Penerapan *Generalized Poisson Regression* I Untuk Mengatasi Overdispersi Pada Regresi Poisson (Studi Kasus: Pemodelan Jumlah Kasus Kanker Serviks di Provinsi Kalimantan Timur)

Application Generalized Poisson Regression I to Handle Overdispersion on Poisson Regression Models (Case Study: Modeling Number of Cases of Cervical Cancer in East Kalimantan Province)

Iim Masfian N[1], Desi Yuniarti[2], dan Memi Nor Hayati[3]
[1]Mahasiswa Program Studi Statistika FMIPA Universitas Mulawarman
[2,3]Dosen Program Studi Statistika FMIPA Universitas Mulawarman
Email: iimmasfian@gmail.com[1]

Abstract

Poisson Regression model is commonly used to analyze count data is assumed to have Poisson distribution where the mean and variance values are equal or also called equdispersion. In fact, this assumption is often violated, because the value of variance is greater than the mean value, this condition is called overdispersion. Poisson regression which is applied to the data that contains overdispersion will imply the value of standard error becomes underestimates, so the conclusion is not valid. One of the models that can be used for overdispersion data is Generalized Poisson Regression I (GPR I). This research discuss the handling of overdispersion on Poisson regression using GPR I, with case study modeling the number of cervical cancer cases in East Kalimantan in 2013. In this research GPR I models meet the criteria for suitability of regression compared Poisson regression models because it has a smaller AIC value.

Keywords: AIC, GPR I, Overdispersion, Poisson Regression.

Pendahuluan

Analisis regresi yang umumnya digunakan adalah analisis regresi klasik, dimana variabel responnya merupakan data kontinu yang mengikuti distribusi normal. Namun dalam perkembangannya, model regresi klasik ini tidak mampu mengatasi permasalahan-permasalahan dimana variabel respon berupa data diskrit dan tidak berdistribusi normal. Data diskrit yaitu data yang nilainya nonnegatif dan menyatakan banyaknya kejadian dalam interval waktu, ruang, atau volume tertentu. Ketika variabel respon berupa data diskrit, analisis regresi yang biasa digunakan adalah analisis regresi Poisson (Berk dan MacDonald, 2008).

Regresi Poisson merupakan metode regresi nonlinier yang digunakan untuk menganalisis data yang variabel responnya berupa data diskrit dan berasal dari distribusi Poisson. Pada regresi Poisson terdapat asumsi yang harus dipenuhi, yaitu nilai variansi dan rata-rata dari variabel respon tersebut sama. Jika nilai variansi lebih besar dari rata-ratanya atau disebut dengan overdispersi, maka hal itu akan berdampak pada nilai *standard error* yang menjadi *underestimate* (McCullagh dan Nelder, 1989).

Terdapat beberapa metode yang dapat digunakan untuk mengatasi overdispersi salah satunya yaitu dengan metode *Generalized Poisson Regreession* I (GPR I) yang merupakan perluasan dari regresi Poisson (Ismail dan Jemain, 2007). Penelitian tentang GPR I pernah dilakukan oleh Nohe (2011), mengenai faktor-faktor yang mempengaruhi kematian ibu melahirkan (*maternal mortality*) di Kalimantan Timur tahun 2009, namun dalam penelitian tersebut tidak menggunakan nilai AIC sebagai penentuan model terbaik.

Berdasarkan uraian di atas maka penulis tertarik untuk melakukan penelitian tentang cara mengatasi overdispersi dengan menggunakan GPR I. Kasus yang digunakan dalam penelitian ini adalah jumlah kasus kanker serviks di Provinsi Kalimantan Timur tahun 2013.

Uji Kecocokan Distribusi (*Kolmogorov-Smirnov*)

Uji *Kolmogorov-Smirnov* digunakan untuk memutuskan apakah sampel berasal dari populasi dengan distribusi tertentu. Uji *Kolmogorov-Smirnov* merupakan pengujian statistik nonparametrik (Djarwanto, 2003).

$$D = \sup_{Y} \left| F_0(Y) - S_N(Y) \right| \quad (1)$$

Dengan daerah penolakan adalah tolak H_0 jika $D > D_{tabel}$ dimana $D_{tabel} = D_{(\sim;n)}$

Distribusi Poisson

Distribusi Poisson termasuk distribusi teoritis yang menggunakan variabel random diskrit. Distribusi Poisson disebut juga distribusi peristiwa yang jarang terjadi, ditemukan oleh S.D. Poisson (1781 – 1841), seorang ahli matematika bangsa Perancis (Hasan, 2003).

