

# MODEL REGRESI LOGISTIK BINER DENGAN METODE *PENALIZED MAXIMUM LIKELIHOOD*

Edi Susilo , Anna Islamiyati , Muh. Saleh AF.

## ABSTRAK

Analisis regresi logistik biner dengan metode *penalized maximum likelihood* digunakan jika terjadi pemisahan pada data yaitu terdapat satu atau lebih sel yang kosong pada table silang. Skripsi ini bertujuan untuk mengestimasi parameter regresi logistik biner dengan metode *penalized maximum likelihood* yang memiliki fungsi  $(Y_i)^* = L(Y_i)|I(Y_i)|^{1/2}$ , diaplikasikan pada data pengaruh pemberian ASI eksklusif untuk bayi umur 0 sampai 6 bulan. Berdasarkan model regresi logistik biner dengan metode *penalized maximum likelihood* diperoleh taksiran model yang menunjukkan bahwa faktor yang berpengaruh terhadap pemberian ASI eksklusif untuk bayi umur 0 sampai 6 bulan adalah pekerjaan ibu, pendapatan keluarga, dan tanggapan ibu akan diberikannya ASI eksklusif. Seorang ibu cenderung tidak memberikan ASI eksklusif pada bayinya pada saat ibu tersebut bekerja meskipun memiliki tanggapan yang baik tentang pemberian ASI eksklusif tetapi akibat penghasilan keluarga yang belum mencukupi menyebabkan ibu tersebut bekerja dan tidak sempat memberikan ASI eksklusif kepada bayinya.

Kata kunci : regresi logistik biner, *Penalized maximum likelihood* (PML), pemisahan, data ASI eksklusif

## 1. Pendahuluan

Analisis regresi logistik biner digunakan untuk menjelaskan hubungan antara variabel respon yang berupa data dikotomik/biner dengan variabel bebas yang berupa data berskala interval dan atau kategorik (Hosmer dan Lemeshow, 1989). Variabel yang dikotomik/biner adalah variabel yang hanya mempunyai dua kategori saja, yaitu kategori yang menyatakan kejadian sukses ( $Y=1$ ) dan kategori yang menyatakan kejadian gagal ( $Y=0$ ).

Beberapa metode yang sudah banyak digunakan dalam menduga parameter pada regresi logistik yaitu metode kemungkinan maximum (*Maximum Likelihood Method*), Metode kuadrat terkecil tertimbang noniterasi (*Noniterative Weight Least Square Method*), dan Analisis fungsi diskriminan (*Discriminant Fuction Analysis*) (Nachrowi dan Usman, 2002). Metode lain yang digunakan untuk menduga parameter model regresi logistik biner metode *penalized maximum likelihood* (Firth, 1993). Metode *penalized maximum likelihood* digunakan apabila hasil estimasi parameter dari *maximum likelihood* kurang bagus.

Adapun tujuan dari penulisan ini adalah untuk mengestimasi parameter pada regresi logistik biner dengan metode *penalized maximum likelihood* pada data motivasi pemberian ASI eksklusif pada bayi usia 0-6 bulan.

## 2. Tinjauan Pustaka

### 2.1 Distribusi Binomial

Untuk membentuk suatu distribusi binomial diperlukan dua hal:

1. Banyaknya/jumlah percobaan/kegiatan;

2. Probabilitas suatu kejadian baik sukses maupun gagal.

Dalam sebuah percobaan Bernoulli, dimana  $p$  adalah probabilitas “sukses” dan  $q = 1 - p$  adalah probabilitas gagal, dan jika  $X$  adalah variabel acak yang menyatakan sukses, maka dapat dibentuk sebuah distribusi probabilitas Bernoulli sebagai fungsi probabilitas sebagai berikut:

$$X(e) = \begin{cases} 1, & \text{jika kejadian sukses} \\ 0, & \text{jika kejadian gagal} \end{cases}$$

Fungsi kepadatan probabilitas distribusi Bernoulli adalah

$$f(x) = p^x q^{1-x}, \quad x = 0, 1, \dots \quad (2.1)$$

dengan  $p$  adalah probabilitas sukses dan  $q$  adalah probabilitas gagal. Nilai harapan  $E(X) = p$ , dan variansi  $var(x) = pq$ .

