

## Metode Pemilihan Variabel pada Model Regresi Poisson Menggunakan Metode Nordberg

Muhammad Bangkit Riksa Utama\*, Nusar Hajarisman

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Islam Bandung, Indonesia.

\*muhammadbangkitriksautama@gmail.com, nusarhajarisman@unisba.ac.id

**Abstract.** In various experiments, data interactions take the form of discrete numbers or counts. The model that can be used for these data is the Poisson regression model. Poisson regression is included in the Generalized Linear Model (GLM). Poisson regression in general is very important in various fields and agreed to receive special attention. Often this model needs many independent variables. Then there needs to be a selection of poisson regression model variables. Due to the number of independent variables that exist, the selection of variables is carried out. Variable selection techniques that are commonly known are the forward, backward method, akaike information criteria and several other methods. In this paper, we will discuss one method of selecting variables in the Poisson regression model that has been made in the algorithm created by Famoye and Rothe. The algorithm created will be compared with the algorithm made by Nordberg. In this study data were used on Infant Mortality Rate (IMR) in West Java Province.

**Keywords:** Infant Mortality, Poisson Regression, Famoye and Rothe.

**Abstrak.** Dalam berbagai eksperimen, seringkali data berupa bilangan diskrit atau cacah. Model yang dapat digunakan untuk data tersebut diantaranya adalah model regresi poisson. Regresi poisson termasuk kedalam Generalized Linear Model (GLM). Regresi poisson secara umum sangat penting dalam berbagai bidang dan karenanya patut mendapat perhatian khusus. Seringkali model ini melibatkan banyak variabel independen. Maka perlu adanya cara untuk mempertimbangkan pemilihan variabel model regresi poisson. Dikarenakan banyaknya variabel independen yang ada maka dilakukan penyeleksian variabel. Teknik pemilihan variabel yang sudah biasa dikenal yaitu metode forward, backward, akaike information criterion dan beberapa metode lainnya. Pada makalah ini akan dibahas mengenai salah satu metode pemilihan variabel dalam model regresi poisson yang telah dibentuk dalam algoritma yang dibuat oleh Famoye dan Rothe. Algoritma yang dibuat ini akan dibandingkan dengan algoritma yang telah dibuat oleh Nordberg. Pada penelitian ini digunakan data mengenai Angka Kematian Bayi (AKB) di Provinsi Jawa Barat.

**Kata Kunci:** Angka Kematian Bayi, Regresi Poisson, Famoye dan Rothe.

## A. Pendahuluan

Analisis regresi merupakan salah satu metode statistika yang digunakan untuk menganalisis hubungan antara satu variabel dependen dengan satu atau lebih variabel independen. Salah satu tujuan analisis regresi adalah untuk tujuan prediksi. Semakin banyak variabel yang ada dalam model maka akan semakin baik model tersebut dalam melakukan fungsi prediksinya. Namun, banyaknya variabel yang masuk memberikan permasalahan dalam sulitnya mengumpulkan data dan kontrol setiap variabel. Sehingga diperlukan adanya seleksi variabel yaitu memilih variabel yang benar-benar memberikan informasi dalam keakuratan prediksi. Pemilihan variabel independen yang tepat dalam pembuatan model regresi khususnya untuk tujuan prediksi merupakan satu hal yang sangat penting. (Breiman & Friedman, 1997; Bilodeau & Brenner, 1999). Umumnya seleksi model dalam pembentukan model regresi terbaik untuk regresi univariate dilakukan dengan metode maju (forward selection), metode mundur (backward selection) dan juga metode bertahap (stepwise selection).

