Jurnal Sains dan Matematika Vol. 19 (3): 91-95 (2011)

Pemodelan Regresi Logistik dalam Penentuan Faktor-Faktor yang Berpengaruh Terhadap Penyakit Jantung Koroner

Abdul Hoyyi
Jurusan Statistika Fakultas MIPA Universitas Diponegoro Semarang
Email: ahy_stat@undip.ac.id

ABSTRACT

Coronary heart disease (CHD) is a health problem and is one of the leading causes of the death in both developed and developing countries. Thereby it needs to analyze the factors which influence it. Some factors supposedly influence the CHD's patients are cholesterol levels, low density lipoprotein levels, high density lipoprotein levels, triglyceride levels, systolic blood pressure, diastolic blood pressure and gender. The data used are secondary data from the medical records of CHD's patients and general check-up records of the healthy people in a hospital in Yogyakarta. The analysis used in this research is binary logistic regression response with dichotomous dependent variables. Of the seven suspected factors, only four significant influence factors namely cholesterol levels, high-density lipoprotein levels, systolic blood pressure and diastolic blood pressure.

Keywords: CHD, logistic regression

PENDAHULUAN

Penyakit Jantung Koroner (PJK) merupakan problema kesehatan utama di negara maju. Di Indonesia telah terjadi pergeseran kejadian penyakit jantung dan pembuluh darah dari urutan ke-l0 tahun 1980 menjadi urutan ke-8 tahun 1986. Sedangkan penyebab kematian tetap menduduki peringkat ke-3. Banyak faktor yang mempengaruhi terjadinya Penyakit Jantung Koroner sehingga usaha pencegahan harus bentuk multifaktorial juga. Pencegahan harus diusahakan sedapat mungkin dengan cara pengendalian faktor-faktor risiko PJK dan merupakan hal yang cukup penting dalam usaha pencegahan PJK, baik primer maupun sekunder. Pencegahan primer lebih ditujukan pada mereka yang sehat tetapi mempunyai risiko tinggi, sedangkan sekunder merupakan upaya memburuknya penyakit yang secara klinis telah diderita[1]. Beberapa faktor risiko terpenting Penyakit Jantung Koroner : kadar kolesterol total dan LDL (Low Density Lipoprotein) tinggi, kadar kolesterol HDL (High Density Lipoprotein) rendah, tekanan darah tinggi (Hipertensi), merokok, diabetes mellitus, kegemukan, riwayat keturunan penyakit jantung dalam keluarga, kurang olah raga, stress[4]. Dalam pendeteksian penyakit jantung koroner ini, dibatasi hanya 7 faktor risiko penyebab timbulnya penyakit jantung koroner ini dengan variabel sebagai berikut : kadar kolesterol, kadar LDL, kadar HDL, kadar trigliserida, tekanan darah sistolik, tekanan darah diastolic dan jenis kelamin[2].

TINJAUAN PUSTAKA

Model Regresi Respon Biner

Model Regresi Logistik Biner digunakan untuk menganalisa hubungan antara satu variabel respon dan beberapa variabel bebas, dengan variabel responnya berupa data kualitatif dikotomi yaitu bernilai 1 untuk menyatakan keberadaan sebuah karakteristik dan bernilai 0 untuk menyatakan ketidakberadaan sebuah karakteristik. Model regresi logistiknya[3]:

$$\pi(x_i) = \frac{e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}}}{1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}}}$$

Logit dari $\pi(x_i)$ adalah

$$\ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi}$$

Jika $g(x_i) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi}$, maka

$$\pi(x_i) = \frac{e^{g(x_i)}}{1+e^{g(x_i)}}$$

Estimasi Parameter

Untuk menentukan estimasi parameter digunakan metode iterasi Newton Raphson yang membutuhkan turunan pertama dan turunan kedua dari fungsi likelihood.

y_i berdistribusi binomial, maka fungsi kepadatan peluangnya

$$p(y_i=1)=\binom{n}{1}\{\pi(x_i)\}^{y_i}\{1-\pi(x_i)\}^{n-y_i}$$

$$= \frac{n!}{1!(n-1)!}\{\pi(x_i)\}^{y_i}\{1-\pi(x_i)\}^{n-y_i}$$

untuk n =1 maka

$$p(y_i = 1) = \{\pi(x_i)\}^{y_i}\{1-\pi(x_i)\}^{1-y_i}$$

Karena observasi-observasi saling bebas maka fungsi likelihoodnya adalah

$$l(\beta) = \prod_{i=1}^{n}\{\pi(x_i)\}^{y_i}\{1-\pi(x_i)\}^{1-y_i}$$

Selanjutnya fungsi log likelihoodnya adalah :

