

PENERAPAN REGRESI POISSON DAN BINOMIAL NEGATIF DALAM MEMODELKAN JUMLAH KASUS PENDERITA AIDS DI INDONESIA BERDASARKAN FAKTOR SOSIODEMOGRAFI

AULIA SAFITRI, IZZATI RAHMI HG, DODI DEVIANTO

*Program Studi Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia,
aulia.safitri14@gmail.com*

Abstrak. Penelitian ini dilakukan untuk menentukan model terbaik untuk memodelkan jumlah kasus AIDS di Indonesia berdasarkan faktor sosiodemografi dengan menggunakan model regresi Poisson dan model regresi Binomial Negatif. Ketika model telah didapatkan, dilakukan perbandingan antara model regresi Poisson dan model regresi Binomial Negatif untuk mencari model terbaik yang dapat digunakan. Berdasarkan nilai Log-likelihood, uji Likelihood ratio, AIC dan BIC, model regresi Binomial Negatif lebih baik digunakan dibandingkan dengan regresi Poisson untuk kasus AIDS menurut provinsi di Indonesia tahun 2011 dengan satu variabel yang signifikan, yaitu kepadatan penduduk. Model untuk regresi Binomial Negatif yang dihasilkan adalah:

$$y = \exp(6.444 + 0.00018x_1).$$

Kata Kunci: AIDS, Sosiodemografi, Regresi Poisson, Regresi Binomial Negatif

1. Pendahuluan

AIDS telah menjadi masalah internasional, karena dalam waktu yang relatif singkat terjadi peningkatan jumlah penderita dan semakin melanda banyak negara. Perkembangan AIDS di Indonesia dari tahun ke tahun terus meningkat.

Meningkatnya penyebaran AIDS ini sangat dipengaruhi oleh faktor sosiodemografi seperti kepadatan penduduk, kemiskinan dan tingkat pengangguran. Dalam model hubungan sebab akibat, maka jumlah kasus AIDS dapat didefinisikan sebagai peubah tak bebas dan faktor-faktor yang mempengaruhi jumlah kasus AIDS sebagai peubah bebas. Karena jumlah kasus AIDS adalah peubah diskrit, maka salah satu cara untuk cara memodelkannya adalah dengan menggunakan model regresi Poisson.

Model regresi Poisson adalah salah satu model regresi yang sering digunakan untuk menganalisis data diskrit. Model regresi Poisson mempunyai asumsi equidispersi, yaitu kondisi dimana nilai rata-rata dan ragam dan variabel respon bernilai sama.

Dalam berbagai kasus, seringkali dijumpai pelanggaran asumsi di atas, dimana ragam peubah respons lebih besar daripada nilai rata-ratanya yang dikenal dengan istilah overdispersi. Untuk menangani masalah overdispersi, dapat dilakukan pemodel-

an dengan regresi Binomial Negatif. Pada penelitian ini akan dicari pemodelan jumlah kasus AIDS di Indonesia berdasarkan faktor-faktor sosiodemografi. Model yang digunakan adalah model regresi Poisson dan model regresi Binomial Negatif.

2. Tinjauan Pustaka

2.1. Distribusi Poisson

Distribusi Poisson adalah distribusi nilai bagi suatu variabel acak X (X diskrit), yaitu banyaknya hasil percobaan yang terjadi dalam suatu interval waktu tertentu atau di suatu daerah tertentu. Sebaran peluang bagi peubah acak Poisson X , yang menyatakan banyaknya hasil percobaan yang terjadi selama suatu selang waktu atau daerah tertentu adalah [1]:

$$P(X = x) = p(x; \mu) = \frac{(e^{-\mu} \mu^x)}{x!}; \quad x = 0, 1, 2, \dots, \quad (2.1)$$

dengan μ menyatakan rata-rata banyaknya sukses yang terjadi dalam selang waktu atau daerah tertentu dan $e = 2.71828 \dots$. Distribusi Poisson mempunyai varians yang sama dengan μ .

2.2. Uji Kolmogorov-Smirnov

Untuk mengetahui distribusi data apakah berdistribusi Poisson atau tidak, maka dilakukan uji Kolmogorov-Smirnov. Berikut ini adalah hipotesis uji Kolmogorov-Smirnov:

H_0 : Data mengikuti distribusi tertentu,

H_1 : Data tidak mengikuti distribusi tertentu.

