

Handling of Overdispersion in the Poisson Regression Model with Negative Binomial for the Number of New Cases of Leprosy in Java*

Penanganan Overdispersi pada Model Regresi Poisson dengan Binomial Negatif untuk Jumlah Kasus Baru Kusta di Jawa

Yopi Ariesia Ulfa¹, Agus M Soleh^{2‡}, and Bagus Sartono³

¹Badan Pusat Statistik, Indonesia

^{2,3}Department of Statistics, IPB University, Indonesia

[‡]corresponding author: agusms@apps.ipb.ac.id

Copyright © 2021 Yopi Ariesia Ulfa, Agus M Soleh, and Bagus Sartono. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Based on data from the Directorate General of Disease Prevention and Control of the Ministry of Health of the Republic of Indonesia, in 2017, new leprosy cases that emerged on Java Island were the highest in Indonesia compared to the number of events on other islands. The purpose of this study is to compare Poisson regression to a negative binomial regression model to be applied to the data on the number of new cases of leprosy and to find out what explanatory variables have a significant effect on the number of new cases of leprosy in Java. This study's results indicate that a negative binomial regression model can overcome the Poisson regression model's overdispersion. Variables that significantly affect the number of new cases of leprosy based on the results of negative binomial regression modeling are total population, percentage of children under five years who had immunized with BCG, and percentage of the population with sustainable access to clean water.

Keywords: leprosy, negative binomial regression, overdispersion, poisson regression.

* Received: Jan 2020; Reviewed: Jan 2021; Published: March 2021

1. Pendahuluan

Kusta adalah penyakit yang disebabkan oleh infeksi *Mycobacterium leprae*. Penyakit ini menyerang berbagai bagian tubuh diantaranya kulit dan saraf. Bila tidak ditangani, kusta dapat sangat progresif menyebabkan kerusakan pada kulit, saraf-saraf, anggota gerak, dan mata. Tidak seperti mitos yang beredar di masyarakat, kusta tidak menyebabkan pelepasan anggota tubuh yang begitu mudah seperti pada penyakit *tzaraath* yang digambarkan dan sering disamakan dengan kusta.

Berdasarkan *Weekly Epidemiological Report* WHO pada September 2016, jumlah kasus baru kusta di dunia paling banyak terdapat di regional Asia Tenggara, diikuti regional Amerika dan Afrika. Jumlah kasus baru kusta pada negara-negara di regional Asia Tenggara semakin meningkat sejak 2011. Walaupun secara global terjadi penurunan jumlah kasus baru kusta, namun pada beberapa negara seperti Bangladesh, Kongo, Ethiopia, India, dan Indonesia justru mengalami peningkatan kasus baru. Selain itu, Mozambique yang pada tahun 2013 dan 2014 sudah tidak memiliki kasus baru, kembali memiliki jumlah kasus baru kusta di tahun 2015 sebanyak 1.335 kasus.

Berdasarkan data dari Ditjen Pencegahan dan Pengendalian Penyakit Kementerian Kesehatan Republik Indonesia, pada tahun 2017 kasus baru kusta yang muncul di Pulau Jawa merupakan yang paling tinggi di Indonesia apabila dibandingkan dengan jumlah kejadian di pulau-pulau lainnya. Untuk itu peneliti memiliki keingintahuan untuk menelaah masalah faktor-faktor apa saja yang mempengaruhi munculnya kasus baru kusta di Pulau Jawa (meliputi seluruh provinsi yang ada di Pulau Jawa yakni DKI Jakarta, Jawa Barat, Jawa Tengah, DIY, Jawa Timur, dan Banten).

Untuk meneliti data kasus baru kusta dibutuhkan teknik analisis yang sesuai. Ketika peubah respon berupa data cacahan yaitu data yang nilainya non negatif dan menyatakan banyaknya kejadian dalam interval waktu, ruang, atau volume tertentu, analisis regresi yang biasa digunakan adalah analisis regresi Poisson (Cameron & Trivedi, 2013).

Beberapa asumsi yang harus dipenuhi jika hendak menggunakan model Regresi Poisson yaitu peubah respon merupakan data cacah dan asumsi kedua yaitu terpenuhinya kondisi ekuidispersi atau rata-rata peubah respon harus sama dengan ragamnya (Dobson & Barnett, 2002). Namun dalam analisis data diskrit dengan menggunakan model Regresi Poisson terkadang terjadi pelanggaran asumsi ekuidispersi tersebut. Dalam model regresi linier klasik pelanggaran asumsi tersebut dinamakan pelanggaran asumsi homoskedastisitas (Cameron & Trivedi, 2013).

Ada kalanya nilai rata-rata peubah respon tidak sama dengan ragamnya. Apabila dugaan ragam lebih besar dari rata-ratanya maka kondisi tersebut disebut overdispersi. Sebaliknya jika dugaan ragam lebih kecil dari rata-ratanya maka kondisi tersebut dinamakan underdispersi. Menurut Giuffrè et al. (2011), kondisi yang lebih sering terjadi pada data Poisson adalah kondisi overdispersi. Kondisi underdispersi sangat jarang dijumpai.

Munculnya kondisi overdispersi menjadikan regresi Poisson kurang cocok diterapkan untuk analisis. Salah satu teknik analisis regresi yang dapat diterapkan yaitu regresi binomial negatif karena pada regresi ini tidak membutuhkan asumsi ekuidispersi (Mann et al., 2014). Metode regresi binomial negatif dapat memodelkan

data yang mengalami overdispersi tanpa harus menghilangkan kondisi overdispersi dari data tersebut.

Penelitian mengenai data kusta dengan regresi Poisson antara lain dilakukan oleh Aini (2013), Dzikrina & Purnami (2013), Shovalina & Atok (2016), serta Arfidina et al. (2017). Tujuan penelitian ini adalah untuk membandingkan model regresi Poisson dan regresi binomial negatif yang lebih tepat diterapkan pada data jumlah kasus baru kusta dan mengetahui peubah penjelas apa saja yang berpengaruh secara nyata terhadap jumlah kasus baru kusta di Pulau Jawa.

