

PENANGANAN OVERDISPERSI PADA REGRESI POISSON (Studi Kasus: Pengaruh Faktor Iklim Terhadap Jumlah Penderita Penyakit Demam Berdarah di Kota Bogor)

Marta Sundari ¹, Pardomuan Robinson Sihombing*²
¹Universitas Indraprasta PGRI, ²Badan Pusat Statistik, Jakarta
sundarimarta@gmail.com¹, robinson@bps.go.id*²

Abstract

Dengue hemorrhagic fever (DHF) has become an endemic disease in major cities in Indonesia. The incidence of DHF is discrete data (count), which follows the Poisson distribution—in this study, modeling the influence of climate on DHF incidence the Poisson regression method. Overdispersion problems are often found in Poisson regression, where the expected value does not equal the variance value as required in the Poisson regression. Furthermore, to deal with the overdispersion problem in Poisson regression, an approach is carried out using several other regression models, namely Negative Binomial, Zero-Truncated Poisson, and Zero-Truncated Negative Binomial models. The model chosen with the best criteria is the Negative Binomial. In the Negative Binomial, only the average humidity variable per year significantly affects the number of dengue fever sufferers in Bogor City in 2013-2019 at the 5% real level. The variable average temperature and the varying amount of rainfall have a significant effect on the number of dengue fever sufferers in Bogor City at the whole level of 15%.

Keywords: dengue, negative binomial, overdispersion, poisson

Abstrak

Penyakit demam berdarah dengue (DBD) telah menjadi penyakit endemik di kota-kota besar di Indonesia. Angka kejadian DBD merupakan data diskrit (count) yang mengikuti distribusi Poisson. Dalam penelitian ini dilakukan pemodelan pengaruh iklim terhadap angka kejadian DBD menggunakan metode regresi Poisson. Masalah overdispersi sering ditemukan dalam regresi poisson, dimana nilai ekspektasi tidak sama dengan nilai variannya seperti yang disyaratkan dalam regresi poisson. Selanjutnya untuk menangani masalah overdispersi yang terjadi pada regresi poisson dilakukan pendekatan dengan menggunakan beberapa model regresi lain yaitu model Regresi Negative Binomial, Zero-Truncated Poisson dan Zero-Truncated Negative Binomial. Model terpilih dengan kriteria terbaik adalah Negative Binomial. Pada Negative Binomial hanya peubah rata-rata kelembapan per tahun yang memiliki pengaruh yang nyata terhadap jumlah penderita penyakit DBD di Kota Bogor pada tahun 2013-2019 pada taraf nyata 5%. Peubah rata-rata suhu rata-rata dan peubah jumlah curah hujan memiliki pengaruh yang nyata terhadap jumlah penderita penyakit DBD di Kota Bogor pada taraf nyata 15%.

Kata kunci: demam berdarah, negative binomial, overdispersi, poisson

Received: January 22, 2021 / Accepted: April 9, 2021 / Published Online: April 14, 2021

PENDAHULUAN

Proyeksi kecenderungan aktivitas ekonomi dan dampak emisi gas rumah kaca yang dihasilkan dari kegiatan manusia di masa datang diperkirakan akan berpengaruh terhadap pergeseran pola curah hujan dan suhu rata-rata bumi yang diperkirakan naik 1 hingga 3,5 0C. Perubahan pada komponen lingkungan ini akan mempengaruhi spesies-spesies pada kelompok ekosistem dan pola penyebaran vektor serta virus penyakit (McMichael, Anthony, & Woodruff, 1996) . Iklim dapat berpengaruh terhadap pola penyakit infeksi karena agen penyakit (virus, bakteri, atau parasit lainnya) dan vektor (serangga atau rodensia) bersifat sensitif terhadap suhu, kelembaban dan kondisi lingkungan ambien lainnya. Cuaca dan iklim berpengaruh terhadap penyakit yang berbeda dengan cara yang berbeda (Gubler, Suharyono, Lubis, Eram, & Saroso, 1979)