Permasalahan yang sering dihadapi dalam pembentukan model regresi adalah menentukan himpunan variabel independen terbaik sehingga dapat memberikan hasil prediksi yang paling akurat. Dalam kaitan pemilihan variabel independen terdapat dua kriteria yang saling bertentangan satu dengan yang lainnya. Pertama, agar persamaan bermanfaat bagi tujuan prediksi, peneliti biasanya ingin memasukkan sebanyak mungkin variabel independen ke dalam model sehingga diperoleh nilai prediksi yang handal. Kedua, oleh karena untuk memperoleh informasi dari banyak variabel serta pememonitorannya seringkali diperlukan biaya yang tinggi, maka salah satu caranya adalah dengan memasukkan sedikit mungkin variabel independen ke dalam model. Atau dengan kata lain, dalam pembentukan model selalu diinginkan model yang paling sederhana. Kompromi kedua ekstrim tersebut yang biasanya disebut pemilihan model regresi terbaik.

Dalam model regresi linear biasa dibentuk dengan asumsi bahwa bahwa unsur-unsur dari residu atau galat adalah berdistribusi normal identik dan saling bebas,  $\varepsilon \sim N(0, \sigma^2)$ , merupakan basis dari kebanyakan analisis untuk data-data kontinu. Namun demikian, dengan adanya berbagai kelebihan dalam teori statistik dan perangkat lunak komputer, maka dapat dikembangkan suatu model linear dalam beberapa situasi sebagai berikut: variabel dependen mempunyai distribusi selain distribusi normal (dapat berbentuk kategori daripada kontinu, serta hubungan antara variabel dependen dengan independennya tidak perlu berbentuk linear sederhana. Salah satu dari kelebihan ini yang telah banyak dikenal juga sebagai sifat-sifat yang 'baik' dari distribusi normal dibagi ke dalam kelas yang lebih luas dari suatu distribusi yang disebut juga sebagai distribusi dari keluarga eksponensial.

Model ini dikenal dengan Generalized Linear Model (GLM). Istilah GLM merujuk pada kelas model yang lebih besar yang dipopulerkan oleh McCullagh dan Nelder (1989). Dalam model ini, variabel dependen diasumsikan mengikuti distribusi dari keluarga eksponensial dengan rata-rata  $\mu$ , yang biasanya diasumsikan sebagai suatu fungsi (seringkali bentuknya nonlinear) dari  $X^T \beta$ . Beberapa penulis mengatakan bentuknya adalah nonlinear karena  $\mu$  seringkali merupakan fungsi nonlinear dari kovariat, tetapi McCullagh dan Nelder (1983) mempertimbangkan fungsi tersebut sebagai bentuk yang linear, karena kovariat ini mempengaruhi distribusi dari variabel dependen hanya melalui kombinasi linear dari  $X^T \beta$ . Salah satu bentuk model dari GLM yang cukup populer, selain model regresi logistik, adalah model regresi poisson. Model regresi poisson digunakan untuk menganalisis hubungan antara beberapa variabel, dimana variabel responnya berbentuk diskrit tapi tidak biner.

Sebagaimana dalam analisis regresi biasa, masalah pemilihan variabel juga merupakan isu yang penting dalam GLM ini. Secara umum, di dalam GLM, perbandingan antara nilai pengamatan dari variabel respon terhadap nilai dugaan yang diperoleh dari model dengan atau tanpa variabel independen yang diamati yang berdasarkan pada fungsi log kemungkinannya merupakan suatu ukuran untuk mengidentifikasi variabel yang dianggap penting di dalam model. Selanjutnya, Tujuan dari penelitian ini adalah untuk memperoleh variabel-variabel yang signifikan didalam model regresi poisson menggunakan metode forward sebagaimana yang pertama kali diusulkan oleh Nordberg yang dikembangkan oleh Famoye dan Rothe.

Adapun manfaat dari penulisan makalah ini adalah untuk menambah wawasan

perkembangan dalam pemodelan statistika, khususnya mengenai pemilihan variabel independen pada konteks regresi poisson.