2. Metodologi

2.1 Bahan dan Data

Sumber data yang digunakan dalam penelitian ini berasal dari Profil Kesehatan Tahun 2017 dari Dinas Kesehatan seluruh provinsi di Pulau Jawa dan publikasi Daerah Dalam Angka Tahun 2018 dari BPS Provinsi di Pulau Jawa. Unit pengamatan yang digunakan adalah seluruh kabupaten/kota di Pulau Jawa tahun 2017 yang ada sebanyak 119 kabupaten/kota. Peubah-peubah yang digunakan dalam penelitian ini disajikan pada Tabel 1. Pemilihan peubah penjelas didasarkan pada penelitian-penelitian yang telah dilakukan sebelumnya yaitu peubah-peubah penjelas yang berpengaruh terhadap jumlah kasus kusta.

Tabel 1: Peubah-peubah yang digunakan dalam penelitian.

Peubah	Satuan	Sumber Data
Jumlah kasus baru kusta (Y)	Kasus	Publikasi Profil Kesehatan Tahun 2017
Jumlah penduduk (X_1)	Ribu Jiwa	Publikasi Daerah Dalam Angka 2018
Persentase balita yang diimunisasi BCG (X_2)	Ribu Jiwa/Km ²	Publikasi Daerah Dalam Angka 2018
Persentase rumah tangga berperilaku hidup bersih dan sehat (X_3)	Ribu Rumah tangga	Publikasi Profil Kesehatan Tahun 2017
Persentase rumah sehat (X_4)	Ribu Rumah	Publikasi Profil Kesehatan Tahun 2017
Persentase jumlah penduduk dengan akses air minum layak (X_5)	Persen	Publikasi Profil Kesehatan Tahun 2017
Persentase jumlah penduduk dengan sanitasi layak (X_6)	Persen	Publikasi Profil Kesehatan Tahun 2017
Rata-rata sarana kesehatan per desa/kelurahan (X_7)	Sarana/ desa	Publikasi Daerah Dalam Angka 2018
Rata-rata tenaga kesehatan per desa/kelurahan (X_8)	Orang/ desa	Publikasi Daerah Dalam Angka 2018

2.2 Metode Penelitian

Analisis dan pemodelan dilakukan menggunakan program statistika R 3.6.2 dengan paket MASS. Adapun tahapan analisis data yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Melakukan pengumpulan data dari berbagai sumber.
2. Melakukan eksplorasi data untuk mendapatkan gambaran dan informasi tentang perubahan respon maupun peubah penjelas yang digunakan. Untuk memudahkan dalam mendeskripsikannya, dilakukan pengelompokan kabupaten/kota. Pengelompokan dilakukan dengan membagi menjadi dua kelas yaitu rendah dan tinggi dengan menggunakan batas nilai rata-rata setiap peubah.
3. Memeriksa asumsi multikolinearitas antara peubah penjelas. Jika di antara peubah penjelas tidak saling bebas atau memiliki hubungan linier maka terindikasi adanya multikolinearitas. Multikolinearitas salah satunya dapat dideteksi dari nilai *Variance Inflation Factor* (VIF). Nilai VIF dihitung dengan rumus:

$$VIF_j = \frac{1}{(1-R_j^2)}$$

dengan R_j^2 adalah nilai koefisien determinasi (ukuran keeratan antar peubah penjelas) atau R^2 ketika X_j diregresikan dengan peubah penjelas selain X_j .

Akibat dari munculnya masalah multikolinear ini adalah ragam estimasi menjadi besar.

4. Melakukan pemodelan regresi Poisson dengan persamaan model:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k = \mathbf{x}'_i \boldsymbol{\beta}$$

dimana $g(\mu_i) = \ln(\mu_i) = \mathbf{x}'_i \boldsymbol{\beta}$ dan $g(\mu_i)$ merupakan fungsi hubung log. Pendugaan parameter regresi Poisson dilakukan menggunakan metode Pendugaan Kemungkinan Maksimum (PKM) dengan fungsi log-likelihood sebagai berikut:

$$\ln L(\boldsymbol{\beta}; \mathbf{y}) = \sum_{i=1}^n y_i \ln(\mu_i) - \sum_{i=1}^n \mu_i - \sum_{i=1}^n \ln(y_i!)$$

Untuk menduga koefisien parameter dilakukan dengan melakukan turunan parsial fungsi logaritma natural likelihood terhadap parameter yang akan diduga dan diiterasikan dengan menggunakan metode iterasi *Newton-Raphson*.

5. Menghitung nilai koefisien dispersi untuk mengidentifikasi overdispersi. Pemeriksaan overdispersi ini bisa didapat dari nilai devian atau dispersi *Pearson Chi-Square* yang dibagi dengan derajat bebasnya. Jika nilainya > 1 maka terindikasi adanya overdispersi. Jika terjadi overdispersi kemudian dilanjutkan dengan pemodelan regresi binomial negatif.

Nilai *Pearson Chi-Square* didapatkan dari persamaan berikut:

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i}$$

dengan χ^2 : nilai *Pearson Chi-Square*, y_i : nilai aktual respon amatan ke- i , dan \hat{y}_i : nilai dugaan respon amatan ke- i .

Sedangkan untuk menghitung nilai devian digunakan persamaan:

$$D = 2 \sum_{i=1}^n \left(y_i \ln \left(\frac{y_i}{\hat{y}_i} \right) - (y_i - \hat{y}_i) \right)$$

dengan D : nilai devian, y_i : nilai aktual respon amatan ke- i , \hat{y}_i : nilai dugaan respon amatan ke- i .