World Health Organization (2018) merilis laporan yang menyatakan bahwa penyakit yang ditularkan melalui nyamuk seperti demam berdarah dengue (DBD), malaria dan demam kuning berhubungan dengan kondisi cuaca yang hangat. Sebaliknya, influenza berhubungan dengan kondisi cuaca yang dingin dan meningitis berhubungan dengan kondisi lingkungan yang kering. Banyak yang menduga bahwa Kejadian Luar Biasa DBD yang terjadi setiap tahun hampir seluruh di Indonesia terkait erat dengan pola cuaca di Asia Tenggara. Tingkat penyebaran virus diperkirakan mengalami peningkatan pada peralihan musim yang ditandai oleh curah hujan dan suhu udara yang tinggi. Selain itu, perubahan gaya hidup ikut berperan menambah population at risk. Penggunaan barang non biodegradable seperti plastik yang sangat tinggi, menyebabkan plastik menjadi komposisi sampah terbesar saat ini sehingga berpotensi menjadi penampung air hujan, tempat perkembangbiakan vektor.

Penyakit demam berdarah dengue (DBD) telah menjadi penyakit endemik di kota-kota besar di Indonesia. Ramalan International Panel on Climate Change (IPCC) tahun 1996 menyebutkan insidens DB di Indonesia dapat meningkat tiga kali lipat pada tahun 2070. Proyeksi kecenderungan aktivitas ekonomi dan dampak emisi gas. Angka kejadian Demam Berdarah Dengue merupakan data diskrit (count) yang mengikuti distribusi Poisson. Dalam penelitian ini dilakukan pemodelan pengaruh iklim terhadap angka kejadian Demam Berdarah Dengue menggunakan metode regresi Poisson.

Regresi Poisson digunakan untuk menganalisis hubungan antara peubah tak bebas (Y) yang berupa data diskrit dengan satu atau lebih peubah bebas (X) dengan nilai ekspektasi (rata-rata) dan ragam diasumsikan sama (equidispersi) yaitu $E[Y]=var[Y]=\mu$. Namun seringkali dalam penerapannya terjadi pelanggaran asumsi tersebut dimana ketika nilai ragamnya lebih

besar rataan (overdispersi) atau ketika nilai ragamnya lebih kecil dari rataan (underdispersi). Beberapa hal yang menyebabkan terjadinya masalah overdispersi adalah terdapat sumber keragaman yang belum teramati (unobserved heterogeneity), adanya pengamatan yang hilang (missing) pada peubah X, adanya pencilan pada data sehingga diperlukan interaksi dalam model, kebutuhan untuk mentransformasi peubah X atau kesalahan spesifikasi fungsi penghubung.

METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data jumlah penderita demam berdarah per tahun di Kotamadya Bogor yang bersumber dari Jawa Barat Dalam Angka 2011-2020 (Badan Pusat Statistik Jawa Barat). Faktor iklim yang diduga mempengaruhi jumlah penderita demam berdarah merupakan data harian yang bersumber dari website BMKG (2020). Adapun variabel yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Variabel Penelitian

Y	Jumlah penderita demam berdarah per tahun
X1	Kepadatan penduduk per km ² per tahun
X2	Rataan suhu minimum per tahun
X3	Rataan suhu rata-rata per tahun
X4	Rataan suhu maksimum per tahun
X5	Rataan kelembaban rata-rata per tahun
X6	Jumlah curah hujan per tahun

Sebaran Poisson

Menurut Cameron dan Trivedi (2013), suatu peubah Y yang tipe datanya diskrit akan mengikuti sebaran Poisson jika μ adalah rataan suatu kejadian per satuan waktu dan I adalah periode waktu tertentu maka rataan dari y menjadi μ_i . Fungsi massa peluang sebaran Poisson diberikan dalam persamaan berikut:

$$P(y_i; \mu_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}, y_i = 0, 1, 2, \dots, \infty$$

(1)

Persamaan ini digunakan untuk menghitung peluang peubah acak Y dengan rataan dan ragam sebaran Poisson adalah sama atau $E[Y] = \mu_{y_i} = var[Y]$.