Menurut Walpole dan Myers (1990), percobaan yang menghasilkan nilai- nilai bagi variabel random Y yang bernilai numerik, yaitu banyaknya sukses selama interval waktu tertentu atau dalam daerah tertentu, disebut percobaan Poisson. Panjang interval waktu tersebut dapat berapa saja, semenit, sehari, seminggu atau setahun. Daerah yang dimaksud dapat berupa sepotong garis, suatu luas, suatu isi, ataupun barangkali sepotong benda.

Variabel random Y yang menyatakan banyaknya kejadian dalam interval waktu, ruang atau volume tertentu dikatakan berdistribusi Poisson jika memiliki fungsi probabilitas sebagai berikut (Walpole dan Myers,1990):

$$f(y;\text{«}) = \frac{\exp(-\text{«})\text{«}^{y}}{y!}, \qquad y = 0,1,2,....;\text{«} > 0 \quad (2)$$

dengan $E(Y) = Var(Y) = \text{«}$

Generalized Linear Models

Analisis regresi yang responnya termasuk salah satu anggota keluarga eksponensial disebut Generalisasi Model linier atau lebih dikenal dengan *Generalized Linear Models* (GLM). GLM memperluas model regresi biasa yang mencakup variabel respon berdistribusi tidak normal dan fungsi model untuk rata-rata. Ada tiga komponen utama dalam analisis GLM, yaitu sebagai berikut (Agresti, 1990):

1. Komponen Random
   variabel respon Y dengan observasi bebas $(Y_1, Y_2,..., Y_n)$ dari sebuah distribusi dalam keluarga eksponensial. Bentuk fungsi densitas probabilitas dari distribusi keluarga eksponensial adalah sebagai berikut:
   $$f(y;\text{¯},\text{¤}) = \exp\left\{\frac{y\text{¯} - b(\text{¯})}{a(\text{¤})} + c(y;\text{¤})\right\} \quad (3)$$
2. Komponen sistematis
   Komponen sistematis dari GLM adalah hubungan dari sebuah vektor untuk menjelaskan variabel-variabel yang berhubungan dalam sebuah model linier.
   $$\text{¦}_i = \;_0 + \sum_{k}^{p} \;_k x_{ki} \quad (4)$$
   Atau dalam matriks dituliskan dalam bentuk:
   $$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$$
   dimana:

$$\boldsymbol{\eta} = \begin{bmatrix} \text{¦}_1 \\ \text{¦}_2 \\ \vdots \\ \text{¦}_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} _0 \\ _1 \\ _2 \\ \vdots \\ _p \end{bmatrix} \quad (5)$$

3. Fungsi *Link*
   Fungsi *link* adalah fungsi yang menghubungkan ekspektasi variabel respon dengan komponen sistematis dan dituliskan sebagai berikut:
   $$g(\text{«}_i) = \text{¦}_i = \sum_{k}^{p} \;_k x_{ki}$$

Regresi Poisson

Regresi Poisson merupakan analisis regresi yang biasanya digunakan untuk data dengan respon berupa variabel diskrit tetapi tidak biner. Dalam hal ini respon data tersebut berdistribusi Poisson dengan parameter μ. Hal yang sangat penting untuk dicatat bahwa parameter μ ini sangat bergantung pada beberapa unit tertentu atau periode dari waktu, jarak, luas area, volume, dan sebagainya. Distribusi ini kemudian digunakan untuk memodelkan suatu peristiwa yang keberadaannya relatif jarang atau langka untuk terjadi pada satuan unit tertentu. Regresi Poisson memiliki beberapa asumsi sebagai berikut (Pateta, 2005):

1. Variabel respon berupa data diskrit.
2. Distribusi bersyarat dari variabel respon mengikuti distribusi Poisson.
3. Nilai rata-rata akan sama dengan variansinya, yakni $E(Y) = Var(Y)$.
4. Tidak terjadi masalah Overdispersi.