6. Melakukan pemodelan regresi binomial negatif dengan persamaan model:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k = \mathbf{x}'_i \boldsymbol{\beta}$$

dimana $g(\mu_i) = \ln(\mu_i) = \mathbf{x}'_i \boldsymbol{\beta}$ dan $g(\mu_i)$ merupakan fungsi hubung log. Pendugaan parameter pada model regresi binomial negatif menggunakan metode Pendugaan Kemungkinan Maksimum, yaitu dengan memaksimumkan fungsi likelihood dari distribusi binomial negatif. Fungsi likelihood dari distribusi binomial negatif adalah:

$$\ln L(y; \mu, \alpha) = \sum_{i=1}^n y_i \ln \left(\frac{\alpha \mu}{1 + \alpha \mu} \right) - \frac{1}{\alpha} \ln(1 + \alpha \mu) + \ln \Gamma \left(y_i + \frac{1}{\alpha} \right) - \ln \Gamma \left(\frac{1}{\alpha} \right)$$

Untuk menyelesaikan persamaan ini melalui iterasi menggunakan metode *Newton-Raphson* untuk mendapatkan dugaan parameter regresinya.

7. Menentukan model terbaik di antara model yang terbentuk dari regresi Poisson dan regresi Binomial Negatif dengan kriteria nilai *Akaike's Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC).

AIC dirumuskan sebagai berikut:

$AIC = 2k - 2\log(L)$, dengan k adalah banyaknya parameter penduga pada model dan L merupakan fungsi Likelihood dari model.

Sedangkan nilai BIC dihitung dengan menggunakan rumus:

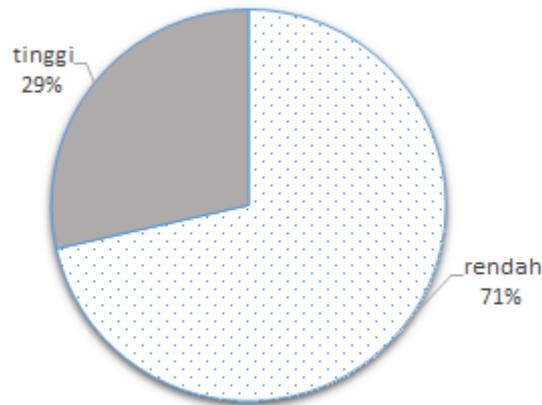
$BIC = -2\log(L) + k \ln(n)$, dengan L merupakan fungsi Likelihood dari model, k adalah banyaknya parameter dalam model, dan n adalah banyaknya observasi. Model dengan nilai AIC dan BIC lebih kecil dipilih sebagai model terbaik untuk data.

3. Hasil dan Pembahasan

3.1 Gambaran Umum Data

Demi mempermudah dalam mendeskripsikan data dilakukan pengelompokan. Kementerian Kesehatan Republik Indonesia mengelompokkan provinsi berdasarkan bebannya menjadi dua kelompok yaitu beban kusta tinggi (*high burden*) dan beban kusta rendah (*low burden*). Suatu provinsi disebut memiliki beban kusta tinggi jika jumlah kasus baru kusta di provinsi tersebut >1000 kasus. Namun jika jumlah kasus baru kusta <1000 maka disebut memiliki beban kusta rendah (PUSDATIN, 2018). Provinsi Jawa Barat, Jawa Tengah, dan Jawa Timur merupakan provinsi dengan beban kusta tinggi di Pulau Jawa, dengan jumlah kasus baru kusta masing-masing sebanyak 1994, 1918, dan 3880 kasus. Sedangkan Provinsi DKI Jakarta, DI Yogyakarta, dan Banten masuk kategori provinsi dengan beban kusta rendah.

Selain pengelompokan provinsi, pada peubah jumlah kasus kusta juga dilakukan pengelompokan kabupaten/kota. Suatu kabupaten/kota yang memiliki jumlah kasus kusta yang lebih kecil dari nilai rata-rata akan dimasukkan ke dalam kelas rendah dan sebaliknya akan dimasukkan ke kelas tinggi (jika jumlah kasus kusta lebih dari nilai rata-ratanya). Pembagian kelas rendah dan tinggi pada peubah respon (jumlah kasus baru kusta) menggunakan batas nilai rata-rata jumlah kasus baru kusta di Pulau Jawa yakni sebesar 76,59 kasus. Hasil pengelompokan ditunjukkan pada Gambar 1. Dari Gambar 1 disimpulkan bahwa sebagian besar kabupaten/kota di Pulau Jawa memiliki jumlah kasus baru kusta rendah. Bahkan ada 5 kabupaten/kota yang tidak memiliki kasus baru kusta yakni Kabupaten Kulon Progo, Kabupaten Sleman, dan Kota Yogyakarta di Provinsi DI Yogyakarta, Kota Banjar di Jawa Barat serta Kota Madiun di Jawa Timur. Pada tahun 2017, jumlah kasus baru kusta di Pulau Jawa paling banyak terjadi di Kabupaten Sumenep, Jawa Timur yakni sebanyak 463 kasus.



Gambar 1: Pengelompokan kabupaten/kota ke dalam kelas rendah dan tinggi berdasarkan jumlah baru kusta.

Pengelompokan juga dilakukan pada peubah penjelas yang akan digunakan dalam pemodelan. Setiap peubah penjelas dikelompokkan juga menjadi kelas rendah (lebih kecil dari nilai rata-rata peubah tersebut) dan kelas tinggi (jika nilainya lebih dari nilai rata-ratanya). Hasil pengelompokan delapan peubah penjelas ditunjukkan pada Gambar 2.

Sebagaimana tersaji dalam Gambar 2 tepatnya dalam batang X1, terlihat bahwa sebagian besar kabupaten/kota memiliki jumlah penduduk yang tergolong rendah. Hal ini diakibatkan penggunaan nilai rata-rata sebagai batas kelas, dimana nilai rata-rata rentan terhadap adanya pencilan. Peubah jumlah penduduk (X_1) memiliki nilai minimum sebesar 23,90 ribu orang (Kab. Kep. Seribu) dan nilai maksimum sebesar 5715 ribu orang (Kabupaten Bogor).

