

Pemodelan *Regresi Poisson Inverse Gaussian* Studi Kasus: Jumlah Kasus Baru HIV di Provinsi Jawa Tengah Tahun 2015

Andriana Y. Herindrawati, I Nyoman Latra, dan Purhadi

Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111

e-mail: i_nyoman_1@statistika.its.ac.id, purhadi@statistika.its.ac.id, andrianayoshinta7@gmail.com

Abstrak—Jumlah kasus baru HIV adalah salah satu contoh data *count* (data cacahan). Pemodelan data *count* dapat menggunakan regresi poisson. Terdapat asumsi yang harus dipenuhi jika menggunakan regresi poisson yaitu *mean* dan *varians* harus sama, sedangkan pada kasus data cacahan asumsi ini sering tidak terpenuhi. Hal ini terjadi karena adanya *overdispersi*, yaitu varians lebih besar dari *mean*. Oleh karena itu dalam memodelkan data cacahan tersebut tidak cukup dengan regresi poisson sederhana. Regresi poisson inverse gaussian (PIG) merupakan salah satu bentuk regresi dari *mixed poisson* yang dirancang untuk data cacahan dengan kasus *overdispersi* dan telah digunakan pada beberapa penelitian yang menggunakan data cacahan. Penaksiran parameter dilakukan dengan metode MLE dan pengujian hipotesis dengan menggunakan metode MLTR. Jumlah kasus baru HIV merupakan salah satu data cacahan yang berpotensi terjadi *overdispersi*. Oleh karena itu, dalam memodelkan jumlah kasus baru HIV di Provinsi Jawa Tengah tahun 2015 dapat digunakan pemodelan dengan regresi PIG. Berdasarkan model tersebut, variabel prediktor yang memberikan pengaruh signifikan terhadap jumlah kasus baru HIV di Provinsi Jawa Tengah adalah persentase PUS yang menggunakan kondom, rasio fasilitas kesehatan, persentase daerah perkotaan, dan persentase penduduk usia 25-34 tahun.

Kata Kunci— Regresi PIG, MLE, MLRT, HIV

I. PENDAHULUAN

Sustainable Development Goals (SDGs) dapat diartikan sebagai lanjutan dari pembangunan MDGS dan merupakan pembangunan global yang di deklarasikan PBB sejak tahun 2015 [1]. Salah satu tujuan dari SDGs adalah untuk mengakhiri epidemi HIV di tahun 2030. HIV merupakan penyakit menular yang disebabkan oleh infeksi *Human Immunodeficiency Virus*. Virus ini menyebar melalui cairan tubuh, dan menyerang sistem kekebalan tubuh, khususnya sel CD4 atau yang sering disebut sel T [2]. Di Indonesia HIV termasuk ke dalam tiga besar penyakit dengan kasus terbanyak setelah TBC dan malaria. Provinsi yang termasuk dalam urutan 5 besar dengan jumlah kasus HIV terbanyak adalah Jawa Tengah.

Jumlah kasus baru HIV di Jawa Tengah yang ditemukan pada tahun 2015 mencapai 1.467 kasus, jumlah ini mengalami peningkatan jika dibandingkan dengan tahun 2014 sebanyak 1352 kasus [3]. Jumlah kasus baru HIV di suatu wilayah merupakan salah satu bentuk data cacahan sehingga dalam pemodelannya bisa menggunakan regresi Poisson. Pada regresi poisson untuk data *count* terkadang ditemukan kasus *overdispersi* yang membuat hasil penelitian tidak valid. Salah

satu *mixed poisson distribution* yang sering digunakan dalam penelitian untuk mengatasi kasus *overdispersi* adalah distribusi Poisson Inverse Gaussian (PIG)[4]. Berdasarkan uraian di atas, maka dilakukan penelitian mengenai pemodelan kasus baru HIV di Provinsi Jawa Tengah tahun 2015 dengan menggunakan regresi Poisson Inverse Gaussian. Sehingga dapat diketahui faktor apa saja yang berpengaruh terhadap kasus baru HIV di Provinsi Jawa Tengah, hal tersebut menjadi masukan untuk Dinas Kesehatan Provinsi Jawa Tengah.

II. TINJAUAN PUSTAKA

A. *Distribusi Poisson Inverse Gaussian*

Distribusi PIG merupakan salah satu distribusi *mixed poisson*. Distribusi PIG ditentukan oleh dua parameter yaitu rata-rata (μ) sebagai parameter lokasi dan parameter dispersi (τ) sebagai parameter bentuk. Probabilitas distribusi PIG dapat dihitung dengan rumus sebagai berikut[5]:

$$P(Y = y | \mu) = \frac{\mu^y e^{-\frac{1}{\tau}}}{y! \left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}}} (2\mu\tau + 1)^{-\frac{1}{2}} K_{y-\frac{1}{2}}\left(\frac{1}{\tau}\sqrt{2\mu\tau+1}\right) \quad (2.1)$$

Rata-rata untuk distribusi PIG adalah:

$$E(Y) = E\{E(Y | \mu\tau)\} = E(\mu\tau) = \mu$$

Varians untuk distribusi PIG adalah:

$$Var(Y) = Var\{E(Y | \mu\tau)\} + E\{Var(Y | \mu\tau)\} = \mu + \tau\mu^2$$

B. *Overdispersi*

Overdispersi pada regresi Poisson terjadi ketika varians dari variabel respons lebih besar dari rata-rata.