Model Regresi Poisson

Pada model regresi Poisson, fungsi penghubung yang digunakan adalah fungsi penghubung *log* karena fungsi *log* menjamin bahwa nilai variabel yang diharapkan dari variabel responnya akan bernilai nonnegatif (Myers, 1996).

$$\text{¦}_i = \ln(\text{«}_i) = \;_0 + \;_1 x_{1i} + \;_2 x_{2i} + ... + \;_p x_{pi}, \; i = 1,2,.., p \quad (6)$$

Persamaan di atas dapat pula dituliskan sebagai:

$$E(y_i) = \text{«}_i = \exp\,(\mathbf{X}_i'\boldsymbol{\beta}) \quad (7)$$

Penaksiran Parameter Regresi Poisson

Untuk menaksir parameter $\boldsymbol{\beta}$ digunakan metode *Maximum Likelihood Estimation* (MLE). Prinsip dari metode ini adalah mencari taksiran maksimum *likelihood* dari parameter $\boldsymbol{\beta}$, yaitu taksiran dari parameter $\boldsymbol{\beta}$ yang memaksimumkan fungsi *likelihood*. Berikut adalah turunan pertama dari fungsi *likelihood* berdasarkan persamaan :

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} f(y_i, \boldsymbol{\beta})$$

$$= \prod_{i=1}^{n} \left\{ \frac{\exp(-\mu)\mu^{y}}{y_i!} \right\}$$

$$= \prod_{i=1}^{n} \left\{ \frac{[\exp(-\exp(\beta_0 + \sum_{k=1}^{p} \beta_k x_{ki}))][\exp(\beta_0 + \sum_{k=1}^{p} \beta_k x_{ki})^{y_i}}{y_i!} \right\} \quad (8)$$

Untuk mempermudah perhitungan dalam memperoleh taksiran maksimum likelihood dari parameter $\boldsymbol{\beta}$, digunakan bentuk logaritma dari fungsi *likelihood* pada persamaan di atas sebagai berikut:

$$\ln L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \begin{array}{l} [-\exp(\beta_0 + \sum_{k=1}^{p} \beta_k x_{ki})] \\ + y_i(\beta_0 + \sum_{k=1}^{p} \beta_k x_{ki}) - \ln(y_i)! \end{array} \right\} \quad (9)$$

Proses untuk menemukan solusi dari turunan fungsi log *likelihood* tidak dapat dilakukan secara langsung karena fungsi log *likelihood* tidak linier dalam parameter yang ingin ditaksir sehingga membutuhkan metode numerik Newton-Raphson untuk menyelesaikannya (Myers, 1996).

Uji Rasio *Likelihood*

Uji Rasio *likelihood* atau disebut juga uji devians adalah uji yang digunakan untuk mengetahui signifikansi parameter yang dibentuk secara menyeluruh oleh keseluruhan variabel prediktor yang signifikan berpengaruh terhadap variabel respon.

$$G = -2\ln\left(\frac{L(\hat{\mu})}{L(\hat{\Omega})}\right) \quad (10)$$

H_0 ditolak pada taraf signifikansi α jika $G > \chi^2_{(\alpha;(n-k-1))}$, dengan n adalah banyak pengamatan dan k adalah banyak variabel prediktor (Myers, 1996).

Uji *t*

Uji t menunjukkan seberapa jauh pengaruh satu variabel prediktor secara individual dalam menerangkan variasi variabel respon (Widarjono, 2007).

$$t = \left(\frac{\hat{\beta}_k}{SE(\hat{\beta}_k)}\right) \quad (11)$$

Kriteria pengujian yaitu H_0 ditolak jika $t_{hitung} > t_{tabel}$ dimana $t_{tabel} = t_{(\alpha/2;n-k-1)}$ atau tolak H_0 jika *p-value* < α.

Overdispersi

Menurut McCullagh dan Nelder (1989), regresi Poisson dikatakan mengandung overdispersi apabila nilai variansnya lebih besar dari nilai rata-ratanya. Overdispersi memiliki dampak yang sama dengan pelanggaran asumsi jika pada data diskrit terjadi overdispersi namun tetap digunakan regresi Poisson, 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 dapat dituliskan *Var(Y) > E(Y)*.

