

Solusi Overdispersi Menggunakan Generalized Poisson Regression (Studi Kasus : Penderita HIV di Provinsi Riau)

Rahmadeni¹, Nurmla Sari²

^{1,2} Jurusan Matematika, Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau
Jl. HR. Soebrantas No. 155 Simpang Baru, Panam, Pekanbaru, 28293
Email: r4dieni@gmail.com, nurmalasarinadapdap@gmail.com

ABSTRAK

Regresi poisson umumnya digunakan untuk menganalisis data dengan asumsi rata-rata dan variansinya sama (equidispersi). Tetapi secara umum data, seringkali terjadi variansi melebihi nilai rata-ratanya atau *overdispersi*. Model regresi poisson tergeneralisir digunakan untuk mengatasi masalah *overdispersi*. Pada penelitian ini, peneliti menggunakan data kasus HIV di Provinsi Riau tahun 2015. Estimasi parameter menggunakan metode *maximum log likelihood estimation* (MLE). Hasil analisis dalam penelitian ini menunjukkan bahwa variabel yang berpengaruh terhadap data kasus HIV adalah tenaga kesehatan per 100, fasilitas kesehatan per 100, persentase penduduk usia 25-34 tahun.

Kata Kunci: Overdispersi, *Maximum log likelihood*, Regresi Poisson tergeneralisir.

ABSTRACT

Poisson regression is generally used to analyze the data with assumption of the average and its variance are same (equidispersion). But generally in the data, often occur the variance which more than the value of its average or overdispersion. The model of generalized Poisson regression (GPR) is used to solve this problem overdispersion. In this study, researchers used the HIV case data in Riau Province in 2015. Parameter estimation using maximum log likelihood method. The analysis result in this study showed that the variables which are influenced to HIV, is the health workers per 100, health facilities per 100, percentage of population aged 25-34.

Keywords: *Overdispersion, Generalized Poisson regression, Maximum log likelihood.*

Pendahuluan

Analisis regresi merupakan analisis statistika yang bertujuan untuk memodelkan hubungan antara variabel respon Y dengan satu atau lebih variabel prediktor X . Asumsi variabel responnya adalah data kontinu yang mengikuti distribusi normal. Namun dalam aplikasinya, banyak ditemukan penelitian yang menggunakan variabel respon yang berupa data cacah (*count*). Metode yang tepat digunakan untuk peubah respon berupa data cacah adalah analisis regresi Poisson. Dalam model regresi Poisson terdapat beberapa asumsi yang harus dipenuhi, salah satunya adalah nilai variansi dari variabel respon Y yang diberikan oleh $X = x$ harus sama dengan nilai rata-ratanya yaitu $Var(Y|x) = E(Y|x) = \mu$. Namun dalam analisis data diskrit dengan menggunakan model regresi poisson terkadang terjadi pelanggaran asumsi tersebut, dimana nilai variansinya lebih besar dari nilai rata-rata yang disebut *overdispersi* atau varian lebih kecil dari nilai rata-rata yang disebut *underdispersi* [5].

Penanganan *overdispersi* atau *underdispersi* pada regresi Poisson dapat ditangani dengan berbagai pilihan model regresi diantaranya yaitu model *generalized Poisson regression (GPR)*. GPR merupakan perluasan dari regresi Poisson yang dapat mengatasi keadaan *overdispersi/ underdispersi* [9]. Hubungan nilai rata-rata dan varian dalam model GPR dapat dikondisikan sebagai (1) Jika nilai varian sama dengan nilai rata-rata $Var(Y|x) = E(Y|x) = \mu$, maka nilai parameter disperse $\phi = 0$, sehingga fungsi densitas peluang GPR, akan diturunkan ke regresi Poisson, (2) Jika nilai varian lebih besar dari nilai rata-rata $Var(Y|x) > E(Y|x)$, maka nilai parameter dispersi $\phi > 0$, sehingga dapat dikatakan bahwa data

terjadi *overdispersi*, (3) jika nilai varian lebih kecil dari nilai rata-rata $Var(Y|x) < E(Y|x)$, maka nilai parameter disperse $\phi < 0$, sehingga dapat dikatakan bahwa data terjadi *underdispersi*.

