BAB II

TINJAUAN PUSTAKA

2.1 Kolinieritas

Pendeteksian adanya kasus kolinieritas menurut (*Hocking, 1996; Setyorini, 2006*) dapat dilihat melalui :

a) Koefisien korelasi pearson (r_{ij}) antar variabel indeoenden > 0.95

b) VIF (*Variance Inflation Factors*) > 10

Variance Inflation Factors (VIF) dinyatakan dalam :

$$VIF_j = \frac{Var(\beta_j)}{\sigma^2} = r^{jj} \tag{2.1}$$

Var (β_j) menunjukkan nilai varian koefisien estimasi variabel predictor ke-j dan σ^2 menunjukkan nilai varian jika variabel-variabel prediktornya ortogonal. Nilai r^{jj} merupakan elemen diagonal ke-j invers matrik korelasi variabel predictor. Nilai VIF_j yang bernilai 1 menunjukkan bahwa variabel-variabel independent tidak saling berkorelasi, jika nilainya lebih dari 10 menunjukkan adanya kolinieritas antara variabel-variabel independen.

2.2 Model Regresi Poisson

Model regresi Poisson merupakan model standar untuk data diskrit dan termasuk dalam model regresi nonlinier (*Cameron,et.a.l*, *1998*). Regresi Poisson digunakan pada data yang berdistribusi poisson. Probabilitas distribusi Poisson diberikan oleh *Myers (1990).*

$$p(y;\mu) = \frac{e^{-\mu}\mu^{y}}{y!} \quad (y = 0,1,2, \dots.) \tag{2.2}$$

dimana μ adalah mean distribusi Poisson. Parameter μ sangat bergantung pada beberapa unit yang ditetapkan atau periode waktu, jarak, luas, volume, dan lain-lain. Distribusi Poisson digunakan untuk memodelkan peristiwa yang jarang terjadi selama periode tertentu. Probabilitas banyak kejadian y dalam periode waktu t yaitu :

$$p(y;\mu) = \frac{e^{-\mu t}(\mu t)^{y}}{y!} \quad (y = 0,1,2, \dots.) \tag{2.3}$$

Persamaan tersebut digunakan untuk probabilitas kejadian y, dan rata-rata jumlah kejadian, berdasarkan asumsi bahwa rata-rata jumlah kejadian per periode waktu adalah konstan. Pengujian kesesuaian distribusi untuk variabel y adalah dengan uji Kolmogorov-Smirnov. Berikut ini adalah hipotesis uji Kolmogorov-Smirnov :

$H_0 : F(y) = F^{*}(y)$ (variabel random y mengikuti distribusi tertentu)

$H_1 : F(y) \neq F^{*}(y)$ (variabel random y tidak mengikuti distribusi tertentu)

Daerah penolakan untuk pengujian ini adalah tolak H_0 jika α pada taraf signifikansi $T > w_{1-\alpha/2}$, dimana $w_{1-\alpha/2}$ adalah nilai kuantil dari statistik uji Kolmogorov-Smirnov pada uji dua sisi. *Baharuddin (2005)* menyatakan bahwa metode regresi Poisson biasanya digunakan pada penelitian kesehatan masyarakat, biologi, dan teknik diaman variabel responnya (y) berupa cacahan objek yang merupakan fungsi dari sejumlah karakteristik tertentu (x). Misal, apabila terdapat sekumpulan data dengan struktur sebagai berikut :

$$\begin{bmatrix} y_1 & x_{11} \cdots & x_{k1} \\ \vdots & \vdots & \vdots \\ y_n & x_{1n} \cdots & x_{kn} \end{bmatrix}$$

maka, model regresi Poissonnya dapat ditulis sebagai berikut *Myers, (1990)* :

$$y_i = \mu_i + \varepsilon_i \, (i = 1,2, \ldots n) \tag{2.4}$$

dimana y_i adalah jumlah kejadian, dan μ_i adalah rata-rata jumlah kejadian dalam periode t_i . μ_i diasumsikan tidak berubah dari data ke data. Persamaan distribusi Poisson dinyatakan dengan persamaan sebagai berikut :

