

**PENERAPAN REGRESI *ZERO INFLATED POISSON* UNTUK
MENGATASI OVERDISPERSI PADA REGRESI POISSON
(Studi Kasus: Ketidaklulusan Siswa SMA/MA
dalam Ujian Nasional di Buleleng)**

**LUH KOMANG MARDIANI¹, I KOMANG GDE SUKARSA²,
I GUSTI AYU MADE SRINADI³**

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

Abstract

The Poisson regression analysis is one of the regression methods used for count data and has the assumption of equidispersion. However, it is the overdispersion and then underestimate standard errors will be obtained. If the data are overdispersed and more data is zero then ZIP (Zero Inflated Regression) regression is used. ZIP regression model is more appropriate to be used to analyze the amount of Senior High School/Madrasah Aliyah who do not pass the exam with five independent variables, because a lot of data failure is zero. In this paper, data are overdispersed on Poisson regression, so ZIP regression are used. ZIP regression models obtained are only influenced by the proportion of Senior High School/Madrasah Aliyah classroom were damaged (X_3), is $\ln(\mu) = -1,7237 + 0,0911 X_3$ and $\text{logit}(\omega) = -0,4867 + 0,0465 X_3$.

Keywords: *Poisson Regression, Overdispersion, ZIP Regression, National Examination Test Failure.*

1. Pendahuluan

Regresi Poisson merupakan model standar untuk data diskret (Jansakul & Hinde, [2]). Regresi Poisson mewajibkan kondisi equidispersi yaitu nilai *mean* dan varians dari variabel respons bernilai sama (Wulandari, [5]). Namun, kadangkala terjadi overdispersi yaitu nilai varians lebih besar dari *mean*. Overdispersi pada regresi Poisson dapat menghasilkan galat baku yang *underestimate* lebih kecil dari nilai sesungguhnya (Pamungkas, [4]). Salah satu penyebab overdispersi adalah banyak observasi bernilai nol. Jika lebih banyak observasi bernilai nol untuk model regresi Poisson, yang dapat menyebabkan overdispersi, maka dapat digunakan model regresi *zero inflated Poisson* (ZIP) (Jansakul & Hinde, [2]). Untuk itu, model regresi ZIP lebih tepat digunakan untuk menganalisis jumlah siswa SMA/MA yang tidak lulus UN di Buleleng dengan lima faktor eksternal yang dapat memengaruhi ketidaklulusan, karena banyak data ketidaklulusan yang bernilai nol.

¹ Mahasiswa Jurusan Matematika FMIPA Universitas Udayana

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

Adapun tujuan dari penelitian ini adalah menentukan model regresi ZIP dari faktor-faktor yang memengaruhi ketidakkulusan UN SMA/MA di Buleleng tahun 2012 dan mengetahui faktor-faktor yang signifikan terhadap ketidakkulusan UN SMA/MA di Buleleng tahun 2012 dengan model regresi ZIP.

Overdispersi pada regresi Poisson dapat dilihat dari nilai taksiran dispersi yaitu nilai Devians/db dan Pearson *chi-square*/db yang lebih besar dari 1. Sedangkan untuk mendapatkan model terbaik dilihat dari nilai Devians, *Pearson Chi-square* dan AIC (Ismail & Jemain, [1]). Pada tabel berikut disajikan perbandingan antara regresi Poisson dengan regresi ZIP.