Uji statistik yang bisa digunakan untuk mendeteksi *overdispersi* pada suatu data adalah uji *overdispersi* yang dapat menggunakan *package* AER dari *software* R. Hipotesis yang digunakan adalah [6]

$$H_0 : \text{var}(Y) = \mu_i$$

$$H_1 : \text{var}(Y) = \mu_i + a.g(\cdot)$$

Dimana $g(\cdot)$ merupakan suatu fungsi tertentu. Secara sederhana, bila nilai $a = 0$ mak dapat dikatakan equidispersi, sebaliknya bila $a > 0$ maka dapat dikatakan *overdispersi*. Nilai koefisien a dapat diperkirakan oleh regresi OLS.

C. *Regresi Poisson Inverse Gaussian*

Model Regresi PIG seperti pada persamaan berikut[7]:

$$\mu_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}} \text{ atau } \ln(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \quad (2.2)$$

dengan

$$\mathbf{x}_i^T = [1 \quad x_{i1} \quad x_{i2} \quad \dots \quad x_{ik}]$$

$$\boldsymbol{\beta} = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \dots \quad \beta_k]^T$$

dimana $i = 1, 2, \dots, n$ menunjukkan nomor observasi.

Dengan fungsi kepadatan peluang sebagai berikut:

$$P(Y = y_i | \mathbf{x}_i; \boldsymbol{\beta}; \tau) = \left\{ \frac{e^{x_i^T \boldsymbol{\beta} y_i} e^{\frac{1}{\tau}}}{y_i!} \left(\frac{2}{\pi \tau} \right)^{\frac{1}{2}} \left(2e^{x_i^T \boldsymbol{\beta}} \tau + 1 \right)^{-\frac{(y_i - \frac{1}{2})}{2}} K_{s_i}(z_i) \right\}$$

D. Estimasi Parameter

Parameter $\boldsymbol{\beta}$ pada regresi PIG ditaksir dengan metode *maximum likelihood*, dengan menentukan fungsi *likelihood* dari distribusi PIG. Fungsi *likelihood* adalah sebagai berikut:

$$L(\boldsymbol{\beta}; \tau) = \prod_{i=1}^n P(Y = y_i | \mathbf{x}_i; \boldsymbol{\beta}; \tau),$$

$$L(\boldsymbol{\beta}; \tau) = \prod_{i=1}^n \left\{ \frac{\mu_i^{y_i} e^{\frac{1}{\tau}}}{y_i!} \left(\frac{2}{\pi \tau} \right)^{\frac{1}{2}} \left(2\mu_i \tau + 1 \right)^{-\frac{(y_i - \frac{1}{2})}{2}} K_{s_i}(z_i) \right\}$$

Fungsi *likelihood* tersebut diubah kedalam bentuk logaritma natural (\ln), sehingga menjadi persamaan sebagai berikut.

$$l(\boldsymbol{\beta}; \tau) = \ln L(\boldsymbol{\beta}; \tau)$$

$$= \sum_{i=1}^n y_i \mathbf{x}_i^T \boldsymbol{\beta} + \frac{n}{\tau} - \ln \left(\sum_{i=1}^n y_i! \right) + \frac{n}{2} \ln \left(\frac{2}{\pi} \right) - \frac{n}{2} \ln \tau - \sum_{i=1}^n \left(\frac{2y_i - 1}{4} \right) \ln(2\mathbf{x}_i^T \boldsymbol{\beta} + 1) + \sum_{i=1}^n \ln K_{s_i}(z_i)$$

Fungsi dimaksimumkan dengan menggunakan *Fisher Scoring Algorithm*, dengan persamaan sebagai berikut[8]:

$$\hat{\boldsymbol{\theta}}_{(r+1)} = \hat{\boldsymbol{\theta}}_{(r)} + \mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}_{(m)}) \mathbf{D}(\hat{\boldsymbol{\theta}}_{(m)}), \quad (2.3)$$