Hubungan parameter dispersi (ϕ) dengan varians dan rata-rata dalam regresi Poisson adalah:

$$Var(Y) = \phi\mu \quad (12)$$

Perhitungan nilai dispersi dengan menggunakan *Pearson Chi-Square* adalah:

$$\phi = \frac{\chi^2}{df} = \frac{Pearson\ Chi-Square}{df} \quad (13)$$

Distribusi *Generalized Poisson*

Menurut Wang dan Famoye dalam Ismail dan Jemain (2007), fungsi kepadatan peluang distribusi GP adalah:

$$f(y_i) = \left(\frac{\mu_i}{1 + a\mu_i}\right)^{y_i} \frac{(1 + ay_i)^{y_i - 1}}{y_i!} \quad (14)$$

$$\exp\left(-\frac{\mu_i(1 + ay_i)}{1 + a\mu_i}\right) \qquad y_i = 0,1,2...$$

dengan rata-ratanya adalah $E(Y_i) = \mu_i$,variansi $Var(Y_i) = \mu_i(1 + a\mu_i)^2$ dan a adalah parameter dispersi.

Generalized Poisson Regression I

Generalized Poisson Regression I (GPR I) merupakan suatu model regresi yang digunakan untuk menganalisis hubungan antara sebuah variabel respon yang berupa data diskrit dengan satu atau lebih variabel prediktor. Model GPR I dapat digunakan baik dalam keadaan underdispersi, equidispersi, ataupun overdispersi. Model ini juga memiliki parameter dispersi yang berguna untu menggambarkan variasi dari data dan dinotasikan dengan a. Fungsi *likelihood* untuk GPR I dapat dituliskan sebagai berikut (Ismail dan Jemain, 2007):

$$L(\boldsymbol{\beta}, a) = \prod_{i=1}^{n} \left(\frac{\exp({}_0 + \sum_{k=1}^{p} {}_k x_{ki})}{1 + a\exp({}_0 + \sum_{k=1}^{p} {}_k x_{ki})} \right)^{y_i} \quad (15)$$

$$\frac{(1+ay_i)^{y_i - 1}}{y_i\,!} \exp\left(- \frac{\exp({}_0 + \sum_{k=1}^{p} {}_k x_{ki})(1+ay_i)}{1 + a\exp({}_0 + \sum_{k=1}^{p} {}_k x_{ki})} \right)$$

Untuk melakukan perhitungan, digunakan bentuk logaritma dari persamaan untuk mencari taksiran $\boldsymbol{\beta}$ dan a yaitu:

$$\ln L(\boldsymbol{\beta}, a) = \ln \prod_{i=1}^{n} \left\{ \left(\frac{\exp({}_0 + \sum_{k=1}^{p} {}_k x_{ki})}{1 + a\exp({}_0 + \sum_{k=1}^{p} {}_k x_{ki})} \right)^{y_i} \quad (16)$$

$$\frac{(1+ay_i)^{y_i - 1}}{y_i\,!} \exp\left(- \frac{\exp({}_0 + \sum_{k=1}^{p} {}_k x_{ki})(1+ay_i)}{1 + a\exp({}_0 + \sum_{k=1}^{p} {}_k x_{ki})} \right) \right\}$$

karena fungsi log *likelihood* pada persamaan (16) tidak linier terhadap masing-masing parameternya maka untuk mencari nilai estimasi dari ${}_0,..,{}_k$ dan a digunakan metode numerik yaitu metode Newton Raphson dengan menggunakan nilai turunan pertama dan kedua dari fungsi log *likelihood* GPR I sebagai dasar estimasi.

Kanker Serviks

Kanker Serviks (*Cervical Cancer*) 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. 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 di antara 2 siklus menstruasi, menoragia, atau bercak/perdarahan *postmenopause*.

Ciri-ciri penderita penyakit kanker serviks adalah (Maysaroh, 2013):

1. Pada saat menstruasi, darah yang keluar dalam jumlah yang banyak dan berlebih.
2. Mengalami sakit saat buang air kecil
3. Sering merasakan sakit pada daerah pinggul
4. Saat berhubungan intim selalu merasakan sakit, bahkan sering diikuti oleh adanya perdarahan.
5. Saat perempuan mengalami stadium lanjut akan mengalami rasa sakit pada bagian paha atau salah satu paha akan mengalami bengkak, nafsu makan menjadi berkurang, berat badan tidak stabil, susah buang air kecil,dan mengalami perdarahan spontan.

