ISBN:1858-1382

J M S K
Jurnal Matematika, Statistika, & Komputasi
http://journal.unhas.ac.id/index.php/jmsk
Vol. 6, No.2, 59-71, Januari 2010

Menentukan Model Terbaik dalam Regresi Poisson dengan Menggunakan Koefisien Determinasi

Darnah*

Abstrak

Analisis regresi merupakan analisis statistika yang bertujuan untuk memodelkan hubungan antara variabel respon dengan sejumlah variabel prediktor. Apabila variabel respon berdistribusi Poisson, maka model regresi yang digunakan adalah regresi Poisson. Salah satu kriteria penentuan model terbaik dalam regresi Poisson adalah dengan menggunakan koefisien determinasi atau ukuran R^2. Ukuran R^2 dalam regresi Poisson ada tiga, yaitu R^2 berdasarkan sekumpulan residual, R^2 berdasarkan residual Pearson, dan R^2 berdasarkan residual devians (R^2_{DEV}). Penelitian ini mengkaji tentang ukuran R^2_{DEV} untuk memperoleh model terbaik. Ukuran R^2_{DEV} tersebut kemudian diaplikasikan pada data jumlah *maternal mortality* ditiap kabupaten/kota di Jawa Timur pada tahun 2003. Dari hasil analisis diperoleh bahwa faktor-faktor yang mempengaruhi jumlah *maternal mortality* ditiap kabupaten/kota di Jawa Timur adalah rata-rata pengeluaran biaya kesehatan perkapita sebulan, persentase penduduk miskin, dan jumlah tenaga medis dan paramedis.

Kata Kunci: *Koefisien determinasi, maternal mortality, model regresi Poisson, residual devians.*

1. Pendahuluan

Analisis regresi merupakan analisis statistika yang bertujuan untuk memodelkan hubungan antara variabel respon Y dengan variabel prediktor X. Apabila variabel respon Y berdistribusi Poisson, maka model regresi yang digunakan adalah regresi Poisson. Regresi Poisson didapatkan dari distribusi Poisson, yaitu suatu distribusi untuk peristiwa yang probabilitas kejadiannya kecil, dimana kejadiannya tergantung pada interval waktu tertentu atau di suatu daerah tertentu dengan hasil pengamatan berupa variabel diskrit dan antar variabel saling independen.

Dalam analisis regresi, disamping penaksiran parameter dan keberartian variabel prediktor, juga diperlukan ukuran proporsi keragaman variabel respon yang dapat diterangkan oleh variabel prediktor. Ukuran ini biasa disebut ukuran R^2 atau koefisien determinasi (Schemper, 2003). Koefisien determinasi (R^2) dalam analisis regresi linear didasarkan pada pemakaian jumlah kuadrat (*sums of squares*) dengan metode kuadrat terkecil *(least square methods*). Sedangkan ukuran R^2 pada regresi Poisson didasarkan pada proporsi reduksi dalam log-likelihood yang dimaksimumkan.

Beberapa penelitian tentang ukuran R^2 dalam model regresi Poisson telah dilakukan diantaranya oleh Cameron dan Windmeijer (1996) yang mengembangkan ukuran R^2 berdasarkan sekumpulan residual, ukuran R^2 berdasarkan residual Pearson, dan ukuran R^2 berdasarkan residual

* *Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman, email: darnah@statistika.its.ac.id*

devians. Ketiga ukuran R^2 tersebut kemudian diaplikasikan pada data *Doctorco* untuk mengetahui faktor-faktor yang mempengaruhi orang melakukan konsultasi kedokter pada dua minggu terakhir setiap bulannya. Berdasarkan hasil yang diperoleh, Cameron dan Windmeijer menyarankan penggunaan ukuran R^2 berdasarkan residual devians (R^2_{DEV}) karena lebih realistik dalam memberikan informasi tambahan dan dapat dijadikan salah satu kriteria penentu model terbaik pada model regresi Poisson.

Sama halnya R^2 dalam analisis regresi linear, adanya penambahan suatu variabel prediktor ke dalam model regresi Poisson akan menaikkan R^2_{DEV} meskipun variabel prediktor tersebut tidak berpengaruh nyata terhadap respon. Sehingga Waldhor, Haidinger, dan Schober (1998) mengusulkan untuk memberikan koreksi terhadap R^2_{DEV} dengan menggunakan derajat bebas yaitu $R^2_{DEV,db}$.

Koreksi lain terhadap R^2_{DEV} dilakukan oleh Mittlbock dan Waldhor (2000), yaitu didasarkan pada statistik rasio likelihood yang disebut $R^2_{DEV,adj,1}$ dan $R^2_{DEV,adj,2}$ yang telah diaplikasikan pada penelitian untuk mengetahui faktor-faktor yang mempengaruhi seseorang melakukan bunuh diri di Denmark, dan pada data kesehatan untuk mengetahui faktor-faktor yang menyebabkan seorang penderita penyakit jantung meninggal dunia di *British Medical Register*.

Di Indonesia sendiri, penelitian tentang ukuran R^2 berdasarkan residual devians dalam model regresi Poisson telah dilakukan oleh Baharuddin (2005). Baharuddin membandingkan nilai ukuran $R^2_{DEV}, R^2_{DEV,db}, R^2_{DEV,adj,1}$ dengan nilai ukuran R^2 dalam model regresi linear, yaitu R^2_{JK}, dan $R^2_{JK,db}$ untuk menentukan faktor-faktor yang mempengaruhi pencemaran air sungai Cikao Kabupaten Purwakarta dan menentukan faktor-faktor yang berpengaruh terhadap jumlah kematian ibu akibat bersalin di Sulawesi Tenggara. Penelitian tersebut tidak sampai pada pembentukan model regresi Poisson.