$$p(y_i; \hat{\beta}) = \frac{e^{-[\mu(x_i;\hat{\beta})]}[t_i\mu(x_i;\hat{\beta})]^{y_i}}{y_i!} \tag{2.5}$$

dimana $\mu(x_i; \widehat{\beta})$ adalah rata-rata Poisson dan vektor $\hat{\beta}$ menunjukkan parameter yang ditaksir. Mean dan varians untuk model regresi Poisson adalah sebagai berikut :

$$\mu_i = t_i\mu_i(x_i; \hat{\beta}) = t_i \, exp(x_i^T \hat{\beta})$$

dan

$$Var(y_i) = t_i\mu_i(x_i;\hat{\beta}) = t_i\, exp(x_i^T\hat{\beta})$$

Selanjutnya model regresi Poisson dapat ditulis sebagai berikut (*Myers,1990)* :

$$y = t_i\, exp(x_i^T\hat{\beta}) + \varepsilon_i \tag{2.6}$$

2.2.1 Estimasi Parameter Model Regresi Poisson

Berdasarkan persamaan distribusi Poisson yang ditujukkan pada persamaan (2.5), maka fungsi likelihoodnya adalah sebagai berikut (*Myers,1990*) :

$$L(y,\hat{\beta}) = \prod_{i=1}^{n} p(y_i;\hat{\beta})$$

$$= \prod_{i=1}^{n} \left\{\frac{[t_i\mu(x_i;\hat{\beta})]^{y_i} e^{-[t_i\mu(x_i;\hat{\beta})]}}{y_i!}\right\}$$

$$L(y,\hat{\beta}) = \prod_{i=1}^{n} \frac{[t_i\mu(x_i;\hat{\beta})]^{y_i} e^{-\sum_{i=1}^{n} t_i\mu(x_i;\hat{\beta})}}{\prod_{i=1}^{n} y_i!} \tag{2.7}$$

Persamaan diatas dimaksimalkan dengan menggunakan teknik iteratif yang menghasilkan penaksir maximum likelihood untuk koefisien regresi dalam $\hat{\beta}$. Prosedur yang disarankan oleh *Myers,(1990)* untuk menemukan penaksir maximum likelihood adalah pendekatan *Iteratively Reweighted Least Squares (IRWLS).*

Menurut *Cameron,et.al., (1998),* IRWLS menggunakan metode Nweton-Raphson, yang umumnya pada iterasi ke-s, metode Newton-Raphson memperbaiki taksiran $\hat{\beta}$, yang biasa dipakai dengan rumus :

$$\hat{\beta}_{s+1} = \hat{\beta}_s - \hat{H}_s^{-1}\hat{g}_s \tag{2.8}$$

dimana $g = \frac{\partial lnL(y;\widehat{\beta})}{\partial \widehat{\beta}}$ dan $H = \frac{\partial^2 lnL(y;\widehat{\beta})}{\partial (\widehat{\beta})^2}$.

Metode Newton-Raphson yang digunakan untuk menyelesaikan persamaan berikut adalah :

$$\frac{\partial lnL(y;\widehat{\beta})}{\partial \widehat{\beta}} = 0 \tag{2.9}$$

dimana

$$\ln L(y; \hat{\beta}) = \sum_{i=1}^{n} y_i \ln t_i \, \mu(x_i; \hat{\beta}) - \sum_{i=1}^{n} t_i \, \mu(x_i; \hat{\beta}) - \sum_{i=1}^{n} \ln(y_i)\,! \tag{2.10}$$

Persamaan likelihood untuk mencari $\hat{\beta}$ adalah sebagai berikut :

$$\sum_{i=1}^{n} \frac{y_i}{\mu(x_i;\widehat{\beta})} \frac{\partial \mu(x_i;\widehat{\beta})}{\partial \widehat{\beta}} - \sum_{i=1}^{n} t_i \frac{\partial \mu(x_i;\widehat{\beta})}{\partial \widehat{\beta}} = 0$$

$$\sum_{i=1`}^{n} \left[\frac{y_i}{\mu(x_i;\widehat{\beta})} - t_i\right] \left[\frac{\partial \mu(x_i;\widehat{\beta})}{\partial \widehat{\beta}}\right] = 0 \tag{2.11}$$