Regresi Poisson

Model regresi Poisson ditulis sebagai berikut:

$$y_i = \mu_i + \varepsilon_i = \exp(x_i^T \beta + \varepsilon_i) = \exp(\beta_0 + \beta_1 X_{1i} + \dots + \beta_{pi} X_{pi}) + \varepsilon_i, i = 1, 2, \dots, n$$

(2)

dengan y_i adalah jumlah kejadian dan μ_i adalah rata-rata jumlah kejadian dengan μ_i diasumsikan tidak berubah dari data ke data. Selanjutnya dilakukan uji *Goodness of Fit* terhadap model. Terdapat beberapa metode yang digunakan dalam menentukan model terbaik, salah satunya adalah AIC (*Akaike's Information Criterion*). AIC didefinisikan sebagai berikut:

$$AIC = -2 \ln L(\hat{\beta}) + 2k$$

(3)

dimana $L(\hat{\beta})$ adalah nilai *likelihood*, dan k adalah jumlah parameter. Model terbaik adalah model yang mempunyai nilai AIC terkecil.

Parameter Dispersi dan Overdispersi

Parameter dispersi (ϕ) diperoleh dari rumus $\phi = \frac{\text{nilai deviance}}{df}$ dengan df adalah derajat bebas. Menurut Rashwan dan Kamel (2011) nilai *deviance* didefinisikan sebagai berikut:

$$G^2 = 2 \sum_{i=1}^n y_i \ln \left(\frac{y_i}{\lambda_i} \right) \quad (4)$$

Jika nilai $\phi > 0$ maka terjadi overdispersi dan sebaliknya jika $\phi < 0$ maka terjadi underdispersi. Salah satu solusi dalam menangani masalah overdispersi dapat menggunakan pendekatan dengan sebaran binomial negatif, *zero inflation poisson (ZIP)* dan lain sebagainya

Regresi Zero-Truncated Poisson

Regresi Zero-Truncated Poisson dan Regresi Zero-Truncated Negative Binomial merupakan salah satu model yang dapat digunakan dalam masalah overdispersi. Model ini mengakomodasi jika data yang digunakan tidak terdapat nilai nol atau dengan kata lain data yang digunakan merupakan data yang selalu ada dalam rentang waktu tertentu (Liu, Saat, Qin, & Barkan, 2013).

Regresi Binom Negatif

Misalkan Y_i berdistribusi binomial negatif, maka Y_i akan memiliki fungsi peluang sebagai berikut:

$$Pr^{NB}(Y_i = y_i | \mu_i) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1}) y_i!} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i} \right)^{\alpha^{-1}} \left(\frac{\mu_i}{\alpha^{-1} + \mu_i} \right)^{y_i}$$

(5)

dengan $\Gamma(\cdot)$ adalah fungsi gamma dan $\alpha (\alpha > 0)$ merupakan parameter overdispersi.

Zero-Truncated Negative Binomial Regression

Dari model binomial negatif diatas, nilai y_i nya masih bisa mengandung nol. Jika dalam suatu observasi nilai $y_i = 0$ tidak muncul maka lebih baik digunakan distribusi *Zero-Truncated Negative Binomial* (Liu, Saat, Qin, & Barkan, 2013). Jika Y_i berdistribusi *Zero-Truncated Negative Binomial* maka Y_i akan memiliki model sebagai berikut:

$$Pr_{ZT}^{NB}(Y_i = y_i | y_i > 0, \mu_i, \alpha) = \frac{Pr^{NB}(Y_i = y_i | \mu_i)}{1 - Pr^{NB}(Y_i = 0 | \mu_i)}$$

(6)

dengan $Pr^{NB}(Y_i = 0 | \mu_i) = (1 + \alpha\mu_i)^{-\alpha^{-1}}$ dan nilai harapan bersyarat dari y_i adalah $(y_i | y_i > 0, \mu_i, \alpha) = \frac{\mu_i}{1 - (1 + \alpha\mu_i)^{-\alpha^{-1}}}$.

