

Perbandingan Analisis Generalized Poisson Regression (GPR) dan Regresi Binomial Negatif Untuk Mengatasi Overdispersi

Studi Kasus: Pemodelan Jumlah Kasus Kanker Serviks di Jawa Timur

Brianika Irawati

Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS)

e-mail: brianika11@gmail.com

Purhadi

Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS)

e-mail: purhadi@statistika.its.ac.id

Abstract: Data jumlah kasus kanker serviks di Provinsi Jawa Timur pada penelitian ini merupakan salah satu contoh data diskrit (*count*) dimana pada umumnya menggunakan regresi poisson dalam analisisnya. Karakteristik penting dalam pemodelan ini yaitu mean harus sama dengan varians. Namun, kondisi seperti ini sulit dipenuhi sedangkan pada umumnya sering ditemui data diskrit dengan varians lebih besar dibandingkan dengan rata-ratanya atau disebut dengan overdispersi seperti yang terjadi pada kasus jumlah kanker serviks di Jawa timur ini. Untuk menangani masalah overdispersi, dapat dilakukan pemodelan dengan Generalized Poisson Regression (GPR) dan Regresi Binomial Negatif. Model ini dapat mengatasi masalah overdispersi karena tidak mengharuskan nilai mean yang sama dengan nilai varians seperti pada model regresi poisson. Model terbaik menggunakan GPR menghasilkan 8 variabel prediktor yang signifikan mempengaruhi jumlah kasus kanker serviks di Jawa Timur antara lain persentase tenaga medis (X_2), persentase penduduk perempuan yang umur kawin pertama ≤ 16 tahun (X_3), persentase penduduk miskin (X_7), persentase penduduk perempuan yang menggunakan kondom (X_8), persentase penduduk perempuan yang berstatus kawin (X_{10}), persentase penduduk perempuan usia ≥ 35 tahun (X_{11}), persentase rata-rata pengeluaran untuk konsumsi makanan per bulan (X_{12}), dan persentase penduduk yang merokok (X_{13}). Sedangkan model terbaik menggunakan regresi binomial negatif menghasilkan 2 variabel prediktor yang signifikan yaitu persentase penduduk yang tinggal di daerah berstatus desa (X_5) dan persentase rata-rata pengeluaran untuk konsumsi makanan per bulan (X_{12}). Model GPR menghasilkan nilai AIC sebesar 349,90. Sedangkan model regresi binomial negatif menghasilkan nilai AIC sebesar 327,65. Maka model terbaik diperoleh dari model regresi binomial negatif karena menghasilkan nilai AIC terkecil.

Keywords: AIC, Generalized Poisson Regression, Kanker Serviks, Regresi Binomial Negatif.

1. Pendahuluan

Regresi poisson merupakan analisis regresi nonlinier dari distribusi poisson, dimana analisis ini sangat cocok digunakan dalam menganalisis data diskrit (*count*) jika mean

proses sama dengan variansnya. Dalam kenyataannya, asumsi mean sama dengan variansnya (*equidispersion*) jarang terpenuhi sedangkan pada umumnya sering ditemui data diskrit dengan varians lebih besar dibandingkan mean (*overdispersi*) seperti yang terjadi pada data kasus jumlah kanker serviks di Jawa timur dalam penelitian ini. Analisis yang sesuai dalam menangani kasus over dispersi ini adalah *Generalized Poisson Regression* (GPR) dan regresi binomial negatif. Kajian yang digunakan dalam penelitian ini adalah membandingkan model terbaik GPR dan regresi binomial negatif serta mengetahui faktor-faktor terhadap jumlah kasus kanker serviks di Jawa Timur.

Dewi [5] melakukan penerapan Regresi Binomial Negatif pada kasus overdispersi pada regresi poisson studi kasus ketidاكلulusan siswa SMA dalam ujian nasional di DKI Jakarta. Hasil yang diperoleh menunjukkan bahwa model Regresi Binomial Negatif merupakan salah satu solusi yang dapat digunakan untuk mengatasi permasalahan dalam data yang mengandung overdispersi. Selanjutnya, Listiani [9] melakukan penerapan analisis regresi poisson dan GPR pada pemodelan faktor-faktor yang mempengaruhi angka kematian bayi di Provinsi Jawa Timur tahun 2007. Hasil penelitian menunjukkan bahwa regresi poisson tidak dapat diaplikasikan untuk pemodelan faktor-faktor yang mempengaruhi angka kematian bayi di Provinsi Jawa Timur tahun 2007 karena variabel respon mengalami overdispersi sehingga digunakan model GPR.

