

PENERAPAN REGRESI POISSON UNTUK MENGETAHUI FAKTOR-FAKTOR YANG MEMENGARUHI JUMLAH SISWA SMA/SMK YANG TIDAK LULUS UN DI BALI

KOMANG AYU YULIANINGSIH¹, KOMANG GDE SUKARSA², LUH PUTU
SUCIPTAWATI³

^{1,2,3}Jurusan Matematika, Fakultas MIPA Universitas Udayana,
e-mail:¹worldoflola@yahoo.com, ²sukarsakomang@yahoo.com
³putusuciptawati@yahoo.co.id

Abstract

This study aims to determine factors that influence the number of high/vocational school students that do not pass the examination in Bali in 2011. Factors that used in this study are the external factors include the proportion of public high/vocational school, proportion of the classrooms high/vocational school have damaged, the proportion of high/vocational school teachers have certified, and the number of participants high/vocational school national examination. This study uses Poisson regression analysis. Poisson regression is included in the Generalized Linear Models and this is one of the regression analysis for count data models with Poisson distributed response variables. Poisson regression analysis methods assume the mean and variance of the response variables are equal. Response variable is said to be Poisson distributed and overdispersion not happen if the value of the ratio between the variance and the mean close to one and less than 2.5. Test results show the response variable in this study Poisson distributed and there is no overdispersion. The analysis showed that all four factors are independent variables in this study affect the response variable is the number of high school/vocational school students which do not pass the examination.

Keywords: Poisson Regression, Generalized Linear Models, Overdispersion

1. Pendahuluan

Pendidikan merupakan salah satu hal yang sangat penting selain kesehatan. Dalam upaya peningkatan mutu pendidikan, baik pemerintah maupun masyarakat melakukan berbagai upaya mulai dari pelaksanaan pendidikan gratis hingga dilaksanakannya Ujian Nasional. Ujian Nasional merupakan salah satu alat yang digunakan untuk mengukur keberhasilan pada suatu jenjang pendidikan.

Faktor-faktor yang digunakan adalah faktor eksternal yaitu proporsi SMA/SMK negeri per kabupaten, proporsi ruang kelas SMA/SMK yang rusak per kabupaten, proporsi guru SMA/SMK tersertifikasi, dan jumlah peserta UN SMA/SMK. Ketidaklulusan siswa SMA/SMK merupakan kejadian yang jarang terjadi. Metode regresi Poisson merupakan salah satu metode regresi yang digunakan untuk menganalisa data cacahan dengan variabel respon berdistribusi Poisson atau menyatakan kejadian yang relatif jarang terjadi. Oleh karena itu untuk menganalisa faktor apa saja yang berpengaruh

¹ Mahasiswa Jurusan Matematika FMIPA Universitas Udayana

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

terhadap jumlah siswa SMA/SMK yang tidak lulus UN di Bali tahun 2011 digunakanlah regresi Poisson.

Analisis regresi adalah kumpulan teknik statistika untuk memodelkan dan menyelidiki hubungan antara sebuah variable respon misalkan y dan sebuah himpunan variabel bebas atau penjelas x_1, x_2, \dots, x_k [1]. Secara umum model regresi linear dapat dimodelkan sebagai berikut:

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i \quad (1)$$

Dengan $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ adalah parameter atau koefisien regresi, $x_{i1}, x_{i2}, \dots, x_{ik}$ adalah variable beba satau *predictor*, dan ε_i = galat yang bersifat acak dan saling bebas yang menyebar normal ($\varepsilon_i \sim N(0, \sigma^2)$) serta $i = 1, 2, \dots, n$.