Berdasarkan uraian di atas, penelitian ini akan mengkaji cara menentukan ukuran R^2 berdasarkan residual devians dan mengaplikasikannya pada data *maternal mortality* di Jawa Timur untuk mendapatkan model terbaik. *Maternal mortality* adalah kematian seorang wanita yang sedang hamil atau dalam periode 42 hari setelah terminasi kehamilannya, disebabkan oleh sebab apapun yang berhubungan atau diperberat oleh kehamilan atau penanganannya, dan tidak disebabkan oleh suatu kebetulan atau akibat kecelakaan (Royston *et al.*, 1994).

Tinggi rendahnya angka *maternal mortality* dapat dipakai mengukur taraf program kesehatan di suatu negara khususnya program kesehatan ibu dan anak (Sukarni, 1994). Semakin rendah angka kematian ibu di suatu negara menunjukkan tingginya taraf kesehatan negara tersebut. Di Indonesia, tiap tahunnya sekitar 14.180 wanita meninggal karena hamil dan melahirkan atau dalam satu jam terdapat dua orang ibu meninggal saat melahirkan. Jika dikalkulasikan, angka kematian ibu saat melahirkan akibat komplikasi kehamilan, persalinan, dan nifas mencapai 20 ribu orang per tahun. Angka ini masih merupakan angka yang tertinggi di Asia Tenggara (Sahrudin, 2008).

Oleh karena itu, berbagai upaya harus dilakukan untuk menurunkan angka kematian ibu. Salah satu cara yang dapat dilakukan untuk menurunkan angka *maternal mortality* adalah dengan mengetahui penyebabnya. Faktor-faktor penyebab tersebut akan dimodelkan dalam regresi Poisson.

Penelitian tentang faktor-faktor yang mempengaruhi jumlah *maternal mortality* di Jawa Timur pada tahun 2003 dengan menggunakan model regresi Poisson telah dilakukan oleh Setyorini (2006). Penentuan model terbaik dilakukan berdasarkan nilai devians terkecil, dimana model terbaik yang diperoleh menunjukkan bahwa faktor-faktor yang berpengaruh pada jumlah *maternal mortality* di tiap kabupaten/kota di Jawa Timur pada tahun 2003 adalah jumlah sarana kesehatan dan persentase penolong proses persalinan yang dilakukan oleh tenaga non medis (dukun bayi).

2. Distribusi Poisson

Distribusi Poisson merupakan suatu distribusi untuk peristiwa yang probabilitas kejadiannya kecil, dimana kejadiannya tergantung pada interval waktu tertentu atau di suatu daerah tertentu dengan hasil pengamatan berupa variabel diskrit dan antar variabel saling independen. Interval waktu tersebut dapat berapa saja panjangnya, misalnya semenit, sehari, seminggu, sebulan atau bahkan setahun. Daerah tertentu yang dimaksudkan dapat berupa suatu garis, suatu luasan, suatu volume, atau mungkin sepotong bahan (Walpole, 1982).

Menurut Cameron dan Trivedi (1998), suatu variabel random Y yang bertipe diskrit akan mengikuti distribusi Poisson jika μ adalah rata-rata suatu kejadian per unit waktu dan t adalah periode waktu tertentu, maka rata-rata dari y menjadi μt. Jadi, peluang terjadinya kejadian y pada periode waktu ke-t diberikan oleh persamaan berikut:

$$P(y;\mu)=\frac{[\exp(\mu t)][\mu t]^y}{y!} \quad , y = 0, 1, 2, \ldots ;\ \mu > 0 \qquad (1)$$

Bila selang waktu kejadian adalah sama, maka fungsi distribusi peluang untuk variabel random Poisson Y dengan parameter μ diberikan oleh:

$$P(y,\mu)=\frac{[\exp(-\mu)]\mu^y}{y!}, y = 0,1, 2, \ldots ;\ \mu > 0,\ E(Y)=\mu,\ \mathrm{var}(Y)=\mu \qquad (2)$$

3. Regresi Poisson

Regresi Poisson sering kali digunakan untuk menganalisis data diskrit (*count data*), dalam hal ini respon data tersebut berdistribusi Poisson dengan parameter μ. 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.

Misalkan data diambil dari bentuk $\begin{matrix} y_1 & x_{11} & x_{21} & \cdots & x_{k1} \\ y_2 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_n & x_{1n} & x_{2n} & \cdots & x_{kn} \end{matrix}$ dengan y_i adalah observasi ke-*i* dari variabel respon *y*, dan x_{ji} adalah nilai variabel prediktor X_j $(j=1,2,\ldots,k)$, maka model regresi Poisson dapat dinyatakan sebagai berikut:

$$E(y_i)=\mu(\boldsymbol{x}_i\ ,\boldsymbol{\beta})=\exp(\boldsymbol{x}_i^T\boldsymbol{\beta});\ i=1,2,\ldots,n\ .$$

dengan $\mathbf{x}_\mathrm{i}^\mathrm{T}=[x_{1i},x_{2i},\ldots,x_{ki}]$ dan $\boldsymbol{\beta}=[\beta_1,\beta_2,\ldots,\beta_k]^\mathrm{T}$. Fungsi $\mu(\boldsymbol{x}_i\ ,\boldsymbol{\beta})$ merupakan model regresi Poisson yang merupakan fungsi dari x_i sebagai variabel prediktor dan $\boldsymbol{\beta}$ sebagai parameter regresi yang akan ditaksir.