## B. Landasan Teori

### Model Linier Tergeneralisir

Misalkan  $y_i$  ( $i = 1, 2, 3, \dots, n$ ) merupakan variabel acak yang saling bebas sedemikian rupa sehingga fungsi peluang  $f_i$  untuk  $y_i$  bergantung pada vektor parameter  $\beta = (\beta_0, \beta_1, \dots, \beta_k)$  dan parameter skala  $\varphi$  melalui struktur persamaan sebagai berikut:

$$f_i(y; \beta, \theta) = \exp \left\{ \frac{y\theta - b\theta}{a(\varphi)} + c(y; \varphi) \right\} \quad (1)$$

dimana  $\theta_i$ ,  $i = 1, 2, \dots, n$  bergantung pada  $\beta$  melalui fungsi sebagai berikut:

$$\theta_i = f(\eta_i) \quad (2)$$

dan

$$\eta_i = \sum_{m=0}^k \beta_m x_{mi} \quad (3)$$

Fungsi peluang yang dinyatakan dalam persamaan 2.1 harus memenuhi beberapa asumsi, diantaranya yaitu:

1.  $(x_{mi}; m = 0, 1, 2, \dots, k; i = 1, 2, \dots, n)$  merupakan variabel independen yang diketahui.
2. Fungsi  $a(\cdot)$ ,  $b(\cdot)$ , dan  $c(\cdot)$  merupakan fungsi yang diketahui dan merupakan turunan, serta mengikuti 3 sifat sebagai berikut:

### Model Regresi Poisson

Model regresi poisson diterapkan untuk variabel dependen yang berbentuk cacahan atau frekuensi, seperti banyaknya kejadian yang terjadi dalam suatu waktu atau ruang tertentu. Diasumsikan bahwa  $y \sim \text{Poisson}(\lambda_i)$ ,  $i = 1, \dots, N$ , kemudian akan dihubungkan nilai tengah  $\lambda_i$  pada suatu kovariat tertentu. Suatu distribusi merupakan anggota dari keluarga eksponensial apabila mempunyai bentuk sebagai berikut:

$$f(y, \theta) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\varphi)} + c(y, \varphi) \right\} \quad (4)$$

$$= \exp \left\{ -\frac{b(\theta)}{a(\varphi)} \right\} \exp \left\{ \frac{\theta}{a(\varphi)} y \right\} c(y, \theta) \quad (5)$$

dimana  $\theta$  disebut juga sebagai parameter kanonik dan  $\varphi$  disebut juga sebagai parameter dispersi. Pada variabel acak  $y$  yang berdistribusi poisson dengan parameter  $\lambda$  sebagai berikut:

$$f(y_i, \lambda_i) = \frac{e^{-\lambda} \lambda^{y_i}}{y_i!} \quad (6)$$

$$= \exp[y_i \log \lambda_i - \lambda_i - \log y_i!] \quad (7)$$

Dimana  $\theta_i = \log \lambda_i$ . Bentuk fungsi densitas di atas akan memberikan  $b\theta_i = \exp(\theta_i)$ ,  $a(\varphi) = 1$ , dan  $c(y, \varphi) = -\log y_i!$

### Pencocokan Model

Rata-rata dalam regresi poisson dimodelkan sebagai fungsi dari sejumlah Variabel independen. Pertama, kita perlu menentukan fungsi kemungkinan dan persamaan kemungkinan yang digunakan untuk menaksir parameter-parameter dalam regresi poisson. Untuk memaksimalkan fungsi kemungkinan dengan menggunakan teknik iteratif akan menghasilkan penaksir kemungkinan maksimum untuk koefisien regresi dalam  $\beta$ .

Penaksir kemungkinan maksimum bagi  $\beta$  dapat ditentukan melalui metode *Fisher scoring*:

$$\beta^{(t+1)} = \beta^{(t)} + [-E l''(\beta^{(t)})]^{-1} l'(\beta^{(t)}) \quad (8)$$

dengan mengabaikan konstanta, maka fungsi log-likelihoodnya adalah

$$l(\theta; y) = \frac{y\theta - b'(\theta)}{a(\phi)} \quad (9)$$

Menurut McCullagh dan Nelder (1983) untuk sembarang fungsi hubung, dapat ditunjukkan bahwa vektor skor dan matriks informasinya adalah