Kanker Serviks (*Cervical Cancer*), atau kanker leher rahim adalah kanker yang terjadi pada serviks uterus, yaitu bagian organ reproduksi wanita yang merupakan pintu masuk ke dalam rahim yang terletak antara rahim (uterus) dengan liang senggama. Menurut WHO pada tahun 2010 terdapat 500000 kasus kanker serviks di dunia. Indonesia merupakan negara dengan kasus kanker serviks terbanyak kedua setelah negara Cina yaitu sebesar 15000 kasus kanker serviks. Provinsi Jawa Timur merupakan salah satu provinsi penyumbang terbesar jumlah kasus kanker serviks di Indonesia. Jumlah kasus kanker serviks di Jawa Timur pada tahun 2010 adalah 1688 atau sebesar 11,25% dari kasus kanker serviks di Indonesia. Penelitian tentang kanker serviks yaitu Hartini (2007) meneliti faktor-faktor apa saja yang mempengaruhi terjadinya kasus kanker serviks. Hasil yang diperoleh adalah faktor-faktor yang signifikan mempengaruhi kanker serviks antara lain usia, penggunaan kontrasepsi, frekuensi melahirkan, usia menstruasi pertama, dan riwayat keguguran.

Penelitian ini dilakukan untuk mendapatkan model terbaik dengan membandingkan model yang diperoleh dari hasil analisis GPR dan Regresi Binomial Negatif pada studi kasus jumlah kanker serviks tiap kabupaten/kota di Provinsi Jawa Timur tahun 2010. Hasil penelitian ini juga diharapkan dapat memberikan informasi tambahan tentang faktor-faktor apa saja yang secara signifikan mempengaruhi terjadinya kanker serviks.

2. Tinjauan Pustaka

Tinjauan yang digunakan terdiri atas tinjauan kasus yang mendeskripsikan tentang penyakit kanker serviks dan tinjauan statistika yang meliputi teori-teori yang digunakan dalam analisis.

2.1. Statistika Deskriptif

Statistika deskriptif adalah analisis statistika yang membahas tentang metode-metode untuk menyajikan data sehingga menarik dan informatif. Secara umum statistika deskriptif dapat diartikan sebagai metode-metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga memberikan informasi yang berguna. Mean (Rataan) didefinisikan sebagai jumlah seluruh data dibagi dengan banyaknya data (Walpole, [14]).

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (1)$$

Varians adalah suatu nilai yang menunjukkan ukuran variabilitas yang dihitung dengan cara mengkuadratkan standar deviasi (Walpole, [14]). Rumus yang digunakan untuk varians ditunjukkan pada persamaan (2) sebagai berikut.

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} \quad (2)$$

2.2. Regresi Poisson

Regresi poisson merupakan analisis regresi nonlinier dari distribusi poisson, dimana analisis ini sangat cocok digunakan dalam menganalisis data diskrit (*count*). Model regresi poisson merupakan *Generalized Linier Model* (GLM) yang data respon diasumsikan berdistribusi poisson. Model regresi poisson diberikan sebagai berikut.

$$y_i = \text{Poisson}(\mu_i)$$

$$\mu_i = \exp(x_i^T \beta)$$

maka

$$\ln(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \quad (3)$$

Estimasi parameter model regresi poisson menggunakan metode *Maximum Likelihood Estimator*. Fungsi log-likelihood poisson sebagai berikut.

$$\ln L(\beta) = - \sum_{i=1}^n \exp(x_i^T \beta) + \sum_{i=1}^n y_i x_i^T \beta - \sum_{i=1}^n \ln(y_i!) \quad (4)$$

Untuk memperoleh nilai taksiran β maka persamaan (3) diturunkan terhadap β dan disamadengankan nol menggunakan metode newton raphson.

Salah satu metode yang digunakan untuk menentukan statistik uji dalam pengujian parameter model regresi poisson adalah dengan menggunakan metode *Maximum Likelihood Ratio Test* (MLRT) dengan hipotesis.

$$H_0 : \beta_1 = \beta_2 = \dots - \beta_k = 0$$

$$H_1 : \text{paling sedikit ada satu } \beta_i \neq 0; i = 1, 2, \dots, k$$

Statistik uji untuk kelayakan model regresi poisson adalah sebagai berikut.

$$D(\hat{\beta}) = -2 \ln \left[\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right] = 2[\ln(L(\hat{\Omega})) - \ln(L(\hat{\omega}))] \quad (5)$$

Keputusan yang akan diambil adalah tolak H_0 jika $D(\hat{\beta}) > \chi_{v;\alpha}^2$ dengan v adalah banyaknya parameter model dibawah populasi dikurangi dengan banyaknya parameter dibawah H_0 . Parameter model regresi poisson yang telah dihasilkan dari estimasi parameter belum tentu mempunyai pengaruh yang signifikan terhadap model. Untuk itu perlu dilakukan pengujian terhadap parameter model regresi poisson secara individu. Dengan menggunakan hipotesis sebagai berikut.