2.2.2 Pengujian Parameter Model Regresi Poisson

Pengujian kesesuaian model dengan *goodness of fit* disebut devians *(Kleinbaum,et.al.,1988).* Berikut ini adalah hipotesis pengujian kesesuaian model regresi Poisson.

$$H_0 : \; \mu_i = t_i \, \mu(x_i; \; \hat{\beta}), i = 1,2, \dots n$$

$$H_1 : \; \mu_i \neq t_i \, \mu(x_i; \; \hat{\beta})$$

Statistik uji yang digunakan adalah sebagai berikut:

$$G = -2l\left[\frac{L(y;\hat{\beta})}{L(y;\mu)}\right] \tag{2.12}$$

Nilai $\mu_i = t_i\mu(x_i;\beta)$, sehingga Persamaan (2.9) dapat ditulis menjadi persamaan berikut :

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

$$= \sum_{i=1}^{n} y_i \ln \hat{y}_i - \sum_{i=1}^{n} \hat{y}_i - \sum_{i=1}^{n} \ln(y_i)\,! \tag{2.13}$$

Sedangkan nilai $L(y;\mu)$ dapat ditulis dalam persamaan berikut :

$$L(y,\mu) = \prod_{i=1}^{n} \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} = \frac{\left(\prod_{i=1}^{n} \mu_i^{y_i}\right) exp\left(-\sum_{i=1}^{n} \mu_i\right)}{\prod_{i=1}^{n} y_i!} \tag{2.14}$$

Nilai $\hat{\mu}_i = y_i$, sehingga Persamaan (2.14) dapat juga ditulis menjadi sebagai berikut :

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

Nilai $\ln(y;\hat{\mu})$ dapat ditulis dalam persamaan berikut :

$$lnL(y;\hat{\mu}) = \sum_{i=1}^{n} y_i \ln y_i - \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \ln(y_i)\,! \tag{2.15}$$

Persamaan (2.13) dapat ditulis menjadi berikut :

$$G = -2ln\left[L(y;\hat{\beta}) - L(y,\hat{\mu})\right]$$

$$= -2ln\left[L(y;\hat{\mu}) - L(y,\hat{\beta})\right]$$

$$= -\sum_{i=1}^{n}[(y_i \ln y_i - y_i - ln(y_i)!) - (y_i \ln \hat{y}_i - \hat{y}_i - ln(y_i)!)]$$

$$G = 2\sum_{i=1}^{n}\left[y_i ln\left(\frac{y_i}{\hat{y}_i}\right) - (y_i - \hat{y}_i)\right] \quad (2.16)$$

Nilai G disebut devians untuk model regresi Poisson. Model yang sesuai devians mendekati distribusi *Chi-Square* dengan derajat bebas $(n-k-1)$, dimana n adalah jumlah pengamatan dan $k+1$ adalah jumlah parameter *(Ismail,et.al.,(2005).*

Kleinbaum,et.al.,(1998) mengatakan bahwa devians seperti Sum Square error pada analisis regresi berganda yaitu nilai data pengamatan sama dengan prediksi $(y_i = \hat{y}_i)$ maka nilai $G = 0$. Besarnya nilai devians terjadi jika semakin besar selisih antara respon pengamatan dan respon taksiran, maka nilai devians juga akan semakin besar. Taksiran diharapkan dapat mendekati pengamatan atau tingkat kesalahan diharapkan bernilai kecil, sehingga nilai devians akan bernilai kecil, dan nilai devians yang diharapkan adalah nilai devians yang kecil. Parameter model regresi Poisson yang telah dihasilkan dari estimasi parameter belum tentu mempunyai pengaruh yang signifikan terhadap model, maka perlu dilakukan pengujian terhadap parameter model regresi Poisson secara individu.