Tabel 1.1. Perbandingan antara Regresi Poisson dengan Regresi ZIP

Kriteria	Regresi Poisson	Regresi ZIP
Variabel respons	$Y_i \stackrel{iid}{\sim} Poi(\mu_i)$	$Y_i \stackrel{iid}{\sim} ZIP(\mu_i, \omega_i)$
Fungsi Distribusi Peluang	$p(y; \mu) = \frac{e^{-\mu} \mu^y}{y!}$ (Myers, [3])	$P(Y_i = y_i) = \begin{cases} \omega_i + (1 - \omega_i)e^{-\mu_i}, & \text{untuk } y_i = 0 \\ (1 - \omega_i) \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, & \text{untuk } y_i = 1, 2, \dots; 0 < \omega_i < 1 \end{cases}$ (Jansakul & Hinde, [2])
Mean dan Varians	$E[y] = var[y] = \mu.$	$E(Y) = (1 - \omega)\mu$ $var(Y) = \mu(1 - \omega)(1 + \mu\omega)$
Penaksir Parameter	Maksimum Likelihood Estimator	Maksimum Likelihood Estimator
Model Regresi	$\mu_i = \exp(x_i^T \beta)$	$\ln(\mu) = X\beta$ dan $logit(\omega) = X\gamma$

Berdasarkan Tabel 1.1 dapat dilihat perbandingan antara regresi Poisson dengan regresi ZIP mengenai variabel respons, fungsi distribusi peluang, mean, varians, penaksir parameter, dan model regresi.

2. Metode Penelitian

Sumber data pada penelitian ini adalah data sekunder yaitu diperoleh dari Dinas Pendidikan Pemuda dan Olahraga Kabupaten Buleleng. Data tersebut diolah dengan bantuan *software* SAS 9.2. Variabel pada penelitian ini adalah: Variabel respons (Y): jumlah siswa SMA/MA yang tidak lulus UN tahun 2012 di Buleleng. Variabel bebas (X): jumlah peserta UN (X_1), proporsi guru SMA/MA sudah sertifikasi (X_2), proporsi ruang kelas SMA/MA yang rusak (X_3), status sekolah (X_4), dan akreditasi sekolah (X_5).

Langkah-langkah analisisnya adalah: melakukan pengujian parameter model regresi Poisson, melakukan uji kesesuaian model regresi Poisson, melakukan uji overdispersi, menganalisis devians model regresi ZIP, melakukan pengujian parameter model regresi ZIP, dan melakukan uji kesesuaian model regresi ZIP.

3. Hasil dan Pembahasan

Berikut adalah statistik deskriptif data ketidakkulusan UN siswa SMA/MA:

Tabel 3.1. Statistik Deskriptif dari Variabel Respon (Y) dan Variabel Bebas (X₁, X₂, X₃)

Variabel	N	Mean	Std. dev	Minimum	Maksimum	Varians
Y	32	0,375	1,601	0	9	2,565
X ₁	32	124,4	78,2	14	241	6121,5
X ₂	33	31,49	28,33	0	78,46	802,42
X ₃	32	7,44	16,34	0	50	266,86

Sumber: Data diolah tahun 2013

Tabel 3.2. Statistik Deskriptif dari Variabel Bebas (X₄, X₅)

Variabel	N	Kategori	Modus
X ₄	32	0 = 15 1 = 17	1 = 17
X ₅	32	0 = 2 1 = 2 2 = 6 3 = 22	3 = 22

Sumber: Data diolah tahun 2013

Dari Tabel 3.1 dan Tabel 3.2 dapat dilihat bahwa data pada penelitian ini ada sebanyak 32 SMA/MA. Pada Tabel 3.1 nilai *mean* untuk variabel Y sebesar 0,375 artinya bahwa dari 1000 siswa peserta UN terdapat 375 siswa yang tidak lulus UN. Variabel X₁, X₂, dan X₃ merupakan variabel bebas yang berskala rasio, sedangkan untuk variabel X₄ dan X₅ adalah variabel bebas berskala nominal. Data yang diperoleh kemudian dianalisis menggunakan model regresi Poisson dan model regresi ZIP.