$H_0 : \beta_i = 0$ (pengaruh variabel ke-i tidak signifikan)

$H_0 : \beta_i \neq 0$ (pengaruh variable ke-i signifikan)

Statistik uji yang digunakan adalah:

$$z = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)} \quad (6)$$

dengan $se(\hat{\beta}_i)$ adalah nilai *standar error* atau tingkat kesalahan dari parameter β_i . Keputusan yang akan diambil adalah tolak H_0 jika $|z_{hit}| > z_{\alpha/2}$ dimana adalah tingkat signifikansi.

Regresi poisson dikatakan mengandung overdispersi apabila nilai variansnya lebih besar dari nilai meannya. Overdispersi memiliki dampak yang sama dengan pelanggaran asumsi jika pada data diskrit terjadi overdispersi namun tetap digunakan regresi poisson, anak dugaan dari parameter koefisien regresinya tetap konsisten namun tidak efisien. Hal ini berdampak pada nilai *standar error* yang menjadi *under estimate*, sehingga kesimpulannya menjadi tidak valid. Fenomena overdispersi (McCullagh dan Nelder [11]) dapat dituliskan $var(Y) > E(Y)$.

2.3. Generalized Regression Poisson (GPR)

Penanganan pelanggaran asumsi equidispersi pada regresi poisson dilakukan pengembangan model menggunakan GPR. Pada model GPR selain terdapat parameter juga terdapat θ sebagai parameter dispersi. Model GPR mirip dengan regresi poisson yaitu pada persamaan (4) akan tetapi model GPR mengasumsikan bahwa komponen randomnya berdistribusi *general poisson*. Dalam analisis GPR, jika θ sama dengan 0 maka model GPR akan menjadi model poisson. Jika θ lebih dari 0 maka model GPR merepresentasikan data count yang mengandung kasus overdispersi dan jika θ kurang dari 0 merepresentasikan data *count* yang mengandung underdispersi.

Penaksiran parameter model GPR menggunakan metode *Maximum Likelihood Estimator* (MLE). Fungsi log-likelihood untuk model GPR adalah.

$$\ln L(\beta, \theta) = \sum_{i=1}^n \left\{ y_i (X_i^T \beta) - y_i \ln(1 + \theta \exp(x_i^T \beta)) + (y_i - 1) \ln(1 + \theta y_i) - \ln(y_i!) - \exp(x_i^T \beta) (1 + \theta y_i) (1 + \theta \exp(x_i^T \beta))^{-1} \right\} \quad (7)$$

Untuk mendapatkan taksiran parameter β dan θ maka persamaan (7) diturunkan terhadap β dan θ menggunakan metode numerik, iterasi Newton-Raphson.

Pengujian parameter model GPR dilakukan sama seperti regresi poisson dengan menggunakan metode MLRT dan uji parsial menggunakan statistik uji z .

2.4. Regresi Binomial Negatif

Selain GPR, penanganan overdispersi pada regresi poisson juga dapat dilakukan menggunakan pendekatan model binomial negatif. Dalam regresi binomial negatif, jika θ menuju nom maka $\text{var}(Y_i)$ menuju μ_i sehingga binomial negatif akan konvergen menuju poisson. Model regresi binomial negatif memiliki bentuk yang sama dengan model regresi poisson yaitu persamaan (4).

Penaksiran parameter regresi binomial negatif dilakukan menggunakan metode *Maximum Likelihood Estimator*. Persamaan log-likelihood untuk binomial negatif adalah.

$$\ln L(\theta, \beta) = \sum_{i=1}^n \left\{ \left(\sum_{j=0}^{y_i-1} \ln(y_j + \theta^{-1}) \right) - \ln y_i! - (y_i + \theta^{-1}) \times \right. \\ \left. \ln(1 + \theta \exp(x_i^T \beta)) + y_i \ln \theta + y_i x_i^T \beta \right\} \quad (8)$$

Estimasi parameter $(\hat{\theta}, \hat{\beta})$ diperoleh dengan menurunkan persamaan (8) terhadap β dan θ . Pengujian parameter yang dilakukan sama dengan pengujian pada regresi poisson. Uji serentak menggunakan statistik uji $D(\theta)$ dan statistik uji parsial menggunakan uji z .

2.5. Kanker Serviks

Kanker Serviks (*Cervical Cancer*), atau kanker leher rahim adalah kanker yang terjadi pada servik uterus, yaitu bagian organ reproduksi wanita yang merupakan pintu masuk ke dalam rahim yang terletak antara rahim (uterus) dengan liang senggama/vagina ([13]). Gejala paling umum dari kanker serviks adalah pendarahan abnormal dari vagina atau flek (bercak) vagina. Pendarahan abnormal ini terutama terjadi setelah berhubungan seksual, namun dapat muncul juga perdarahan diantara 2 siklus menstruasi, menoragia, atau bercak/perdarahan postmeno-pause. Bila perdarahan berlangsung dalam jangka waktu lama maka pasien akan mengeluh lelah dan lemas karena anemia yang dialaminya. Bercak kekuningan yang encer diikuti dengan bau amis dapat merupakan tanda-tanda keganasan. Gejala biasanya baru muncul ketika sel yang abnormal berubah menjadi keganasan dan menyusup ke jaringan sekitar. Kanker serviks dapat menyebar ke berbagai macam organ. Diantaranya ke kelenjar getah bening, vagina, kandung kemih, rektum, endometrium (selaput dinding rahim), dan ovarium (indung telur). Masing-masing mem-berikan gejala yang berbeda-beda. Penyebaran kanker ser-viks pada umumnya melalui peredaran kelenjar getah bening, peredaran melalui darah jarang terjadi.