Faktor-faktor yang mempengaruhi kanker serviks diantaranya (Rasjidi, 2010):

1. Paritas
   Paritas adalah banyaknya kelahiran hidup yang dipunyai oleh seorang wanita atau jumlah kelahiran yang menghasilkan janin yang mampu hidup diluar rahim (28 minggu). Pada wanita yang sering melahirkan per vagina, dimana melahirkan anak lebih dari tiga kali akan mempertinggi risiko kanker serviks.
2. Kontrasepsi Hormon dan IUD
3. Kondom dan diafragma dapat memberikan perlindungan. Kontrasepsi hormon dan IUD yang dipakai dalam jangka panjang yaitu lebih dari 4 tahun dapat meningkatkan risiko 1,5-2,5 kali.
4. Petugas Kesehatan dan Sarana Kesehatan
5. Dengan adanya dukungan petugas kesehatan seperti dokter dan bidan serta adanya fasilitas kesehatan diharapkan dapat menghambat pertumbuhan kanker serviks dengan memberikan penyembuhan atau penyuluhan.
6. Sosial Ekonomi
7. Kanker mulut rahim salah satunya disebabkan karena ketidaktahuan atau rendahnya pengetahuan tentang pencegahan terhadap terjadinya kanker serviks. Hal itu diakibatkan oleh faktor sosial ekonomi yang rendah.

Faktor lain yang berhubungan dengan kanker mulut rahim ini adalah aktivitas seksual yang terlalu muda (< 16 tahun), jumlah pasangan seksual yang banyak (> 4 orang), dan adanya riwayat pernah menderita kandiloma.

Metodologi Penelitian

Adapun variabel yang digunakan dalam penelitian ini adalah:

1. Variabel respon (Y) adalah data jumlah kasus kanker serviks
2. Variabel prediktor (X) adalah:
   X_1 = Persentase sarana kesehatan (Rumah Sakit dan Puskesmas)
   X_2 = Persentase tenaga medis
   X_3 = Persentase penduduk perempuan yang umur kawin pertama ≤ 16 tahun
   X_4 = Persentase penduduk miskin
   X_5 = Persentase penduduk perempuan yang menggunakan kondom
   X_6 = Persentase rata-rata pengeluaran untuk konsumsi makanan perbulan

X_7 = Persentase perempuan dengan jumlah anak yang dilahirkan > 4.

Adapun tahap-tahap dalam analisis sebagai berikut:

1. Melakukan analisis statistika deskriptif.
2. Melakukan pengujian distribusi Poisson pada variabel respon dengan menggunakan uji *Kolmogorov-Smirnov.*
3. Menentukan model awal regresi Poisson.
4. Melakukan estimasi parameter regresi Poisson. Metode yang digunakan untuk mengestimasi parameter model regresi Poisson adalah metode maksimum *likelihood*.
5. Melakukan pengujian parameter regresi poisson secara simultan dengan uji rasio *likelihood*.
6. Melakukan pemilihan variabel prediktor (X) yang berpengaruh terhadap variabel prediktor (Y) untuk menentukan model terbaik regresi Poisson dengan menggunakan metode *backward*.
7. Mendeteksi adanya kasus overdispersi.
8. Menentukan model awal GPR I.
9. Melakukan estimasi parameter GPR I. Metode yang digunakan untuk mengestimasi parameter model GPR I adalah metode maksimum *likelihood*.
10. Melakukan pengujian parameter GPR I secara simultan dengan uji rasio *likelihood*
11. Melakukan pemilihan variabel prediktor (X) yang berpengaruh terhadap variabel prediktor (Y) untuk menentukan model terbaik regresi Poisson dengan menggunakan metode *backward*.

Hasil dan Pembahasan

Deskripsi Variabel Respon

Sebelum melakukan analisis regresi Poisson, hal yang perlu dilakukan terlebih dahulu ialah analisis statistika deskriptif yang bertujuan menampilkan karakteristik dari data berdasarkan ukuran-ukuran nilai statistik.

Tabel 1 Analisis Statistik Deskriptif

Analisis Statistika Deskriptif					
Var	N	Min	Max	Rata-Rata	Variansi
Y	14	4	27	12,57	60,571

Pengujian Distribusi Poisson

Pengujian distribusi Poisson pada variabel respon dilakukan dengan menggunakan uji *Kolmogorov-Smirnov* dengan menggunakan persamaan (1) sebagai berikut:

$$D = \sup_{Y} \left| F_0(Y) - S_N(Y) \right|$$
$$= 0{,}30398$$

Karena nilai $D > D_{(0{,}05;14)}$ = 0,349 maka dapat disimpulkan bahwa data jumlah kasus kanker serviks berdistribusi Poisson.

