G_{tabel}$ sehingga keputusannya H_0 ditolak, yang artinya model regresi *ZINB* dengan dua variabel layak digunakan.

Lebih lanjut lagi dilakukan pengujian signifikansi parameter Regresi *ZINB* secara individu. Dari hasil perhitungan diperoleh koefisien Regresi *ZINB* dan nilai uji Wald seperti tampak pada Tabel 3.3.

Tabel 3.3 Koefisien Regresi *ZINB* dengan dua variabel prediktor X_3 dan X_5

Parameter	Koefisien	Uji Wald	P-Value
β_0	0,2724	0,02	0,8846
β_3	0,0061	0,07	0,7919
β_5	-0,0089	1,34	0,2473
γ_0	-49,1936	1,84	0,1753
γ_3	0,6367	1,89	0,1695
γ_5	-0,2412	1,76	0,1852

Sumber: Data diolah tahun 2013

Berdasarkan kriteria uji parameter Regresi *ZINB* memiliki $P - value > \alpha = 0.05$, dan $W_j < \chi_{\alpha;1}^2 = 3.841$, yang artinya ketika pengujian secara individu menggunakan uji Wald tidak ada variabel prediktor yang berpengaruh secara signifikan terhadap variabel respon.

Berikutnya dilakukan kembali pengujian overdispersi. Berdasarkan hasil perhitungan pada regresi *ZINB* diperoleh nilai Devians/ DB sebesar 2,1395 dan

nilai Pearson Chi-square/ DB sebesar 0,9655. Hal ini mengindikasikan tidak terjadi overdispersi.

3.4 Perbandingan Model Regresi Poisson dengan Model Regresi ZINB

Tabel 3.4 Nilai devians, *Pearson Chi-Square* dan *AIC*.

Kriteria	Regresi Poisson	Regresi ZINB
Devians	94,0535	109,1146
<i>Pearson Chi-Square</i>	135,3078	49,2389
AIC	150,94	123,1146

Sumber: Data diolah tahun 2013

Dari nilai *Pearson Chi-Square* dan nilai AIC, regresi ZINB lebih tepat digunakan untuk memodelkan kematian anak balita dibandingkan dengan model regresi Poisson.

4. Kesimpulan

Untuk kematian anak balita di Provinsi Bali tahun 2011 berdasarkan nilai AIC terkecil diperoleh model regresi ZINB dengan dua variabel prediktor yaitu

Model data diskrit untuk μ_i yaitu:

$$\ln \mu_i = 0,2724 + 0,0061X_3 - 0,0089X_5$$

Model *zero-inflated* untuk p_i yaitu:

$$\text{logit}(p_i) = -49,1936 + 0,6367X_3 - 0,2412X_5$$

Dari semua variabel prediktor yang digunakan pada model regresi ZINB tidak ada yang berpengaruh secara signifikan terhadap penyebab kematian anak balita. Dalam penelitian ini regresi ZINB dapat mengatasi masalah overdispersi yang disebabkan oleh *excess zeros*.

Daftar Pustaka

- [1] Ariawan, B., Suparti & Sudarno. 2012. "Pemodelan Regresi Zero Inflated Negative Binomial (ZINB) Untuk Data Respon Diskrit dengan Excess Zeros". *Gaussian*. Vol. 1, No. 1 , 55-64.
- [2] Sundari, I. 2012. "Regresi Poisson dan Penerapannya untuk Memodelkan Hubungan Usia dan Perilaku Merokok Terhadap Jumlah Kematian Penderita Penyakit Kanker Paru-Paru". *Jurnal Matematika Unad*. Vol. 1, no. 1, Oktober, 75-80.
- [3] Garay, A.M. and Hashimoto, E.M. 2011. "On Estimation And Influence Diagnostics for Zero Inflated Negative Binomial Regression Models". *Computational Statistics and Data Analysis*. Vol. 55 , 1304-1318.