Metode Penelitian

Distribusi Poisson

Dalam eksperimen Poisson, probabilitas memperoleh dengan tepat peristiwa Y sebanyak y kejadian untuk setiap satu satuan unit (waktu dan ruang) yang ditentukan membentuk sebuah distribusi yang fungsi probabilitasnya adalah [5]

$$f(y, \lambda) = \Pr(Y = y; \lambda) = \frac{\lambda^y e^{-\lambda}}{y!} \quad (1)$$

Regresi Poisson

Regresi Poisson merupakan salah satu regresi nonlinier yang sering digunakan untuk memodelkan hubungan antara variabel respon yang berupa data diskrit dengan variabel prediktor yang berupa data diskrit atau kontinu [13]. Regresi Poisson merupakan penerapan dari *Generalized Linear Model (GLM)*. *GLM* merupakan perluasan dari model regresi umum untuk variabel respon yang memiliki sebaran eksponensial. Regresi Poisson digunakan untuk menganalisis data *count* (berjenis diskrit). Pada regresi Poisson diasumsikan variabel respon Y berdistribusi Poisson dan tidak terjadi multikolinearitas diantara masing-masing variabel prediktor (X).

Parameter Dispersi

Rumus dari parameter dispersi (ϕ) adalah [2]

$$\phi = \frac{\text{nilai deviance}}{df}$$

Apabila nilai $\phi > 0$ maka terjadi *overdispersi* dan apabila $\phi < 0$ maka terjadi *underdispersi*.

Overdispersi dan Underdispersi

Overdispersi ataupun *underdispersi* akan menghasilkan nilai devians model menjadi sangat besar sehingga model yang dihasilkan kurang tepat. Nilai devians diperoleh dari nilai *deviance* dibagi dengan derajat kebebasan (dilihat pada output SPSS). Salah satu model yang dapat digunakan untuk mengatasi masalah *overdispersi* dan *underdispersi* adalah dengan menggunakan model regresi Poisson tergeneralisasi. Model regresi ini merupakan perluasan dari regresi Poisson dan baik digunakan dalam keadaan *equidispersi*, *overdispersi*, dan *underdispersi*.

Multikolinearitas

Mendeteksi ada tidaknya multikolinearitas dapat dengan melihat nilai *tolerance* dan *VIF (Variance Inflation Factor)* [7]. Jika nilai *tolerance* lebih dari 0,1 dan *VIF* kurang dari 10 maka tidak terjadi multikolinearitas.

Model Regresi Poisson Tergeneralisasi (*Generalized Poisson Regression*)

Dalam *Generalized Poisson Regression (GPR)* fungsi probabilitas bersyarat dari Y_i diberikan nilai $x_{1i}, x_{2i}, \dots, x_{pi}$ adalah:

$$f_i(y_i | x_{1i}, x_{2i}, \dots, x_{pi} : \mu_i, \alpha) = \left(\frac{\mu_i}{1 + \alpha \mu_i} \right)^{y_i} \frac{(1 + \alpha y_i)^{y_i - 1}}{y_i!} \exp \left(- \frac{\mu_i (1 + \alpha y_i)}{1 + \alpha \mu_i} \right) \quad (2)$$

Rata-rata dan variansi bersyarat dari Y_i diberikan untuk $X_{1i} = x_{1i}, X_{2i} = x_{2i}, \dots, X_{pi} = x_{pi}$ regresi Poisson tergeneralisir adalah [4]

$$E(Y_i | X_{1i} = x_{1i}, X_{2i} = x_{2i}, \dots, X_{pi} = x_{pi}) = \mu_i$$

$$Var(Y_i | X_{1i} = x_{1i}, X_{2i} = x_{2i}, \dots, X_{pi} = x_{pi}) = \mu_i (1 + \alpha \mu_i)^2$$