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= \frac{y - \mu}{\text{var}(y)} \left( \frac{\partial \mu}{\partial \eta} \right) x_{ij} \\ -E \left( \frac{\partial^2 l}{\partial \beta_j \partial \beta_k} \right) &= \frac{1}{\text{var}(y)} \left( \frac{\partial \mu}{\partial \eta} \right)^2 x_{ij} x_{ik}. \end{aligned}$$

Vektor skor dapat juga dituliskan sebagai:

$$\frac{\partial l}{\partial \beta} = X^T A (y - \mu), \quad (10)$$

dimana  $\mathbf{X} = (x_1, \dots, x_N)^T$ ,  $\mathbf{y} = (y_1, \dots, y_N)^T$ , dan  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_N)^T$ , serta

$$A = \text{Diag} \left[ [\text{var}(y_i)]^{-1} \left( \frac{\partial \mu_i}{\partial \eta_i} \right) \right] \quad (11)$$

$$= \text{Diag} \left[ \text{var}(y_i) \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^{-1} \right] \quad (12)$$

Nilai harapan dari matriks Hessian atau disebut juga matriks informasinya menjadi:

$$-E \left( \frac{\partial^2 l}{\partial \beta \partial \beta^T} \right) = X^T W X \quad (13)$$

dimana

$$W = \text{Diag} \left[ [\text{var}(y_i)]^{-1} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 \right] \quad (14)$$

$$= \text{Diag} \left[ \text{var}(y_i) \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^{-2} \right] \quad (15)$$

dengan demikian suatu iterasi dari metode *Fisher scoring* diberikan oleh

$$\beta^{(t+1)} = \beta^{(t)} + (X^T W X)^{-1} X^T A (y - \mu) \quad (16)$$

dimana  $\mathbf{W}$ ,  $\mathbf{A}$ , dan  $\boldsymbol{\mu}$  dievaluasi pada  $\beta^{(t)}$ .  $\mathbf{A}$  &  $\mathbf{W}$  dapat dihubungkan melalui:

$$A = W \left( \frac{\partial \eta}{\partial \mu} \right) \quad (17)$$

dimana:

$$\frac{\partial \eta}{\partial \mu} = \text{Diag} \left( \frac{\partial \eta_i}{\partial \mu_i} \right) \quad (18)$$

dengan demikian iterasi *Fisher scoring* dapat juga dituliskan sbb:

$$\beta^{(t+1)} = (X^T W X)^{-1} X^T w_z \quad (19)$$

dimana

$$z = \eta + \left( \frac{\partial \eta_i}{\partial \mu_i} \right) (y - \mu) = (z_1, z_2, \dots, z_N)^T \quad (20)$$

dan

$$z_i = \eta_i + \left( \frac{\partial \eta_i}{\partial \mu_i} \right) (y_i - \mu_i) \quad (21)$$

Dalam konteks GLM, variabel  $z$  ini sering juga disebut sebagai *working variate*. Dengan demikian *Fisher-scoring* dapat dianggap sebagai metode *iteratively re-weighted least square* (IRWLS). Dalam model poisson, dengan fungsi hubung log, maka prosedur IRWLS atau

*Fisher-scoring* juga merupakan prosedur *Newton-Raphson*. Untuk model poisson dengan fungsi hubung log, diketahui bahwa  $\eta_i = \log \mu_i$  sehingga diperoleh

$$\frac{\partial \eta_i}{\partial \mu_i} = \frac{1}{\mu_i} \quad (22)$$

$$W = \text{Diag } \mu_i \quad (23)$$

$$z_i = \eta_i + \frac{(y_i - \mu_i)}{\mu_i} \quad (24)$$

Teori tersebut dapat dilakukan dengan algoritma Sebagai berikut:

1. Hitung:

$$\eta_i = x_i \beta \quad (25)$$

$$\mu_i = \exp(\eta_i) \quad (26)$$

$$z_i = \eta_i + \frac{(y_i - \mu_i)}{\mu_i} \quad (27)$$

2. Regresikan  $z$  pada  $X$  dengan menggunakan  $\mu_i$  sebagai pembobot untuk mendapat  $\beta$  yang baru.
3. Ulangi sampai langkah 1 dan 2 sampai memperoleh nilai  $\beta$  yang konvergen.