## 2.2 Regresi Logistik

Model regresi logistik yang dipengaruhi oleh  $p$  variabel prediktor dapat dinyatakan sebagai nilai harapan dari  $Y$  dengan diberikan nilai  $x$

$$E(Y|x) = \frac{\exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)}{1 + \exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)} \quad (2.2)$$

dengan  $0 \leq E(Y|x) \leq 1$  dan  $Y$  mempunyai nilai 0 atau 1. Nilai 1 merupakan probabilitas  $E(Y|x)$  sukses, sehingga dapat dinyatakan dengan  $\pi(x)$ , sehingga persamaan di atas menjadi

$$\pi(x) = \frac{\exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)}{1 + \exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)} \quad (2.3)$$

dengan  $\beta_k$  menyatakan parameter-parameter regresi,  $x_k$  adalah pengamatan variabel prediktor ke- $k$  dari sejumlah  $p$  variabel predictor (Hosmer dan Lemeshow (1989) dalam Tiro (2000))

Transformasi logit diterapkan pada model regresi logistik,

$$\text{Logit}(\pi(x)) = g(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \sum_{k=1}^p \beta_k x_k \quad (2.4)$$

Keterangan :

- $\pi(x)$  adalah peluang kejadian.
- $g(x)$  adalah nilai estimasi logit.
- $\beta_0, \beta_1, \beta_2, \text{ dan } \beta_3$  berturut-turut adalah nilai koefisien untuk variabel-variabel konstan, usia, jenis kelamin, dan berat badan yang diperoleh menggunakan metode *Maximum Likelihood Estimation*.

Transformasi logit bertujuan untuk membuat fungsi linear dari parameter-parameternya. Fungsi  $g(x)$  linear terhadap parameter dan memiliki range  $(-\infty, \infty)$ , tergantung dari range variabel prediktor  $X$ .

### 2.3 Maximum Likelihood

Misal data observasi bersifat bebas maka likelihood dari data  $Y_1, Y_2, \dots, Y_n$  adalah  $p_i$  dan  $1 - p_i$ . Jika untuk setiap  $Y_i = 1$ , dengan probabilitas  $p_i$  dan untuk setiap  $Y_i = 0$  dengan probabilitas  $1 - p_i$ , bentuk umum dari likelihood (L):

$$L = \prod_{i=1}^n p_i^{Y_i} (1 - p_i)^{1 - Y_i} \quad (2.6)$$

Nilai parameter  $\beta$  dapat diperoleh dengan memaksimumkan fungsi *likelihood*-nya. Hal tersebut dilakukan dengan metode turunan pertama fungsi *likelihood*-nya terhadap setiap parameternya yang disamakan dengan nol. Terkadang sulit menemukan turunan dari fungsi *likelihood*-nya sehingga yang dilakukan adalah menemukan nilai maksimum dari logaritma natural fungsi *likelihood* tersebut atau fungsi *log-likelihood*.

Fungsi *log-likelihood* adalah bentuk logaritma dari fungsi *likelihood*, yang dituliskan dalam bentuk:

$$\ln L(p_i | y_1, y_2, \dots, y_n) = \sum_{i=1}^n \ln(p_i^{y_i} (1 - p_i)^{1 - y_i}) \quad (2.7)$$

### 2.4 Penalized Maximum Likelihood

*Penalized maximum likelihood* yang diberikan oleh Firth (1993) dapat diberikan sebagai berikut:

$$L(\beta)^* = L(\beta) |I(\beta)|^{1/2} \quad (2.8)$$

Fungsi tereduksi  $|I(\beta)|^{1/2}$  dikenal juga sebagai invarian Jeffrey.