HASIL DAN PEMBAHASAN

Rata-rata penderita DBD per tahun di Kota Bogor sebanyak 898 penderita dengan jumlah penderita paling sedikit terjadi pada tahun 2014 dan paling tinggi pada tahun 2016. Kecenderungan tingkat kepadatan penduduk meningkat setiap tahun dengan rata-rata kepadatan penduduk per tahun sebanyak 8.977 penduduk/km². Rataan dari rata-rata suhu minimum per tahun di Kota Bogor mencapai 21,29 °C dengan rata-rata suhu minimum terendah terjadi pada tahun 2015 mencapai 20,05 °C dan rata-rata dari rata-rata suhu maksimum mencapai 33,38 °C dengan rata-rata suhu maksimum terjadi pada tahun 2019 mencapai 36,10 °C.

Kecenderungan tingkat rata-rata kelembapan per tahun di Kota Bogor menurun setiap tahun dengan rata-rata dari rata-rata kelembapan per tahun di Kota Bogor mencapai 82,12% dengan rata-rata kelembapan minimum mencapai 79% pada tahun 2019 dan maksimum pada tahun 2016 mencapai 85,34%. Kecenderungan rata-rata curah hujan per tahun di Kota Bogor menurun setiap tahun dengan rata-rata dari jumlah curah hujan per tahun di Kota Bogor mencapai 3.747 mm dengan jumlah curah hujan paling sedikit mencapai 3.045 mm pada tahun 2018 dan paling banyak pada tahun 2016 mencapai 4.832 mm.

Tabel 2. Deskriptif Variabel Penelitian

Peubah	Rataan	Simpangan Baku	Minimum	Q1	Median	Q3	Maximum
Y	898,00	221,20	670,00	687,00	855,00	1107,00	1229,00
X1	8.977,00	301,00	8.549,00	8.698,00	8.985,00	9.256,00	9.385,00

X2	21,29	1,84	18,20	20,04	20,99	22,96	23,30
X3	27,02	2,38	25,95	25,95	26,10	26,50	32,40
X4	33,28	1,68	31,60	31,68	33,80	34,16	36,10
X5	82,12	2,22	79,00	79,90	82,55	83,87	85,34
X6	3.747,00	645,00	3.045,00	3.215,00	3.657,00	4.231,00	4.832,00

Korelasi Antar Peubah Bebas

Peubah Y memiliki korelasi yang kuat terhadap peubah X1, X2, X5 dan X6 (diatas 0.4) dan memiliki korelasi yang cukup rendah terhadap peubah X3 dan X4 (dibawah 0.3). Korelasi yang kuat juga terlihat pada hubungan antara X5 dengan seluruh peubah bebas kecuali pada peubah X4 (diatas 0.5). Korelasi antar peubah lain dapat dilihat pada tabel 4.2 berikut:

Tabel 3. Analisis Korelasi Antar Peubah

	Y	X1	X2	X3	X4	X5
X1	-0,447					
X2	0,442	-0,562				
X3	-0,271	0,615	-0,769			
X4	-0,204	0,282	-0,519	0,409		
X5	0,713	-0,594	0,760	-0,545	-0,270	
X6	0,469	-0,193	0,233	0,007	-0,000	0,747

Regresi Poisson

Model regresi poisson pertama yang melibatkan seluruh peubah bebas yang digunakan dapat dilihat pada Tabel 4. Pada model ini terdapat indikasi terjadinya multikolinieritas dengan adanya nilai VIF pada peubah rata-ran kelembaban rata-ran per tahun yang lebih besar dari 10. Kemudian dilakukan pemodelan ulang dengan tidak menyertakan peubah X2 karena meskipun X5 memiliki nilai VIF tertinggi namun X5 memiliki korelasi paling kuat terhadap peubah Y sehingga dicari peubah X lain yang memiliki korelasi paling kuat terhadap peubah X5 namun memiliki korelasi tidak terlalu tinggi terhadap peubah Y.