Dimana

$$\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^T, \hat{\tau})^T$$

$$\mathbf{D}(\hat{\boldsymbol{\theta}}) = \left(\frac{\partial l}{\partial \hat{\tau}}, \frac{\partial l}{\partial \hat{\boldsymbol{\beta}}^T} \right)^T$$

$$\mathbf{I}(\hat{\boldsymbol{\theta}}_{(m)}) = -E \left[\mathbf{H}(\hat{\boldsymbol{\theta}}_{(m)}) \right]$$

$$\mathbf{H}(\hat{\boldsymbol{\theta}}_{(m)})_{(k+1)(k+1)} = \left[\begin{array}{cc} \frac{\partial^2 l}{\partial \hat{\tau}^2} & \frac{\partial^2 l}{\partial \hat{\tau} \partial \hat{\boldsymbol{\beta}}^T} \\ \frac{\partial^2 l}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\tau}} & \frac{\partial^2 l}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\boldsymbol{\beta}}^T} \end{array} \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_m}$$

$$\text{Sehingga } \mathbf{I}(\hat{\boldsymbol{\theta}}_{(m)}) = -E \left[\begin{array}{cc} \frac{\partial^2 l}{\partial \hat{\tau}^2} & \frac{\partial^2 l}{\partial \hat{\tau} \partial \hat{\boldsymbol{\beta}}^T} \\ \frac{\partial^2 l}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\tau}} & \frac{\partial^2 l}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\boldsymbol{\beta}}^T} \end{array} \right]$$

Adapun langkah-langkah *Fisher scoring Algorithm* sebagai berikut:

1. Menentukan vektor awal parameter $\hat{\boldsymbol{\theta}}_0$ dengan mengasumsikan data memenuhi model regresi linier berganda:

$\mathbf{Y}^* = \beta_{0(i)} + x_{i1}\beta_{1(i)} + \dots + x_{ip}\beta_{p(i)} + \varepsilon_i$ dimana $i=1,2,K,n$ dan dengan metode kuadrat terkecil diperoleh $\hat{\boldsymbol{\beta}}_{(0)} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}^*$

2. Membentuk vektor gradien $\mathbf{D}(\hat{\boldsymbol{\theta}}_0)$.
3. Membentuk matriks hessian $\mathbf{H}(\hat{\boldsymbol{\theta}}_{(0)})$.
4. Membentuk matriks informasi Fisher $\mathbf{I}(\hat{\boldsymbol{\theta}}_0)$.
5. Memasukkan nilai $\hat{\boldsymbol{\theta}}_{(0)}$ sehingga diperoleh vektor gradien $\mathbf{D}(\hat{\boldsymbol{\theta}}_{(0)})$ dan matriks hessian $\mathbf{H}(\hat{\boldsymbol{\theta}}_{(0)})$
6. Mulai dari $m = 0$ dilakukan iterasi pada $\hat{\boldsymbol{\theta}}_{(r+1)} = \hat{\boldsymbol{\theta}}_{(r)} + \mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}_{(m)}) \mathbf{D}(\hat{\boldsymbol{\theta}}_{(m)})$, nilai $\hat{\boldsymbol{\theta}}_{(m)}$ merupakan sekumpulan penaksir parameter yang konvergen saat iterasi ke- m .
7. Jika belum diperoleh penaksiran parameter yang konvergen saat iterasi ke- m , maka dilanjutkan kembali ke langkah 5 hingga iterasi ke- $m+1$. Iterasi akan berhenti apabila nilai dari $\|\hat{\boldsymbol{\theta}}_{(m+1)} - \hat{\boldsymbol{\theta}}_{(m)}\| \leq \varepsilon$ dan $\varepsilon > 0$ adalah bilangan yang sangat kecil.

E. Pengujian Parameter

Pengujian secara serentak mencakup seluruh parameter $\boldsymbol{\beta}$ secara bersama-sama dengan hipotesis sebagai berikut:

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

$$H_1 : \text{minimal ada satu } \beta_l \neq 0 \text{ dengan } l = 1, 2, K, k$$

$$G = -2 \ln \left(\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right)$$

$$= 2 \left(\ln(L(\hat{\Omega})) - \ln(L(\hat{\omega})) \right)$$

Statistik G adalah pendekatan dari distribusi *chi square* dengan derajat bebas ν sehingga kriteria pengujiannya adalah tolak H_0 apabila $G_{hit} > \chi^2_{(\alpha, \nu)}$.

Pengujian secara parsial untuk parameter β dan τ adalah sebagai berikut. Hipotesis untuk menguji signifikansi parameter $\boldsymbol{\beta}$.

$$H_0 : \beta_l = 0$$

$$H_1 : \beta_l \neq 0 \text{ dengan } l = 1, 2, K, k$$

Statistik uji yang digunakan dalam pengujian signifikansi parameter β adalah sebagai berikut:

$$Z = \frac{\hat{\beta}_l}{SE(\hat{\beta}_l)}$$

Kriteria uji tolak H_0 apabila $|Z_{hit}|$ lebih besar dari nilai $Z_{\alpha/2}$.

Hipotesis untuk menguji signifikansi parameter τ .

$$H_0 : \tau = 0$$

$$H_1 : \tau \neq 0$$

Statistik uji yang digunakan:

$$Z = \frac{\hat{\tau}}{SE(\hat{\tau})}$$

Kriteria pengujianya adalah tolak H_0 apabila $|Z_{hit}|$ lebih besar dari nilai $Z_{\alpha/2}$

F. Uji Korelasi

Korelasi merupakan suatu indikator yang digunakan dalam hubungan linear antar dua variabel [9]. Koefisien korelasi didefinisikan sebagai berikut.

$$r_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Nilai koefisien korelasi berkisar antara -1 sampai 1.