$$L(\beta) = \ln\{l(\beta)\}$$

$$= \ln\left\{\prod_{i=1}^{n}\{\pi(x_i)\}^{y_i}\{1-\pi(x_i)\}^{1-y_i}\right\}$$

$$= \sum_{i=1}^{n}\left[y_i g(x_i) - \ln(1+e^{g(x_i)})\right]$$

dengan $g(x_i) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi}$

Sehingga turunan pertamanya :

$$\frac{\partial L(\beta)}{\partial \beta_0} = \sum_{i=1}^{n}\left[y_i - \frac{e^{g(x_i)}}{1+e^{g(x_i)}}\right]$$

$$= \sum_{i=1}^{n}\left[y_i - \pi(x_i)\right]$$

$$\frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^{n}\left[y_i x_{1i} - \frac{x_{1i}e^{g(x_i)}}{1+e^{g(x_i)}}\right]$$

$$= \sum_{i=1}^{n} x_{1i}\left[y_i - \pi(x_i)\right]$$

.

.

.

$$\frac{\partial L(\beta)}{\partial \beta_p} = \sum_{i=1}^{n}\left[y_i x_{pi} - \frac{x_{pi}e^{g(x_i)}}{1+e^{g(x_i)}}\right]$$

$$= \sum_{i=1}^{n} x_{pi}\left[y_i - \pi(x_i)\right]$$

dalam bentuk matriks

$$\begin{bmatrix} 1 & 1 & . & . & . & 1 \\ x_{11} & x_{12} & . & . & . & x_{1n} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ x_{p1} & x_{p2} & . & . & . & x_{pn} \end{bmatrix} \begin{bmatrix} y_1 - \pi(x_1) \\ y_2 - \pi(x_2) \\ . \\ . \\ . \\ y_n - \pi(x_n) \end{bmatrix} = \mathbf{X'(Y - \pi(x))}$$

Selanjutnya akan dicari turunan keduanya.

$$\frac{\partial^2 L(\beta)}{(\partial \beta_0)^2} = -\sum_{i=1}^{n}\left[\frac{e^{g(x_i)}\left(1+e^{g(x_i)}\right)-\left(e^{g(x_i)}\right)^2}{\left(1+e^{g(x_i)}\right)^2}\right]$$

$$= -\sum_{i=1}^{n}\pi(x_i)\left[1-\pi(x_i)\right]$$

dan

$$\frac{\partial^2 L(\beta)}{\partial \beta_0 \partial \beta_j} = -\sum_{i=1}^{n}\left[\frac{x_{ji}e^{g(x_i)}\left(1+e^{g(x_i)}\right)-x_{ji}\left(e^{g(x_i)}\right)^2}{\left(1+e^{g(x_i)}\right)^2}\right]$$

$$= -\sum_{i=1}^{n} x_{ji}\,\pi(x_i)\left[1-\pi(x_i)\right]$$

JSM

Jurnal Sains dan Matematika Vol. 19 (3): 91-95 (2011)

Misal turunan parsial pertama dari L (β) terhadap β_j , $j \leq p$ adalah

$$\frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=1}^{n}\left[y_i - \frac{x_{ji}e^{g(x_i)}}{1+e^{g(x_i)}}\right]$$

$$= \sum_{i=1}^{n} x_{ji}\left[y_i - \pi(x_i)\right]$$

maka turunan parsial kedua terhadap β_u , $u \leq p$ adalah :

$$\frac{\partial^2 L(\beta)}{\partial \beta_u \partial \beta_j} = -\sum_{i=1}^{n}\left[\frac{x_{ui}x_{ji}e^{g(x_i)}\left(1+e^{g(x_i)}\right) - x_{ui}x_{ji}\left(e^{g(x_i)}\right)^2}{\left(1+e^{g(x_i)}\right)^2}\right]$$

$$= -\sum_{i=1}^{n} x_{ui}x_{ji}\pi(x_i)\left[1-\pi(x_i)\right]$$

untuk u, j = 1, 2, . . . , p

dan

$$\frac{\partial^2 \mathrm{L}(\beta)}{(\partial \beta_j)^2} = -\sum_{i=1}^{n}\left[\frac{x_{ji}x_{ji}e^{g(x_i)}\left(1+e^{g(x_i)}\right) - x_{ji}x_{ji}\left(e^{g(x_i)}\right)^2}{\left(1+e^{g(x_i)}\right)^2}\right]$$

$$= -\sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{x}_{\mathrm{ji}})^2 \pi(\mathrm{x}_\mathrm{i})\left[1-\pi(\mathrm{x}_\mathrm{i})\right]$$