TABLE 1
Variabel Penelitian

Variabel	Keterangan	Skala Pengukuran
Y	Jumlah kasus kanker serviks	Rasio
x_1	Persentase sarana kesehatan	Rasio
x_2	Persentase tenaga medis.	Rasio
x_3	Persentase penduduk perempuan yang umur kawin pertama ≤ 16 tahun	Rasio
x_4	Persentase penduduk dan Rumah Tangga (RT) perempuan.	Rasio
x_5	Persentase penduduk yang tinggal di daerah berstatus desa.	Rasio
x_6	Persentase penduduk perempuan yang tamat SMA.	Rasio
x_7	Persentase penduduk miskin.	Rasio
x_8	Persentase penduduk perempuan yang menggunakan kondom.	Rasio
x_9	Persentase perempuan dengan jumlah anak yang dilahirkan lebih dari 4.	Rasio
x_{10}	Persentase penduduk perempuan yang berstatus kawin.	Rasio
x_{11}	Persentase penduduk perempuan usia ≥ 35	Rasio
x_{12}	Persentase rata-rata pengeluaran untuk konsumsi makanan per bulan	Rasio
x_{13}	Persentase penduduk yang merokok.	Rasio

3. Metodologi Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yaitu data variabel respon (jumlah kasus kanker serviks tiap kabupaten/kota di Provinsi Jawa Timur tahun 2010). diperoleh dari Dinas Kesehatan Provinsi Jawa Timur, sedangkan data variabel prediktor (X) diperoleh dari hasil Survei Sosial Ekonomi Nasional (SUSENAS) BPS Jawa Timur tahun 2010.

Langkah-langkah analisis dalam penelitian ini adalah sebagai berikut.

1. Menganalisis statistika deskriptif untuk variabel respon dan variabel prediktor.
2. Menganalisis korelasi antar variabel-variabel prediktor untuk mendeteksi adanya kasus multikolinieritas.
3. Mendapatkan model terbaik untuk regresi poisson.
4. Mendeteksi adanya kasus overdispersi pada data dengan melihat nilai *Pearson Chi-squares* dan *Deviance* yang dibagi derajat bebasnya.
5. Mendapatkan model terbaik untuk GPR dengan menaksir parameter GPR, menguji signifikansi parameter secara serentak dan parsial serta menghitung nilai AIC dari model GPR.
6. Mendapatkan model terbaik untuk regresi binomial negatif dengan menaksir parameter, menguji signifikansi parameter secara serentak dan parsial serta menghitung nilai AIC.
7. Membandingkan model terbaik hasil GPR dan regresi binomial negatif menggunakan nilai AIC.

4. Analisis dan Pembahasan

4.1. Statistika Deskriptif

Rata-rata dari jumlah kasus kanker serviks tiap kabupaten/kota di Jawa Timur adalah 46 kasus. Kota Malang merupakan kota yang memiliki jumlah kasus kanker serviks tertinggi yaitu sebesar 479 kasus pada tahun 2010, sedangkan daerah yang tidak ditemukan

terjadi kasus kanker serviks diantaranya adalah kabupaten Pacitan, Jember, Situbondo, Pasuruan, Mojokerto, Jombang, Nganjuk, Tuban, Pamekasan, dan kota Probolinggo. Nilai varians dari variabel respon yaitu 7493,52. Hal ini menunjukkan varians kasus kanker serviks di Jawa Timur sangat besar karena terdapat daerah dengan jumlah kasus kanker serviks sampai ratusan namun ada daerah yang tidak ditemukan adanya kasus kanker serviks.

Nilai varians tertinggi dari beberapa variabel prediktor yang diduga mempengaruhi jumlah kasus kanker serviks di Jawa Timur terdapat pada variabel X_5 (persentase penduduk yang tinggal di daerah berstatus desa) yaitu sebesar 1064,29 dengan nilai minimum sebesar 0 dan nilai maksimum sebesar 93,55. Artinya bahwa setiap kabupaten/kota di Jawa Timur memiliki kondisi tempat tinggal yang berbeda-beda. Semakin besar range data, maka semakin besar nilai varians.