Pengujian hipotesisnya sebagai berikut:

H_0 : Tidak ada hubungan antara kedua variabel.

H_1 : Terdapat hubungan antar kedua variabel.

Statistik uji yang digunakan adalah

$$t = \frac{r_{x,y} \sqrt{n-2}}{\sqrt{1-r^2}}$$

Keputusan tolak H_0 jika $|t_{hit}| > t_{(\frac{\alpha}{2}, n-2)}$

G. Multikolinearitas

Pada pemodelan regresi, korelasi antara variabel-variabel prediktor disebut dengan multikolinearitas. Salah satu cara untuk mendeteksi terjadinya multikolinearitas adalah dengan melihat nilai *Variance Inflation Factor* (VIF).

$$VIF = \frac{1}{1 - R_j^2} \tag{2.4}$$

Apabila nilai VIF lebih dari 10 maka dapat dikatakan terjadi kasus multikolinearitas[10].

H. Pemilihan model terbaik

Akaike Information Criterion (AIC) merupakan salah satu sarana dalam pemilihan model. AIC memperkirakan kualitas masing-masing model, relatif terhadap model lain. Misalkan L adalah nilai maksimum dari fungsi *likelihood* suatu model, dan k adalah jumlah parameter yang diestimasi dalam model tersebut, maka nilai AIC dari model tersebut adalah sebagai berikut:

$$AIC = 2k - 2 \ln(L(\hat{\theta})) \tag{2.5}$$

Apabila diberikan beberapa model untuk sebuah set data, maka model yang lebih baik adalah model dengan AIC kecil [10]

I. HIV

HIV merupakan penyakit menular yang disebabkan oleh infeksi *Human Immunodeficiency Virus*. Virus ini menyebar melalui cairan tubuh, dan menyerang sistem kekebalan tubuh, khususnya sel CD4 atau yang sering disebut sel T [2]. Infeksi tersebut menyebabkan beberapa sel tubuh hancur sehingga penderita mengalami penurunan ketahanan tubuh dan tidak dapat melawan infeksi maupun penyakit lain. Jumlah HIV

positif yang ada di masyarakat dapat diketahui melalui layanan *Voluntary, Counseling, and Testing* (VCT)[12].

III. METODOLOGI PENELITIAN

A. Sumber Data

Penelitian ini menggunakan data sekunder tentang jumlah kasus baru HIV dan faktor-faktor yang mempengaruhi jumlah kasus baru HIV. Data diambil dari Profil Kesehatan Provinsi Jawa Tengah dan Buku Saku Kesehatan 2015 yang dikeluarkan dinas Kesehatan Provinsi Jawa Tengah [3] dan Publikasi hasil Survey Sosial Ekonomi Nasional (Susenas) yang dikeluarkan oleh Badan Pusat Statistik (BPS) Provinsi Jawa Tengah[13]. Data tersebut merupakan data pada tahun 2015 dengan unit pengamatan yang diambil pada tingkat Kabupaten/ Kota di Provinsi Jawa Tengah dengan 35 Kabupaten/ Kota.

B. Variabel Penelitian

Adapun variabel yang digunakan pada praktikum ini adalah sebagai berikut.

Tabel 1. Variabel Penelitian

Notasi	Variabel
Y	Jumlah Kasus Baru HIV
X ₁	Persentase Penduduk Miskin
X ₂	Persentase penduduk dengan pendidikan tertinggi SLTA
X ₃	Persentase PUS yang sedang menggunakan alat KB kondom
X ₄	Rasio jumlah tenaga kesehatan per 100.000
X ₅	Rasio fasilitas kesehatan per 100.000 penduduk
X ₆	Persentase Daerah Perkotaan
X ₇	Persentase penduduk usia 25-34 tahun

C. Langkah Analisis Data

Langkah analisis yang digunakan dalam penelitian sebagai berikut.

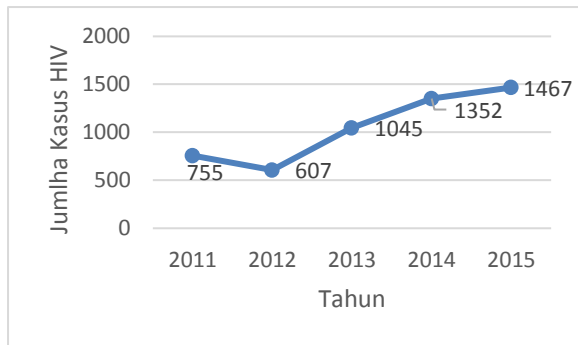
1. Mendeskripsikan karakteristik data dengan statistika deskriptif.
2. Menguji korelasi antara variabel respon dengan variabel prediktor.
3. Melakukan pemeriksaan kasus multikolinieritas dengan menggunakan kriteria VIF.
4. Melakukan uji overdispersi.
5. Menentukan nilai penaksir parameter model Regresi PIG dengan langkah-langkah sebagai berikut:
 - a. Mendapatkan penduga parameter dengan menggunakan *Maximum Likelihood Estimation* (MLE).
 - b. Melakukan pengujian hipotesis untuk Regresi PIG.
 6. Membandingkan nilai AIC untuk mencari model terbaik.
 7. Melakukan interpretasi model PIGR yang didapatkan.
 8. Membuat kesimpulan dari hasil analisis tersebut.