Dengan mensubstitusikan $\mu_i = \exp(\beta_0 + \sum_{j=i}^p \beta_j x_{ji})$ ke dalam (2) maka diperoleh:

$$f_i(y_i | x_{1i}, x_{2i}, \dots, x_{pi} : \mu_i, \alpha) = \left(\frac{\exp(\beta_0 + \sum_{j=i}^p \beta_j x_{ji})}{1 + \alpha \exp(\beta_0 + \sum_{j=i}^p \beta_j x_{ji})} \right)^{y_i} \frac{(1 + \alpha y_i)^{y_i - 1}}{y_i!} \exp \left(- \frac{\exp(\beta_0 + \sum_{j=i}^p \beta_j x_{ji} (1 + \alpha y_i))}{1 + \alpha \exp(\beta_0 + \sum_{j=i}^p \beta_j x_{ji})} \right) \quad (3)$$

Estimasi Parameter Model Regresi Poisson Tergeneralisir

Dalam penggunaan regresi Poisson tergeneralisir, nilai dari parameter-parameter yang tidak diketahui, yaitu $\beta_0, \beta_1, \dots, \beta_p, \alpha$ harus ditaksir. Untuk menaksir parameter tersebut digunakan metode *maximum likelihood*. Jika n pasang pengamatan $\{(X_{1i}, X_{2i}, \dots, X_{pi}, Y_i), i = 1, 2, 3, \dots, n\}$ diasumsikan saling bebas, maka fungsi *likelihood* diperoleh dengan mengalikan semua fungsi probabilitas bersyarat dari Y_i diberikan nilai $X_{1i}, X_{2i}, \dots, X_{pi}$ pada (3), yaitu:

$$L(\beta^*) = \prod_{i=1}^n f(Y_i | x_{1i}, x_{2i}, \dots, x_{pi}; \beta_0, \beta_1, \dots, \beta_p, \alpha)$$

$$= \prod_{i=1}^n \left(\frac{\exp(\beta_0 + \sum_{j=i}^p \beta_j x_{ji})}{1 + \alpha \exp(\beta_0 + \sum_{j=i}^p \beta_j x_{ji})} \right)^{y_i} \frac{(1 + \alpha y_i)^{y_i - 1}}{y_i!} \exp \left(- \frac{\exp(\beta_0 + \sum_{j=i}^p \beta_j x_{ji} (1 + \alpha y_i))}{1 + \alpha \exp(\beta_0 + \sum_{j=i}^p \beta_j x_{ji})} \right)$$