Estimasi *Penalized Maximum Likelihood* akan diperoleh dengan cara membagi atau memilih banyaknya observasi  $i$  ke dalam dua observasi baru yang memiliki nilai respon  $y_i$  dan  $1 - y_i$  dengan pembobot masing-masing  $1 + \frac{h_i}{2}$  dan  $\frac{h_i}{2}$ .

$$\begin{aligned} U(\beta_r)^* &= \sum_{i=1}^n \left\{ (y_i - \pi_i) \left( 1 + \frac{h_i}{2} \right) + (1 - y_i - \pi_i) \frac{h_i}{2} \right\} x_{ir} \\ &= \sum_{i=1}^n \left\{ y_i - \pi_i + h_i \left( \frac{1}{2} - \pi_i \right) \right\} x_{ir} = 0 \quad (r = 1, \dots, k) \end{aligned} \quad (2.9)$$

dimana  $h_i$  merupakan elemen diagonal ke- $i$  dari matriks topi  $H = W^{1/2} X (X^T W X)^{-1} X^T W^{1/2}$ , dengan  $W = \text{diag}\{\pi_i(1 - \pi_i)\}$ . Kini metode *Penalized Maximum Likelihood* (PMLE) estimasi  $\hat{\beta}$  dapat diperoleh dari proses iterasi seperti biasa hingga konvergensinya diperoleh sebagai berikut

$$\beta^{(s+1)} = \beta^{(s)} + I^{-1}(\beta^{(s)}) U(\beta^{(s)})^* \quad (2.10)$$

dimana  $s$  merupakan banyaknya iterasi.

### 2.5 Uji Signifikansi Parameter

Statistik uji G yaitu uji rasio kemungkinan maksimum (*maximum likelihood ratio test*) yang digunakan untuk menguji peranan variabel prediktor di dalam model secara bersama-sama dengan rumusnya sebagai berikut:

$$G = -2 \ln \left[ \frac{L_o}{L_p} \right] \quad (2.11)$$

Dengan:

$L_o$  = *likelihood* tanpa variabel prediktor

$L_p$  = *likelihood* dengan p variabel prediktor

tolak  $H_0$  jika  $G > \chi^2_{(\alpha, v)}$

Selanjutnya dengan menggunakan uji Wald, akan dilakukan pengujian secara individu terhadap signifikansi parameter model. Menurut Hosmer dan Lemeshow (2000: 16), statistik Uji Wald didefinisikan sebagai:

$$W = \left[ \frac{\hat{\beta}_i}{se(\hat{\beta}_i)} \right]^2 \quad (2.12)$$

Dengan:

$\hat{\beta}_i$  = penaksir dari  $\beta_i$

$SE(\hat{\beta}_i)$  = penaksir galat baku dari  $\beta_i$

Tolak  $H_0$  jika  $|W| > \chi^2_{(\alpha, 1)}$

## 2.6 Uji Kecocokan Model

Statistik uji Hosmer dan Lemeshow  $\hat{C}$  yang dihitung berdasarkan nilai  $y = 1$  dirumuskan

$$\hat{C} = \sum_{r=1}^g \frac{(O_r - n'_r \bar{p}_{1r})^2}{n'_r \bar{p}_{1r} (1 - \bar{p}_{1r})} \quad (r = 1, 2, \dots, g) \quad (2.13)$$

Dengan

$\bar{p}_{1r}$  = rata-rata taksiran probabilitas sukses kelompok ke- $r$

$O_r$  = jumlah sampel kejadian sukses dalam kelompok ke- $r$

$n'_r$  = total sampel kelompok ke- $r$

$$\sum_{r=1}^g n'_r = n$$

## 3. Metode Penelitian

### 3.1 Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yaitu data pengaruh pemberian ASI eksklusif pada bayi umur 0 sampai 6 bulan Tahun 2014 wilayah Puskesmas Palangga Kecamatan Palangga Kabupaten Gowa. Variabel-variabel yang diteliti adalah pemberian ASI eksklusif (Y), pekerjaan ibu ( $X_1$ ), pendapatan keluarga ( $X_2$ ), dan tanggapan ibu akan diberikannya ASI eksklusif ( $X_3$ ).