Model Linear Terampat (Generalized Linear Models =GLMs)

GLMs dapat dinyatakan sebagai berikut [2]:

- a. y_1, y_2, \dots, y_n adalah n respon bebas yang mengikuti sebuah distribusi peluang anggota keluarga eksponensial, dengan $E\{y_i\} = \mu_i$.
- b. Sebuah prediktor linear berdasarkan pada variable bebas $x_{i1}, x_{i2}, \dots, x_{ik}$ yang digunakan, dinotasikan dengan $x_i' \beta$:

$$x_i' \beta = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} \quad (2)$$

- c. Fungsi pemetaan yang menghubungkan prediktor linear dan mean respon:

$$x_i' \beta = g(\mu_i) \quad (3)$$

Regresi Poisson

Regresi Poisson termasuk dalam GLMs dan merupakan salah satu bentuk analisis regresi yang digunakan untuk model data cacah (*count data*). Dalam memodelkan regresi Poisson digunakan pemetaan log sebagai berikut:

$$g(\mu_i) = \ln \mu_i = x_i' \beta \quad (4)$$

Dari persamaan (3) dan (4), maka hubungan antara mean respon dengan prediktor linear dalam regresi Poisson adalah

$$\mu_i = g^{-1}(x_i' \beta) = e^{\beta_0 + \sum_{j=1}^k \beta_j x_{ij}} \quad (5)$$

Pendugaan parameter dalam regresi Poisson menggunakan metode *Maximum Likelihood Estimation (MLE)*.

Koefisien determinasi (R^2) dapat menggambarkan keeratan hubungan antara peubah respon Y dengan peubah bebas X. Semakin besar nilai R^2 ($0 \leq R^2 \leq 1$) maka semakin akurat taksiran dari model regresi. Salah satu ukuran R^2 yang digunakan adalah R^2 yang berdasarkan pada nilai sisa devian [3].

$$R_{Dev}^2 = 1 - \frac{\ln L(y) - \ln L(\hat{\mu})}{\ln L(y) - \ln L(\bar{y})} \quad (6)$$

Dengan $\ln L(y)$ adalah log likelihood ketika semua parameter β_j ($j = 0, 1, \dots, p$) tidak disertakan dalam model dan y_i adalah nilai pengamatan peubah respon ke- i . $\ln L(\hat{\mu})$ adalah log likelihood ketika semua parameter β_j disertakan dalam model dan $\hat{\mu}_i$ adalah nilai cocokan (*predicted value*) untuk obsevel ke- i . $\ln L(\bar{y})$ adalah log likelihood ketika hanya β_0 yang disertakan dalam model dan \bar{y} adalah rata- rata respon.

Uji parameter secara parsial dilakukan untuk mengetahui apakah suatu variabel bebas berpengaruh terhadap variabel respon apabila di dalam model terdapat variabel lain. Hipotesis untuk uji parsial adalah:

$$H_0: \beta_j = 0; j = 1, 2, \dots, k \text{ (pengaruh variabel ke- } j \text{ tidak signifikan)}$$

$$H_1: \beta_j \neq 0; j = 1, 2, \dots, k \text{ (pengaruh variabel ke- } j \text{ signifikan)}$$

Statistik uji yang digunakan adalah

$$W = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2 \quad (7)$$

H_0 ditolak apabila nilai $W > \chi^2_{(\alpha,1)}$ dengan α adalah tingkat signifikansi dan derajat bebas satu.

Dalam metode regresi Poisson diasumsikan bahwa mean dan varian responnya sama. Namun adakalanya nilai varian lebih besar dari nilai mean (*overdispersi*) atau varian lebih kecil dari mean respon (*underdispersi*). Indraswari [4] menyatakan bahwa overdispersi terjadi apabila nilai rasio antara varian dan mean lebih besar dari 2,5.

Perbandingan Regresi Poisson dan Regresi Logistik Biner

Regresi Logistik biner adalah salah satu model regresi dengan nilai variabel responnya adalah 0 dan 1. Umumnya ketika variabel responnya biner, ini menunjukkan bahwa bentuk fungsinya non linear. Fungsi yang demikian disebut dengan fungsi respon logistik dengan bentuk sebagai berikut:

$$E(y) = \frac{1}{1 + \exp(-x'_i \beta)} \quad (8)$$

Bentuk linear dari persamaan (10) dengan $\eta = x'_i \beta$ adalah prediktor linear di mana η didefinisikan oleh transformasi $\eta = \ln \frac{\pi}{1-\pi}$ yang disebut dengan transformasi logit peluang π .

