

## Penerapan Metode *Penalized Maximum Likelihood Estimation* pada Regresi Multinomial

Erviana\*, Yayat Karyana

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

\*ervianapdy@gmail.com, yayatkaryana@gmail.com

**Abstract.** Multinomial regression models are often used to analyze the relationship between categorical response variables that have more than two categories. Parameter estimation in multinomial logistic regression using the Maximum Likelihood Estimation (MLE) method followed by iteration using the Newton-Raphson method. However, sometimes the Maximum Likelihood Estimation (MLE) method cannot be used because it produces an estimator that does not converge, this is due to data separation. To overcome this problem, the *Penalized Maximum Likelihood Estimation* (PMLE) approach is used. The PMLE method is the result of modifying the probability score function into a penalty probability score. The data used is the 2012 IDHS data for the province of Aceh using data on the desire to have children, after analyzing the data on the desire to have children there are factors that most influence the desire to have children, namely the number of children who are still alive.

**Keywords:** *Fertilization, Multinomial Regression, Penalized Maximum Likelihood Estimation, Desire to Have Children, Newton-Raphson.*

**Abstrak.** Model regresi multinomial sering digunakan untuk menganalisis hubungan antara peubah respon yang bersifat kategorik yang memiliki lebih dari dua kategorik. Penduga parameter pada regresi logistik multinomial menggunakan metode *Maximum Likelihood Estimation (MLE)* yang dilanjutkan dengan iterasi menggunakan metode *Newton-Raphson*. Namun, terkadang jika menggunakan metode *Maximum Likelihood Estimation (MLE)* tidak dapat digunakan karena menghasilkan penduga yang tidak konvergen hal ini disebabkan karena adanya pemisahan pada data. Untuk mengatasi permasalahan ini maka digunakan pendekatan *Penalized Maximum Likelihood Estimation (PMLE)*. Metode PMLE merupakan hasil modifikasi fungsi skor *likelihood* menjadi skor *penalized likelihood*. Data yang digunakan adalah data SDKI tahun 2012 untuk wilayah Provinsi Aceh dengan menggunakan data keinginan memiliki anak, setelah dianalisis pada data keinginan memiliki anak terdapat faktor yang paling mempengaruhi keinginan memiliki anak yaitu jumlah anak hidup.

**Kata Kunci:** *Fertilisasi, Regresi Multinomial, Penalized Maximum Likelihood Estimation, Keinginan Memiliki Anak, Newton-Raphson.*

## A. Pendahuluan

Analisis regresi suatu metode yang terkenal untuk mencari hubungan antar dua peubah atau lebih. Regresi juga diterapkan dalam menggambarkan hubungan antar peubah respon dan prediktor dan melakukan prediksi. Pada umumnya regresi digunakan untuk menganalisis data yang bertipe kuantitatif. Namun, banyak ditemukan data yang berupa kualitatif pada peubah respon yang mengakibatkan regresi biasa tidak bisa digunakan untuk menganalisis data tersebut. Pada kondisi ini regresi yang cocok untuk menganalisis data kategorik yaitu menggunakan analisis regresi logistik. Dalam kehidupan sehari-hari sering ditemukan data pada peubah respon yang bersifat kualitatif yang berbentuk nominal, untuk mendapatkan model regresi maka perlu di analisis menggunakan pemodelan regresi logistik multinomial.

Pada pemodelan regresi multinomial yaitu perluasan dari pemodelan regresi biner dimana peubah respon bersifat kategorik namun hanya memiliki dua kategorik yaitu “ya” atau “tidak” pada peubah respon, sedangkan model regresi multinomial merupakan jenis regresi dimana peubah respon bersifat kategorik dan memiliki lebih dari dua kategori dalam sebuah peubah respon. Dalam kehidupan sehari-hari terdapat permasalahan yang hanya bisa di analisis menggunakan regresi logistik.

Pemodelan pada regresi multinomial adalah perluasan dari model regresi biner yang memiliki dua kategorik untuk peubah respon. Regresi logistik multinomial digunakan jika variabel respon memiliki lebih dari dua kategori yang berskala nominal (multinomial) yang diberi kode 1,2 dan 3 maka fungsi yang terbentuk dari persamaan tersebut yaitu  $(q-1)$  atau jika memiliki tiga kategori maka logit yang dihasilkan akan menjadi dua fungsi logit. Karena fungsi logit yang dihasilkan menjadi dua fungsi maka harus menentukan kategori peubah respon mana yang akan menjadi pembanding.