Statistik uji yang digunakan adalah:

$$D = \max |F_0(x) - S_N(x)|. \quad (2.2)$$

Kriteria untuk pengujian ini adalah tolak H_0 jika nilai D_{hitung} lebih besar dari nilai D_{tabel} .

2.3. Distribusi Binomial Negatif

Bila ulangan yang bebas dan berulang-ulang dapat menghasilkan keberhasilan dengan peluang p dan kegagalan dengan peluang $q=1-p$, maka sebaran peluang bagi peubah acak X , yaitu banyaknya ulangan sampai terjadinya k keberhasilan, diberikan menurut [1]:

$$b^*(x; k, p) = \binom{x-1}{k-1} p^k q^{(x-k)}, \quad \text{untuk } x = k, k+1, k+2, \dots$$

2.4. Generalized Linear Model (GLM)

Untuk mengatasi kondisi dengan variabel respon yang tidak berdistribusi normal, tetapi masih saling bebas, maka para statistikawan yang dipelopori oleh Nelder telah mengembangkan model linier yang dikenal dengan istilah *Generalized Linear Model* (GLM). Ada tiga komponen utama dalam analisis GLM, yaitu:

- (1) Komponen Random.
- (2) Komponen Sistematis.
- (3) Fungsi link.

2.5. Regresi Poisson

Dalam model regresi Poisson terdapat beberapa asumsi yang harus dipenuhi, salah satunya adalah nilai varians dari variabel respon Y yang diberikan oleh $X = x$ harus sama dengan nilai meannya, yaitu:

$$\text{Var}(Y|x) = E(Y|x) = \mu.$$

Regresi Poisson merupakan pilihan yang tepat, ketika variabel respon Y merupakan bilangan bulat positif. Model regresi Poisson ditulis sebagai berikut:

$$y_i = \mu_i + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (2.3)$$

dimana y_i adalah jumlah kejadian dan ε_i adalah rata-rata jumlah kejadian dalam periode t_i . Diasumsikan ε_i tidak berubah dari titik data ke titik data. Dalam *generalized linier model* (GLM), terdapat sebuah fungsi g yang menghubungkan rata-rata dari variabel responnya dengan sebuah prediktor linier, yaitu:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k.$$

Fungsi g disebut fungsi penghubung (*link function*).

2.6. Overdispersi

Overdispersi adalah kondisi pada saat ragam dari peubah respon lebih besar dari nilai tengah peubah respon. Overdispersi dapat terjadi karena banyaknya jumlah pengamatan yang bernilai nol pada peubah respon.

Ada atau tidaknya overdispersi dapat dilihat dari nilai *Deviance* atau *Pearson Chi-square* yang dibagi dengan derajat bebasnya. Apabila nilai *Pearson Chi-square* dibagi dengan derajat bebas lebih besar daripada 1, ini menunjukkan nilai varians yang lebih besar daripada nilai meannya. Hal ini menunjukkan overdispersi pada data [4].

2.7. Regresi Binomial Negatif

Regresi binomial negatif merupakan salah satu solusi untuk mengatasi masalah varians peubah respon yang melebihi nilai harapan atau dikenal dengan overdispersi. Model regresi binomial negatif mengasumsikan bahwa variabel respon ke- i mengikuti distribusi binomial negatif dengan parameter μ_i dan κ yang dinotasikan oleh $Y_i \sim BN(\mu_i, \kappa)$ [4].

2.8. Pemilihan Model Terbaik

Setelah model regresi Poisson dan model regresi Binomial Negatif diperoleh, selanjutnya adalah membandingkan model tersebut untuk mencari model terbaik untuk memodelkan jumlah kasus AIDS di Indonesia. Pengukuran yang sering digunakan adalah nilai Log-likelihood, Likelihood ratio tests, Akaike Information Criteria (AIC), Bayesian Schwartz Information Criteria (BIC) [4].

Nilai Log-likelihood yang lebih besar menunjukkan model tersebut lebih baik digunakan daripada model yang lain. Uji Likelihood ratio (T) biasa digunakan untuk uji perbandingan model. Berikut ini adalah hipotesis uji Likelihood ratio.