4.2. Pemodelan Jumlah Kasus Kanker Serviks tiap Kabupaten/Kota di Jawa Timur Menggunakan Regresi Poisson.

Salah satu syarat dalam regresi yang melibatkan beberapa variabel prediktor adalah antara variabel prediktornya saling bebas. Jika terdapat adanya hubungan antara variabel prediktor maka terjadi adanya kasus multikolinieritas. Iden-tifikasi kasus multikolinieritas dalam penelitian ini diguna-kan 2 kriteria yaitu nilai korelasi dan VIF. Berdasarkan hasil analisis diperoleh nilai korelasi lebih besar 0,95 dan nilai VIF lebih besar dari 10 yaitu pada variabel X_1 dan X_6 . Maka kedua variabel tersebut tidak diikutsertakan dalam analisis selanjutnya yaitu pemodelan dengan regresi poisson, GPR, dan regresi binomial negatif. Data jumlah kasus kanker serviks adalah data diskrit yang mengikuti distribusi poisson. Pemodelan menggunakan analisis regresi poisson ini dilakukan untuk mengetahui faktor-faktor yang mempengaruhi jumlah kasus kanker serviks. Dengan menggunakan metode MLE diperoleh esti-masi parameter model regresi poisson yang disajikan dalam Tabel 2 dan nilai AIC yang dihasilkan sebesar 1226,6.

Tabel 2 Estimasi Parameter Model Regresi Poisson

Parameter	Estimasi	Standard Error	Z	P-Value
β_0	11,2909	3,1317	3,61	0,009
β_2	0,2045	0,01735	11,79	< 0,0001
β_3	0,04927	0,005438	9,06	< 0,0001
β_4	-0,02922	0,05028	-0,58	0,5647
β_5	0,02022	0,002644	7,65	< 0,0001
β_7	-0,1036	0,008928	-11,60	< 0,0001
β_8	-0,3134	0,09652	-3,25	0,0025
β_9	0,09394	0,01363	6,89	< 0,0001
β_{10}	-0,04298	0,01766	-2,43	0,0199
β_{11}	0,04362	0,01366	3,19	0,0029
β_{12}	-0,1883	0,01007	-18,71	< 0,0001
β_{13}	0,1021	0,009495	10,76	< 0,0001

Nilai devians $D(\hat{\beta})$ yaitu sebesar 1202,6. Kemudian nilai devians dibandingkan dengan nilai Chi-Square dengan taraf signifikansi (α) sebesar 15%. Nilai $\chi^2 = 19,270$ maka

$D(\hat{\beta}) > \chi_{12,0,15}^2$. Jadi keputusannya yaitu Tolak H_0 artinya bahwa minimal terdapat satu parameter yang berpengaruh signifikan terhadap model. Selanjutnya untuk mengetahui variabel prediktor mana saja yang memberikan pengaruh secara signifikan, dilakukan pengujian parameter secara parsial. Berdasarkan Tabel 2 dengan taraf signifikansi 15% dapat diketahui bahwa p-value dari semua parameter lebih kecil 0,15 kecuali parameter β_4 . Selain itu, jika nilai dari $|Z_{hit}| > Z_{\alpha/2}$ maka tolak H_0 yang artinya bahwa parameter tersebut signifikan. Nilai Z_{hit} dari semua parameter lebih besar dari $Z_{\alpha/2} = 1,44$ kecuali $|Z_{hit}|$ dari β_4 sehingga parameter $\beta_0, \beta_2, \beta_3, \beta_5, \beta_7, \beta_8, \beta_9, \beta_{10}, \beta_{11}, \beta_{12}$, dan β_{13} signifikan berpengaruh terhadap model. Jadi model regresi poisson yang dihasilkan adalah.

$$\hat{\mu} = \exp \left(\begin{array}{l} 11,209 + 0,2045X_2 + 0,04927X_3 + 0,02022X_5 - 0,1036X_7 + \\ -0,3134X_8 + 0,09394X_9 - 0,04298X_{10} + 0,04362X_{11} \\ -0,1883X_{12} + 0,1021X_{13} \end{array} \right)$$

Variabel prediktor yang berpengaruh terhadap jumlah kasus kanker serviks di Jawa Timur yaitu persentase tenaga medis (X_2), persentase penduduk perempuan yang umur kawin pertama ≤ 16 tahun (X_3), persentase penduduk yang tinggal di daerah berstatus desa (X_5), persentase penduduk miskin (X_7), persentase penduduk perempuan yang menggunakan kondom (X_8), persentase perempuan dengan jumlah anak yang dilahirkan lebih dari 4 (X_9), persentase penduduk perempuan yang berstatus kawin (X_{10}), persentase penduduk perempuan usia ≥ 35 tahun (X_{11}), persentase rata-rata pengeluaran untuk konsumsi makanan per bulan (X_{12}), dan persentase penduduk yang merokok (X_{13}). Peningkatan maupun penurunan jumlah kasus kanker serviks tiap kabupaten/kota di Jawa Timur tergantung dari nilai koefisien masing-masing variabel yang berpengaruh. Selanjutnya dilakukan pemeriksaan kasus overdispersi pada model regresi poisson yang disajikan pada Tabel 3.