4. Penaksiran Parameter Model Regresi Poisson

Regresi Poisson sering kali digunakan untuk menganalisis data. Penaksiran parameter regresi Poisson dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Taksiran maksimum likelihood untuk parameter $\boldsymbol{\beta}$ dinyatakan dengan $\hat{\boldsymbol{\beta}}$ yang merupakan penyelesaian dari turunan pertama dari fungsi likelihoodnya, yaitu:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} P(y_i, \boldsymbol{\beta}) = \frac{\{\prod_{i=1}^{n} [\mu(\mathbf{x}_i, \boldsymbol{\beta})]^{y_i}\} \exp(-\sum_{i=1}^{n} \mu(\mathbf{x}_i, \boldsymbol{\beta}))}{\prod_{i=1}^{n} y_i\,!} \tag{3}$$

Fungsi log natural-likelihoodnya adalah :

$$\ln L(\boldsymbol{\beta}) = \sum_{i=1}^{n} y_i \mathbf{x}_{\mathrm{i}}^{\mathrm{T}} \boldsymbol{\beta} - \sum_{i=1}^{n} \exp(\mathbf{x}_{\mathrm{i}}^{\mathrm{T}} \boldsymbol{\beta}) - \sum_{i=1}^{n} \ln y\,! \tag{4}$$

Turunan pertamanya adalah :

$$\frac{\partial \ln(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} y_i \mathbf{x}_{\mathrm{i}}^{\mathrm{T}} - \sum_{i=1}^{n} \mathbf{x}_{\mathrm{i}}^{\mathrm{T}} \exp(\mathbf{x}_{\mathrm{i}}^{\mathrm{T}} \boldsymbol{\beta}) = 0 \tag{5}$$

Karena fungsi pada (5) berbentuk implisit, maka digunakan suatu prosedur iterasi numerik yaitu metode Newton-Raphson yang diawali dengan menentukan nilai taksiran awal parameter $\hat{\boldsymbol{\beta}}_{(0)}$. Misalkan $\hat{\boldsymbol{\beta}}_{(0)}[1] = \bar{y}$. Selanjutnya nilai taksiran untuk setiap *i* dapat diperoleh dari $\hat{\mu}_i = \exp(\mathbf{x}_{\mathrm{i}}^{\mathrm{T}} \hat{\boldsymbol{\beta}})$. Dan vektor gradien $\mathbf{g}(\boldsymbol{\beta})$ diperoleh dari $\mathbf{g}(\boldsymbol{\beta}) = \left[\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]^T$ dimana $\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ seperti pada persamaan (5). Setelah diperoleh $\mathbf{g}(\boldsymbol{\beta})$ kemudian diuraikan menurut deret Taylor pada $\boldsymbol{\beta} = \boldsymbol{\beta}_{(m)}$ yaitu:

$$\mathbf{g}(\boldsymbol{\beta}) = \mathbf{g}(\boldsymbol{\beta}_{(m)}) + \left.\frac{\partial \mathbf{g}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{(m)}} (\boldsymbol{\beta} - \boldsymbol{\beta}_{(m)}) + \cdots \tag{6}$$

Jika $\boldsymbol{\beta}_{(m+1)}$ adalah solusi dari $\mathbf{g}(\boldsymbol{\beta}) = \mathbf{0}$, maka $\mathbf{g}(\boldsymbol{\beta}_{(m+1)}) = \mathbf{0}$. Sehingga

$$\mathbf{g}(\boldsymbol{\beta}_{(m+1)}) = \mathbf{g}(\boldsymbol{\beta}_{(m)}) + \mathbf{H}(\boldsymbol{\beta}_{(m)})(\boldsymbol{\beta}_{(m+1)} - \boldsymbol{\beta}_{(m)}) + \cdots \tag{7}$$

Selanjutnya dari (7) jika diambil sampai suku yang kedua, maka diperoleh :

$$-\mathbf{g}(\boldsymbol{\beta}_{(m)}) = \mathbf{H}(\boldsymbol{\beta}_{(m)})(\boldsymbol{\beta}_{(m+1)} - \boldsymbol{\beta}_{(m)}) \tag{8}$$

Persamaan (8) dapat juga ditulis sebagai :

$$\boldsymbol{\beta}_{(m+1)} - \boldsymbol{\beta}_{(m)} = -\mathbf{H}^{-1}(\boldsymbol{\beta}_{(m)})\mathbf{g}(\boldsymbol{\beta}_{(m)}) \text{ atau } \boldsymbol{\beta}_{(m+1)} = \boldsymbol{\beta}_{(m)} - \mathbf{H}^{-1}(\boldsymbol{\beta}_{(m)})\mathbf{g}(\boldsymbol{\beta}_{(m)}) \tag{9}$$

Proses iterasi pada (9) akan berhenti jika $\| \boldsymbol{\beta}_{(m+1)} - \boldsymbol{\beta}_{(m)} \| < \delta$, dimana δ adalah bilangan yang sangat kecil, dan $\| \boldsymbol{\beta}_{(m+1)} - \boldsymbol{\beta}_{(m)} \|$ dapat diperoleh dari $\| \boldsymbol{\beta}_{(m+1)} - \boldsymbol{\beta}_{(m)} \| = \| \boldsymbol{\beta}_{(m+1)} - \boldsymbol{\beta}_{(m)} \|_2 = \sqrt{[\boldsymbol{\beta}_{(m+1)} - \boldsymbol{\beta}_{(m)}]^T \; [\boldsymbol{\beta}_{(m+1)} - \boldsymbol{\beta}_{(m)}]}$.

5. Pengujian Parameter Model Regresi Poisson

Pengujian kelayakan model yang diperoleh dari estimasi parameter dilakukan dengan menggunakan metode *Maximum Likelihood Ratio Test* (MLRT) dengan melakukan pengujian hipotesis-hipotesis berikut:

1. $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$.
   H_1 : Paling sedikit ada satu $\beta_j \neq 0$.