Model Awal Regresi Poisson

$$\ln(\hat{\mu}) = 4{,}437160 + 0{,}057966\,X_1 - 0{,}003478X_2$$
$$+0{,}079881X_3 - 0{,}188818X_4 - 0{,}161972X_5$$
$$-0{,}036401X_6 - 0{,}063498X_7$$

Pengujian Parameter Regresi Poisson

Pengujian Parameter Secara Simultan (Uji Rasio *Likelihood*)

Dengan menggunakan persamaan (10), nilai rasio likelihood yang diperoleh yaitu:

$$G = -2\ln\left(\frac{L(\hat{\mu})}{L(\hat{\Omega})}\right) = 13{,}4504$$

Karena nilai $G = 13{,}4504 > \chi^2_{tabel} = 12{,}592$ maka dapat disimpulkan bahwa minimal terdapat satu variabel yaitu sarana kesehatan (rumah sakit dan puskesmas), tenaga medis, penduduk perempuan yang umur kawin pertama ≤ 16 tahun, penduduk miskin, penduduk perempuan yang menggunakan kondom, rata-rata pengeluaran untuk konsumsi makanan perbulan, dan perempuan dengan jumlah anak yang dilahirkan > 4 berpengaruh terhadap jumlah kasus kanker serviks.

Pengujian Parameter Secara Parsial (Uji *t*)

Pengujian parameter secara parsial ini bertujuan untuk melihat apakah ada tidaknya pengaruh signifikan variabel prediktor terhadap variabel respon. Adapun hasil pengujian terdapat pada Tabel 2.

Tabel 2 Hasil Pengujian Secara Parsial

Variabel	Koefisien	*P-value*
Konstanta	4,6442	0,0282
X_1	4,437160	0,1417
X_2	0,057966	0,8233
X_3	-0,003478	0,0285
X_4	0,079881	0,0203
X_5	-0,188818	0,2642
X_6	-0,161972	0,3268
X_7	-0,036401	0,1793

Berdasarkan Tabel 2, dapat dilihat bahwa variabel X_1 , X_2 , X_5 , X_6 , dan X_7 tidak berpengaruh terhadap Y karena memliki *p-value* > α = 0,05. Oleh karena itu harus dikeluarkan dengan menggunakan teknik pemilihan variabel. Dalam penelitian ini menggunakan metode *backward*. Setelah dikeluarkan satu persatu maka diperoleh hasil pada Tabel 3.

Tabel 3 Hasil Setelah Variabel Dikeluarkan

Variabel	Koefisien	*P-value*
X_1	0,060947	0,0419
X_3	0,085021	0,0118
X_4	-0,207351	0,0021

Berdasarkan Tabel 3, dapat dilihat bahwa variabel X_1 , X_3 dan X_4 berturut-turut memilik *p-value* sebesar 0,0419, 0,0118, dan 0,0021 < α = 0,05 sehingga dapat diputuskan untuk menolak H_0 dan dapat disimpulkan bahwa persentase sarana kesehatan (Rumah Sakit dan Puskesmas), persentase penduduk perempuan yang umur kawin pertama ≤ 16 tahun dan persentase penduduk miskin berpengaruh terhadap jumlah kasus kanker serviks. Model yang diperoleh dari estimasi parameter regresi Poisson adalah:

$$\ln(«) = 2{,}174751 + 0{,}060947X_1 + 0{,}085021X_3 - 0{,}207351X_4$$

Pada model terlihat nilai taksiran parameter model untuk sarana kesehatan (rumah sakit dan puskesmas) dan penduduk perempuan yang umur kawin pertama ≤ 16 tahun bernilai positif artinya hubungan antar variabel tersebut dengan log rata-rata dari jumlah kasus kanker serviks berbanding lurus. Untuk setiap kenaikan sarana kesehatan (rumah sakit dan puskesmas) sebanyak 1 persen, dengan asumsi nilai variabel lainnya tetap, maka rata-rata jumlah kasus kanker serviks akan bertambah sebesar $\exp(0{,}060947) \approx 1$ orang. Untuk setiap kenaikan penduduk perempuan yang umur kawin pertama ≤ 16 tahun sebanyak 1 persen, dengan asumsi nilai variabel lainnya tetap, maka rata-rata jumlah kasus kanker serviks akan bertambah sebesar $\exp(0{,}085021) \approx 1$ orang. Untuk taksiran parameter model pada variabel penduduk miskin bernilai negatif. Hal ini menunjukkan bahwa hubungan antar variabel tersebut dengan log rata-rata dari jumlah kasus kanker serviks berbanding terbalik. Untuk setiap kenaikan penduduk miskin sebanyak 1 persen, dengan asumsi nilai variabel lainnya tetap, maka rata-rata jumlah kasus kanker serviks akan menurun sebesar $\exp(-0{,}207351) \approx 1$ orang.