Jika dinyatakan dalam bentuk matriks adalah sebagai berikut :

$$= -\begin{bmatrix} 1 & 1 & . & . & . & 1 \\ x_{11} & x_{21} & . & . & . & x_{n1} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ x_{1p} & x_{2p} & . & . & . & x_{np} \end{bmatrix}
\begin{bmatrix} \pi(x_1)[1-\pi(x_i)] & 0 & . & . & . & 0 \\ 0 & \pi(x_2)[1-\pi(x_i)] & . & . & . & 0 \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ 0 & 0 & . & . & . & \pi(x_n)[1-\pi(x_i)] \end{bmatrix}
\begin{bmatrix} 1 & x_{11} & . & . & . & x_{1p} \\ 1 & x_{21} & . & . & . & x_{2p} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ 1 & x_{n1} & . & . & . & x_{np} \end{bmatrix}$$

= **X'VX**

Pendugaan parameter β dengan metode iterasi Newton-Raphson :

1. Dipilih taksiran awal untuk β, misal $\hat{\beta} = 0$
2. Dihitung **X'(Y - π(x))** dan **X'VX**, selanjutnya dihitung invers dari **X'VX**
3. Pada setiap i+1 dihitung taksiran baru yaitu $\hat{\beta}_{i+1} = \hat{\beta}_i + \{\mathbf{X'VX}\}^{-1}$ { **X'(Y - π(x))**}
4. Iterasi berakhir jika diperoleh $\hat{\beta}_{i+1} \cong \hat{\beta}_i$

UJI PARAMETER

1. Uji rasio likelihood

Hipotesis

$H_0 : \beta_1 = \beta_2 = . . . = \beta_p = 0$

H_1 : paling sedikit satu $\beta_j \neq 0$, j = 1, 2, … , p

Statistik uji :

$$G = -2 \ln\left(\frac{l_0}{l_1}\right) = -2\ (\ln l_0 - \ln l_1) = -2\ (L_0 - L_1)$$

dengan L_0 = log likelihood tanpa variabel bebas.

L_1 = log likelihood dengan variabel bebas

Kriteria uji : Tolak H_0 jika $G > \chi^2_{\alpha,p}$

2. Uji Wald

Hipotesis

$H_0 : \beta_j = 0$, j = 1, 2, … , p

$H_1 : \beta_j \neq 0$, j = 1, 2, … , p

Statistik uji :

$$W_j = \left[\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right]^2$$

Kriteria uji : Tolak H_0 jika $W_j > \chi^2_{\alpha,1}$

HASIL DAN PEMBAHASAN

Model Awal Regresi Logistik Biner

Data diperoleh secara sekunder dari salah satu rumah sakit di Yogyakarta. Hasil pengolahan data diperoleh nilai estimasi parameter untuk regresi logistik biner sebagai berikut:

Tabel 1. Variabel dalam Persamaan

Variabel	Parameter β
X1	-0.456
X2	0.427
X3	0.469
X4	0.094
X5	-0.066
X6	0.102
X7(1)	0.246
Konstanta	4.167

Model awal regresi logistik biner yang terbentuk adalah

$$\text{Logit}\ [\pi(x)] = \log\left(\frac{\pi(x)}{1-\pi(x)}\right)$$
$$= 4.167 - 0.456X1 + 0.427X2 + 0.469X3 + 0.094X4 - 0.066X + 0.102X6 + 0.246X7(1)$$

Uji Kecocokan Model

Uji kecocokan model digunakan uji Rasio Likelihood.

1. Hipotesis :
   $H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$ (model tidak cocok)
   H_1 : salah satu dari $\beta_{jk} \neq 0$ dengan j = 1, 2, . . .,p (model cocok)
2. Taraf signifikansi $\alpha = 0.05$
3. Statistik uji :

Untuk menentukan nilai statistik G, maka ditentukan nilai -2Log likelihood, output estimasi nilai tersebut didapat melalui iterasi dengan hasil sebagai berikut :

Tabel 2. Iterasi Block 0

Iterasi		-2 Log likelihood
Step 0	1	68.994
	2	68.994

Tabel 3. Iterasi Block 1

Iterasi		-2 Log likelihood
Step 1	1	43.451
	2	39.196
	3	36.771
	4	35.714
	5	35.475
	6	35.469
	7	35.469
	8	35.469

Nilai statistik G sebagai berikut :

$$G = -2 \ln\left(\frac{likelohood\ tanpa\ variabel\ bebas}{likelihood\ dengan\ variabel\ bebas}\right)$$

= (-2 ln likelihood tanpa variabel bebas) – (-2 ln likelihood dengan variabel bebas)
= 68,994 – 35,469
= 33,525

Nilai G ini dapat dilihat dari nilaiChi square pada Tabel 4 sebagai berikut :

Tabel 4. Nilai Chi-Kuadrat

		Chi-square
Step 1	Step	33.526
	Block	33.526
	Model	33.526

4. Kriteria Uji
   H_0 ditolak jika nilai $G > \chi^2_{(0.05,7)}$ dengan $\chi^2_{(0.05,7)} = 14,07$
5. Kesimpulan
   H_0 ditolak karena $G > \chi^2_{(0.05,7)}$ yaitu 33,525 > 14,07.