Himpunan parameter dibawah populasi adalah $\Omega = \{\beta_0, \beta_1, \ldots, \beta_k\}$, dan fungsi likelihoodnya:

$$L(\Omega) = \prod_{i=1}^{n} f(y_i; \boldsymbol{\beta}) = \frac{\exp[-\sum_{i=1}^{n} \mu(\mathbf{x}_i, \boldsymbol{\beta})]\{\prod_{i=1}^{n} [\mu(\mathbf{x}_i, \boldsymbol{\beta})]^{y_i}\}}{\prod_{i=1}^{n} y_i!}, \tag{10}$$

dimana $\mu(\mathbf{x}_i, \boldsymbol{\beta}) = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki})$.

Himpunan parameter jika H_0 benar adalah $\omega = \mu(\mathbf{x}_i, \beta_0) = \exp(\beta_0)$ dan fungsi likelihoodnya:

$$L(\omega) = \prod_{i=1}^{n} f\left(y_i; \beta_0\right) = \frac{\exp[-\sum_{i=1}^{n} \exp(\beta_0)]\{\prod_{i=1}^{n} [\exp(\beta_0)]^{y_i}\}}{\prod_{i=1}^{n} y_i!},$$

$$\ln L(\omega) = -n\exp(\beta_0) + \sum_{i=1}^{n} y_i \beta_0 - \sum_{i=1}^{n} \ln y_i!\,. \tag{11}$$

Memaksimumkan $L(\Omega)$ untuk menentukan $\hat{\Omega}$, dari (10) diperoleh:

$$L(\hat{\Omega}) = \max_{\Omega} L(\Omega) = \frac{\exp[-\sum_{i=1}^{n} \exp(\mathbf{x}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}})]\{\prod_{i=1}^{n} \left[\exp(\mathbf{x}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}})\right]^{y_i}\}}{\prod_{i=1}^{n} y_i!}.$$

Parameter $\hat{\boldsymbol{\beta}}$ diperoleh dengan menyelesaikan persamaan (5). Memaksimumkan $L(\omega)$ untuk menentukan $(\hat{\omega})$, dari (11) diperoleh:

$$\frac{\partial \ln L(\omega)}{\partial \beta_0} = -n\exp(\beta_0) + \sum_{i=1}^{n} y_i = 0$$

$$\hat{\mu}_i = \exp(\beta_0) = \bar{Y} \tag{12}$$

sehingga $\hat{\beta}_0 = \ln(\bar{Y})$,

maka

$$L(\hat{\omega}) = \max_{\omega} L(\omega) = \frac{\exp\left[-n\bar{Y}\right]\left\{\prod_{i=1}^{n}\left[\bar{Y}\right]^{y_i}\right\}}{\prod_{i=1}^{n} y_i\,!} .$$

Rasio antara $L(\hat{\Omega})$ dan $L(\hat{\omega})$ dapat dituliskan sebagai:

$$\Lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \frac{\exp\left[-n\bar{Y}\right]\{\prod_{i=1}^{n}\left[\bar{Y}\right]^{y_i}\}}{\exp\left[-\sum_{i=1}^{n}\exp(\mathbf{x}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}})\right]\left\{\prod_{i=1}^{n}\left[\exp(\mathbf{x}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}})\right]^{y_i}\right\}} ,$$

kriteria pengujiannya adalah tolak H_0 jika $\frac{L(\hat{\omega})}{L(\hat{\Omega})} < \Lambda_0 < 1$, dimana $0 < \Lambda_0 < 1$.

$$D(\hat{\boldsymbol{\beta}}) = -2\ln\left[\frac{L(\hat{\omega})}{L(\hat{\Omega})}\right] = 2[\ln L(\hat{\Omega}) - \ln L(\hat{\omega})] == 2\sum_{i=1}^{n}(y_i\mathbf{x}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}} - \exp(\mathbf{x}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}) - (y_i\hat{\beta}_{\mathbf{0}} - \exp(\hat{\beta}_{\mathbf{0}})) ,$$

kriteria pengujiannya adalah tolak H_0 jika $D(\hat{\boldsymbol{\beta}}) > \chi^2_{v1}$, dimana v adalah derajat bebas yang diperoleh dari banyaknya parameter model dibawah populasi dikurangi banyaknya parameter dibawah H_0 .

2. H_0 : $\beta_j = 0$.
   H_1 : Paling sedikit ada satu $\beta_j \neq 0$.

Dalam pengujian hipotesis di atas dapat digunakan statistik uji $t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$. Statistik uji tersebut dikenal sebagai statistik uji *Student* dengan kriteria pengujiannya adalah tolak H_0 jika $|t_{hit}| > t_{\alpha/2;n-(p+1)}$, hal ini berarti bahwa pengaruh variabel prediktor ke-*i* terhadap variabel respon (Y) signifikan dalam model (Kleinbaum, 1988).

6. Ukuran R^2_{DEV} pada Model Regresi Poisson

Pengujian besarnya proporsi keragaman variabel respon yang dapat diterangkan oleh variabel prediktor dalam model regresi Poisson dapat ditentukan dari nilai R^2 atau koefisien determinasi. Nilai R^2 yang didasarkan pada residual devians juga dapat dijadikan sebagai salah satu penentu kriteria kebaikan model. Semakin besar nilai R^2_{DEV} ($0 \leq R^2_{DEV} \leq 1$), semakin akurat dari taksiran model regresi.