2.3 Overdispersi

Khoshgoftaar dkk. (2004) mengatakan bahwa metode regresi Poisson mewajibkan equidispersi, yaitu kondisi dimana nilai mean dan varians dari variabel respon bernilai sama. Namun, adakalanya terjadi fenomena overdispersi dalam data yang dimodelkan dengan distribusi Poisson. Overdispersi berarti varians lebih besar dari pada mean. Taksiran dispersi diukur dengan devians atau Pearson Chi-square yang dibagi derajat bebas.

Data overdispersi jika taksiran dispersi lebih besar dari 1 dan underdispersi jika taksiran dispersi kurang dari 1.

2.4 Model Regresi Zero-Inflated Poisson

Jansakul,et.al.,(2001) mengatakan bahwa salah satu penyebab terjadinya overdisersi adalah lebih banyak observasi yang bernilai nol daripada yang ditaksir untuk model regresi Poisson. Salah satu metode analisis yang diusulkan untuk kasus seperti ini adalah model regresi *Zero-Inflated Poisson (ZIP).* Jika y_i adalah variabel random independen yang mempunyai distribusi ZIP, observasi diduga muncul dalam dua cara yang sesuai untuk state yang terpisah. State pertama terjadi dengan probabilitas ω_i dan menghasilkan hanya observasi bernilai nol, sementara state kedua terjadi dengan probabilitas $(1-\omega_i)$ dan berdistribusi Poisson dengan mean μ_i. Proses dua state ini memberikan distribusi campuran dua komponen dengan fungdi probabilitas sebagai berikut :

$$Pr(Y_i = y_i) = \begin{cases} \omega_i + (1-\omega_i)e^{-\mu_i}, y_i = 0 \\ (1-\omega_i)\frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, y_i = 1,2, \dots 0 \le \omega_i \le 1 \end{cases} \qquad (2.17)$$

Lambert dalam *Jansakul,et.al.,(2001)* menunjukkan model golongan untuk μ dan ω sebagai berikut :

$$log(\mu) = X\beta \text{ dan } logit\,(\omega) = log\left(\frac{\omega}{1-\omega}\right) = X\gamma \qquad (2.18)$$

Dimana X adalah matriks variabel prediktor sedangkan β dan γ adalah parameter yang akan ditaksir.

2.4.1 Estimasi Parameter Model Regresi *Zero-Inflated Poisson*

Menurut *Khoshgoftaar,et.al.,(2004),* estimasi parameter regresi ZIP dengan menggunakan metode Maximum Likelihood Estimation (MLE). Fungsi Likelihood ZIP adalah sebagai berikut :

$$L(\beta,\gamma|y_i,x_i) = \begin{cases} \prod_{i=1}^{n} \frac{exp(x_i^T\gamma)+exp\left(-exp(x_i^T\beta)\right)}{1+exp(x_i^T\gamma)}, y_i = 0 \\ \prod_{i=1}^{n} \frac{\frac{1}{1+exp\left(x_i^T(\gamma)\right)}\left(exp\left((-exp(x_i^T\beta)+(x_i^T\beta)y_i)\right)\right)}{y_i!}, y_i > 0 \end{cases} \qquad (2.19)$$

dimana x_i adalah variabel prediktor, y_i adalah variabel respon, serta β dan γ adalah parameter yang akan ditaksir.

Fungsi Log-Likelihood gabungan untuk model regresi ZIP diberikan oleh :

$$lnL(\beta,\gamma|y_i,x_i) = \sum_{\substack{i=1 \\ y_i=0}}^{n} ln\left(exp(x_i'\gamma)\right) - \sum_{i=1}^{n} ln\left(1+exp(x_i'\gamma)\right) + \sum_{\substack{i=1 \\ y_i>0}}^{n} \left((y_i x_i^T\beta) - exp(x_i^T\beta)\right) - \sum_{\substack{i=1 \\ y_{i>0}}}^{n} ln(y_i)! \qquad (2.20)$$

Estimasi maximum likelihood untuk β dan γ dapat di peroleh dengan menggunakan pendekatan standart untuk model campuran, yaitu Algoritma EM. Algoritma EM memberikan prosedur sederhana yang dapat diimplementasi dalam software standar, atau metode estimasi langsung seperti metode Newton-Raphson.