Dari hasil pengolahan data, model regresi Poisson yang diperoleh yaitu : $\mu = \exp(-2,0615 + 0,0576 X_3)$. Berdasarkan uji signifikansi parameter model regresi Poisson, dari 5 variabel bebas hanya satu variabel bebas yang signifikan terhadap variabel respon (Y), yaitu variabel X₃. Hipotesis untuk pengujian model regresi Poisson yaitu: H₀: model regresi Poisson cocok dengan data, H₁: model regresi Poisson tidak cocok dengan data. Model regresi Poisson memiliki nilai devians sebesar 44,5841 lebih besar dari nilai tabel $\chi^2_{0.05;30} = 43,773$. Sehingga H₀ ditolak yang berarti model regresi Poisson tidak cocok dengan data. Artinya model regresi Poisson tidak dapat menjelaskan hubungan antara variabel respons (Y) dengan variabel-variabel bebas (X), sehingga digunakan metode regresi lain untuk membentuk model hubungan variabel respons (Y) dengan variabel-variabel bebas (X). Untuk menginterpretasikan model regresi Poisson digunakan odd rasio, yaitu dilihat dari nilai $\exp(\beta)$. Interpretasi model regresi Poisson yaitu untuk setiap penambahan proporsi ruang kelas SMA/MA yang rusak sebesar 1% akan meningkatkan rata-rata jumlah siswa SMA/MA yang tidak lulus UN sebesar

$100(e^{0,0567(1)} - 1)\% = 100(1,0593-1)\% = 5,93\%$. Model regresi Poisson memiliki nilai devians dan Pearson *chi-square* seperti yang disajikan pada Tabel 3.3:

Tabel 3.3 Taksiran Dispersi pada Regresi Poisson

Kriteria	DB	Nilai	Nilai/DB
Devians	30	44,5841	1,4861
<i>Pearson Chi-square</i>	30	65,4634	2,1821

Sumber: Data diolah tahun 2013

Pada Tabel 3.3 dapat dilihat dari nilai devians/db sebesar 1,4861 dan Pearson *chi-square*/db sebesar 2,1821 masing-masing lebih besar dari 1, artinya bahwa terjadi overdispersi pada model regresi Poisson. Maka dilakukan analisis menggunakan model regresi ZIP.

Dari hasil analisis devians diperoleh model regresi ZIP dengan kelima variabel bebas (X_1, X_2, X_3, X_4, X_5), namun semua parameter dalam model regresi ZIP tidak signifikan sehingga model dengan lima variabel (X_1, X_2, X_3, X_4, X_5) tidak layak digunakan. Tidak signifikannya semua parameter disebabkan oleh adanya multikolinearitas, yaitu adanya korelasi/hubungan antarvariabel bebas (X). Untuk itu dilakukan seleksi model menggunakan metode *stepwise*. Metode *stepwise* dilakukan untuk mengetahui variabel mana yang dimasukkan ke dalam model sehingga diperoleh model terbaik untuk regresi ZIP. Dari metode *stepwise* diperoleh hanya variabel X_3 yang dimasukkan ke dalam model. Model regresi ZIP yang diperoleh yaitu:

$$\ln(\mu) = -1,7237 + 0,0911 X_3$$

dan

$$\text{logit}(\omega) = -0,4867 + 0,0465 X_3$$

Dari model regresi ZIP yang diperoleh menunjukkan bahwa ketidاكلulusan UN siswa SMA/MA hanya dipengaruhi oleh proporsi ruang kelas SMA/MA yang rusak. Selanjutnya dilakukan interpretasi model regresi ZIP. Sama halnya dengan model regresi Poisson, untuk menginterpretasikan model regresi ZIP digunakan odd rasio, yaitu dilihat dari nilai $\exp(\beta)$. Interpretasi model regresi ZIP untuk model $\ln(\mu)$ adalah untuk setiap peningkatan proporsi ruang kelas SMA/MA yang rusak sebesar 1% akan meningkatkan rata-rata jumlah siswa SMA/MA yang tidak lulus UN sebesar $100(e^{0,0911(1)} - 1)\% = 100(1,0954 - 1)\% = 9,54\%$. Sedangkan interpretasi model regresi ZIP untuk model $\text{logit}(\omega)$ adalah untuk setiap peningkatan proporsi ruang kelas SMA/MA yang rusak sebesar 1% akan meningkatkan rata-rata jumlah siswa SMA/MA yang tidak lulus UN sebesar $100(e^{0,0465(1)} - 1)\% = 100(1,0476 - 1)\% = 4,76\%$.