**Catatan:** Penduga matriks kovarians bagi  $\beta$  diberikan oleh  $(X^T W X)^{-1}$

### Kriteria Pemilihan Variabel

#### Information Criterion (AIC)

AIC merupakan model terbaik dari beberapa model yang disajikan, dipilih berdasarkan kriteria yang digunakan. Adapun pilihan kriteria sebagai perbandingan model untuk memperoleh model terbaik, antara lain *Akaike's Information Criterion* (AIC), *Bayesian Information Criterion* (BIC) dan *Root Mean Square Error* (RMSE).

AIC adalah suatu kriteria yang menyeimbangkan *goodness of fit* model berdasarkan nilai likelihood dengan banyaknya parameter dari model. Sehingga, model yang terbaik adalah model dengan nilai (AIC) terkecil. Nilai AIC diperoleh dari persamaan sebagai berikut:

$$AIC = -2 \log L_{fit} + 2k \quad (28)$$

#### Determinan Merkle dan Zimmermann

$R^2$  merupakan nilai persentase jumlah data dari variabel independen yang secara bersama-sama mempengaruhi variabel dependen, kali ini metode determinan dimodifikasi untuk menjadi suatu nilai yang dapat menyeleksi variabel, dimana variabel tersebut bisa dihilangkan karna tidak memiliki pengaruh yang signifikan terhadap model, Merkle dan Zimmermann mendefinisikan:

$$R_D^2 = \frac{l(\hat{\mu}_i) - l(\bar{y})}{l(y_i) - l(\bar{y})} \quad (29)$$

Kuantitas  $R_D^2$  *good-of-fit* berkaitan peningkatan yang dijelaskan dalam log-kemungkinan untuk kenaikan maksimum yang mungkin  $R_D^2$  yang tinggi menunjukkan kecocokan model Cameron dan Windmeijer (1996) menganalisis  $R^2$  langkah-langkah untuk menghitung data. Mereka membangun lima kriteria untuk menilai berbagai tindakan  $R^2$ . Diantara semua tindakan  $R^2$  yang dipertimbangkan, hanya tindakan  $R_D^2$  yang didefinisikan oleh Merkle dan Zimmermann (1992) memenuhi semua lima kriteria.

#### Uji Wald

Tiga cara yang biasa digunakan untuk menggunakan fungsi kemungkinan adalah untuk membentuk inferensi pada sampel besar. Pada bagian ini akan diperkenalkan inferensi itu untuk menguji hipotesis nol ( $H_0: \beta = \beta_0$ ), kemudian akan dibahas hubungannya dengan pendugaan selang untuk galat baku dari  $\hat{\beta}$  maka statistik uji:

$$z = (\hat{\beta} - \beta_0)/SE \quad (30)$$

Mengikuti pendekatan distribusi normal baku pada saat  $\beta = \beta_0$  nilai  $z$  ini kemudian digunakan untuk memperoleh  $p$ -value satu pihak atau dua pihak dari tabel normal baku. Selain itu, dapat juga digunakan alternatif dua pihak,  $z^2$  yang mengikuti distribusi chi-kuadrat dengan ( $db = 1$ ). Statistik uji seperti ini, disebut juga sebagai Statistik *Wald*. Perluasan secara multivariat dari Statistik *Wald* ini adalah untuk menguji hipotesis nol  $H_0: \beta = \beta_0$ , dimana statistik ujinya adalah

$$w = (\hat{\beta} - \beta_0) \sqrt{[cov(\hat{\beta})]^{-1}} (\hat{\beta} - \beta_0) \quad (31)$$

Tujuan penggunaan metode fungsi kemungkinan adalah melalui rasio dari dua proses pemaksimalan, yaitu (1) maksimum menurut nilai parameter yang mungkin dibawah  $H_0$ , serta (2) maksimum menurut segugus nilai parameter yang lebih besar pada saat  $H_0$  atau  $H_1$  yang dianggap besar.