Untuk peubah persentase balita yang telah diimunisasi BCG (X_2), dapat dilihat dalam Gambar 2 bahwa sebagian besar balita di Pulau Jawa telah diimunisasi BCG. Nilai rata-rata peubah ini pun tergolong tinggi yaitu sebesar 92,93 persen. Diharapkan dengan tingginya persentase balita yang telah diimunisasi BCG ini akan memberikan kekebalan komunal terhadap penyakit kusta di kemudian hari, sehingga jumlah kasus kusta dapat menurun.

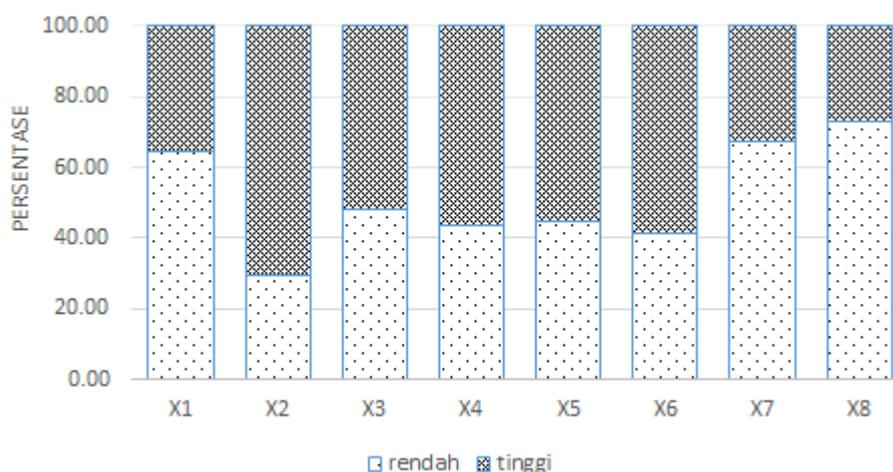
Peubah persentase rumah tangga yang berperilaku hidup bersih dan sehat (PHBS) didapatkan dari jumlah rumah tangga yang melaksanakan 10 indikator PHBS dibagi dengan rumah tangga yang dipantau. Sepuluh indikator tersebut adalah pertolongan persalinan oleh tenaga kesehatan; bayi diberi ASI eksklusif; balita ditimbang setiap bulan; menggunakan air bersih; mencuci tangan dengan air bersih dan sabun; menggunakan jamban sehat; memberantas jentik di rumah sekali seminggu; makan sayur dan buah setiap hari; melakukan aktifitas fisik setiap hari; dan tidak merokok di dalam rumah. Peubah persentase rumah tangga berperilaku hidup bersih dan sehat (X_3) menggunakan batas kelas 62,97. Pada batang X3 di Gambar 2 menampilkan kelas tinggi dan rendah hampir seimbang. Kabupaten/kota yang masuk ke kelas rendah sebanyak 47,90 persen dan kelas tinggi sebesar 52,10 persen. Persentase rumah tangga berperilaku hidup bersih dan sehat paling banyak dimiliki oleh Kota Magelang 97,25%, dan terendah dimiliki oleh Kab. Probolinggo sebesar 24,22%.

Rumah Sehat adalah bangunan rumah tinggal yang memenuhi syarat kesehatan yang terdiri dari komponen rumah, sarana sanitasi dan perilaku antara lain yaitu

memiliki jamban sehat, tempat pembuangan sampah, sarana air bersih, sarana pembuangan air limbah, ventilasi baik, kepadatan hunian rumah sesuai dan lantai rumah tidak dari tanah. Peubah persentase rumah sehat (X_4) menggunakan batas kelas sebesar 72,90. Sebagian besar kabupaten/kota di Pulau Jawa masuk dalam kategori jumlah rumah sehat yang tinggi. Kota Magelang memiliki persentase rumah sehat yang paling tinggi di Pulau Jawa, sedangkan persentase rumah sehat paling sedikit dimiliki oleh Kab. Sampang.

Salah satu target dalam tujuan pembangunan berkelanjutan (*Sustainable Development Goals/SDGs*) pada sektor lingkungan hidup adalah memastikan masyarakat mencapai akses universal air bersih dan sanitasi yang layak. Pemenuhan kebutuhan air minum di rumah tangga dapat diukur dari akses air minum layak, beberapa faktor yang berpengaruh terhadap akses air minum layak, diantaranya adalah jenis sumber air utama yang digunakan untuk diminum; jenis sumber air utama yang digunakan untuk memasak, mandi, dan mencuci; dan jarak sumber air ke penampungan limbah/kotoran/tinja terdekat ≥ 10 meter. Peubah persentase penduduk yang memiliki akses air bersih (X_5) menggunakan batas kelas sebesar 77,05. Sebagian besar kabupaten/kota digolongkan dalam kelas yang tinggi.

Definisi sanitasi menurut WHO merujuk kepada penyediaan sarana dan pelayanan pembuangan limbah kotoran manusia seperti *urine* dan *faeces*. Definisi rumah tangga yang memiliki akses sanitasi layak adalah apabila fasilitas sanitasi yang digunakan memenuhi syarat kesehatan, antara lain dilengkapi dengan jenis kloset leher angsa atau plengsengan dengan tutup dan memiliki tempat pembuangan akhir tinja tangki (*septic tank*) atau Sistem Pengolahan Air Limbah (SPAL), dan merupakan fasilitas buang air besar yang digunakan sendiri atau bersama. Persentase penduduk yang memiliki akses terhadap fasilitas sanitasi layak (jamban sehat) paling banyak berada di Kota Bekasi dan paling sedikit berada di Kabupaten Lebak. Peubah persentase penduduk dengan sanitasi layak (X_6) menggunakan batas kelas sebesar 75,15. Sebagian besar kabupaten/kota masuk dalam kategori tinggi.