Pada tabel 3.4 dapat dilihat perbandingan untuk model regresi Poisson dengan model regresi ZIP :

Tabel 3.4 Perbandingan Hasil Model Regresi Poisson dengan Model Regresi ZIP

Kriteria	Regresi Poisson	Regresi ZIP
Devians	44,5841	28,9556
Pearson <i>chi-square</i>	65,4634	27,6622
Devians / DB	1,4861	1,0341
Pearson <i>chi-square</i> / DB	2,1821	0,9879
AIC	58,6377	36,9556

Sumber: Data diolah tahun 2013

Berdasarkan Tabel 3.4, untuk nilai devians, Pearson *chi-square* dan AIC pada model regresi ZIP memiliki nilai yang lebih kecil dibandingkan nilai pada model regresi Poisson. Nilai devians/db dan Pearson *chi-square*/db pada model regresi ZIP masing-masing lebih mendekati 1 dibandingkan nilai devians/db dan Pearson *chi-square*/db pada model regresi Poisson. Hal ini menunjukkan bahwa model regresi ZIP lebih baik dibandingkan model regresi Poisson.

4. Kesimpulan

Berdasarkan pemaparan sebelumnya, diperoleh beberapa kesimpulan berikut:

1. Ketidakkulusan siswa SMA/MA dalam UN tahun 2012 dengan variabel bebas jumlah peserta UN SMA/MA (X_1), proporsi guru SMA/MA sudah sertifikasi (X_2), proporsi ruang kelas SMA/MA yang rusak (X_3), status sekolah (X_4), dan akreditasi sekolah (X_5) dalam penelitian ini diperoleh model regresi Poisson yaitu :

$$\mu = \exp(-2,0615 + 0,0576 X_3)$$

dan model regresi ZIP yaitu :

$$\ln(\mu) = -1,7237 + 0,0911 X_3$$

dan

$$\text{logit}(\omega) = -0,4867 + 0,0465 X_3$$

2. Dari model regresi Poisson dan model regresi ZIP diperoleh faktor proporsi ruang kelas SMA/MA yang rusak signifikan terhadap jumlah siswa SMA/MA yang tidak lulus UN.
3. Berdasarkan kriteria nilai devians, Pearson *chi-square*, dan AIC dengan membandingkan model regresi Poisson dan model regresi ZIP dapat disimpulkan bahwa model regresi ZIP lebih tepat digunakan dibandingkan model regresi Poisson untuk data yang mengandung banyak observasi nol yang cenderung menimbulkan terjadinya overdispersi.

Daftar Pustaka

- [1] Ismail, N. & A.A. Jemain. 2007. Handling Overdispersion with Binomial Negative and Generalized Poisson Regression Models. *Casualty Actuarial Society Forum*, Winter, 103-158.
- [2] Jansakul, N & J. P. Hinde. 2002. Score Tests for Zero-Inflated Poisson Models. *Computational Statistics & Data Analysis*, Vol. 40, 75-96.
- [3] Myers, R.H.. 1990. *Classical and Modern Regression with Applications, second edition*. Boston: PWS-KENT Publishing Company.
- [4] Pamungkas, D.H.. 2003. Kajian Pengaruh Overdispersi dalam Regresi Poisson. [Online]. Tersedia : <http://repository.ipb.ac.id.pdf>. Diakses 6 Desember 2012.
- [5] Wulandari, S.P.. 2010. Pemodelan Resiko Penyakit Kaki Gajah (Filariasis) di Provinsi Papua dengan Regresi Zero-Inflated Poisson. *Forum Statistika dan Komputasi*, Vol. 15, No.1, 8-16.