Jadi disimpulkan bahwa model cocok artinya keserempakan koefisien tersebut mempunyai pengaruh yang nyata terhadap model.

Uji Wald

1. Hipotesis :
   $H_0 : \beta_j = 0$
   $H_1 : \beta_j \neq 0$ untuk
   j = 1,2,3,4,5,6,7

2. Taraf signifikansi α = 0.05
3. Statistik uji :
   Nilai estimasi parameter β, nilai Wald dan nilai signifikansi dapat dilihat pada Tabel 5.

Tabel 5. Nilai Estimasi Parameter, Nilai Wald dan Signifikansi

Variabel	β	S.E.	Wald	Df	Sig.
X1	-0.456	0.226	4.051	1	0.044
X2	0.427	0.220	3.758	1	**0.053**
X3	0.469	0.214	4.823	1	0.028
X4	0.094	0.047	3.953	1	0.051
X5	-0.066	0.026	6.508	1	0.011
X6	0.102	0.050	4.071	1	0.044
X7(1)	0.246	1.192	0.043	1	**0.836**
Konstanta	4.167	5.394	0.597	1	0.440

Berdasarkan nilai Sig yang terdapat pada Tabel 5 dengan α = 5% terdapat tiga variabel bebas yaitu X2 (Low Density Lipoprotein), X4 (Trigliserid), dan X7(1) (Jenis Kelamin) yang tidak signifikan, hal ini dapat dilihat dengan nilai Sig > 0.05. Sehingga dapat disimpulkan bahwa X1 (Kadar Kolesterol), X3 (High Density Lipoprotein), X5 (Tekanan Darah Sistolik), dan X6 (Tekanan Darah Diatolik) adalah signifikan terhadap terjangkit atau tidak penyakit jantung koroner.

Model Akhir Regresi Respon Biner

Berdasarkan uji wald maka variabel yang dipakai dalam model akhir adalah variabel yang signifikan berpengaruh terhadap status penderita jantung koroner, yaitu X1 (Kadar Kolesterol), X3 (High Density Lipoprotein), X5 (Tekanan Darah Sistolik), dan X6 (Tekanan Darah Diastolik). Nilai estimasi parameternya dapat dilihat pada Tabel 6.

Tabel 6. Nilai Estimasi Parameter, Nilai Wald dan Signifikansi untuk Model Akhir

Variabel	B	S.E.	Wald	Df	Sig.	Exp(B)
X1	-0.021	0.007	8.623	1	0.003	0.979
X3	0.055	0.026	4.331	1	0.037	1.057
X5	-0.053	0.021	6.355	1	0.012	0.949
X6	0.125	0.043	8.512	1	0.004	1.133
Konstanta	-0.458	2.406	0.036	1	0.849	0.633

Model akhirnya adalah

$$\text{Logit}\,[\pi(x)] = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = -0.458 - 0.021X1 + 0.055X3 - 0.0563X5 + 0.125X6$$

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan dapat disimpulkan :

1. Model akhir regresi logistik dalam kasus ini adalah
   $$\text{Logit}\,[\pi(x)] = \log\left(\frac{\pi(x)}{1-\pi(x)}\right)$$
   $=- 0.458 - 0.021X1 + 0.055X3 - 0.0563X5 + 0.125X6$
2. Pada kasus ini, hanya empat faktor yang signifikan berpengaruh terhadap penyakit jantung koroner yaitu kadar kolesterol, high density lipoprotein, tekanan darah sistolik, dan tekanan darah diastolik .

DAFTAR PUSTAKA

[1] Djohan, T.B.A. 2004. *Penyakit Jantung Koroner Dan Hypertensi*. Sumatera : USU. http://library.usu.ac.id/download/fk/gizi-bahri10.pdf. [diakses tanggal 10 Mei 2012].

[2] Effendy, N, Subagja, Faisal, A. 2008. *Prediksi Penyakit Jantung Koroner (PJK) Berdasarkan Faktor Risiko Menggunakan Jaringan Syaraf Tiruan Backpropagation*. Prosiding Seminar Nasional Aplikasi Teknologi Informasi 2008 (SNATI 2008) Yogyakarta, 21 Juni 2008 . ISSN: 1907-5022. http://journal.uii.ac.id/index.php/Snati/article/view/725/680 [diakses tanggal 10 Mei 2012].

[3] Hosmer, D. W., Lemeshow, S. 1989. *Applied Logistic Regression*. New York : John Wiley & Sons, Inc..

[4] RS MEDISTRA. 2012. *Apakah Penyebab Dari Penyakit Jantung Koroner ?*. Jakarta. http://www.medistra.com/index.php?option=com_content&view=article&id=76 [diakses tanggal 10 Mei 2012].