IV. HASIL DAN PEMBAHASAN

A. Statistika Deskriptif

Provinsi Jawa Tengah merupakan provinsi kelima setelah Provinsi Jawa Barat dengan jumlah kasus HIV terbanyak ditemukan. Perkembangan jumlah kasus baru HIV di Provinsi

Jawa Tengah dari tahun 2011 hingga tahun 2015 dapat dilihat pada Gambar 1.



Gambar 1 Perkembangan Jumlah Kasus Baru HIV di Provinsi Jawa Tengah Tahun 2011-2015

Berdasarkan Gambar 1 di atas dapat dilihat perkembangan kasus baru HIV dari tahun 2009 hingga tahun 2015. Pada tahun 2012 jumlah kasus HIV di Provinsi Jawa Tengah menurun sebanyak 148 kasus, namun pada tahun 2013 jumlah kasus baru HIV meningkat menjadi 1045 kasus. Di tahun 2013 hingga tahun 2015 jumlah kasus HIV di Provinsi Jawa Tengah mengalami peningkatan setiap tahunnya. Sebelum melakukan analisis Regresi Poisson Inverse Gaussian adalah mengidentifikasi masing-masing variabel dengan analisis Statistika Deskriptif, didapatkan hasil sebagai berikut.

Tabel 1. Statistika Deskriptif

Var	Mean	Stdev	Variance	Min	Max
Y	41,91	26,45	699,85	3,00	116,00
X ₁	13,031	4,315	18,618	4,970	21,450
X ₂	12,340	5,309	28,182	6,680	25,230
X ₃	2,224	1,635	2,673	0,350	7,410
X ₄	33,89	42,76	1828,63	5,13	222,72
X ₅	3,863	1,591	2,533	2,200	9,940
X ₆	21,74	36,25	1314,27	1,23	100,00
X ₇	7,468	0,667	0,446	5,820	9,090

Pada Tabel 1 dapat dilihat bahwa rata-rata pada tahun 2015 di provinsi Jawa Tengah terdapat 41,91≈42 kasus. Jumlah kasus tertinggi sebanyak 116 kasus di Kota Semarang dan terendah sebanyak 3 kasus di Kota Tegal.

B. Pemeriksaan Korelasi dan Multikolinearitas

Pemeriksaan hubungan antara variabel respon dengan variabel prediktor dilakukan terlebih dahulu sebelum melakukan pemodelan dengan regresi Poisson Inverse Gaussian (PIG).

Berdasarkan Tabel 2 dapat dilihat jika variabel yang berpengaruh dengan variabel respon adalah X₄ dan X₅, namun dikarenakan variabel yang lain secara teori mempengaruhi respon maka variabel yang lain tetap digunakan dalam penelitian ini. Selanjutnya deteksi multikolinearitas dilakukan dengan melihat nilai VIF

Tabel 2.

Koefisien Korelasi Variabel Respon dan Variabel Prediktor

Korelasi	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
X ₁	-0,001						
	0,997						
X ₂	-0,088	-0,692					
	0,614	0,000					
X ₃	-0,037	-0,420	0,766				
	0,834	0,012	0,000				
X ₄	-0,315	-0,467	0,772	0,782			
	0,066	0,004	0,000	0,000			
X ₅	-0,460	-0,336	0,596	0,718	0,850		
	0,005	0,048	0,000	0,000	0,000		
X ₆	-0,229	-0,565	0,799	0,826	0,814	0,767	
	0,186	0,000	0,000	0,000	0,000	0,000	
X ₇	0,184	-0,523	0,613	0,395	0,325	0,123	0,574
	0,290	0,001	0,000	0,019	0,057	0,480	0,000

Nilai VIF dari variabel prediktor dapat dilihat pada Tabel 3.

Tabel 3.

Koefisien Korelasi Variabel Respon dan Variabel Prediktor

Variabel	VIF
X ₁	2,150
X ₂	5,743
X ₃	4,201
X ₄	6,119
X ₅	5,621
X ₆	7,703
X ₇	2,795

Pada Tabel 3 dapat dilihat bahwa tidak terjadi kasus multiko, dikarenakan nilai VIF tidak ada yang melebihi nilai 10.

C. Overdispersi

Pada penelitian ini diuji apakah mengalami overdispersi atau tidak dengan hipotesis sebagai berikut:

$$H_0 : \text{var}(Y) = \mu_i$$

$$H_1 : \text{var}(Y) = \mu_i + a.g(.)$$

Dengan menggunakan package AER pada software R, diperoleh nilai $a = 15,22$ dan $p\text{-value}$ sebesar 0,0001791 lebih kecil dari tingkat signifikansi 10% sehingga tolak H₀ yang dapat disimpulkan bahwa varians tidak sama dengan rata-rata dan berarti bahwa data tersebut mengalami overdispersi.