Tabel 4. Nilai VIF Variabel Independen

Peubah	Koefisian	Standar Error	VIF
Konstanta	-29,38	1,91	
X1	0,000214	0,000057	2,31
X2	-0,2548	0,0195	8,13

X3	0,0787	0,0110	3,17
X4	-0,02603	0,00332	1,53
X5	0,5029	0,0244	21,26

Selanjutnya dilakukan regresi poisson tanpa melibatkan variabel X2. Masalah overdispersi ditemukan dalam regresi poisson model kedua, hal ini terlihat dari nilai parameter dispersi yang lebih dari satu. Selanjutnya untuk menangani masalah overdispersi yang terjadi pada regresi poisson model kedua dilakukan pendekatan dengan menggunakan beberapa model regresi lain yaitu model regresi *negative binomial*, *quasi-poisson*, *Zero-Truncated Poisson* dan *Zero-Truncated Negative Binomial*.

Pemilihan Model Terbaik

Berdasarkan Tabel 5, Regresi *Negative Binomial* memiliki nilai parameter dispersi paling kecil dibandingkan model lain yaitu sebesar 2.51. Model *Zero Truncated Negative Binomial* juga memiliki nilai parameter dispersi kecil yaitu sebesar 3.35, dan dengan menggunakan uji Log-Likelihood Ratio diketahui bahwa nilai parameter dispersi model *Zero Truncated Poisson* lebih besar dari model *Zero Truncated Negative Binomial*. Model *Negative Binomial* memiliki nilai AIC yang sama dengan model *Zero Truncated Negative Binomial* sebesar 145.68 jauh lebih rendah dari dua model lainnya. Berdasarkan kedua hal ini, dapat disimpulkan bahwa pemodelan pengaruh faktor iklim terhadap jumlah penderita DBD di Kota Bogor paling baik menggunakan model Regresi *Negative Binomial* dibandingkan tiga model lainnya.

Tabel 5. Pemilihan Model Terbaik

Model Regresi	Log-Likelihood	Parameter Dispersi	Log-Likelihood Ratio Test (p-value)	AIC
Poisson	-427,01	84,64		437,07
<i>Negative Binomial</i>	-131,68	2,51		145,68
<i>Zero Truncated Poisson (ZTP)</i>	-212,53		Model ZTNB lebih baik dari ZTP	437,07
<i>Zero Truncated Negative Binomial (ZTNB)</i>	-65,84	3,35		145,68

Dugaan Parameter Model Regresi

Overdispersi pada model regresi Poisson dan *Zero Truncated Poisson* menyebabkan galat

baku menjadi terlalu kecil sehingga banyak peubah X yang menjadi signifikan padahal sebenarnya tidak signifikan sehingga menyebabkan pengambilan keputusan menjadi tidak valid. Kondisi ini dapat terlihat dari banyaknya peubah X yang signifikan pada model regresi Poisson dan *Zero Truncated Poisson* jika dibandingkan dengan model *Negative Binomial* dan model *Zero Truncated Negative Binomial*. Tabel 6 memperlihatkan bahwa pada model *Negative Binomial* hanya peubah rata-rata kelembaban per tahun yang memiliki pengaruh yang nyata terhadap jumlah penderita penyakit DBD di Kota Bogor pada tahun 2013-2019 pada taraf nyata 5%. Peubah rata-rata suhu rata-rata dan peubah jumlah curah hujan memiliki pengaruh yang nyata terhadap jumlah penderita penyakit DBD di Kota Bogor pada taraf nyata 15%.

Tabel 6. Signifikansi Parameter Regresi

Peubah	Poisson		Negative Binomial		Zero-Truncated Poisson		Zero-Truncated Neg Binom	
	Koef	p-Value	Koef	p-Value	Koef	p-Value	Koef	p-Value
Intersep 1	-1,47E+04	<2e-16 ***	-1,44E+04	0.10440	-1,47E+04	<2e-16 ***	-1,44E+04	0.10440
Intersep 2							3,36E+03	2.56e-13 ***
X1	-8,22E-02	0.1026	5,08E-02	0.86578	-8,22E-02	0.1026	5,08E-02	0.86576
X3	1,08E+02	<2e-16 ***	8,67E+01	0.12892	1,08E+02	<2e-16 ***	8,67E+01	0.12893
X4	-6,12E+00	0.0407 *	-4,98E+00	0.77345	-6,12E+00	0.0407 *	-4,98E+00	0.77345
X5	2,55E+02	<2e-16 ***	2,42E+02	0.00615 **	2,55E+02	<2e-16 ***	2,42E+02	0.00616 **
X6	-3,69E-01	<2e-16 ***	-3,47E-01	0.13210	-3,69E-01	<2e-16 ***	-3,47E-01	0.13211