### Prosedur Pemilihan Variabel Dalam GLM

#### 1. Metode Algoritma Nordberg

Apabila variabel  $x$  dan  $y$  dalam GLM dapat ditransformasi kedalam bentuk variabel  $w_i$  dan  $z_i$  sebagai berikut:

$$w_i = \text{Diag}(\mu_i) \quad (32)$$

dan

$$z_i = \eta_i + \left( \frac{y_i - \mu_i}{\mu_i} \right), i = 1, 2, \dots, n \quad (33)$$

dimana

$$\eta_i = x\beta \quad (34)$$

dan

$$\mu_i = e^{xb} \quad (35)$$

dengan demikian masalah pemilihan variabel dalam GLM dapat dilakukan sebagaimana pemilihan variabel dalam model regresi linier biasa dengan model sebagai berikut:

$$\hat{z}_i = b_0 + b_1 w_{mi} \quad (36)$$

dimana  $\varepsilon_i$  merupakan variabel acak tidak berkorelasi yang mempunyai rata-rata tidak sama dengan nol dan bervariansi sama (homogen).

Hosmer dan Lemeshow (1989) telah memberikan suatu strategi pemilihan variabel bebas ke dalam model. Proses pemilihan variabel ini dimulai dengan analisis variabel tunggal (*univariat*) untuk masing masing variabel yang ada kemudian dilakukan analisis variabel (*multivariat*).

#### 2. Faktor Penyebab Kematian Ibu

Data SDKI khusus dirancang untuk mengumpulkan berbagai informasi mengenai tingkat kelahiran, mortalitas, prevalensi keluarga berencana dan kesehatan khususnya kesehatan reproduksi. Sejumlah variabel yang akan digunakan dalam penelitian ini merupakan variabel yang dianggap mempengaruhi Angka Kematian Bayi. Variabel dependen yang diamati adalah variabel BM, yaitu jumlah kematian bayi (berumur kurang 1 tahun) pada satu tahun tertentu di daerah (kabupaten/kota) tertentu. Sedangkan variabel LH merupakan variabel yang menyatakan jumlah kelahiran hidup pada satu tahun tertentu di daerah tertentu. Kemudian, terdapat 10 variabel independen yang akan diamati dalam penelitian ini, yaitu:

**Tabel 1.** Faktor Angka Kematian Bayi

Variabel	Keterangan
NAKES	Persentase persalinan yang ditolong bukan oleh tenaga kesehatan.
KN1	Perentase ibu yang tidak melakukan kunjungan bayi. Kunjungan anak usia kurang dari satu tahun(29 hari-11 bulan) untuk mendapatkan pelayanan
MSKN	Persentase penduduk miskin.
ASI	Persentase bayi yang tidak mendapat ASI eksklusif bayi berusia 6 bulan.
POSY	Persentase rasio ketersediaan POSYANDU terhadap penduduk.
RSHT	Persentase rumah tidak sehat. Bangunan rumah tinggal yang memenuhi syarat kesehatannya itu memiliki jamban sehat, tempat pembuangan sampah, sarana air bersih, sarana pembuangan air limbah, ventilasi baik, kepadatan hunian rumah sesuai dan lantai rumah tidak dari tanah.
K1	Persentase ibu hamil yang tidak melakukan kunjungan untuk mendapatkan pelayanan kesehatan oleh tenaga kesehatan profesional selama masa kehamilannya.
RTKS	Rasio ketersediaan tenaga kesehatan profesional dan tenaga kesehatan masyarakat terhadap penduduk.