D. Pemodelan Regresi Poisson Inverse Gaussian

Berdasarkan tujuh variabel yang signifikan pada model regresi PIG, menghasilkan empat kombinasi kemungkinan model regresi PIG yang sudah konvergen, kemudian dicari model terbaiknya. Berikut merupakan empat kemungkinan model PIG adalah sebagai berikut.

$$\mu = e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7)}$$

$$\mu = e^{(\beta_0 + \beta_1 x_1 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7)}$$

$$\mu = e^{(\beta_0 + \beta_1 x_1 + \beta_3 x_3 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7)}$$

$$\mu = e^{(\beta_0 + \beta_3 x_3 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7)}$$

Berikut merupakan estimasi parameter dari model-model yang mungkin menjadi model terbaik dalam regresi PIG yang ditunjukkan pada Tabel 4 sebagai berikut.

Tabel 4

Estimasi Parameter Kemungkinan Model Regresi PIG

Variabel dari Model	β_0	β_1	β_2	β_3
$X_1, X_2, X_3, X_4, X_5, X_6, X_7$	0,742	-0,006	-0,000312	0,392
$X_1, X_3, X_4, X_5, X_6, X_7$	0,765	-0,008		0,395
X_1, X_3, X_5, X_6, X_7	0,847	-0,007		0,386
X_3, X_5, X_6, X_7	0,629			0,385

Tabel 4 (Lanjutan)

Variabel dari Model	β_4	β_5	β_6	β_7	τ
$X_1, X_2, X_3, X_4, X_5, X_6, X_7$	-0,0003	-0,210	-0,021	0,452	-1,360
$X_1, X_3, X_4, X_5, X_6, X_7$	-0,0015	-0,209	-0,021	0,457	-1,324
X_1, X_3, X_5, X_6, X_7		-0,232	-0,022	0,452	-1,314
X_3, X_5, X_6, X_7		-0,230	-0,021	0,466	-1,316

Langkah selanjutnya setelah didapatkan nilai estimasi dari masing-masing perkiraan model yang ditampilkan pada Tabel 4 adalah pengujian hipotesis untuk regresi Poisson Inverse Gaussian.

E. Pengujian Hipotesis

- Pengujian Parameter Secara Serentak

Pengujian parameter secara serentak dilakukan pada kemungkinan model yang sesuai dengan model regresi Poisson Inverse Gaussian. Pengujian parameter secara serentak dapat dilihat dari nilai statistik G dengan hipotesis sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$$

$$H_1 : \text{paling sedikit ada } \beta_i \neq 0 \text{ dengan } i=1,2,K,7$$

$$\alpha=0,1$$

Tabel 5

Pengujian Parameter Regresi PIG Secara Serentak

Variabel dari Model	Statistik G	v	$\chi^2_{(\alpha,v)}$	Keputusan
$X_1, X_2, X_3, X_4, X_5, X_6, X_7$	299,146	27	36,741	Tolak H_0
$X_1, X_3, X_4, X_5, X_6, X_7$	299,1545	28	37,916	Tolak H_0
X_1, X_3, X_5, X_6, X_7	299,2379	29	39,087	Tolak H_0
X_3, X_5, X_6, X_7	299,3138	30	40,456	Tolak H_0

Tabel 5 menunjukkan bahwa dari semua kemungkinan model, didapatkan hasil statistik G lebih dari $\chi^2_{(\alpha,v)}$ maka keputusannya adalah tolak H_0 , yang berarti minimal terdapat satu parameter yang berpengaruh signifikan terhadap model. Untuk mengetahui variabel prediktor yang berpengaruh signifikan terhadap model, maka dilanjutkan pada pengujian parameter secara individu.

- Pengujian Parameter Secara Individu

Pengujian parameter secara individu digunakan untuk mencari variabel prediktor yang berpengaruh signifikan terhadap jumlah kasus baru HIV di Provinsi Jawa Tengah dengan hipotesis sebagai berikut.

- Parameter β

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

$$\alpha=0,1$$

- Parameter τ

$$H_0 : \tau = 0$$

$$H_1 : \tau \neq 0$$

$$\alpha = 0,1$$

Pengujian parameter secara individu dapat dilihat pada Tabel 6 sebagai berikut.