Penentuan ukuran R^2_{DEV} pada model regresi Poisson diawali dari definisi fungsi distribusi Poisson pada (2). Dengan menggunakan MLE, diperoleh $\hat{\mu}_i = y_i$ sehingga fungsi log-likelihood ketika semua parameter disertakan dalam model dapat ditulis menjadi:

$$\ln L(\hat{\Omega}) = \sum_{i=1}^{n} y_i \ln y_i - \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \ln y_i\,! \quad . \tag{13}$$

dan fungsi log-likelihood ketika tidak semua parameter disertakan dalam model adalah:

$$\ln L(\hat{\omega}) = \sum_{i=1}^{n} y_i \ln \hat{\mu}_i - \sum_{i=1}^{n} \hat{\mu}_i - \sum_{i=1}^{n} \ln y_i!, \qquad (14)$$

Devians didefinisikan sebagai:

$$D(\hat{\boldsymbol{\beta}}) = -2\ln\left[\frac{L(\hat{\omega})}{L(\hat{\Omega})}\right] = 2\ln[L(\hat{\Omega}) - L(\hat{\omega})] = \sum_{i=1}^{n} 2\left[(y_i \ln(y_i / \hat{\mu}_i)) - (y_i - \hat{\mu}_i)\right], \qquad (15)$$

dan residual deviansnya adalah

$$d_i = \pm\left[2y_i \ln(y_i / \hat{\mu}_i) - 2(y_i - \hat{\mu}_i)\right]^{\frac{1}{2}}. \qquad (16)$$

Berdasarkan (12), diperoleh

$$\hat{\mu}_i = \bar{Y} \qquad (17)$$

Dengan mensubtitusi (17) pada (14) akan diperoleh persamaan berikut:

$$\ln L(\bar{Y}) = \sum_{i=1}^{n} y_i \ln \bar{Y} - \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} y_i! \;, \qquad (18)$$

yang merupakan fungsi log-likelihood ketika hanya intersep (β_0) yang disertakan dalam model. Maka devians ketika hanya β_0 yang disertakan dalam model adalah:

$$D(\hat{\beta}_0) = -2\ln\left[\frac{L(\bar{Y})}{L(\hat{\Omega})}\right] = 2\left[\ln L(\hat{\Omega}) - \ln L(\bar{Y})\right] = 2\sum_{i=1}^{n} y_i \ln(y_i / \bar{Y}). \qquad (19)$$

Dari (15) dan (19) akan menghasilkan R^2 yang didasarkan pada residual devians dalam model regresi Poisson sebagai berikut:

$$R^2_{DEV} = 1 - \frac{\ln L(\hat{\Omega}) - \ln L(\hat{\omega})}{\ln L(\hat{\Omega}) - \ln L(\bar{Y})} = 1 - \frac{D(\hat{\boldsymbol{\beta}})}{D(\bar{Y})}$$

Penambahan suatu variabel prediktor kedalam model regresi dapat menaikkan nilai R^2_{DEV} meskipun variabel prediktor tersebut tidak berpengaruh nyata terhadap respon, sehingga Waldhor, dkk (1998) mengusulkan untuk memberikan koreksi terhadap R^2_{DEV} dengan menggunakan derajat bebas, yang dirumuskan sebagai berikut:

$$R^2_{DEV,db} = 1 - \frac{(n-k-1)^{-1}[\ln L(\hat{\Omega}) - \ln(\hat{\omega})]}{(n-1)^{-1}[\ln L(\hat{\Omega}) - \ln(\hat{\omega})]} = 1 - \frac{(n-k-1)^{-1} D(\hat{\boldsymbol{\beta}})}{(n-1)^{-1} D(\bar{Y})}.$$

Koreksi lain terhadap R^2_{DEV} adalah didasarkan pada statistik rasio likelihood, yaitu:

$$R^2_{DEV,adj,1} = 1 - \frac{\ln(\hat{\Omega}) - [\ln L(\hat{\omega}) - (k/2)]}{\ln(\hat{\Omega}) - \ln L(\bar{Y})} \quad \text{dan} \quad R^2_{DEV,adj,2} = 1 - \frac{\ln(\hat{\Omega}) - [\ln L(\hat{\omega}) - (k+1)/2]}{\ln(\hat{\Omega}) - [\ln L(\bar{Y}) - 1/2]}.$$

7. Sumber Data

Data yang digunakan adalah data yang berasal dari penelitian sebelumnya yang berjudul Pemodelan Regresi Poisson pada *Maternal Mortality* di Jawa Timur oleh Setyorini (2006). Data yang digunakan sebagai variabel respon diperoleh dari Dinas Kesehatan Jawa Timur (Dinkes Jatim). Sedangkan data yang digunakan sebagai variabel kovariat berasal dari data hasil Survei Sosial Ekonomi Nasional (SUSENAS) tahun 2003 yang dilakukan oleh BPS Jawa Timur di tiap kabupaten/kota di Jawa Timur. Variabel yang digunakan adalah sebagai berikut:

y = Jumlah *maternal mortality*.

x_1 = Persentase sarana kesehatan (rumah sakit dan puskesmas).

x_2 = Rata-rata pengeluaran biaya kesehatan perkapita sebulan.

x_3 = Persentase wanita yang menikah dibawah umur (kurang dari 17 tahun).

x_4 = Persentase penolong proses persalinan yang dilakukan oleh tenaga nonmedis (dukun bayi).

x_5 = Persentase penduduk miskin.

x_6 = Persentase wanita dengan pendidikan paling tinggi Sekolah Dasar.

x_7 = Persentase tenaga medis dan paramedis.