2.4.2 Pengujian Peremeter Model Zero-Inflated Poisson

Pengujian kesesuaian model regresi ZIP adalah dengan menggunakan *Likelihood Ratio (LR) test.* Hipotesis untuk pengujian kesesuaian model adalah sebagai berikut :

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = \gamma_1 = \gamma_2 = \cdots \gamma_k = 0$$

$$H_1: paling\ sedikit\ ada\ satu\ \beta_r \neq 0\ atau\ \gamma_r \neq 0, 0 < r < k - 1$$

dimana k+1 adalah jumlah parameter, β_r adalah parameter model log ke-r, dan γ_r adalah parameter model logit ke-r. *Lestari* (2008) telah melakukan perhitungan statistik uji untuk pengujian kesesuaian model sebagai berikut :

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

$$= \left(2 \sum_{i=1}^{n} \left(z_i x_i^T \hat{\gamma} - ln\left(1 + exp(x_i^T \hat{\gamma})\right)\right) + 2 \sum_{i=1}^{n} (1 - z_i) \left(y_i x_i^T \hat{\beta} - exp\left(x_i^T \hat{\beta}\right)\right) - 2 \sum_{i=1}^{n} z_i \hat{\gamma}_0 - ln(1 + x_i^T \hat{\gamma}_0) + 2 \sum_{i=1}^{n} (1 - z_i) \left(y_i \hat{\beta}_0 - exp\left(\hat{\beta}_0\right)\right)\right) \qquad (2.21)$$

Pengujian parameter secara individu ada dua, yaitu pengujian parameter model log dan pengujian parameter model logit. Statistik uji untuk pengujian parameter model log secara individu adalah sebagai berikut (*Lestari,2008*) :

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

$$= \left(2\sum_{i=1}^{n}\left(z_i x_i^T\hat{\gamma} - ln\left(1 + exp(x_i^T\hat{\gamma})\right) + 2\sum_{i=1}^{n}(1 - z_i)\left(y_i x_i^T\hat{\beta} - exp\left(x_i^T\hat{\beta}\right)\right)\right) - 2\sum_{i=1}^{n}(1 - z_i)\left(y_i x_i^T\hat{\beta}_i - exp\left(x_i^T\hat{\beta}_i\right)\right)\right) \quad (2.22)$$

Statistik uji yang digunakan untuk pengujian parameter model logit adalah sebagai berikut *(Lestari,2008)* :

$$G = -2\ln\left[\frac{L(y;\widehat{\omega})}{L(y;\widehat{\Omega})}\right]$$

$$= 2\sum_{i=1}^{n}\left(z_i x_i^T\hat{y} - ln\left(1 + exp(x_i^T\hat{\gamma})\right) + 2\sum_{i=1}^{n}(1 - z_i)\left(y_i x_i^T\hat{\beta} - exp\left(x_i^T\hat{\beta}\right)\right) - 2\sum_{i=1}^{n}(1 - z_i)\,ln(y_i)! - 2\sum_{i=1}^{n}\left(z_i\hat{\gamma}_0 - ln\left(1 + exp(\hat{\gamma}_0)\right)\right)\right) \quad (2.23)$$

Daerah penolakan untuk ketiga pengujian diatas adalah tolak H_0 pada taraf signifikansi α jika $G_{hitung} > \chi^2_{(v,\alpha)}$, dimana v adalah derajat bebas.

2.4.3 Pemilihan Model Terbaik

Metode Akaike's Information Criterion (AIC) merupakan salah satu metode yang digunakan untik memilih model terbaik dari Regresi Poisson dan ZIP. Nilai AIC adalah sebagai berikut *(Dalrymple,et.al.,(2001)* :

$$AIC = G + (k + 1) \quad (2.24)$$

dimana G adalah statistik uji kesesuaian model, dan $k + 1$ adalah jumlah parameter. Model terbaik Regresi Poisson dan ZIP adalah model dengan nilai AIC terkecil.