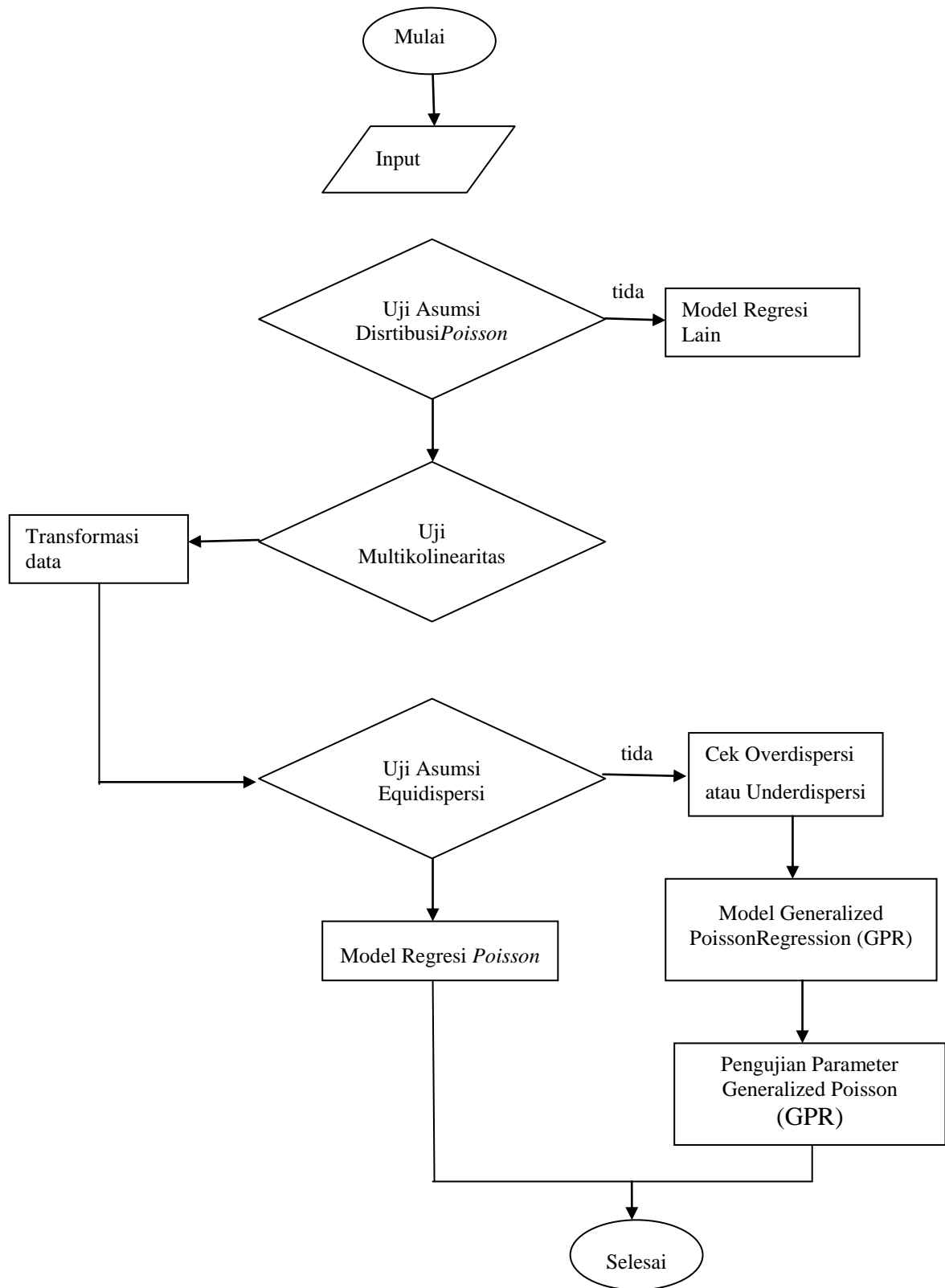
Bahan Penelitian

Data yang diperoleh dari Dinas Kesehatan Provinsi Riau dan Badan Pusat Statistik (BPS) yaitu data banyaknya kasus *HIV* (Y), persentase penduduk miskin (X_1), persentase penduduk dengan pendidikan SLTA (X_2), persentase PUS yang sedang menggunakan alat KB kondom (X_3), rasio jumlah tenaga kesehatan per 100 (X_4), rasio fasilitas kesehatan per 100 (X_5), dan persentase penduduk usia 25-34 (X_6) di Provinsi Riau pada tahun 2015. Data yang diambil adalah data per kabupaten/kota di Provinsi Riau tahun 2015.

Langkah-langkah untuk memperoleh model regresi Poisson tergeneralisasi terbaik adalah sebagai berikut:

- 1) Memeriksa asumsi berdistribusi poisson
- 2) Memeriksa hubungan antar variabel prediktor (kolinearitas).
- 3) Memeriksa kasus *overdispersi/ underdispersi*.
- 4) Menentukan model regresi Poisson tergeneralisasi.

- 5) Pengujian parameter *generalized Poisson regression (GPR)*.
Berikut diagram alur dari langkah-langkah yang dilakukan dalam penelitian ini :



Hasil dan Pembahasan

Regresi Poisson Tergeneralisir pada Kasus HIV

Analisis data dilakukan dengan langkah-langkah sebagai berikut:

Uji Kolmogorov Smirnov

Tabel 4.1 Uji Kolmogorov Smirnov

	<i>Y</i>
<i>N</i>	12
Kolmogorov-Smirnov <i>Z</i>	0,3579
Asymp.siq.(2-tailed)	0,8999

Analisis output untuk uji kolmogorov smirnov adalah sebagai berikut:

- ✓ Hipotesis:
 - H_0 : Sampel berasal dari populasi yang berdistribusi Poisson
 - H_1 : Sampel tidak berasal dari populasi yang berdistribusi Poisson
- ✓ Taraf Nyata:
 - $\alpha = 0,05$
- ✓ Statistik uji :
 - Tolak H_0 jika nilai $Sig < \alpha$ dan sebaliknya jika $Sig > \alpha$ maka H_0 diterima.