Tabel 6

Pengujian Parameter Regresi PIG Secara Individu

Variabel dari Model	Parameter Signifikan
$X_1, X_2, X_3, X_4, X_5, X_6, X_7$	$\beta_3, \beta_6, \beta_7, \tau$
$X_1, X_3, X_4, X_5, X_6, X_7$	$\beta_3, \beta_6, \beta_7, \tau$
X_1, X_3, X_5, X_6, X_7	$\beta_3, \beta_5, \beta_6, \beta_7, \tau$
X_3, X_5, X_6, X_7	$\beta_3, \beta_5, \beta_6, \beta_7, \tau$

Tabel 6 menunjukkan bahwa model dengan parameter yang signifikan adalah model yang memenuhi daerah kritis atau memiliki *p-value* kurang dari taraf signifikansi yaitu $\alpha=0,1$ sehingga keputusan tolak H_0 yang berarti parameter β_i berpengaruh signifikan terhadap model atau dengan melihat nilai $|z_{hitung}|$ yang dibandingkan dengan $z_{\alpha/2} = 1,64$.

F. Pemilihan Model Terbaik

Metode *backward elimination* dilakukan berdasarkan nilai *Akaike Information Criterion* (AIC). Pada pemodelan regresi poisson inverse gaussian didapatkan nilai AIC di bawah ini

Tabel 7

Nilai AIC dari Model Regresi PIG

Model	AIC
$X_1, X_2, X_3, X_4, X_5, X_6, X_7$	317,15
$X_1, X_3, X_4, X_5, X_6, X_7$	315,15
X_1, X_3, X_5, X_6, X_7	313,24
X_3, X_5, X_6, X_7	311,31

Pada Tabel 6 terlihat bahwa nilai AIC yang paling kecil yaitu model yang mengandung variabel $X_3, X_5, X_6,$ dan X_7 . Sehingga model yang digunakan adalah model ke 4. Berdasarkan model tersebut, dengan menggunakan *package* *gamlss* yang tersedia pada *software* R didapatkan hasil yang disajikan pada Tabel 8.

Tabel 8

Penaksiran Parameter Model Regresi PIG pada Jumlah Kasus Baru HIV di Provinsi Jawa Tengah Tahun 2015

Parameter	Taksiran	Standard Error	Z Hitung	P-value
β_0	0,629965	1,788147	0,352	0,72718
β_3	0,385849	0,108118	3,569	0,00127*
β_5	-0,23026	0,116140	-1,983	0,05695*
β_6	-0,02185	0,007278	-3,003	0,00546*
β_7	0,466977	0,212157	2,201	0,03584*
τ	-1,3161	0,3079	-4,274	0,000189*

*) Signifikan dengan taraf signifikansi $\alpha = 0,1$

Berdasarkan Tabel 8 Hasil dari penaksiran parameter diperoleh model regresi PIG sebagai berikut

$$\hat{\mu} = \exp(a + bX_3 - cX_5 - dX_6 + eX_7)$$

dimana

$$a = 0,629965$$

$$b = 0,385849$$

$$c = 0,230265$$

$$d = 0,021858$$

$$e = 0,466977$$

Berdasarkan model di atas, maka dapat diinterpretasikan sebagai berikut ini. Setiap penambahan 1 persen variabel X_3 maka akan meningkatkan rata-rata variabel respon Y sebesar $\exp(0,385849)=1,47086$ kali dari rata-rata variabel respon semula bila variabel lain tetap. Jika pengurangan 1 persen variabel X_3 akan menurunkan rata-rata variabel respon sebesar $\exp(-0,385849)=0,679873$ kali.

Terlihat bahwa kasus kenaikan HIV dengan kenaikan pengguna kondom sejalan, hal ini memperlihatkan juga bahwa penderita HIV telah menggunakan kondom pada saat terkena HIV agar tidak tertular pada pasangan. Dimungkinkan hal tersebut dilakukan atas saran dokter, karena tujuan dari pasangan usia subur menggunakan kondom sebenarnya untuk tujuan kelahiran. Fluktuasi yang terjadi merupakan akibat dari elastisitas μ . Oleh karena alasan tersebut maka interpretasi pada model ini tidak dapat dilihat sebagai sebab akibat.

Setiap penambahan 1 persen dari variabel X_5 maka akan meningkatkan rata-rata variabel respon Y sebesar $\exp(-0,230265) = 0,79432$ kali dari rata-rata variabel respon semula bila variabel lain tetap. Dengan kata lain, penambahan 1 rasio persentase fasilitas kesehatan maka akan sebanding dengan penurunan rata-rata jumlah kasus baru HIV sebesar 0,79432 kali dari rata-ratanya semula bila variabel lain tetap.

Setiap penambahan 1 persen variabel X_6 maka akan melipatgandakan rata-rata variabel respon Y sebesar $\exp(-0,021858)=0,97838$ kali dari rata-rata variabel respon semula bila variabel lain tetap. Dengan kata lain, penambahan 1 persen dari persentase daerah perkotaan akan sebanding dengan penurunan rata-rata jumlah kasus baru HIV sebesar 0,97838 kali dari rata-ratanya semula bila variabel lain tetap.

Setiap penambahan 1 persen variabel X_7 maka akan melipatgandakan rata-rata variabel respon Y sebesar $\exp(0,466977)=1,595164714$ kali dari rata-rata variabel respon semula bila variabel lain tetap. Dengan kata lain, penambahan 1 persen dari persentase penduduk usia 25-34 tahun maka akan sebanding dengan kenaikan rata-rata jumlah kasus baru HIV sebesar 1,595164714 kali dari rata-ratanya semula bila variabel lain tetap.