Model regresi Poisson dan regresi Logistik biner hampir sama. Baik regresi Poisson dan Logistik biner sama-sama menggunakan metode maksimum *likelihood* dalam menduga parameternya. Yang membedakan regresi Poisson dengan regresi Logistik biner adalah dalam hal variabel responnya. Pada regresi Logistik, responnya bernilai 0 dan 1. Sedangkan pada regresi Poisson, responnya menyatakan jumlah suatu kejadian.

2. Metode Penelitian

Variabel respon pada penelitian ini adalah jumlah siswa SMA/ SMK yang tidak lulus UN di Bali tahun 2011 untuk setiap kabupaten. Sedangkan variabel bebasnya adalah X1 (proporsi SMA/SMK negeri (%)), X2 (proporsi ruang kelas SMA/SMK yang rusak (%)), X3 (proporsi guru SMA/SMK tersertifikasi (%)), dan X4 (jumlah peserta UN SMA/SMK (orang)).

Analisis data pada penelitian ini menggunakan bantuan *software NCSS (Number Cruncher Statistical System)*. Adapun langkah- langkah teknik analisis datanya adalah: (1) Memeriksa distribusi Poisson pada variable respon (Y); (2) Memeriksa terjadinya over disperse dengan menghitung nilai rasio antara mean dan varian respon; (3) Menentukan model regresi Poisson; (4) Melakukan pengujian parameter; dan (5) Interpretasi model.

3. Hasil dan Pembahasan

Pada penelitian ini, dari hasil perhitungan diperoleh nilai mean= 2,6 dan nilai varian= 3,9, sehingga diperoleh rasio dari varian dan meannya adalah $3,9/2,6=1,5$. Karena rasio varian dan mean lebih kecil dari 2,5 maka ini menunjukkan bahwa variabel respon(Y) berdistribusi Poisson dan juga pada data tidak terjadi *overdispersi*, sehingga analisis Regresi Poisson dapat digunakan pada penelitian ini.

Analisis devian merupakan salah satu analisis yang digunakan dalam analisis regresi dalam pembentukan model. Dalam suatu analisis regresi, nilai devian digunakan

untuk menentukan apakah variabel tersebut disertakan dalam model atau tidak. Ringkasan hasil analisis devian disajikan dalam Tabel 2.

Tabel 2. Analisis Devian

Model	DB	Devian	Selisih	P- Value
Tanpa Variabel	1	186854,4942		
Satu Variabel				
X3	1	185698,6122	1155,8820	0,0000*
Dua Variabel				
X3,X4	2	184513,5760	1185,0362	0,0000*
Tiga Variabel				
X3,X4,X1	3	168317,8974	16195,6786	0,0000*
Empat Variabel				
X3,X4,X1,X2	4	153174,4231	15143,4743	0,0000*

Dari hasil analisis devian, model tanpa variabel memiliki nilai devian sebesar 186854,4942. Untuk model satu variabel, proporsi guru SMA/SMK sertifikasi adalah variabel yang pertama kali masuk ke dalam model karena memiliki selisih devian terbesar yaitu 1155,8820. Untuk model dua variabel, setelah proporsi guru SMA/SMK sertifikasi masuk ke dalam model variabel kedua yang masuk ke dalam model adalah jumlah peserta UN SMA/SMK karena memiliki selisih nilai devian terbesar yaitu 1185,0362. Demikian seterusnya hingga diperoleh model empat variabel dengan selisih nilai devian sebesar 15143,4743.

Dari hasil analisis devian, model regresi Poisson yang diperoleh adalah sebagai berikut:

$$\hat{\mu} = \exp(12,57370 - 0,16612 * X_1 + 0,11908 * X_2 - 1,03146 * X_3 - 0,00049 * X_4)$$

dengan nilai R^2 sebesar 0,4451.