Berdasarkan Tabel 4.1 terlihat bahwa nilai $Sig > \alpha$ yaitu $0,89 > 0,05$ maka H_0 diterima, Artinya sampel berasal dari populasi yang berdistribusi Poisson.

Multikolinearitas

Tabel 4.2 Pengujian Multikolinearitas

Variabel	Collinearity statistics	
	Tolerance	Vif
Penduduk Miskin	0,371	2,699
Pendidikan SLTA	0,477	2,096
Alat KB	0,421	2,377
Tenaga Kesehatan	0,371	2,695
Fasilitas Kesehatan	0,459	2,180
Usia 25-34 Tahun	0,591	1,692

Analisis output untuk uji kolmogorov smirnov adalah sebagai berikut:

- a. Hipotesis:
 - H_0 : Model regresi memiliki masalah multikolinieritas
 - H_1 : Model regresi tidak memiliki masalah multikolinieritas
- b. Taraf Nyata:
 - $\alpha = 0,05$
- c. Statistik Uji:

Tolak H_0 jika seluruh variabel prediktor memiliki nilai *VIF* kurang dari 10 dan nilai *tolerance* lebih dari 0,1. Sebaliknya jika seluruh variabel prediktor memiliki nilai *VIF* lebih besar 10 dan nilai *Tolerance* kurang dari 0,1 maka H_0 diterima.

Berdasarkan Tabel 4.2 terlihat bahwa nilai *VIF* pada masing-masing prediktor tidak ada yang lebih dari 10. Dan nilai *tolerance* lebih dari 0,1, hal ini menunjukkan bahwa tidak terdapat kasus multikolinearitas, sehingga layak diikutsertakan dalam pembentukan model regresi.

Pendeteksian Overdispersi

Tabel 4.3 Taksiran Dispersi Regresi Poisson

Kriteria	<i>Db</i>	Nilai	Nilai/db
<i>Deviance</i>	5	93,595	18,719
<i>Person Chi-Square</i>	5	92,778	18,719

Berdasarkan Tabel 4.3, diperoleh bahwa nilai *deviance* dan *pearson chi-square* lebih besar dari 1. Sehingga dapat disimpulkan bahwa terjadi *overdispersi*.

Model Taksiran Data HIV dengan Regresi Poisson Tergeneralisir

Model regresi Poisson tergeneralisir yang dihasilkan adalah:

$$\hat{\mu} = \exp(5,670 - 0,072x_1 - 0,089x_2 + 0,012x_3 - 0,109x_4 - 0,072x_5 + 0,171x_6 + \varepsilon_i)$$

Tabel 4.4 Uji *Deviance* Regresi Poisson Tergeneralisir

Kriteria	Nilai
<i>Deviance</i>	93,595

a. Hipotesis:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$$

$$H_1 : \beta_i \neq 0$$

untuk $i = 1, 2, \dots, 6$

b. Taraf Nyata: $\alpha = 0,05$

c. Statistik uji: H_0 ditolak jika *deviance* $> \chi^2_{\alpha, n-k-1}$ H_0 diterima jika nilai *deviance* $\leq \chi^2_{\alpha, n-k-1}$. Jika H_0 ditolak hal ini menginterpretasi bahwa model sesuai.

Dari Tabel 4.4 diperoleh nilai *deviance* = 93,595 dengan $\chi^2_{0,05, 12-6-1} = 11,070$ karena nilai *deviance* $> \chi^2_{\alpha, n-k-1}$, maka H_0 ditolak hal ini menginterpretasi bahwa model sesuai. Maka persentase penduduk miskin, persentase penduduk dengan pendidikan SLTA, persentase PUS yang sedang menggunakan alat KB kondom, rasio jumlah tenaga kesehatan per 100, rasio fasilitas kesehatan per 100, dan persentase penduduk usia 25-34 berpengaruh terhadap kasus *HIV*.