8. Aplikasi Ukuran R^2_{DEV} pada Data *Maternal Mortality* di Jawa Timur

Ukuran R^2_{DEV} dalam model regresi Poisson akan diaplikasikan pada data *Maternal Mortality* di tiap kabupaten/kota di Jawa Timur untuk menentukan faktor-faktor yang berpengaruh pada jumlah *maternal mortality* di Jawa Timur pada tahun 2003. Sebagai langkah awal, akan dilakukan pengujian distribusi data untuk mengetahui apakah variabel respon (jumlah *maternal mortality*) berdistribusi Poisson atau tidak dengan menggunakan hipotesis berikut:

H_0 : jumlah *maternal mortality* di tiap kabupaten/kota di Jawa Timur pada tahun 2003 berdistribusi Poisson.

H_1 : jumlah *maternal mortality* di tiap kabupaten/kota di Jawa Timur pada tahun 2003 tidak berdistribusi Poisson.

Dalam pengujian distribusi data tersebut tingkat signifikansi (α) yang digunakan adalah sebesar 5%, dan diperoleh nilai *p-value* (0,184) yang berarti terima H_0 sehingga dapat dikatakan bahwa jumlah *maternal mortality* di Jawa Timur pada tahun 2003 berdistribusi Poisson. Selanjutnya dilakukan uji kolinearitas untuk mengetahui apakah antar variabel prediktor telah memenuhi kondisi tidak saling berkorelasi. Kriteria pertama adalah dengan menggunakan koefisien korelasi (*Pearson correlation*).

Tabel 1. Koefisien Korelasi Data *Maternal Mortality*

Variabel	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_2	-0,395						
x_3	0,295	-0,527					
x_4	0,231	-0,712	0,688				

x_5	0,142	-0,632	0,473	0,676			
x_6	0,307	-0,251	0,032	-0,097	0,046		
x_7	0,624	0,130	-0,099	-0,202	-0,268	0,084	
y	0,542	-0,459	0,169	0,400	0,222	0,372	0,253

Koefisien korelasi untuk masing-masing variabel prediktor pada Tabel 1 menunjukkan bahwa semua variabel prediktor mempunyai nilai koefisien korelasi yang lebih kecil dari 0,95, sehingga dapat dikatakan bahwa antar variabel prediktor tidak saling berkorelasi.

Kriteria kedua yang digunakan untuk memeriksa kolinieritas antar variabel prediktor adalah dengan menggunakan nilai *Variance Inflation Factors* (VIF), sebagaimana yang ditunjukkan pada Tabel 2. Nilai-nilai pada tabel tersebut menunjukkan nilai kurang dari 10, sehingga dapat dikatakan bahwa antar variabel prediktor tidak saling berkorelasi.

Tabel 2. Nilai *Variance Inflation Factors* (VIF) Data *Maternal Mortality*

Variabel	x_1	x_2	x_3	x_4	x_5	x_6	x_7
VIF	3,00638	2,29320	2,00595	2,96805	2,32869	1,43115	2,80096

Kriteria ketiga adalah dengan menggunakan nilai *eigen* dari matriks korelasi antar variabel prediktor. Nilai eigen masing-masing variabel prediktor pada Tabel 3 menunjukkan nilai yang lebih besar dari 0,05, sehingga dapat dikatakan bahwa antar variabel prediktor tidak saling berkorelasi.

Tabel 3. Nilai Eigen Data *Maternal Mortality*

Variabel	x_1	x_2	x_3	x_4	x_5	x_6	x_7
Nilai Eigen	3,11214	1,70404	0,95278	0,5229	0,37135	0,19083	0,14593

Langkah selanjutnya adalah melakukan estimasi parameter model regresi Poisson dengan metode maksimum likelihood, dimana fungsi likelihoodnya dimaksimumkan dengan metode Newton-Raphson. Nilai estimasi parameter mencapai konvergen setelah iterasi ke-21. Dengan menggunakan paket program SAS diperoleh hasil sebagaimana yang diberikan pada Tabel 4 berikut.

Pengujian parameter secara serentak untuk mengetahui ada tidaknya pengaruh variabel prediktor terhadap variabel respon dengan dilakukan dengan menggunakan hipotesis berikut:

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$ dan H_1 : Paling sedikit ada satu $\beta_i \neq 0$.

Diperoleh nilai $D(\hat{\boldsymbol{\beta}})$ = 1424,0 sedangkan $\chi^2_{0,5;1}$ = 3,84 sehingga dapat disimpulkan bahwa setiap variabel prediktor memberikan pengaruh yang berbeda terhadap variabel respon.

Tabel 4. Nilai Estimasi Parameter Model Regresi Poisson

Parameter	**Estimasi**	**SE**	**DF**	**t**	**P-Value**	**Lower**	**Upper**
β_0	6,9497	0,07508	34	92,57	0,0001	6,7971	7,1023
β_1	0,004167	0,001555	34	2,68	0,0113	0,001008	0,007327
β_2	0,006399	0,000212	34	30,17	0,0001	0,005968	0,006830
β_3	-0,00005	8,164E-6	34	-5,51	0,0001	-0,00006	-0,00003
β_4	-0,00579	0,000837	34	-6,92	0,0001	-0,00749	-0,00409

β_5	-0,00589	0,000601	34	-9,80	0,0001	-0,00711	-0,00467
β_6	-0,00713	0,001213	34	-5,88	0,0001	-0,00960	-0,00467
β_7	-0,02169	0,001693	34	-12,81	0,0001	-0,02513	-0,01825

Untuk mengetahui pengaruh yang diberikan masing-masing variabel prediktor tersebut, dilakukan pengujian parameter secara parsial dengan menggunakan hipotesis sebagai berikut:

$H_0 : \beta_i = 0$ dan $H_1 : \beta_i \neq 0$.