2.5 Kusta

Penyakit kusta *(Morbus hansen)* adalah suatu penyakit infeksi menahun akibat bakteri tahan asam yaitu *Mycobacterium leprae* yang secara primer menyerang saraf tepi dan secara sekunder menyerang kulit serta organ lainnya (*WHO, 2010; Noto & Schreuder, 2010*). Penyakit kusta adalah penyakit kronis yang dapat menimbulkan masalah kecacatan (*Susanto, 2006*). Masalah yang timbul tidak hanya pada masalah kesehatan fisik saja, tetapi juga masalah psikologis, ekonomi dan sosial bagi penderitanya (*Amiruddin, 2006*).

Penyebab munculnya penyakit kusta adalah bakteri *Mycobacterium leprae* yang ditemukan pertama kalioleh G. H. Armauer Hansen pada tahun 1873. Bakteri ini masuk ke dalam tubuh manusia melalui luka pada permukaan kulit atau bisa juga melalui *droplet* yangdihembuskan dari saluran pernafasan. *Sehgal (dalam Putra, 2012)* mengatakan bahwa *Mycobacterium leprae* memiliki ciri-ciri yaitu tahan asam,bersifat gram positif, berbentuk batang, lebar 0,3-0,4 mikrometer, panjang 2-7mikometer, dan hidup di dalam sel yang banyak mengandung lemak dan lapisanlilin. *Mycobacterium leprae* membelah dalam kurun waktu 21 hari, sehingga menyebabkan masa tunas yang sangat lama yaitu 4 tahun. Munculnya penyakit kusta tersebut ditunjang oleh cara penularan.

2.5.1 Kepadatan Penduduk

Kepadatan penduduk adalah suatu keadaan yang dikatakan semakin padat bila jumlah manusia pada suatu batas ruang tertentu semakin banyak dibandingkan dengan luas ruangannya (*Sarwono, 1992).* Kepadatan penduduk adalah perbandingan antara jumlah penduduk dengan luas wilayah yang dihuni (*Mantra,* *2007).* Kepadatan penduduk merupakan indikator dari pada tekanan penduduk di suatu daerah. Kepadatan di suatu daerah dibandingkan dengan luas tanah yang ditempati dinyatakan dengan banyaknya penduduk per kilometer persegi.

2.5.2 Rumah Tangga Akses Sanitasi Layak

Sanitasi adalah upaya kesehatan dengan cara memelihara dan melindungi kebersihan lingkungan dari subjeknya, misalnya menyediakan air bersih untuk kperluan mencuci tangan, menyediakan tempat sampah agar tidak dibuang sembarangan (*Depkes RI, 2014*). Sanitasi sering juga disebut dengan sanitasi lingkungan dan kesehatan lingkungan, sebagai suatu usaha pengendalian semua faktor yang ada pada lingkungan fisik manusia yang diperkirakan dapat menimbulkan hal – hal yang menggangu perkembanan fisik, kesehatnnya ataupun kelangsungan hidupnya (*adisasmito, 2006*).

2.5.3 Rumah Tangga Sumber Air Minum Layak

Air minum yang berkualitas (layak) adalah air minum yang terlindung meliputi air ledeng (keran), keran umum, hydrant umum, terminal air, penampungan air hujan (PAH) atau mata air dan sumur terlindung, sumur bor atau sumur pompa, yang jaraknya minimal 10 m dari pembuangan kotoran, penampungan limbah dan pembuangan sampah. Tidak termasuk air kemasan, air dari penjual keliling, air yang dijual melalui tanki, air sumur dan mata air tidak terlindung. Proporsi rumah tangga dengan akses berkelanjutan terhadap air minum layak adalah perbandingan antara rumah tangga dengan akses terhadap sumber air minum berkualitas (layak) dengan rumah tangga seluruhnya, dinyatakan dalam persentase (*Susenas, 2016*).