Hasil uji Wald menunjukkan Bahwa keempat variabel berpengaruh signifikan karena *p-value* yang dihasilkan adalah 0,0000 dan ini lebih kecil dari $\alpha = 0,05$ sehingga H_0 ditolak. Dengan demikian dapat disimpulkan bahwa variabel yaitu proporsi SMA/SMK negeri, proporsi ruang kelas SMA/SMK rusak, proporsi guru SMA/SMK sertifikasi, dan jumlah peserta UN SMA/SMK berpengaruh terhadap jumlah siswa SMA/SMK yang tidak lulus UN.

Untuk menginterpretasikan model regresi Poisson yang diperoleh digunakanlah nilai *odds ratio* dari masing-masing koefisien. Untuk setiap penambahan proporsi SMA/SMK negeri sebesar 1% akan menurunkan rata-rata jumlah siswa SMA/SMK yang tidak lulus UN sebesar $100(1 - e^{-0,16612(1)})\% = 15,30\%$. Variabel proporsi ruang kelas SMA/SMK yang rusak memiliki koefisien regresinya bernilai positif maka untuk setiap penambahan proporsi ruang kelas rusak sebesar 1% akan mengakibatkan meningkatnya rata-rata jumlah siswa SMA/SMK yang tidak lulus UN sebesar $100(e^{0,11908(1)} - 1)\% = 12,64\%$. Selanjutnya untuk variabel proporsi guru SMA/SMK sertifikasi, penambahan proporsi sebesar 1% akan menurunkan rata-rata jumlah siswa SMA/SMK yang tidak lulus UN sebesar $100(1 - e^{-1,03146(1)})\% = 64,40\%$. Untuk penambahan jumlah peserta UN SMA/SMK sebanyak satu orang akan menurunkan rata-rata jumlah siswa SMA/SMK yang tidak lulus UN sebesar $100(1 - e^{-0,00049(1)}) = 0,049\%$.

4. Kesimpulan

Dari hasil yang telah dipaparkan sebelumnya, dapat ditarik kesimpulan bahwa:

1. Dari keempat faktor yang merupakan variable bebas pada penelitian ini yaitu proporsi SMA/SMK negeri, proporsi ruang kelas SMA/SMK rusak, proporsi guru SMA/SMK sertifikasi, dan jumlah peserta UN SMA/SMK semuanya secara signifikan berpengaruh terhadap jumlah siswa SMA/SMK yang tidak lulus UN di Bali tahun 2011. Dari keempat variabel tersebut yang paling berpengaruh adalah proporsi guru SMA/SMK sertifikasi dengan penurunan jumlah siswa yang tidak lulus sebesar 64,40% sejalan dengan peningkatan proporsi guru sertifikasi sebesar 1%.

2. Model regresi Poisson yang diperoleh adalah sebagai berikut:

$$\hat{\mu} = \exp(12,5737 - 0,16612X_1 + 0,11908X_2 - 1,03146X_3 - 0,00049X_4)$$

dengan

X_1 = proporsi SMA/SMK negeri;

X_2 = proporsi ruang kelas SMA/SMK yang rusak;

X_3 = proporsi guru SMA/SMK tersertifikasi;

X_4 = jumlah peserta UN SMA/SM

Daftar Pustaka

- [1] Myers, Raymond H., Douglas C. Montgomery, G. Geoffrey Vining, and Timothy J. Robinson. 2010. *Generalized Linear Models with Applications in Engineering and The Sciences*. Ed.2. John Wiley and Sons. New Jersey.
- [2] Kutner, Michael H., Christopher J. Nachtsheim, John Neter, and William Li. 2005. *Applied Linear Statistical Models*. Ed.2. The McGraw-Hill Companies. Newyork.
- [3] Baharuddin. 2005. *Ukuran R^2 dalam Model Regresi Poisson*. Universitas Haluoleo. http://staff.undip.ac.id/matematika/dwi_ispriyanti/files/2010/07/regresi-Poisson2.pdf, diakses tanggal 25 September 2011.
- [4] Indraswari, Made Susy. 2011. "Faktor- faktor Penyebab Kematian Bayi di Kabupaten Karangasem dengan Pendekatan Regresi Poisson". *Skripsi*. Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Udayana. Bukit Jimbaran.