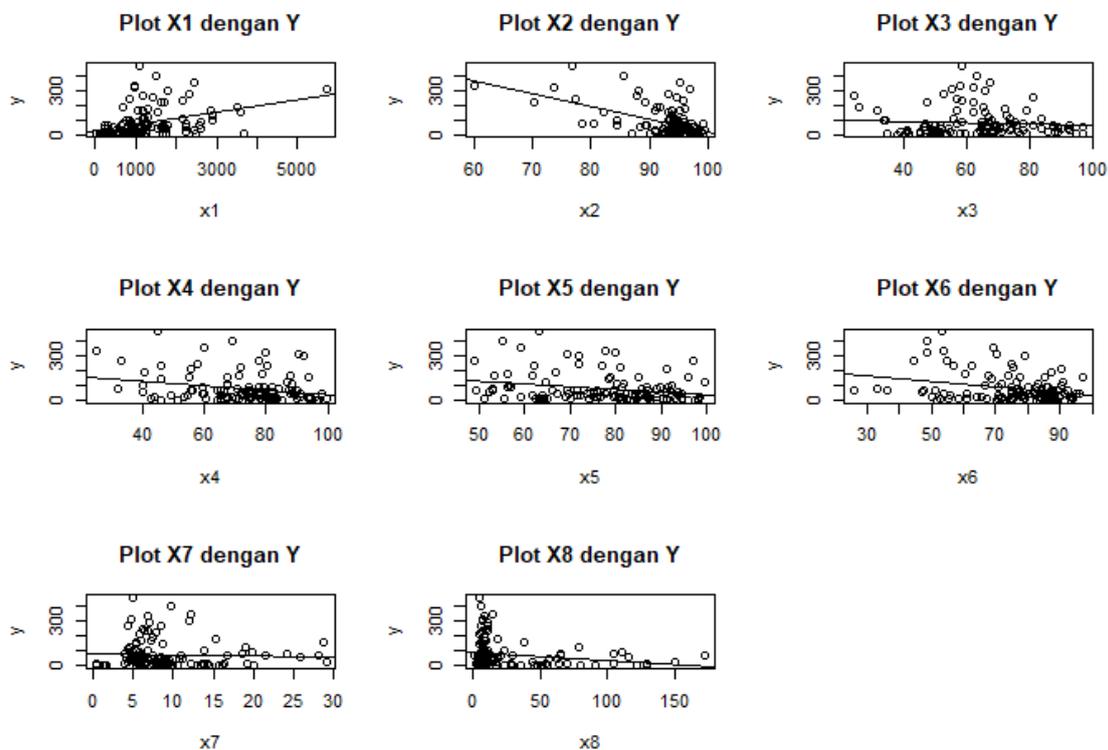


Gambar 2: Pengelompokan kabupaten/kota ke dalam kelas rendah dan tinggi menurut peubah penjelas.

Peubah rataan fasilitas kesehatan per desa (X_7) menggunakan batas kelas sebesar 9,01. Sedangkan peubah rataan tenaga kesehatan per desa (X_8)

menggunakan batas kelas sebesar 25,99. Pada Gambar 2 nampak bahwa pada batang milik X_8 dan X_9 , hanya sedikit kabupaten/kota yang masuk kategori tinggi. Wilayah yang memiliki rataan fasilitas kesehatan paling tinggi adalah Kota Jakarta Pusat dan paling kecil ada di Kabupaten Gunung Kidul. Wilayah dengan rataan tenaga kesehatan paling tinggi adalah Kota Jakarta Utara dan paling kecil ada di Kabupaten Pandeglang.

Gambar 3 menyajikan diagram pencar dari masing-masing peubah penjelas yang akan digunakan dalam pemodelan dengan peubah respon jumlah kasus baru kusta. Dengan melihat diagram tersebut dapat mempermudah dalam menganalisa bagaimana pola hubungan antara peubah respon dengan masing-masing peubah penjelas. Bentuk hubungan yang diharapkan dalam penelitian ini adalah hubungan yang linier. Selain untuk melihat pola hubungan, diagram pencar juga dapat digunakan untuk mengobservasi keberadaan nilai pencilan.

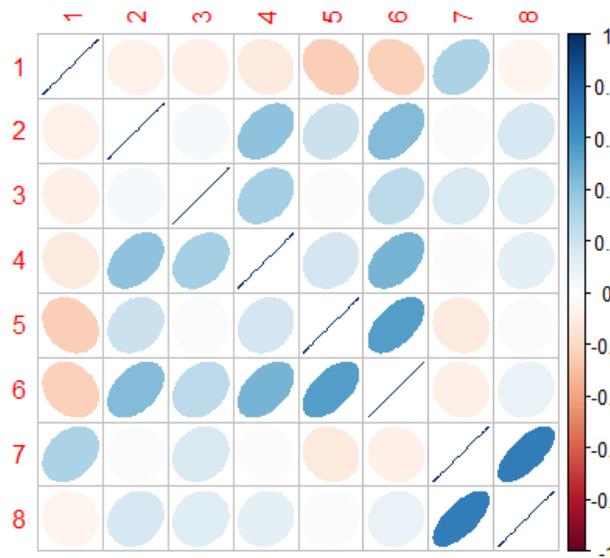


Gambar 3: Diagram pencar jumlah kasus baru kusta dan faktor-faktor yang mempengaruhinya.

3.2 Pengujian Multikolinearitas

Sebelum melakukan analisis menggunakan metode regresi, dilakukan pengujian multikolinearitas untuk mengetahui apakah di antara peubah penjelas memiliki hubungan yang tinggi. Gambar 4 menunjukkan korelasi antar peubah penjelas yang akan diterapkan dalam model. Terlihat bahwa beberapa peubah memiliki korelasi yang cukup tinggi (ditunjukkan dari ketipisan dan kegelapan warna di dalam kotak). Namun tidak ada yang berbentuk sangat tipis seperti garis miring (nilai koefisien korelasi mendekati 1), yang menandakan di antara kesembilan peubah penjelas yang akan digunakan dalam model tidak ada yang memiliki korelasi yang sangat tinggi.