Hasil yang didapat senada dengan penelitian yang dilakukan oleh peneliti lainnya, sebagaimana penularan beberapa penyakit menular sangat dipengaruhi oleh faktor iklim (Brisbois & Ali, 2010). Parasit dan vektor penyakit sangat peka terhadap faktor iklim, khususnya suhu, curah hujan, kelembaban, permukaan air, dan angin (Hopp & Foley, 2001) dan (Ramesh, Sharmila, Dhillon, & Aditya, 2010) .

KESIMPULAN

Model regresi poisson yang didapat mengalami masalah overdispersi, sehingga dilakukan penanganan dengan menyesuaikan distribusi data dengan metode Negative Binomial, Zero Truncated Poisson (ZTP) dan Zero Truncated Negative Binomial (ZTNB). Berdasarkan nilai parameter dispersi dan nilai AIC yang dihasilkan dari ketiga model regresi yang dicobakan, regresi *Negative Binomial* merupakan model terbaik yang dapat digunakan untuk memodelkan

hubungan pengaruh faktor iklim terhadap jumlah penderita DBD di Kota Bogor. Pada taraf nyata 5% terdapat satu peubah penjelas yang memiliki pengaruh nyata terhadap jumlah penderita DBD di Kota Bogor dan pada taraf nyata 15% terdapat tiga peubah bebas yang memiliki pengaruh nyata terhadap jumlah penderita DBD di Kota Bogor. Pada penelitian selanjutnya dapat menambahkan model data panel dan efek spasial terhadap kabupaten-kota disekitran wilayah penelitian.

DAFTAR PUSTAKA

- Badan Pusat Statistik Jawa Barat. (2011-2020). *Jawa Barat dalam Angka*. Bandung: Badan Pusat Statistik Jawa Barat.
- BMKG. (2020, Desember 1). Retrieved from <http://dataonline.bmkg.go.id/>.
- Brisbois, B., & Ali, S. (2010). *Climate Change, VectorBorne Disease and Interdisciplinary Research: Social Science Perspectives on an Environment and Health Controversy*. Heidelberg: Springer.
- Cameron, C., & Trivedi, P. (2013). *Regression Analysis of Count Data (Second Edition)*. New York: Cambridge University Press.
- Gubler, D. J., Suharyono, W., Lubis, I., Eram, S., & Saroso, J. S. (1979). Epidemic dengue hemorrhagic fever in rural Indonesia. I. Virological and epidemiological studies. *The American Journal of Tropical Medicine and Hygiene*, 701-710.
- Hopp, M., & Foley, J. (2001). *Global-Scale Relationships Between Climate and the Dengue Fever Vector, Aedes Aegypti*. New York: Kluwer Academic Publishers.
- Liu, Saat, M., Qin, & Barkan. (2013). Analysis of U.S. freight-train derailment severity using zero-truncated negative binomial regression and quantileregression. *Accident Analysis and Prevention* 59, 87–93.
- McMichael, Anthony, J., & Woodruff, E. R. (1996). *Climate Change and Human Health*. Netherlands: Springer .
- Purwandari, D. (2020). PEMODELAN REGRESI LATEN PADA EFEK PLASEBO MEMAKAI METODE MAXIMUM LIKELIHOOD. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, 1(3), 193-198.
- Ramesh, C., Sharmila, P., Dhillon, G., & Aditya, P. (2010). Climate change and threat of vector-borne diseases in India: Are we prepared? *Parasitol. Res*, 106(4), 763-773.
- WHO. (2018). *Dengue data application*. Jenewa: World Health Organization.