Tabel 3 Nilai Deviance dan Pearson dari Model Regresi Poisson

Kriteria	Nilai	db	Nilai/db
Deviance	1056,621	26	40,639
Pearson Chi-Square	1166,763	26	44,876

Berdasarkan Tabel 3 dapat diketahui bahwa nilai *deviance/db* dan *pearson chi-square/db* lebih besar dari 1 sehingga dapat disimpulkan pada model regresi poisson jumlah kasus kanker serviks di Jawa Timur terjadi overdispersi.

4.3. Pemodelan Jumlah Kasus Kanker Serviks tiap Kabupaten/Kota di Jawa Timur Menggunakan GPR.

GPR merupakan salah satu model yang digunakan untuk mengatasi kasus overdispersi pada regresi poisson. Penaksiran parameter model GPR disajikan pada Tabel 4 dan nilai AIC yang dihasilkan adalah 349,9.

Tabel 4 Estimasi Parameter Model GPR

Parameter	Estimasi	Standard Error	Z	P-Value
β_0	124,92	25,9611	4,81	< 0,0001
β_2	75,4861	13,9540	5,41	< 0,0001
β_3	0,6756	0,1624	4,16	< 0,0002
β_7	3,4561	0,6266	5,52	< 0,0001
β_8	-120,45	23,7836	-5,06	< 0,0001
β_9	0,4818	0,5880	0,82	0,4178
β_{10}	-8,8125	1,2018	-7,33	< 0,0001
β_{11}	17,2680	3,4279	5,04	< 0,0001
β_{12}	-13,8343	2,4783	-5,58	< 0,0001
β_{13}	12,7174	2,6253	4,84	< 0,0001
θ	0,3216	0,05795	5,55	< 0,0001

Nilai devians $D(\hat{\beta})$ yaitu sebesar 327,9. Kemudian nilai devians dibandingkan dengan nilai Chi-Square dengan taraf signifikansi (α) sebesar 15%. Nilai $\chi^2_{(10;0,15)} = 16,711$ maka $D(\hat{\beta}) > \chi^2_{(10;0,15)}$. Jadi keputusannya yaitu Tolak H_0 artinya bahwa minimal terdapat satu parameter yang berpengaruh signifikan terhadap model. Selanjutnya untuk mengetahui variabel prediktor mana saja yang memberikan pengaruh secara signifikan, dilakukan pengujian parameter secara parsial. Berdasarkan Tabel 4 dengan taraf signifikansi 15% dapat diketahui bahwa p-value dari semua parameter lebih kecil 0,15 kecuali parameter β_9 . Selain itu, jika nilai dari $|Z_{hit} > Z_{\alpha/2}|$ maka tolak H_0 yang artinya bahwa parameter tersebut signifikan. Nilai $|Z_{hit}|$ dari semua parameter lebih besar dari $Z_{\alpha/2} = 1,44$ kecuali $|Z_{hit}|$ dari β_9 sehingga parameter $\beta_0, \beta_2, \beta_3, \beta_7, \beta_8, \beta_{10}, \beta_{11}, \beta_{12}$, dan β_{13} signifikan berpengaruh terhadap model. Parameter θ sebesar 0,3216 atau lebih besar 0 menunjukkan terjadinya kasus overdispersi sesuai dengan hasil pengujian yang dilakukan sebelumnya. Jadi model GPR yang dihasilkan adalah sebagai berikut.

$$\hat{\mu} = \exp \left(\begin{array}{l} 124,92 + 75,4861X_2 + 0,6756X_3 + 3,4561X_7 - 120,45X_8 + \\ -8,8125X_{10} + 17,2680X_{11} - 13,8343X_{12} + 12,7174X_{13} \end{array} \right)$$

Variabel prediktor yang berpengaruh terhadap jumlah kasus kanker serviks di Jawa Timur berdasarkan model GPR yang dihasilkan yaitu persentase tenaga medis (X_2), persentase penduduk perempuan yang umur kawin pertama ≤ 16 tahun (X_3), persentase penduduk miskin (X_7), persentase penduduk perempuan yang menggunakan kondom (X_8), persentase penduduk perempuan yang berstatus kawin (X_{10}), persentase penduduk perempuan usia ≥ 35 tahun (X_{11}), persentase rata-rata pengeluaran untuk konsumsi makanan per bulan (X_{12}), dan persentase penduduk yang merokok (X_{13}).

4.4. Pemodelan Jumlah Kasus Kanker Serviks tiap Kabupaten/Kota di Jawa Timur Menggunakan Regresi Binomial Negatif.

Selain menggunakan GPR dalam menangani overdispersi pada regresi poisson, juga bisa menggunakan model regresi binomial negatif. Sepuluh variabel yang signifikan pada regresi poisson menghasilkan 1023 kombinasi kemungkinan model regresi binomial negatif yang disajikan pada Tabel 5.