Untuk mendeteksi adanya multikolinearitas selain dengan menggunakan nilai koefisien korelasi, dapat digunakan pula kriteria nilai VIF. Multikolinearitas menjadi masalah yang serius jika nilai VIF yang diperoleh lebih besar dari sepuluh ($VIF > 10$). Tabel 2 menampilkan nilai VIF setiap peubah penjelasnya. Berdasarkan nilai dalam Tabel 2 dapat disimpulkan bahwa tidak ada pengaruh besar multikolinearitas karena semua nilai VIF dari peubah penjelas X_1 sampai X_8 kurang dari sepuluh ($VIF < 10$). Dengan demikian dalam penelitian ini seluruh peubah penjelas dapat digunakan dalam pemodelan regresi Poisson dan regresi binomial negatif.



Gambar 4: Diagram korelasi antar peubah penjelas.

Tabel 2: Nilai VIF peubah penjelas.

Peubah	VIF
X_1	1,397
X_2	1,383
X_3	1,268
X_4	1,482
X_5	1,530
X_6	2,078
X_7	2,700
X_8	2,424

3.3 Model Regresi Poisson

Data jumlah kasus baru kusta diasumsikan berdistribusi Poisson karena merupakan data cacah. Model regresi Poisson dengan sembilan peubah penjelas menghasilkan dugaan parameter yang disajikan dalam Tabel 3. Persamaan model regresi Poisson yang diperoleh adalah sebagai berikut:

$$\eta_i = 10,360 + 4,021 \times 10^{-4}x_1 - 5,687 \times 10^{-2}x_2 + 2,274 \times 10^{-3}x_3 - 4,924 \times 10^{-3}x_4 - 1,759 \times 10^{-2}x_5 + 5,019 \times 10^{-3}x_6 - 1,664 \times 10^{-2}x_7 - 4,450 \times 10^{-3}x_8 \quad (1)$$

Setelah didapatkan nilai estimasi parameter model regresi Poisson maka selanjutnya dilakukan pengujian parameter secara serentak dan parsial. Pengujian

parameter model regresi Poisson secara serentak bertujuan untuk menguji apakah peubah penjelas secara bersama-sama memiliki pengaruh signifikan terhadap peubah respon. Dari hasil pengolahan, nilai *deviance* yang diperoleh sebesar 6280. Taraf signifikansi yang digunakan yaitu 0,05 sehingga nilai $\chi^2_{(0,05;9)}$ sebesar 16,91898. Hal ini menunjukkan bahwa nilai *deviance* lebih besar dari nilai $\chi^2_{(0,05;9)}$, sehingga diperoleh kesimpulan bahwa minimal terdapat satu peubah penjelas yang berpengaruh signifikan terhadap peubah respon.

Tabel 3: Nilai dugaan parameter model Regresi Poisson.

Parameter	Nilai dugaan	Galat baku	Zhitung	Nilai p
β_0 (Intersep)	10,360	$1,340 \times 10^{-1}$	77,328	$< 2 \times 10^{-16}^*$
β_1	$4,021 \times 10^{-4}$	$1,077 \times 10^{-5}$	37,343	$< 2 \times 10^{-16}^*$
β_2	$-5,687 \times 10^{-2}$	$1,505 \times 10^{-3}$	-37,788	$< 2 \times 10^{-16}^*$
β_3	$2,274 \times 10^{-3}$	$8,147 \times 10^{-4}$	2,792	0,005*
β_4	$-4,924 \times 10^{-3}$	$7,668 \times 10^{-4}$	-6,422	$1,340 \times 10^{-10}^*$
β_5	$-1,759 \times 10^{-2}$	$1,081 \times 10^{-3}$	-16,264	$< 2 \times 10^{-16}^*$
β_6	$5,019 \times 10^{-3}$	$9,894 \times 10^{-4}$	5,073	$3,920 \times 10^{-7}^*$
β_7	$-1,664 \times 10^{-2}$	$3,324 \times 10^{-3}$	-5,006	$5,570 \times 10^{-7}^*$
β_8	$-4,450 \times 10^{-3}$	$5,846 \times 10^{-4}$	-7,612	$2,710 \times 10^{-14}^*$
β_9	10,360	$1,340 \times 10^{-1}$	77,328	$< 2 \times 10^{-16}^*$

Keterangan: *signifikan pada taraf nyata 5%

Selanjutnya dilanjutkan pada pengujian secara parsial untuk mengetahui peubah mana saja yang terbukti secara nyata mempengaruhi jumlah kasus baru kusta. Pengujian parameter secara parsial dapat dilihat dari nilai p. Tabel 3 menunjukkan bahwa seluruh peubah bebas $X_1, X_2, X_3, X_4, X_5, X_6, X_7$, dan X_8 memiliki nilai p yang lebih kecil dari $\alpha = 0.05$, artinya seluruh peubah bebas $X_1, X_2, X_3, X_4, X_5, X_6, X_7$, dan X_8 memiliki pengaruh yang signifikan terhadap jumlah kasus baru kusta.

Untuk menginterpretasikan model regresi Poisson yang diperoleh digunakanlah nilai rasio odd dari masing-masing koefisien, yang dijabarkan sebagai berikut:

- $\hat{\beta}_1 = 4,021 \times 10^{-4}$; artinya untuk setiap penambahan seribu orang penduduk akan meningkatkan rata-rata jumlah kasus baru kusta sebesar $(e^{4,021 \times 10^{-4}} - 1) \times 100\% = 0,04021\%$ dengan asumsi peubah lain tetap.
- $\hat{\beta}_2 = -5,687 \times 10^{-2}$; peubah persentase balita yang diimunisasi BCG memiliki koefien regresinya bertanda negatif maka untuk setiap penambahan persentase balita yang diimunisasi BCG sebesar 1% akan menurunkan rata-rata jumlah kasus baru kusta sebesar $(1 - e^{-5,687 \times 10^{-2}}) \times 100\% = 5,52831\%$ dengan asumsi peubah lain tetap.
- $\hat{\beta}_3 = 2,274 \times 10^{-3}$; artinya untuk setiap penambahan persentase rumah tangga berperilaku hidup bersih dan sehat sebesar 1% akan meningkatkan rata-rata jumlah kasus baru kusta sebesar $(e^{2,274 \times 10^{-2}} - 1) \times 100\% = 2,30005\%$ dengan asumsi peubah lain tetap.
- $\hat{\beta}_4 = -4,924 \times 10^{-3}$; artinya untuk setiap penambahan persentase rumah sehat sebesar 1% akan mengakibatkan penurunan rata-rata jumlah kasus baru kusta sebesar $(1 - e^{-4,924 \times 10^{-3}}) \times 100\% = 0,49119\%$ dengan asumsi peubah lain tetap.