Masalah keterbatasan data yang tersedia menyebabkan beberapa interpretasi dari model yang terbentuk tidak sesuai dengan teori yang berlaku tentang HIV, sehingga interpretasi tersebut tidak dapat dilihat sebagai hubungan sebab akibat. Data jumlah kasus baru HIV yang digunakan dalam penelitian ini merupakan data yang berasal dari hasil laporan klinik-klinik VCT di Provinsi Jawa Tengah dimana pasien datang secara sukarela untuk memeriksakan diri ke klinik tersebut.

V. KESIMPULAN

Berdasarkan hasil analisis yang dilakukan disimpulkan bahwa pada tahun 2015 jumlah kasus baru HIV di Provinsi Jawa Tengah mengalami peningkatan sebesar 115 kasus dibandingkan tahun sebelumnya. Jumlah kasus baru HIV tertinggi adalah di Kota Semarang. Setelah dilakukan uji overdispersi pada data jumlah kasus baru HIV di Provinsi Jawa Tengah tahun 2015 dinyatakan bahwa data mengalami overdispersi. Oleh karena itu penelitian dengan menggunakan metode Regresi Poisson Inverse Gaussian dapat dilakukan. Model regresi Poisson Inverse Gaussian (PIG) yang terbentuk adalah model regresi dengan variabel-variabel prediktor yang signifikan yaitu persentase PUS yang menggunakan kondom (X_3), rasio fasilitas kesehatan (X_5), persentase daerah perkotaan (X_6), dan persentase penduduk usia 25-34 tahun (X_7). Berikut model Regresi Poisson Inverse Gaussian yang terbentuk:

$$\hat{\mu} = \exp(a + bX_3 - cX_5 - dX_6 + eX_7)$$

dimana

$$a = 0,629965$$

$$b = 0,385849$$

$$c = 0,230265$$

$$d = 0,021858$$

$$e = 0,466977$$

DAFTAR PUSTAKA

- [1] UNAIDS. (2016). *Global AIDS Update 2016*. Geneva: WHO.
- [2] CDC. (2016). *HIV/AIDS*. <http://www.cdc.gov/hiv/statistics/index.html>. Tanggal Akses: 5 Oktober 2016.
- [3] Dinas Kesehatan Provinsi Jawa Tengah. (2015). *Data Saku Kesehatan Provinsi Jawa Tengah Tahun 2015*. Semarang: Dinkes Jateng.
- [4] Consul, P.C. dan Famoye, F. (1992). "Generalized Poisson Regression Model", *Communication in Statistics – Theory and Methods*. Vol. 21, No.1, hal. 89-109.
- [5] Widiari, S. M. (2016). *Penaksiran Parameter Dan Statistik Uji Dalam Model Regresi Poisson Inverse Gaussian (PIG) Studi Kasus: Jumlah Kasus Baru HIV di Provinsi Jawa Timur Tahun 2013*. Tesis. Mahasiswa Jurusan Statistika Institut Teknologi Sepuluh Nopember Surabaya.
- [6] Cameron, A. C. Dan Trivedi, P. K. (1990). "Regression-Based Test For Overdispersion In The Poisson Model", *Journal of Econometrics*, Vol. 46, No. 1, hal 347-346.
- [7] Purnamasari, I. (2016). *Penaksiran Parameter Dan Statistik Uji Dalam Model Regresi Geographically Weighted Poisson Inverse Gaussian*. Tesis. Mahasiswa Jurusan Statistika Institut Teknologi Sepuluh Nopember Surabaya.
- [8] Widiari, S. M. (2016). *Penaksiran Parameter Dan Statistik Uji Dalam Model Regresi Poisson Inverse Gaussian (PIG) Studi Kasus: Jumlah Kasus Baru HIV di Provinsi Jawa Timur Tahun 2013*. Tesis. Mahasiswa Jurusan Statistika Institut Teknologi Sepuluh Nopember Surabaya.
- [9] Draper, N. dan Smith, H. (1992). *Analisis Regresi Terapan*. Jakarta: Gramedia.
- [10] Setiawan, dan Kusri, D. E. (2010). *Ekonometrika*. Yogyakarta: C.V. Andi Offset.
- [11] Akaike, H. (1978). *A Bayesian Analysis of The Minimum AIC Procedure*. Annals of the Institute of Statistical Mathematics, Part A Hal. 914. <http://www.ism.ac.jp/editsec/aism/pdf/> Tanggal Akses: 1 Oktober 2016.
- [12] Dinas Kesehatan Provinsi Jawa Tengah. (2013). *Profil Kesehatan Provinsi Jawa Tengah Tahun 2013*. Semarang: Dinkes Jateng.
- [13] BPS Jateng. (2016). *Statistik Sosial dan Kependudukan Jawa Tengah Hasil Susenas 2015*. Semarang: BPS Jateng.