Tabel 4 menunjukkan bahwa semua parameter yang dihasilkan adalah signifikan karena masing-masing nilai *p-value* lebih kecil dari $\alpha = 0{,}5$.

Selanjutnya dilakukan pemodelan dengan menggunakan nilai R^2 berdasarkan residual devians. Tabel 5 menunjukkan bahwa nilai R^2 terbesar diperoleh ketika variabel x_2 (rata-rata pengeluaran biaya kesehatan perkapita sebulan) yang pertama kali dimasukkan kedalam model, yaitu sekitar 40% keragaman jumlah *maternal mortality* dapat dijelaskan oleh rata-rata pengeluaran biaya kesehatan perkapita sebulan. Penambahan variabel x_5 (persentase penduduk miskin) kedalam model memberikan kontribusi terbesar bagi peningkatan nilai R^2 dibanding variabel prediktor lainnya, yaitu 14%. Sekitar 54% keragaman jumlah *maternal mortality* dapat dijelaskan oleh rata-rata pengeluaran biaya kesehatan perkapita sebulan dan persentase penduduk miskin.

Nilai R^2 meningkat sebesar 6% ketika variabel x_7 (jumlah tenaga medis dan non medis) dimasukkan kedalam model. Sekitar 60% keragaman jumlah *maternal mortality* dapat dijelaskan oleh rata-rata pengeluaran biaya kesehatan perkapita sebulan, persentase penduduk miskin dan jumlah tenaga medis dan paramedic. Penambahan variabel prediktor selanjutnya kedalam model, hanya memberikan kontribusi yang sangat kecil bagi peningkatan nilai R^2.

Jadi model regresi Poisson terbaik untuk pemodelan jumlah *maternal mortality* di tiap kabupaten/kota di Jawa Timur pada tahun 2003 adalah:

$$\ln(\hat{\mu}) = 6{,}5834 + 0{,}0073\mathrm{x}_2 - 0{,}0078\mathrm{x}_5 - 0{,}0183\mathrm{x}_7 . \qquad (20)$$

Model pada persamaan (20) menjelaskan bahwa rata-rata jumlah *maternal mortality* di tiap kabupaten/kota di Jawa Timur akan meningkat sebesar exp(0,0073) sama dengan 1 orang jika rata-rata pengeluaran biaya kesehatan perkapita sebulan meningkat satu rupiah dengan syarat persentase penduduk miskin dan jumlah tenaga medis dan paramedis konstan.

Tabel 5. Nilai R^2 (%) dari Model Regresi Poisson

Sumber	R^2_{DEV}	$R^2_{DEV,db}$	$R^2_{DEV,adj,1}$	$R^2_{DEV,adj,2}$
β_0 dan x_1	0,1767246	0,1509972	0,1764107	0,1751001
β_0 dan x_2	0,3993139	0,3805425	0,399	0,3976195
β_0 dan x_3	0,0057409	-0,02533	0,005427	0,0041701
β_0 dan x_4	0,0260953	-0,004339	0,0257814	0,024518
β_0 dan x_5	0,066242	0,037062	0,0659281	0,0646521
β_0 dan x_6	0,1120342	0,0842853	0,1117203	0,11043

β_0 dan x_7	0,005696	-0,025376	0,0053821	0,0041252
β_0 , x_2 , dan x_1	0,3993208	0,3605673	0,3986929	0,3969988
β_0 , x_2 , dan x_3	0,4481813	0,4125801	0,4475534	0,4458439
β_0 , x_2 , dan x_4	0,5259125	0,4953262	0,5252846	0,5235507
β_0 , x_2 , dan x_5	0,5394427	0,5097293	0,5388149	0,5370767
β_0 , x_2 , dan x_6	0,5336375	0,5035496	0,5330096	0,5312733
β_0 , x_2 , dan x_7	0,4289326	0,3920896	0,4283048	0,4266013
β_0 , x_2 , x_5 , dan x_1	0,5399263	0,4939189	0,5389845	0,5369325
β_0 , x_2 , x_5 , dan x_3	0,5408043	0,4948847	0,5398625	0,5378102
β_0 , x_2 , x_5 , dan x_4	0,5638586	0,5202444	0,5629168	0,5608573
β_0 , x_2 , x_5 , dan x_6	0,5613733	0,5175106	0,5604315	0,5583728
β_0 , x_2 , x_5 , dan x_7	0,5974749	0,5572224	0,5965332	0,5944631
β_0 , x_2 , x_5 , x_7 , dan x_1	0,6071893	0,5530085	0,6059336	0,6035468
β_0 , x_2 , x_5 , x_7 , dan x_3	0,6013413	0,5463539	0,6000856	0,5977006
β_0 , x_2 , x_5 , x_7 , dan x_4	0,6133151	0,5599793	0,6120595	0,6096707
β_0 , x_2 , x_5 , x_7 , dan x_6	0,6143905	0,561203	0,6131349	0,6107458

Rata-rata jumlah *maternal mortality* di tiap kabupaten/kota di Jawa Timur akan berkurang sebesar exp(0,0078) sama dengan 1 orang jika persentase penduduk miskin bertambah satu persen dengan syarat rata-rata pengeluaran biaya kesehatan perkapita dan jumlah tenaga medis dan paramedis konstan. Kondisi ini tentunya kontradiksi dengan dugaan umum karena semakin tinggi pengeluaran biaya kesehatan maka semakin tinggi pula jaminan fasilitas dan layanan kesehatan ibu sebelum dan sesudah melahirkan sehingga resiko kematian ibu bisa diminimalisir.

Rata-rata jumlah *maternal mortality* di tiap kabupaten/kota di Jawa Timur akan berkurang sebesar exp(0,00183) sama dengan 1 orang jika jumlah tenaga medis dan paramedis bertambah satu orang dengan syarat persentase penduduk miskin dan rata-rata pengeluaran biaya kesehatan perkapita konstan.