- $\hat{\beta}_5 = -1,759 \times 10^{-2}$; artinya untuk setiap penambahan persentase penduduk dengan akses air bersih berkelanjutan sebesar 1% akan mengakibatkan rata-rata jumlah kasus baru kusta menurun sebesar $(1 - e^{-4,924 \times 10^{-3}}) \times 100\% = 0,49119\%$ dengan asumsi peubah lain tetap.
- $\hat{\beta}_6 = 5,019 \times 10^{-3}$; artinya untuk setiap penambahan persentase penduduk dengan sanitasi layak sebesar 1% akan meningkatkan rata-rata jumlah kasus baru kusta sebesar $(e^{5,019 \times 10^{-3}} - 1) \times 100\% = 0,50316\%$ dengan asumsi peubah lain tetap.
- $\hat{\beta}_7 = -1,664 \times 10^{-2}$; artinya untuk setiap penambahan rata-rata fasilitas kesehatan sebanyak 1 unit per desa akan mengakibatkan penurunan rata-rata jumlah kasus baru kusta sebesar $(1 - e^{-1,664 \times 10^{-2}}) \times 100\% = 1,65023\%$ dengan asumsi peubah lain tetap.
- $\hat{\beta}_8 = -4,450 \times 10^{-3}$; artinya untuk setiap penambahan rata-rata tenaga kesehatan sebanyak 1 orang per desa akan mengakibatkan rata-rata jumlah kasus baru kusta menurun sebesar $(1 - e^{-4,450 \times 10^{-3}}) \times 100\% = 0,44401\%$ dengan asumsi peubah lain tetap.

3.4 Pemeriksaan Overdispersi

Salah satu asumsi yang harus terpenuhi dalam regresi Poisson adalah kondisi nilai tengah sama dengan ragam yang disebut ekuidispersi. Namun dari hasil yang pengolahan nilai rata-rata jumlah kasus baru kusta sebesar 76,59 tidak sama dengan nilai ragamnya yaitu sebesar 9029,261, bahkan nilai ragamnya jauh lebih besar, hal ini menunjukkan adanya overdispersi pada data. Di samping itu besaran nilai uji dispersi = 54,71033 yang lebih dari 1 maka disimpulkan adanya kondisi overdispersi. Nilai *residual deviance* sebesar 6280 dengan derajat bebas 110 sehingga rasio nilai *deviance* dengan derajat bebasnya bernilai 57,61468. Nilai tersebut lebih besar dari 1 yang artinya data jumlah kasus baru kusta mengalami overdispersi. Dengan demikian model regresi Poisson tidak sesuai untuk diterapkan karena menghasilkan dugaan parameter yang tidak efisien.

3.5 Model Regresi Binomial Negatif

Model regresi yang diharapkan dapat mengatasi masalah overdispersi yang terjadi adalah model regresi binomial negatif. Pada Tabel 4 ditampilkan hasil pendugaan parameter model regresi dengan sembilan peubah penjelas. Bentuk persamaan model regresi binomial negatif yang diperoleh adalah:

$$\eta = 11,313 + 0,001 x_1 - 0,076 x_2 + 0,009 x_3 - 0,013 x_4 - 0,019 x_5 + 0,010 x_6 - 0,013 x_7 - 0,007 x_8$$

Pengujian serentak signifikansi parameter model regresi binomial negatif bertujuan untuk mengetahui apakah peubah penjelas secara bersama-sama memiliki pengaruh signifikan terhadap respon. Nilai *deviance* yang diperoleh dari pengolahan adalah sebesar 137,97 sehingga kesimpulannya adalah paling tidak ada satu peubah penjelas yang berpengaruh signifikan terhadap jumlah kasus baru kusta.

Selanjutnya dilanjutkan pada pengujian secara parsial untuk mengetahui peubah mana saja yang terbukti secara nyata mempengaruhi jumlah kasus baru kusta. Hasil pengujian parameter secara parsial dapat dilihat dari nilai p. Dari Tabel 4 diketahui bahwa dari sembilan peubah penjelas, terdapat enam peubah (yaitu $X_1, X_2, X_3, X_6, X_7,$

dan X_8) yang memiliki nilai p yang lebih kecil dari $\alpha = 0,05$ yang berarti keenam peubah penjelas memiliki pengaruh yang signifikan terhadap peubah respon. Sedangkan tiga peubah penjelas yang lain yakni X_4 , X_5 , dan X_9 memiliki nilai p yang lebih besar dari α , sehingga disimpulkan bahwa data yang digunakan dalam penelitian ini belum memiliki cukup bukti untuk menyatakan ketiga peubah tersebut memberi pengaruh yang nyata terhadap jumlah kasus baru kusta.

Tabel 4: Nilai dugaan parameter model regresi binomial negatif.