### C. Hasil Penelitian dan Pembahasan

Dari hasil perhitungan yang dibantu dengan penggunaan aplikasi SAS dengan Sintaks. Didapatkan hasil sebagai berikut:

**Tabel 2.** Hasil perhitungan SAS

Peubah	Penduga Koefisien	Galat Baku	Statistik Wald	Nilai-p	Keterangan
MSKN	0.0036	0.0002	462.20	<.0001	Signifikan
K1	0.0049	0.0007	57.33	<.0001	Signifikan
KN1	0.0094	0.0007	176.33	<.0001	Signifikan
ASI	0.0444	0.0017	692.58	<.0001	Signifikan
RSHT	0.0000	0.0000	90.73	<.0001	Signifikan
RTKS	-0.0004	0.0007	0.35	0.5563	Tidak Signifikan
NAKES	0.0000	0.0001	0.26	0.6078	Tidak Signifikan

Pada tabel 1 didapat nilai P-value yang akan dibandingkan dengan nilai kritis yang pada penelitian ini digunakan nilai sebesar 0,05. Dapat disimpulkan bahwa ada 5 variabel yang memiliki pengaruh signifikan terhadap angka kematian ibu yaitu MSKN, K1, KN1, ASI,RSHT. Setelah menyeleksi variabel yang berpengaruh selanjutnya menyeleksi model terbaik dimana setiap variabel dilakukan pengkombinasian untuk melihat model.

Setelah melakukan pengkombinasian teradap 5 variabel yang signifikan, didapatlah model terbaik yang mana pemilihan model terbaik ini berdasarkan pada nilai Pearson Chi-Square terbesar, nilai Log Likelihood terbesar dan Nilai AIC terkecil. Maka didapatkanlah variabel terbaik sebagai berikut:

**Tabel 3.** Hasil perhitungan SAS

Nama Variabel	Ukuran Kecocokan Model		
	Pearson Chi-Square	Log Likelihood	AIC
MSKN, K1, RSHT	792.2136	7739.7452	901.056

Model terbaik ini merupakan model yang dimana setiap variabelnya memiliki pengaruh yang besar terhadap angka kematian bayi di Provinsi Jawa Barat.

#### D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil sebagai berikut:

1. Hasil perhitungan variabel terbaik menunjukkan bahwa kombinasi MSKN, K1, RSHT memiliki pengaruh yang kuat terhadap angka kematian ibu.
2. ada 5 faktor utama yang berpengaruh secara signifikan terhadap angka kematian ibu di Jawa Barat yaitu MSKN, K1, KN1, ASI, RSHT.

#### Daftar Pustaka

- [1] Ayu Famoye, F. And Rothe, D.E., 2003. Variable Selection For Poisson Regression Model. Nordberg, L. 1982. On variable Selection In Generalized Linier And Related Regression Models.
- [2] Breiman, L. and Friedman, J. H. (1997), "Predicting Multivariate Responses in Multiple Linear Regression", *Journal of the Royal Statistical Society*, 59 (No. 1), 3-54
- [3] Erti dan Suwanda. (2011). Diagram Kontrol VV yang direduksi. Usulan penelitian Tesis. Program Magister Statistika Terapan Universitas Padjajaran Bandung.
- [4] Kusuma, Komalasari, Hadijati, 2013. Model Regresi Zero Inflated Poisson Pada Data Overdispersion
- [5] Kismiantini. 2008. Perbandingan Model Regresi Poisson Dan Model Regresi Binomial Negatif. Universitas Negeri Yogyakarta
- [6] Hajarisman, N., Pengembangan Model Regresi Pada Peubah Acak Respon Diskrit
- [7] Sari, A.N. 2006. Analisis Jalur Faktor Faktor Yang Mempengaruhi Angka Kematian Ibu Di Jawa Timur. Vol:1, No.2
- [8] Sudjana, 2013. Metoda Statistika. Cetakan: 3, Edisi: 7. Tarsito
- [9] Sugiyono. 2012. Metode Penelitian Kuantitatif Kualitatif dan R&D. Alfabeta
- [10] Ardana, I Komang dkk. 2012. Manajemen Sumber Daya Manusia. Yogyakarta: Graha ilmu
- [11] Gourieroux, dkk. (1984) . Estimating Equations for Parameters in Means and Covariances of Multivariate Discrete and Continuous Responses. Vol. 47, No. 3