Pengujian Signifikansi Parameter Model Regresi Poisson Tergeneralisir

Tabel 4.5 Hasil Uji *Log Likelihood* Regresi Poisson Tergeneralisir

	Kriteria Poisson Tergeneralisir
$L(\hat{\Omega})$	2391,846
$L(\hat{\Theta})$	82,933

a. Hipotesis:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$$

$$H_1 : \beta_i \neq 0 \quad \text{untuk } i = 1, 2, \dots, 6$$

b. Taraf Nyata: $\alpha = 0,05$

c. Statistik Uji:

H_0 ditolak jika $G > \chi^2_{\alpha, n-k-1}$ yang berarti minimal ada satu parameter yang berpengaruh secara signifikan terhadap model. Statistik uji yang digunakan untuk pengujian tersebut adalah:

Berdasarkan Tabel 4.6 maka diperoleh nilai uji G :

$$G = 2 \left[\ln(L(\hat{\Theta}) - L(\Theta)) \right]$$

$$= 2 \left[\ln(2391,846 - 82,933) \right]$$

$$= 15,489$$

Berdasarkan tabel *chi-square* dengan tingkat signifikan 0,05 dan derajat bebas 5 diperoleh $\chi^2_{0,05,5} = 11,070$. Nilai $G > \chi^2_{\alpha, n-k-1}$, maka H_0 ditolak pada tingkat signifikansi 0,05. Maka dapat disimpulkan bahwa terdapat minimal salah satu pengaruh variabel. Sehingga, model tersebut dapat digunakan untuk menggambarkan hubungan antara persentase penduduk miskin (x_1), persentase penduduk dengan pendidikan SLTA (x_2), persentase PUS yang sedang menggunakan alat KB kondom (x_3), rasio jumlah tenaga kesehatan per 100 (x_4), rasio fasilitas kesehatan per 100 (x_5), dan persentase penduduk usia 25-34 (x_6) terhadap jumlah kasus *HIV* di Provinsi Riau pada tahun 2015.

Tabel 4.6 Hasil Nilai Uji Signifikansi Regresi Poisson Tergeneralisir

Parameter	Estimasi	Std. Error	Wald χ^2	Tabel χ^2	α	Keputusan
Penduduk Miskin (x_1)	-0,072	0,0110	2,526	3,841	0,05	Tidak signifikansi
Pendidikan SLTA (x_2)	-0,089	0,0504	3,096	3,841	0,05	Tidak signifikansi
Alat KB (x_3)	0,012	0,0096	1,657	3,841	0,05	Tidak signifikansi
Tenaga Kesehatan per 100 (x_4)	-0,109	0,0188	33,917	3,841	0,05	Signifikansi
Fasilitas Kesehatan per 100 (x_5)	-0,072	0,0091	63,566	3,841	0,05	Signifikansi
Persentase Penduduk Usia 25-34 (x_6)	0,171	0,0079	42,526	3,841	0,05	Signifikansi

Perolehan hasil keputusan di atas dengan tingkat signifikansi 0,05 maka dapat diputuskan bahwa variabel yang berpengaruh terhadap jumlah *HIV* adalah tenaga kesehatan per 100 (x_4), fasilitas kesehatan per 100 (x_5), persentase penduduk usia 25-34 (x_6).

Interpretasi Model Regresi Poisson Tergeneralisir

$$\mu = \exp(5,670 - 0,109x_4 - 0,072x_5 + 0,171x_6 + \varepsilon_i)$$

Berdasarkan model di atas, maka dapat diinterpretasikan sebagai berikut ini:

1. Setiap penambahan 1 koefisien tenaga kesehatan (x_4) maka akan melipat gandakan rata-rata variabel respon Y sebesar $\exp(-0,109) = 0,8967$ kali per 100 penduduk dari rata-rata variabel respon semula bila variabel lain tetap. Dengan kata lain, penambahan 1 koefisien tenaga kesehatan (x_4) maka akan mengurangi dari rata-rata jumlah kasus baru *HIV* sebesar 0,8967.
2. Setiap penambahan 1 koefisien fasilitas kesehatan (x_5) maka akan melipat gandakan rata-rata variabel respon Y sebesar $\exp(-0,076) = 0,9268$ kali per 100 penduduk dari rata-rata variabel respon semula bila variabel lain tetap. Dengan kata lain, penambahan 1 koefisien fasilitas kesehatan (x_5) maka akan mengurangi dari rata-rata jumlah kasus baru *HIV* sebesar 0,9268.
3. Setiap penambahan 1 persen penduduk usia 25-34 (x_6) maka akan melipat gandakan rata-rata variabel respon Y sebesar $\exp(0,171) = 1,1864$ kali per dari rata-rata variabel respon semula bila variabel lain tetap. Dengan kata lain, penambahan 1 persen penduduk usia 25-34 maka akan menambah jumlah kasus baru *HIV* sebesar 1,186%.