### 3.2 Metode Analisis

Metode yang digunakan untuk menaksir parameter regresi logistik biner pada data ASI eksklusif adalah *penalized maximum likelihood*, dengan tahapan sebagai berikut:

- Mengasumsikan variabel respon Y berdistribusi bernoulli dalam regresi logistik biner.
- Memeriksa kejadian pemisahan pada regresi logistic biner.
- Menentukan estimasi parameter  $\beta$  dengan menggunakan metode *penalized maximum likelihood*.

- d. Melakukan uji signifikansi parameter  $\beta$  dengan menggunakan uji *maximum likelihood ratio test* dan dilanjutkan dengan menggunakan uji Wald.
- e. Melakukan uji kecocokan model dengan menggunakan Uji *Hosmer and Lemeshow Goodness of Fit Test*.

#### 4. Hasil dan Pembahasan

##### 4.1 Estimasi Parameter

Langkah pertama yaitu memeriksa keberadaan pemisahan dalam data kejadian dengan membuat table silang yaitu:

Tabel 1 : Tabel silang antara variabel respon (y) dan variabel predictor (x)

Variabel Respon		Y		Jumlah
		0	1	
Variabel Prediktor	1	15	17	32
		2	27	21
x <sub>2</sub>	1	20	18	38
	2	22	20	42
x <sub>3</sub>	1	23	38	61
	2	19	0	19
Jumlah		42	38	80

Kejadian pemisahan terjadi apabila terdapat sel yang bernilai nol dalam table silang dan berdasarkan Tabel 1 terlihat bahwa terdapat sel yang bernilai nol yaitu pada x<sub>3</sub> kategori 2 untuk y kategori 1. Hasil tersebut menunjukkan bahwa pada data pemberian ASI terdapat kejadian pemisahan yang dapat diselesaikan dengan regresi logistic biner melalui metode *penalized maximum likelihood*.

Selanjutnya data dianalisis dengan menggunakan regresi logistic biner dengan metode *penalized maximum likelihood* dengan nilai awal  $\beta$  yang digunakan berdasarkan dari hasil estimasi parameter dengan metode *maximum likelihood*, adalah sebagai berikut:

$$\begin{bmatrix} \hat{\beta}_{0(t)} \\ \hat{\beta}_{1(t)} \\ \hat{\beta}_{2(t)} \\ \hat{\beta}_{3(t)} \end{bmatrix} = \begin{bmatrix} -21,415 \\ 0,815 \\ -0,320 \\ 21,757 \end{bmatrix}$$

Hasil yang diperoleh dari penaksiran parameter dari regresi logistik menggunakan metode *Penalized Maximum Likelihood* dengan iterasi Newton-Raphson, yaitu:

Tabel 2: Hasil Penaksiran Parameter Untuk X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub> dengan metode *Penalized Maximum Likelihood*

Parameter	$\hat{\beta}$	SE
Intercept	-19,020	0,160
X <sub>1</sub>	-9,108	0,195
X <sub>2</sub>	11,417	0,200
X <sub>3</sub>	15,460	0,150

Berdasarkan Tabel 2, diperoleh taksiran model regresi logistic biner adalah sebagai

berikut:

$$\hat{\pi}(x_i) = \frac{\exp(-19,020 - 9,108(x_1) + 11,417(x_2) + 15,460(x_3))}{1 + \exp(-19,020 - 9,108(x_1) + 11,417(x_2) + 15,460(x_3))} \quad (4.14)$$

## 4.2 Uji Signifikansi Parameter

Tabel3 : Pengujian signifikansi parameter secara simultan

G	$\chi^2_{(0,05;1)}$
3,912	3,841

Berdasarkan Tabel 3 diperoleh nilai  $G = 3,912 > X^2 = 3,841$ , berarti  $H_0$  ditolak dan  $H_1$  diterima. Artinya terdapat variabel prediktor yang berpengaruh signifikan terhadap variabel respon.