Maka didapatkan bentuk umum model peluang dengan 3 kategori peubah respon dengan persamaan, sebagai berikut:

$$\pi_k(x) = P(Y = k|x) = \frac{\exp(\beta_{k0} + \beta_{k1}x_1 + \dots + \beta_{kp}x_p)}{\sum_{k=1}^3 \exp(\beta_{k0} + \beta_{k1}x_1 + \dots + \beta_{kp}x_p)}$$

Penduga parameter pada regresi multinomial yaitu menggunakan metode *Maximum Likelihood Estimation* (MLE). Namun, pada beberapa kondisi metode MLE menghasilkan dugaan yang bersifat tak hingga atau pada proses iterasi *Newton-Raphson* menghasilkan iterasi yang tidak konvergen mengakibatkan metode MLE tidak dapat digunakan. Hal ini disebabkan oleh adanya kombinasi linier dari beberapa peubah prediktor terpisah secara sempurna atau terpisah secara kurang sempurna sehingga menghasilkan penduga tidak konvergen (Albert dan Anderson, 1984).

Pada saat kondisi metode MLE tidak bisa digunakan maka penaksiran parameter dapat diselesaikan oleh *Penalized Maximum Likelihood Estimation* (PMLE) yang diusulkan oleh Firth (1993). Dasar metode PMLE yaitu menghilangkan bias dengan cara memodifikasi fungsi skor *likelihood* pada metode MLE menjadi fungsi skor *penalized likelihood*. Oleh karena itu, untuk mengatasi pendugaan yang tidak konvergen maka dalam penyelesaiannya menggunakan metode *Penalized Maximum Likelihood Estimation* (PMLE). Metode PMLE akan diterapkan pada data keinginan memiliki anak yang bersumber dari Survei Demografi dan Kesehatan (SDKI) pada tahun 2012.

Prinsip dasar dari metode PLME yaitu memodifikasi fungsi nilai *likelihood* menjadi fungsi *penalized likelihood*. Firth (1993) mengusulkan bahwa bias pada orde pertama dapat dihilangkan dengan memodifikasi persamaan  $u$  sebagai berikut:

$$b(\beta) = (X'WX)^{-1} X'W\xi$$

dari hasil modifikasi tersebut didapatkan fungsi *penalized likelihood* sebagai berikut:

$$L^*(\beta) = L(\beta) |l(\beta)|^{1/2}$$

Maka didapatkan persamaan *penalized log-likelihood* sebagai berikut:

$$l^*(\beta) = l(\beta) + \frac{1}{2} \ln |l(\beta)|$$

## B. Metodologi Penelitian

Pada penelitian ini menggunakan data yang bersifat sekunder. Data sekunder yaitu data yang diperoleh dari sumber yang sudah ada tanpa perlu melakukan wawancara, survei, observasi dan teknik pengumpulan data lainnya. Data sekunder pada penelitian ini yaitu data yang bersumber dari Survei Demografi dan Kesehatan (SDKI) untuk Provinsi Aceh. SDKI adalah salah satu survei sosial kependudukan yang dilakukan secara berkala dan diselenggarakan oleh Badan Pusat Statistik (BPS) bekerja sama dengan Badan Kependudukan dan Keluarga Berencana (BKKBN) dan Kementerian Kesehatan (KemenKes).

Data yang digunakan pada penelitian ini yaitu data keinginan memiliki anak (*Desire for more children*) dengan peubah respon memiliki tiga kategori sebagai berikut:

$Y_1$  = ingin anak lagi (kategori 1)

$Y_2$  = belum menentukan (kategori 2)

$Y_3$  = tidak ingin anak lagi (kategori 3)

Adapun peubah prediktor sebagai berikut:

$X_1$  = Anak Lahir Hidup

$X_2$  = Lamanya Pendidikan

$X_3$  = Usia Kawin Pertama

$Y_4$  = Umur

Adapun langkah-langkah penelitian sebagai berikut:

1. Melakukan analisis deskripsi untuk data keinginan memiliki anak.
2. Melakukan pemodelan data keinginan memiliki anak dengan model regresi multinomial.
3. Melakukan pengujian indikasi keberadaan pemisahan pada data dengan menggunakan nilai peluang ketepatan alokasi.
4. Melakukan pengujian parameter menggunakan metode PMLE.
5. Melakukan uji parameter menggunakan uji secara simultan dan melakukan uji secara parsial untuk melihat peubah prediktor mana yang pengaruhnya besar terhadap peubah respon.
6. Melakukan pengujian kelayakan model.
7. Melakukan interpretasi model menggunakan nilai *Odds Ratio* (OR).