Nilai AIC yang diperoleh dari model terbaik regresi Poisson yaitu sebesar 91,26029.

Overdispersi

Setelah diperoleh model regresi Poisson, tahap selanjutnya adalah mendeteksi apakah terjadi overdispersi pada model. Pengujian overdispersi dilakukan dengan mencari nilai ¤ . Jika nilai ¤ > 1, maka dapat diduga terjadi overdispersi. Hasil analisis dapat dilihat pada Tabel 4

Tabel 4 Nilai *Pearson Chi Square*

DF	*Pearson Chi Square*	Nilai/DF
10	21,0729	2,1073

Berdasarkan Tabel 4 yang diperoleh dari persamaan (13) maka hasil perhitungan sebagai berikut:

$$¤ = \frac{¡^2}{df} = \frac{21{,}0729}{10} = 2{,}1073$$

Karena ¤ = 2,1073 > 1 maka dapat disimpulkan bahwa terjadi overdispersi pada model regresi Poisson.

Model Awal GPR I

Model awal GPR I adalah:

$$\ln(«) = 4{,}6442 + 0{,}05185X_1 - 0{,}00070X_2 + 0{,}08084X_3 - 0{,}1831X_4 - 0{,}1877X_5 - 0{,}04102X_6 - 0{,}06251X_7$$

Pengujian Parameter GPR I

Pengujian Parameter Secara Serentak (Uji Rasio *Likelihood*)

Berdasarkan hasil analisis diperoleh:

$$G = -2\ln\left(\frac{L(\hat{\mu})}{L(\hat{\Omega})}\right) = 76{,}5$$

Dari hasil perhitungan, maka dapat diputuskan untuk menolak H_0 karena $G = 76{,}5 > ¡^2_{tabel} = 12{,}592$ sehingga dapat disimpulkan bahwa minimal terdapat satu variabel yaitu sarana kesehatan (rumah sakit dan puskesmas), tenaga medis, penduduk perempuan yang umur kawin pertama ≤ 16 tahun, penduduk miskin, penduduk perempuan yang menggunakan kondom, rata-rata pengeluaran untuk konsumsi makanan perbulan, dan perempuan dengan jumlah anak yang dilahirkan > 4 berpengaruh terhadap jumlah kasus kanker serviks.

Pengujian Parameter Secara Parsial (uji *t*)

Hasil pegujian diberikan pada Tabel 5.

Tabel 5 Hasil Pengujian Secara Parsial

Variabel	Koefisien	*P-value*
Konstanta	4,6442	0,0282
X_1	0,05186	0,2697
X_2	-0,00070	0,9699
X_3	0,08084	0,0604
X_4	- 0,1831	0,0615
X_5	-0,1877	0,2998
X_6	-0,04102	0,3437
X_7	-0,06251	0,2547

Berdasarkan Tabel 5, dapat dilihat bahwa hanya konstanta yang bepengaruh terhadap *Y* karena memliki *p-value* < α = 0,05. Oleh karena itu variabel yang tidak berpengaruh terhadap *Y* harus dikeluarkan dengan menggunakan metode *backward*. Setelah dikeluarkan satu persatu maka diperoleh hasil sebagai berikut:

Tabel 6 Hasil Setelah Variabel Dikeluarkan

Variabel	Koefisien	*P-value*
X_3	2,43	0,0289
X_4	-3,93	0,0015

Berdasarkan Tabel 6, dapat dilihat bahwa variabel X_3 dan X_4 berturut-turut memiliki *p-value* sebesar 0,0289 dan 0,0015 < α = 0,05 sehingga dapat diputuskan untuk menolak H_0 dan dapat disimpulkan bahwa persentase penduduk perempuan yang umur kawin pertama ≤ 16 tahun dan persentase penduduk miskin berpengaruh terhadap jumlah kasus kanker serviks. Model yang diperoleh dari estimasi parameter regresi Poisson adalah:

$$\ln(\hat{«}) = 2{,}9585 + 0{,}09964 X_3 - 0{,}2823 X_4$$

Pada model terlihat nilai taksiran parameter model untuk penduduk perempuan yang umur kawin pertama ≤ 16 tahun bernilai positif artinya hubungan antar variabel tersebut dengan log rata-rata dari jumlah kasus kanker serviks berbanding lurus. Untuk setiap kenaikan penduduk perempuan yang umur kawin pertama ≤ 16 tahun sebanyak 1 persen, dengan asumsi nilai variabel lainnya tetap, maka rata-rata jumlah kasus kanker serviks akan bertambah sebesar $\exp(0{,}099647) \approx 1$ orang. Untuk taksiran parameter model pada variabel penduduk miskin bernilai negatif. Hal ini menunjukkan bahwa hubungan antar variabel tersebut dengan log rata-rata dari jumlah kasus kanker serviks berbanding terbalik. Untuk setiap kenaikan penduduk miskin sebanyak 1 persen, dengan asumsi nilai variabel lainnya tetap, maka rata-rata jumlah kasus kanker serviks akan menurun sebesar $\exp(-0{,}2823) \approx 1$ orang. Nilai AIC yang diperoleh dari model terbaik regresi Poisson yaitu sebesar 89,5.

Kesimpulan

Berdasarkan hasil analisis dan pembahasan, maka diperoleh kesimpulan sebagai berikut:

1. Model regresi Poisson yang diperoleh adalah $\ln(\hat{«}) = 2{,}174751 + 0{,}060947 X_1 + 0{,}085021 X_3 - 0{,}207351 X_4$ dengan nilai AIC = 91,26029 dan faktor-faktor yang mempengaruhi jumlah kasus kanker serviks di Provinsi Kalimantan Timur dengan regresi Poisson yaitu sarana kesehatan (rumah sakit dan puskesmas), penduduk perempuan yang umur kawin pertama ≤ 16 tahun dan penduduk miskin.
2. Terjadi overdispersi pada model regresi Poisson karena nilai $¤ = 2{,}1073 > 1$
3. Model GPR I yang diperoleh adalah $\ln(\hat{«}) = 2{,}9585 + 0{,}09964 X_3 - 0{,}2823 X_4$ dengan nilai AIC = 89,5 dan faktor-faktor yang mempengaruhi jumlah kasus kanker serviks di Provinsi Kalimantan Timur dengan GPR I yaitu penduduk perempuan yang umur kawin pertama ≤ 16 tahun dan penduduk miskin.

Daftar Pustaka

Agresti, A. 1990. *Categorical Data Analysis*. New York : John Wiley and Sons, Inc.

Berk, R. dan MacDonald, J. 2008. *Overdispersion and Poisson Regression*. Philadelphia: Springer.

Djarwanto, Drs. 2003. *Statistik Nonparametrik*. Yogyakarta: BPFE.

Ismail, N. dan Jemain, A.A. 2007. *Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Model*. Casualty Actuarial Society Forum. Malaysia.

Maysaroh, H. 2013. *Kupas Tuntas Kanker Pada Perempuan Dan Penyembuhannya*. Klaten: Trimedia Pustaka

McCullagh, P. dan Nelder, J. A. 1989. *Generalized Linear Models: 2nd Edition*. London: Chapman and Hall.

Myers, R.H. 1996. *Classical and Modern Regression With Application Second Edition*. USA: PWS Kent Publishing Company.

Nohe, D.A. 2011. *Mengatasi Overdispersi pada Model Regresi Poisson dengan Generalized Poisson Regression I*. Samarinda: Universitas Mulawarman.

Pateta, M. 2005. *Fitting Poisson Regression Models Using the Genmod Procedure*. USA: SAS Institute Inc.

Rasjidi, I. 2010. *Epidemiologi Kanker Pada Wanita*. Jakarta: Sagung Seto.

Walpole, R.E. dan Myers, R.H. 1990. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan Terbitan ke-2*. Bandung: ITB

Widarjono, A. 2007. *Ekonometrika Teori dan Aplikasi Untuk Ekonomi dan Bisnis*. Yogyakarta: Ekonisia.