Tabel 4 : Pengujian signifikansi parameter secara parsial

Parameter	W	$\chi^2_{(0,05;1)}$
Intercept	14125,22	3,841
$X_1$	2170,914	
$X_2$	3266,439	
$X_3$	10553,8	

Dengan melihat Tabel 4 diperoleh  $H_0$  ditolak untuk semua koefisien regresi karena nilai statistik uji  $W > X^2_{(\alpha,1)}$  artinya semua variabel prediktor yaitu pekerjaan (x1), pendapatan keluarga (x2), dan sikap ibu (x3) berpengaruh terhadap pemberian ASI kepada balita. Taksiran model regresi logistik pada data pengaruh pemberian ASI kepada balita, dituliskan sebagai berikut:

$$\hat{\pi}(x_i) = \frac{\exp(-19,020 - 9,108(x_1) + 11,417(x_2) + 15,460(x_3))}{1 + \exp(-19,020 - 9,108(x_1) + 11,417(x_2) + 15,460(x_3))}$$

Tabel 5 : Hasil uji Hosmer and Lemeshow

$\hat{C}$	$\chi^2_{(0,05;8)}$
12,423	15,507

Berdasarkan Tabel 5, diperoleh  $\hat{C} = 12,423 < \chi^2_{(0,05;8)} = 15,507$  berarti  $H_0$  diterima artinya model regresi logistik biner layak digunakan pada data pengaruh pemberian ASI pada bayi usia 0-6 bulan

## 5. Kesimpulan dan Saran

### 5.1 Kesimpulan

Berdasarkan uraian pembahasan mengenai penaksiran parameter model regresi logistik biner dengan metode *penalized maximum likelihood*, maka dapat diambil kesimpulan sebagai berikut :

1. Dari contoh penerapan regresi logistik biner dengan metode *penalized maximum likelihood* diperoleh model regresi sebagai berikut:

$$\hat{\pi}(x_i) = \frac{\exp(-19,020 - 9,108(x_1) + 11,417(x_2) + 15,460(x_3))}{1 + \exp(-19,020 - 9,108(x_1) + 11,417(x_2) + 15,460(x_3))}$$

2. Berdasarkan penerapan regresi logistik biner dengan metode *penalized maximum likelihood* pada data motivasi pemberian ASI eksklusif pada bayi usia 0-6 bulan dipengaruhi oleh beberapa faktor yaitu pekerjaan seorang ibu, penghasilan keluarga (suami dan istri), dan sikap ibu terhadap pemberian ASI eksklusif.

## 5.2 Saran

Regresi logistik biner merupakan model regresi yang respon Y berdistribusi Bernoulli dengan nilai variabel responnya terdiri dari dua kategori yaitu 0 dan 1. Penulis menyarankan untuk mengestimasi parameter dengan metode *penalized maximum likelihood* pada regresi logistic multinomial dan ordinal.

## DAFTAR PUSTAKA

- Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. John Wiley and Sons. New York
- Ecevit Eydurán; 2008; Usage of Penalized Maximum Likelihood Estimation Method in Medical Research: An Alternative to Maximum Likelihood Estimation Method
- Christopher Zorn; 2005; A Solution to Separation in Binary Response Models
- Delbra Andhini Fitrianty, Ni Wayan S. Wardhani, Loekito A. Soehono; Ketepatan Klasifikasi dengan Analisis Regresi Logistik dan *Multivariate Adaptive Regression Splines (MARS)* pada Data dengan Peubah Respon Biner; 3
- Georg Heinze and Michael Schemper; 2002; A solution to the problem of separation in logistic regression
- Ikaro Daniel de Carvalho Barreto, Suzana Leitão Russo, Gutemberg Hespanha Brasil, Vitor Hugo Simon; 2014; Separation Phenomena Logistic Regression
- Kleinbaum, D.G; Kupper, L.L; Muller, K.E, and Nizam, A (1998) *Applied Analysis and Multivariate Methods*. Third Edition. Duxbury Press.
- <http://statistikceria.blogspot.com/2013/01/konsep-regresi-logistik-biner-dikotomi.html>