Tabel 5 Kemungkinan Model Regresi Binomial Negatif dari Kombinasi Variabel

Kemungkinan Model (Y dengan X_i)	AIC	Parameter yang Signifikan
X_{12}	325,93	β_0, β_{12}
$X_7 X_{12}$	326,97	β_0, β_{12}
$X_5 X_{12} X_{13}$	327,65	$\beta_0, \beta_5, \beta_{12}$
$X_2 X_3 X_7 X_9$	333,83	β_0, β_7
$X_5 X_9 X_{10} X_{11} X_{12}$	332,36	$\beta_0, \beta_5, \beta_{12}$
$X_5 X_8 X_9 X_{10} X_{12} X_{13}$	333,24	$\beta_0, \beta_5, \beta_{12}$
$X_5 X_8 X_9 X_{10} X_{11} X_{12} X_{13}$	335,24	$\beta_0, \beta_5, \beta_{12}$
$X_2 X_3 X_5 X_9 X_{10} X_{11} X_{12} X_{13}$	336,68	β_{12}
$X_2 X_3 X_5 X_8 X_9 X_{10} X_{11} X_{12} X_{13}$	338,60	β_{12}

Berdasarkan Tabel 5 dapat diketahui bahwa kombinasi tiga variabel menghasilkan parameter signifikan paling banyak dengan nilai AIC terkecil daripada kombinasi lainnya. Maka kombinasi tiga variabel prediktor ini akan dianalisis lebih lanjut untuk mendapatkan model GPR.

Tabel 6 Estimasi Parameter Model Regresi Binomial Negatif

Parameter	Estimasi	SE	Z-Value	P-Value
β_0	12,313	3,719	3,310	0,000933
β_5	0,024	0,016	1,484	0,137751
β_{12}	-0,217	0,087	-2,497	0,012509
β_{13}	0,076	0,056	1,360	0,173717
θ	0,443	0,111	3,991	-

Nilai devians $D(\hat{\beta})$ yaitu sebesar 317,653. Kemudian nilai devians dibandingkan dengan nilai Chi-Square dengan taraf signifikansi (α) sebesar 15%. Nilai $\chi^2_{(4;0,15)} = 7,359$ maka $D(\hat{\beta} > \chi^2_{(4;0,15)})$. Jadi keputusannya yaitu Tolak H_0 artinya bahwa minimal terdapat satu parameter yang berpengaruh signifikan terhadap model. Selanjutnya untuk mengetahui variabel prediktor mana saja yang memberikan pengaruh secara signifikan, dilakukan pengujian parameter secara parsial. Berdasarkan Tabel 6 dengan taraf signifikansi 15% dapat diketahui bahwa p-value dari semua parameter lebih kecil 0,15 kecuali parameter β_{13} . Selain itu, jika nilai dari $|Z_{hit}| > Z_{\alpha/2}$ maka tolak H_0 yang artinya bahwa parameter tersebut signifikan. Nilai $|Z_{hit}|$ dari semua parameter lebih besar dari $Z_{\alpha/2} = 1,44$ kecuali dari β_{13} sehingga parameter β_0, β_5 , dan β_{12} signifikan berpengaruh terhadap model. Parameter θ sebesar 0,443 atau lebih besar 0 menunjukkan terjadinya kasus overdispersi sesuai dengan hasil pengujian yang dilakukan sebelumnya. Jadi model regresi binomial negatif yang dihasilkan adalah sebagai berikut.

$$\hat{\mu} = \exp(12,313 + 0,024X_5 - 0,217X_{12})$$

Variabel prediktor yang berpengaruh terhadap jumlah kasus kanker serviks di Jawa Timur berdasarkan model regresi binomial negatif yang dihasilkan yaitu persentase penduduk yang tinggal di daerah berstatus desa (X_5) dan persentase rata-rata pengeluaran untuk konsumsi makanan per bulan (X_{12}).

4.5. Pemilihan Model Terbaik

Kriteria pemilihan model terbaik yang digunakan adalah AIC. Model terbaik adalah model yang memiliki nilai AIC terkecil.

Tabel 7 Pemilihan Model Terbaik

Model	Variabel Signifikan	AIC
GPR	$X_2 X_3 X_5 X_7 X_8 X_{10} X_{11} X_{12} X_{13}$	349,90
Regresi Binomial Negatif	$X_5 X_{12}$	327,65

Berdasarkan Tabel 7 dapat diketahui bahwa model terbaik untuk menangani kasus overdispersi pada regresi poisson model jumlah kasus kanker serviks di Jawa Timur adalah regresi binomial negatif meskipun regresi binomial negatif menghasilkan variabel prediktor yang signifikan lebih sedikit dibandingkan hasil dari model GPR.

5. Kesimpulan

Berdasarkan hasil analisis dan pembahasan maka dapat diperoleh kesimpulan sebagai berikut.