Hipotesis:

$H_0 : \kappa = 0$ (Model regresi Poisson lebih baik daripada model regresi Binomial Negatif)

$H_1 : \kappa \neq 0$ (Model regresi Binomial Negatif lebih baik daripada model regresi Poisson).

Statistik uji yang digunakan:

$$T = 2(\widehat{L}_{BN} - \widehat{L}_{Poisson}), \quad (2.4)$$

dimana:

\widehat{L}_{BN} : nilai log-likelihood duga model regresi Binomial Negatif,

$\widehat{L}_{Poisson}$: nilai log-likelihood duga model regresi Poisson.

Kriteria untuk pengujian ini adalah tolak H_0 jika T lebih besar $\chi_{\alpha, db}^2$, dimana α adalah taraf signikan dan $db = 1$. Uji nilai Likelihood ratio (T) yang menunjukkan penolakan H_0 , berarti model tersebut lebih baik, karena mengindikasikan overdispersi pada data.

3. Metode Penelitian

3.1. Jenis dan Sumber Data

Data yang digunakan pada penulisan ini berupa data sekunder tentang kasus AIDS menurut provinsi di Indonesia yang diperoleh dari Survei Sosial Ekonomi Nasional (Susenas) dan Badan Pusat Statistik (BPS) tahun 2011. Data yang digunakan berupa data banyaknya jumlah kasus AIDS menurut provinsi di Indonesia, kepadatan penduduk, jumlah penduduk miskin dan pengangguran dengan total sebanyak 33 data.

3.2. Variabel Data

Variabel data yang digunakan dalam penulisan ini yaitu banyaknya jumlah kasus AIDS menurut provinsi di Indonesia sebagai variabel respon (Y) dan variabel prediktor (X) meliputi:

- (a) Kepadatan penduduk (X_1) diukur dalam satuan jiwa per kilometer persegi.
- (b) Penduduk miskin (X_2) diukur dalam ukuran persentase.
- (c) Pengangguran (X_3) yang diukur dalam persentase.

3.3. Teknik Pengolahan Data

Data yang digunakan diolah dengan menggunakan software SPSS 16.0. Langkah-langkah analisis data yang digunakan dalam penulisan ini adalah sebagai berikut:

- (1) Melakukan uji Kolmogorov-Smirnov untuk menguji apakah variabel respon Y mengikuti distribusi Poisson atau tidak.

- (2) Menentukan model regresi Poisson.
- (3) Menghitung nilai koefisien dispersi untuk mengidentifikasi overdispersi.
- (4) Jika terjadi overdispersi, kemudian dilanjutkan dengan estimasi parameter binomial negatif.
- (5) Menentukan model regresi binomial negatif.
- (6) Melakukan pengujian kesesuaian model regresi binomial negatif.
- (7) Melakukan pemilihan model terbaik.

4. Hasil

4.1. *Pengujian Distribusi Poisson Pada Variabel Respon Y*

Pengujian distribusi Poisson pada variabel respon Y yaitu banyaknya jumlah kasus AIDS, dilakukan dengan uji Kolmogorov-Smirnov. Berikut ini adalah hipotesis uji Kolmogorov-Smirnov:

H_0 : Banyaknya jumlah kasus AIDS berdistribusi Poisson.

H_1 : Banyaknya jumlah kasus AIDS tidak berdistribusi Poisson.

Berdasarkan hasil yang diperoleh, nilai $D_{hitung} = 0,476$ dan nilai $p - value = 0,132$. Sementara nilai D_{tabel} untuk uji Kolmogorov Smirnov adalah $0,515$ dengan taraf uji $\alpha = 10\%$. Karena nilai $D_{hitung} < D_{tabel}$ dan $p - value > \alpha (0,132 > 0,1)$ maka H_0 diterima dan disimpulkan bahwa jumlah kasus AIDS menurut provinsi di Indonesia tahun 2011 mengikuti bentuk distribusi Poisson.

4.2. *Model Regresi Poisson*

Berikut ini adalah parameter estimasi model regresi Poisson dengan menggunakan SPSS 16.0.