Parameter	Nilai dugaan	Galat baku	Z_{hitung}	Nilai p
β_0 (Intersep)	11,313	1,536	7,364	$1,790 \times 10^{-13}^*$
β_1	0,001	0,0001	6,030	$1,640 \times 10^{-9}^*$
β_2	-0,076	0,017	-4,371	$1,230 \times 10^{-5}^*$
β_3	0,009	0,006	1,410	0,159
β_4	-0,013	0,007	-1,905	0,057
β_5	-0,019	0,008	-1,966	0,049*
β_6	0,010	0,008	1,218	0,223
β_7	-0,013	0,026	-0,505	0,614
β_8	-0,007	0,004	-1,789	0,074

Keterangan: *signifikan pada taraf nyata 5%.

Peubah yang signifikan pada taraf nyata 5% yaitu jumlah penduduk (X_1), persentase balita yang diimunisasi BCG (X_2), dan persentase penduduk dengan akses air bersih berkelanjutan (X_5). Interpretasi yang diperoleh berdasarkan persamaan model yang diperoleh yaitu:

- $\hat{\beta}_1 = 0,001$, artinya untuk setiap penambahan jumlah penduduk sebesar seribu orang akan meningkatkan rata-rata jumlah kasus baru kusta sebesar $(e^{0,001} - 1) \times 100\% = 0,10005\%$ dengan asumsi peubah lain tetap.
- $\hat{\beta}_2 = -0,076$, artinya untuk setiap penambahan persentase balita yang diimunisasi BCG sebesar 1% akan menurunkan rata-rata jumlah kasus baru kusta sebesar $(1 - e^{-0,076}) \times 100\% = 7,31838\%$ dengan asumsi peubah lain tetap.
- $\hat{\beta}_5 = -0,019$, artinya untuk setiap penambahan persentase penduduk dengan akses air bersih berkelanjutan sebesar 1% akan menurunkan rata-rata jumlah kasus baru kusta sebesar $(1 - e^{-0,019}) \times 100\% = 1,88206\%$ dengan asumsi peubah lain tetap.

Nilai *residual deviance* sebesar 137,97 dengan derajat bebas 110 sehingga rasio nilai *deviance* dengan derajat bebasnya bernilai 1,25427. Nilai ini mendekati nilai 1 dan jauh lebih kecil dibandingkan dengan nilai rasio dispersi dari model regresi Poisson yang sebesar 57,09091. Hal ini menunjukkan bahwa model regresi binomial negatif telah mampu mengatasi overdispersi pada model regresi Poisson.

3.5 Pemilihan Model Terbaik

Model terbaik adalah model yang memiliki nilai AIC dan BIC yang lebih kecil. Berdasarkan Tabel 5 dapat dilihat bahwa model regresi binomial negatif memiliki nilai AIC dan BIC yang lebih kecil dibandingkan dengan model regresi Poisson. Oleh karena itu disimpulkan bahwa model regresi binomial negatif adalah model terbaik.

Tabel 5: Kriteria pemilihan model terbaik.

Kriteria	Regresi Poisson	Regresi Binomial Negatif
AIC	6280,000	1212,200
BIC	6313,656	1249,483

4. Simpulan

Untuk menganalisis data jumlah kasus baru kusta yang merupakan data cacah digunakan pemodelan regresi Poisson. Namun pemodelan regresi Poisson pada jumlah kasus baru kusta mengalami overdispersi. Hal tersebut menjadikan model menjadi tidak baik. Pemodelan regresi binomial negatif membuat hasil menjadi lebih baik karena nilai rasio dispersi mendekati satu. Dengan demikian model regresi binomial negatif mampu mengatasi overdispersi yang terjadi pada model regresi Poisson. Peubah yang mempunyai pengaruh signifikan terhadap jumlah kasus baru kusta berdasarkan hasil pemodelan regresi binomial negatif yaitu jumlah penduduk (X_1), persentase balita yang diimunisasi BCG (X_2), dan persentase penduduk dengan akses air bersih berkelanjutan (X_5).

Daftar Pustaka

- Aini, A. N. (2013). Regresi Spasial dengan Pendekatan Geographically Weighted Poisson Regression (GWPR): Studi Kasus Banyak Penderita Kusta Kering Tahun 2012 di Beberapa Kota dan Kabupaten di Provinsi Jawa Barat [skripsi]. Bandung (ID): Universitas Pendidikan Indonesia.
- Arfidina, K. I., Susilawati, M., & Srinadi, I. G. A. M. (2017). Regresi Poisson dalam Memodelkan Jumlah Penderita Kusta di Provinsi Bali. Prosiding SEMINAR NASIONAL SAINSTEK 2017, 49–54. Badung (ID): Fakultas MIPA, Univesitas Udayana.
- Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data (Vol. 53). Cambridge (UK): Cambridge university press.
- Dobson, A. J., & Barnett, A. G. (2002). An introduction to generalized linear models. Boca Raton (US): CRC press.
- Dzikrina, A. M., & Purnami, S. W. (2013). Pemodelan angka prevalensi kusta dan faktor-faktor yang mempengaruhi di Jawa Timur dengan pendekatan Geographically Weighted Regression (GWR). Jurnal Sains Dan Seni ITS, 2(2): D275–D281.
- Giuffrè, O., Granà, A., Roberta, M., & Corriere, F. (2011). Handling underdispersion in calibrating safety performance function at urban, four-leg, signalized intersections. Journal of Transportation Safety & Security, 3(3): 174–188.
- Mann, J., Larsen, P., & Brinkley, J. (2014). Exploring the use of negative binomial regression modeling for pediatric peripheral intravenous catheterization. J Med Stat Inform, 2(6): 1–8. <https://doi.org/10.7243/2053-7662-2-6>
- [PUSDATIN] Pusat Data dan Informasi Kementerian Kesehatan. (2018). Hapuskan Stigma dan Diskriminasi Terhadap Kusta. Jakarta (ID): Kementerian Kesehatan RI.
- Shovalina, M. R., & Atok, R. M. (2016). Pemodelan dan Pemetaan Prevalensi Kusta di Kabupaten/Kota Jawa Timur dengan Pendekatan Mixed Geographically Weighted Regression. Jurnal Sains Dan Seni ITS, 5(2): D396–D401.