1. Secara umum, gambaran angka kasus kanker serviks di Jawa Timur pada tahun 2010 menunjukkan tingkat kasus yang masih tinggi dengan rata-rata 46 kasus tiap kabupaten/kota. Kota Malang merupakan kota yang memiliki jumlah kasus kanker serviks tertinggi yaitu sebesar 479 kasus, sedangkan daerah yang tidak ditemukan terjadi kasus kanker serviks diantaranya adalah kabupaten Pacitan, Jember, Situbondo, Pasuruan, Mojokerto, Jombang, Nganjuk, Tuban, Pamekasan, dan kota Probolinggo.
2. Hasil pemodelan jumlah kasus kanker serviks tiap kabupaten/kota di Jawa Timur menggunakan regresi poisson ternyata memberikan hasil adanya pengaruh overdispersi. Dalam menangani kasus tersebut dilakukan pemodelan menggunakan Generalized Poisson Regression (GPR) dan regresi binomial negatif. Model terbaik menggunakan GPR menghasilkan 8 variabel prediktor yang signifikan mempengaruhi jumlah kasus kanker serviks di Jawa Timur antara lain persentase tenaga medis (X_2), persentase penduduk perempuan yang umur kawin pertama ≤ 16 tahun (X_3), persentase penduduk miskin (X_7), persentase penduduk perempuan yang menggunakan kondom (X_8), persentase penduduk perempuan yang berstatus kawin (X_{10}), persentase penduduk perempuan usia ≥ 35 tahun (X_{11}), persentase rata-rata pengeluaran untuk konsumsi makanan per bulan (X_{12}), dan persentase penduduk yang merokok (X_{13}). Sedangkan model terbaik menggunakan regresi binomial negatif menghasilkan 2 variabel prediktor yang signifikan yaitu persentase penduduk yang tinggal di daerah berstatus desa (X_5) dan persentase rata-rata pengeluaran untuk konsumsi makanan per bulan (X_{12}).
3. Perbandingan model GPR dan regresi binomial negatif untuk mengetahui model yang lebih baik dalam penanganan kasus overdispersi pada regresi poisson. Model GPR menghasilkan nilai AIC sebesar 349,90. Sedangkan model regresi binomial negatif menghasilkan nilai AIC sebesar 327,65. Maka model terbaik diperoleh dari model regresi binomial negatif karena menghasilkan nilai AIC terkecil.

Daftar Pustaka

- [1] Agresti, A. (2002). *Categorical Data Analysis Second Edition*, John Wiley & Sons, New York.
- [2] Berk, D. dan MacDonald, J. (2007). *Overdispersion and Poisson regression*. Department of Statistics, Department of Criminology, University of Pennsylvania.
- [3] Bozdogan, H. (2000). Akaike's Information Criterion and Recent Developments in Information Complexity, *Mathematical Psychology*, 44, 62-91.
- [4] Cameron, A.C. dan Trivedi, P.K. (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.
- [5] Dewi, R.P. (2008). *Kajian Overdispersi Pada Regresi Poisson Dengan Menggunakan Regresi Binomial Negatif (Studi Kasus: Ketidاكلulusan Siswa SMA dalam Ujian Nasional di DKI Jakarta)*. Bogor: Program Sarjana, Institut Pertanian Bogor.
- [6] Ferlay J. G (2002). *Cancer Incidence, Mortality and Prevalence Worldwide*. Lyon: IARC CancerBase, 2004.
- [7] Greene, W., *Functional Forms For The Negative Binomial Model For Count Data*. Foundations and Trends in Econometrics. Working Paper, Department of Economics, Stern School of Business, New York University, 2008: 585-590.
- [8] Hardin JW dan Hilbe JM. (2007). *Generalized Linier Models and Extensions*. Texas: A Strata Perss Publication.
- [9] Listiani, Y. (2010). *Pemodelan Regresi Generalized Poisson pada Faktor-Faktor yang mempengaruhi Angka Kematian Bayi di Jawa Timur Tahun 2007*. Surabaya: Program Sarjana, Institut Teknologi Sepuluh Nopember.
- [10] Mansjoer, A, Triyanti, K, dan Savitri, R. (2008). *Kapita Selektta Kedokteran*. Edisi ketiga. Fakultas Kedokteran Universitas Indonesia : Media Aesculapis.
- [11] McCullagh P dan Nelder JA. 1983. *Generalized Linier Models*. London: Chapman and Hall.
- [12] Mood, Alexander M., Graybill, Franklin A., dan Boes, D.C. (1974). *Introduction to the Theory of Statistics*, Third Edition. New York, NY: McGraw-Hill.
- [13] Prawirohardjo, S. (2008). *Ilmu Kandungan*. Edisi Kedua. Jakarta : PT Bina Pustaka.
- [14] Walpole, E Ronald. (1995). *Pengantar Statistik Edisi Ketiga*. Jakarta : PT. Gra-media Pustaka Utama.