Kesimpulan

Data penderita kasus *HIV* di Provinsi Riau yang digunakan diperoleh model regresi Poisson tergeneralisir adalah $\mu = \exp(5,670 - 0,109x_4 - 0,072x_5 + 0,171x_6 + \varepsilon_i)$. Variabel yang berpengaruh terhadap jumlah *HIV* adalah tenaga kesehatan per 100 (x_4), fasilitas kesehatan per 100 (x_5), persentase penduduk usia 25-34 tahun (x_6).

Daftar Pustaka

- [1] Badan Pusat Statistik. “Provinsi Riau dalam Angka 2016”. Riau, Badan Pusat Statistik Provinsi Riau. 2016.
- [2] Darnah. “Mengatasi Overdispersi pada Model Regresi Poisson dengan *Generalized Poisson Regression P*”. *Jurnal Eksponensial*. 2011.
- [3] Dinas Kesehatan Provinsi Riau. “*Profil Kesehatan Provinsi Riau 2016*”. Riau, Dinas Kesehatan Provinsi Riau. 2016.
- [4] Famoye, Wulu&Singh. “*On the Generalized Poisson Regression Model with an Application to Accident Data*”. 2004.
- [5] Harinaldi. Prinsip – Prinsip Statistik Untuk Teknik dan Sains. Penerbit Erlangga. Jakarta. 2005.
- [6] J. Scott Long, “Regression Models for Categorical and Limited Dependent Variables. *Advanced Quantitative Techniques in the Social Sciences Number 7*”. Sage Publications: Thousand Oaks, CA. 1997.
- [7] Melliana, setyorini, at al. “The Comparison of Generalized Poisson Regression and Negative Binomial Regression Methods in Overcoming *Overdispersi*”. *International Journal of Scientific & Thechnology Research*. 2013.
- [8] Mohmoud.M.M, &Alderiny. “On Estimating Parameters of Consored Generalized Poisson Regression Model”. *Aplied Mathematical Sciences*. 2010.
- [9] Priyatno, Duwi. Analisis korelasi, Regresi, dan Multivariate dengan SPSS Yogyakarta, Gava Media. 2013.
- [10] Putra, I.P.Y.E., I.P.E.N. Kencana, & I.G.A.M. Srinadi. Penerapan Regresi Generalized Poisson Untuk Mengatasi Fenomena Overdispersi Pada Kasus Regresi Poisson. *Jurnal Matematika*, 2(2):49-53. 2013.
- [11] Rahmadeni, & Desmita, Zulya. “Perbandingan Model Regresi Generalized Poisson Dan Binomial Negatif Untuk Mengatasi Overdispersi Pada Regresi Poisson (Studi Kasus: Penderita Filariasis di Provinsi Riau Tahun 2011)”. *Jurnal Sains Matematika dan Statistika*. 2016.
- [12] Ruliana. “Pemodelan Generalized Poisson Regression (Gpr) Untuk Mengatasi Pelanggaran Equidispersi Pada Regresi Poisson Kasus Campak Di Kota Semarang”. *FMIPA Universitas Negeri Semarang*. 2015.

- [13] Rusianti. “Penanganan Data Overdispersi Menggunakan Regresi Poisson Tergeneralisir (Kasus Persentase Kematian Ibu Di Provinsi Sulawesi Tenggara 2012)”. *FMIPA Universitas Haluoleo Kendari*. 2016.
- [14] Safrida, N., D. Ispriyanti, & T. Widiharih. Aplikasi Model Regresi Poisson Tergeneralisasi Pada Kasus Angka Kematian Bayi di Jawa Tengah Tahun 2007. *Jurnal Gaussian*, 2(2): 361-368. 2013.