Parameter l	Estimasi	Standar Error	p-value
<i>Intercept</i>	6.365	0.0200	0.000
Kepadatan penduduk	0.00014	1.615×10^6	0.000
Penduduk Miskin	-0.00089	0.0027	0.747
Pengangguran	0.009	0.0006	0.000

Tabel 1. Hasil analisa regresi Poisson dalam pemodelan kasus AIDS:

Ini berarti kepadatan penduduk dan tingkat pengangguran berpengaruh terhadap penyebaran virus AIDS di Indonesia. Sehingga diperoleh model regresi Poisson sebagai berikut:

$$y_i = \exp(6.365 + 0.00014x_1 + 0.009x_3).$$

4.3. *Pengujian Overdispersi*

Berikut ini adalah output goodness of fit SPSS 16 dari data kasus AIDS dengan menggunakan model regresi Poisson.

Deviance = 36.100	df = 29	Value= 1.245
Pearson = 49.570	df = 29	Value = 1.709

Tabel 2. Statistik uji Kecocokan Model Regresi Poisson

Berdasarkan Tabel 2, pada model regresi Poisson yang dihasilkan, nilai Deviance dan nilai Pearson Chi-square dibagi dengan derajat bebas yang lebih besar dari 1, sehingga terjadi overdispersi pada model regresi Poisson yang dihasilkan. Adanya overdispersi menyebabkan model regresi Poisson menjadi kurang baik, karena memiliki tingkat kesalahan yang tinggi. Salah satu cara untuk mengatasi adanya kasus overdispersi dalam regresi Poisson adalah dengan menggunakan model regresi Binomial Negatif.

4.4. Model Regresi Binomial Negatif

Berikut ini adalah output parameter estimates SPSS 16 dari data kasus AIDS dengan menggunakan model regresi Binomial Negatif.

Parameter l	Estimasi	Standar Error	p-value
<i>Intercept</i>	6.229	0.5673	0.000
Kepadatan penduduk	0.0002	0.0001	0.082
Penduduk Miskin	-0.003	0.0707	0.481
Pengangguran	0.014	0.0201	0.967

Tabel 3. Hasil analisa regresi binomial negatif

Model Regresi Binomial Negatif yang terbentuk yaitu:

$$y = \exp(6.229 + 0.0002x_1).$$

Berikut ini adalah output *goodness of fit* SPSS 16 dari data kasus AIDS dengan menggunakan model regresi Binomial Negatif.

Deviance = 80.275	df = 29	Value = 2.768
Pearson = 68.872	df = 29	Value = 2.375

Tabel 4. Statistik uji Kecocokan Model Regresi Binomial Negatif

Berdasarkan Tabel 4 dapat dilihat bahwa nilai Deviance dibagi dengan derajat bebasnya adalah 2.768 dan nilai Pearson Chi-square dibagi dengan derajat bebasnya adalah 2.375.

Walaupun masih ada indikasi terjadi overdispersi, namun jika dibandingkan dengan nilai Deviance dan Pearson Chi-square pada regresi Poisson, terjadi penurunan

koefisien overdispersi setelah menggunakan Binomial Negatif. Sehingga bisa disimpulkan bahwa regresi Binomial Negatif mampu mengendalikan overdispersi pada regresi Poisson, namun karena terdapat dua parameter yang tidak signifikan, maka perlu dilakukan uji signifikansi terhadap parameter yang signifikan, yaitu kepadatan penduduk.

4.5. Uji Signifikansi Parameter

Setelah digunakan model regresi Binomial Negatif, ternyata terdapat parameter yang tidak signifikan, sehingga dilakukan uji signifikansi parameter terhadap satu parameter yang signifikan, yaitu kepadatan penduduk. Berikut ini adalah hasil parameter estimates dari satu parameter yang signifikan.

Parameter l	Estimasi	Standar Error	p-value
<i>Intercept</i>	6.444	0.1880	0.000
Kepadatan penduduk	0.00018	0.0001	0.076

Tabel 5. Nilai Parameter Regresi Binomial Negatif berdasarkan parameter signifikan

Berdasarkan Tabel 5 diperoleh model regresi Binomial Negatif sebagai berikut:

$$y = \exp(6.444 + 0.00018x_1).$$

Parameter yang signifikan pada tingkat signifikansi 0.1 dalam model regresi Binomial Negatif, yaitu kepadatan penduduk .