Tabel 5 juga menunjukkan bahwa nilai $R^2_{DEV,db}$, $R^2_{DEV,adj,1}$, dan $R^2_{DEV,adj,2}$ selalu lebih rendah dari nilai R^2_{DEV}. Hal ini disebabkan ketiga ukuran R^2 tersebut dikoreksi dengan derajat bebasnya. Perbedaan yang cukup signifikan terlihat antara nilai $R^2_{DEV,db}$ dengan nilai R^2 lainnya. $R^2_{DEV,db}$ didalam model regresi Poisson lebih tepat digunakan untuk *mean* yang besar, yaitu ketika regresi Poisson dapat didekati dengan regresi linear (Waldhor, 2000).

Model terbaik dengan menggunakan nilai R^2 berdasarkan residual devians pada penelitian ini menunjukkan bahwa faktor-faktor yang mempengaruhi jumlah *maternal mortality* di tiap kabupaten/kota di Jawa Timur pada tahun 2003 adalah rata-rata pengeluaran biaya kesehatan perkapita, persentase penduduk miskin dan jumlah tenaga medis dan paramedis. Sedangkan pada penelitian sebelumnya, model terbaik yang diperoleh berdasarkan nilai devians terkecil menunjukkan bahwa variabel prediktor yang mempunyai pengaruh yang signifikan pada jumlah *maternal mortality* di tiap kabupaten/kota di Jawa Timur pada tahun 2003 adalah jumlah sarana kesehatan dan persentase penolong persalinan yang dilakukan oleh tenaga non medis (dukun bayi).

9. Kesimpulan

Berdasarkan hasil analisis, maka disimpulkan bahwa ukuran R^2 berdasarkan residual devians (R^2_{DEV}) diperoleh dengan menentukan fungsi log-likelihood dari fungsi distribusi Poisson kemudian menentukan nilai devians dari model regresi Poisson. Ukuran R^2_{DEV} merupakan selisih antara perbandingan nilai devians model lengkap dengan nilai devians ketika hanya intersep didalam model. Nilai R^2_{DEV} dapat dikoreksi dengan $R^2_{DEV,db}$, $R^2_{DEV,adj,1}$, dan $R^2_{DEV,adj,2}$. Dengan menggunakan nilai R^2 berdasarkan residual devians dapat ditentukan model terbaik. Nilai R^2 yang diperoleh dari hasil analisis adalah $R^2_{DEV,db} = 56\%$, dan $R^2_{DEV} = R^2_{DEV,adj,1} = R^2_{DEV,adj,2} = 60\%$, yang berarti sekitar 60% keragaman variabel respon dapat dijelaskan oleh variabel prediktor sedangkan sisanya dipengaruhi oleh variabel lain. Pemodelan regresi Poisson memberikan hasil bahwa faktor-faktor yang berpengaruh pada jumlah *maternal mortality* di Jawa Timur pada tahun 2003 adalah rata-rata pengeluaran biaya kesehatan perkapita sebulan, persentase penduduk miskin, dan jumlah tenaga medis dan paramedis.

Daftar Pustaka

[1] Baharuddin, 2005, Ukuran R^2 dalam model Regresi Poisson, *Integral***,** Vol. 10, No.3.

[2] Cameron, A.C. dan Trivedi, P.K., 1998, *Regresi Analysis of Count Data*, Cambridge University Press, United Kingdom.

[3] Cameron, A.C. dan Windmeijer, F.A.G, 1996, R^2 measures for count data regression models with applications to healthcare utilization, *Journal Business Economic Statistic*, 14, pp.209 – 220.

[4] Hocking, R. R., 1996, *Methods and Application Log Linear Models*, John Wiley and Sons, Inc., New York.

[5] Kleinbaum, D.G., Kupper, L.L., dan Keith E.M., 1978, *Applied Regression Analysis and Other Multivariable Methods*, PWS-Kent Publishing Company, New York.

[6] McCullagh, P. dan Nelder, J.A, 1989, *Generalized Linear Models*, 2nd Edition, Chapman & Hall, London.

[7] Mittlböck, M. dan Waldhör, T., 2000, Adjustments for R^2 measures for Poisson regression models, *Computational Statistics and Data Analysis*, 34, pp. 461–472.

[8] Mittlböck, M., 2002, Calculating adjusted R^2 measures for Poisson regression models, *Computational Statistics and Data Analysis*, 68, pp. 205–214.

[9] Myers, R.H., 1990, *Classical and Modern Regression With Applications*, PWS Kent Publishing Company, USA.

[10] Royston, E. dan Amstrong, S., 1994, *Pencegahan Maternal mortality Hamil*, Binarupa Aksara, Jakarta.

[11] Sahrudin, 2008, *Angka Kematian Ibu Melahirkan Masih Tinggi*, http://www.BeritaJakarta.com, [diakses pada tanggal 21 Agustus 2008].

[12] Setyorini, E., 2006, Pemodelan regresi Poisson pada maternal mortality di Jawa Timur, *Laporan Tugas Akhir FMIPA Statistika ITS*, Surabaya.

[13] Sukarni, M., 1994, *Kesehatan Keluarga dan Lingkungan*, Kanisius, Yogyakarta.

[14] Waldhör, T., Haidinger, G. dan Schober, E., 1998, Comparison of R^2 measures for Poisson regression by simulation, *Journal Epidemiol. Biostatist*., 3, pp. 209-215.

[15] Walpole, R.E, 1982, *Pengantar Statistika*, edisi ketiga, Gramedia Pustaka Utama, Jakarta.