Interpretasi $b_1 = 0.00018$: Untuk setiap kenaikan kepadatan penduduk sebesar 1 jiwa/kilometer persegi, maka diduga rata-rata jumlah kasus AIDS bertambah sebesar $\exp(0.00018) \sim 1$ kasus.

4.6. Pemilihan Model Terbaik

Model terbaik adalah model yang memiliki nilai AIC, BIC yang lebih kecil, nilai Log-likelihood yang lebih besar dan uji nilai Likelihood ratio (T) yang menunjukkan penolakan berarti model tersebut lebih baik.

Kriteria	Regresi Poisson	Regresi Binomial Negatif dengan parameter signifikan
AIC	-3.098×10^5	-3.456×10^5
BIC	-3.098×10^5	-3.456×10^5
Log-likelihood	1.549×10^5	1.728×10^5

Tabel 6. Uji Model Terbaik

Berdasarkan Tabel 6 dapat dilihat bahwa dari kriteria model terbaik yang memiliki nilai AIC, BIC yang lebih kecil dan nilai Log-likelihood yang lebih besar adalah regresi Binomial Negatif .

Selanjutnya akan dilakukan uji Likelihood ratio. Berdasarkan nilai Log-likelihood pada tabel 6. diperoleh nilai Likelihood ratio sebagai berikut:

$$\begin{aligned} T &= 2(\widehat{L}_{BN} - \widehat{L}_{Poisson}) \\ &= 2((1.728 \times 10^5) - (1.549 \times 10^5)) \\ &= 2(17.900) \\ &= 35.800 \end{aligned}$$

Bila dipilih taraf nyata $\alpha = 0,1$, nilai $\chi^2_{(0,1;1)} = 3,841$, karena $T > 3,841$ maka H_0 ditolak yang berarti terjadi overdispersi atau dengan kata lain model regresi Binomial Negatif lebih baik digunakan daripada model regresi Poisson.

Jadi, model terbaik untuk memodelkan jumlah kasus AIDS menurut provinsi di Indonesia tahun 2011 adalah regresi Binomial Negatif.

5. Kesimpulan

Model regresi Poisson dan regresi Binomial Negatif digunakan untuk memodelkan kasus AIDS menurut provinsi di Indonesia tahun 2011. Jika terjadi overdispersi namun tetap digunakan regresi Poisson, maka nilai standard error menjadi naik, sehingga kesimpulannya menjadi tidak valid.

Berdasarkan nilai Log-likelihood, uji Likelihood ratio, AIC dan BIC, model regresi Binomial Negatif lebih baik digunakan dibandingkan dengan regresi Poisson untuk kasus AIDS menurut provinsi di Indonesia tahun 2011. Model untuk regresi Binomial Negatif yang dihasilkan adalah:

$$y = \exp(6.444 + 0.00018x_1). \tag{5.1}$$

6. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Bapak Yudiantri Asdi M.Sc, Ibu Dr. Ferra Yanuar dan Bapak Narwen M.Si yang telah memberikan masukan dan saran sehingga paper ini dapat diselesaikan dengan baik.

Daftar Pustaka

- [1] Bain, L.J dan Engelhardt, M. 1992. *Introduction to Probability and Mathematical Statistics*. Second Edition. PWS-KENT: Boston.
- [2] Cameron, A.C. dan P.K. Trivedi. 1998. *Regression analysis of count data*. Cambridge University Press, New York.
- [3] Kleinbaum, D.G., L.L. Kupeer., A. Nizam., dan K.E. Muller. 1978. *Applied Regression Analysis and Other Multivariabel*. Fourth Edition. Thomson Brooks//Cole, Canada.
- [4] Joseph, M.H. 2011. *Negative Binomial Regression*. Second Edition. Cambridge University Press, New York.
- [5] McCullagh P., and J.A. Nelder. 1989. *Generalized Linier Models*. Second Edition. Chapman and Hall, London.
- [6] Myers, R.H. 1990. *Classical and Modern Regression with Application*. Second Edition. PWS-KENT Publishing Company, Boston.
- [7] Siegel, S. 1992. *Statistik Nonparametrik*. PT Gramedia, Jakarta.