## C. Hasil Penelitian dan Pembahasan

### Pemeriksaan Keberadaan Pemisahan

Tabel 1. Penduga Parameter

Penduga Parameter						
Iterasi	Intersep 1	Intersep 2	X1	X2	X3	X4
1	-4.29046	-0.55279	0.00000	0.00000	0.00000	0.00000
2	-2.03185	-3.53464	0.40560	-0.00449	-0.05070	0.07967
3	-6.07007	-5.60661	0.79281	-0.05140	-0.03221	0.10018
4	-6.68675	-6.56429	0.73253	0.01112	-0.08656	0.14909
5	-6.91496	-6.64226	0.83021	-0.00620	-0.07767	0.13867
6	-6.45639	-6.67428	0.86868	-0.00978	-0.07434	0.13499
7	-5.47412	-6.62976	0.87509	-0.01176	-0.07275	0.13276
<b>8</b>	<b>-4.90710</b>	<b>-6.58576</b>	<b>0.87524</b>	<b>-0.01304</b>	<b>-0.07175</b>	<b>0.13130</b>
9	-5.87485	-6.55074	0.87625	-0.01482	-0.07037	0.12988
10	-8.86569	-6.53017	0.87984	-0.01672	-0.06924	0.12877
11	-8.86568	-6.54044	0.88302	-0.01675	-0.06938	0.12884
12	-8.86568	-6.54324	0.88355	-0.01670	-0.06943	0.12887

13	-8.86568	-6.54336	0.88359	-0.01670	-0.06943	0.12887
14	-8.86568	-6.54336	0.88359	-0.01670	-0.06942	0.12887
15	-8.86568	-6.54336	0.88359	-0.01670	-0.06942	0.12887

Penduga parameter yang terdapat pada tabel diatas menunjukkan indikasi adanya pemisahan, pemeriksaan pemisahan dapat dilakukan dengan melihat banyaknya iterasi yang dihasilkan. Dengan kriteria kekonvergenan =  $|\hat{\beta}^{(t+1)} - \hat{\beta}^{(t)}| < 1 \times 10^{-6}$ . Pada tabel diatas disajikan hasil penduga parameter model dengan MLE menghasilkan pada iterasi ke delapan (t=8) kekonvergenan tidak dipenuhi, sehingga terdapat indikasi adanya pemisahan pada data dan diperlukan pemeriksaan lanjutan dengan mencari nilai peluang alokasi pada setiap penduga parameter yang memiliki iterasi lebih dari delapan (t>8).

Peluang ketepatan alokasi digunakan untuk mengidentifikasi masalah pemisahan baik pemisahan sempurna dan pemisahan tidak sempurna. Pada tabel 4.2 disajikan peluang ketepatan alokasi untuk iterasi yang lebih dari 8 (t>). Untuk menghitung peluang ketepatan alokasi yang digunakan sebagai berikut:

$$P(\hat{Y}_i = y_i) = \frac{1}{1 + \exp[(2y_i - 1)\hat{\beta}x_i]}, y_i = 1,2,3$$

**Tabel 2.** Peluang Ketepatan Alokasi untuk iterasi t>8

iterasi	Peluang Alokasi			
	X1	X2	X3	X4
9	29.33	70.74	<b>97.97</b>	46.79
10	29.24	70.74	<b>97.97</b>	46.79
11	29.24	70.74	<b>97.97</b>	46.79
12	29.24	70.74	<b>97.97</b>	46.79
13	29.24	70.74	<b>97.97</b>	46.79
14	29.24	70.74	<b>97.97</b>	46.79
15	29.24	70.74	<b>97.97</b>	46.79

Setelah melakukan pengecekan indikasi pemisahan pada data untuk iterasi diatas 8 (t > 8), didapatkan pada tabel 4.2 hasil peluang alokasi dari iterasi ke Sembilan (t = 9) sampai iterasi ke lima belas (t =15) menunjukkan hasil alokasi sebesar 0.9797 yang mengindikasikan bahwa terjadinya pemisahan kurang sempurna pada data, karena nilai peluang ketepatan alokasi lebih dari 0.95 dan tidak sama dengan 1 untuk setiap pengamatan di Provinsi Aceh.

#### **Pengujian Parameter Secara Simultan dan Parsial**

Pengujian secara parameter digunakan untuk mengecek adanya pengaruh peubah prediktor terhadap peubah respon. Pada tabel dibawah disajikan hasil pendugaan dan pengujian secara parameter. Berikut hasil dari penduga parameter yang ditampilkan pada tabel dibawah.

**Tabel 3.** Hasil Pendugaan dan Pengujian Parameter

k	Peubah(X <sub>j</sub> )	( $\hat{\beta}_{kj}$ )	SE( $\hat{\beta}_{kj}$ )	W <sub>kj</sub>	nilai-p
2	X0 (intersep 1)	-6.101	6.819	0.147	0.371
	X1 (ALH)	1.113	0.368	2.716	0.0025*
	X2 (Lama Pendidikan )	-0.107	0.215	4.655	0.618
	X3 (UKP)	0.063	0.09	11.108	0.482
	X4 (Umur)	-0.059	0.177	5.659	0.739
3	X0 (intersep 2)	-5.951	2.379	0.42	0.012
	X1 (ALH)	0.776	0.207	4.825	0.0002*
	X2 (Lama Pendidikan )	-0.012	0.071	14.033	0.865

	X3 (UKP)	-0.067	0.037	27.095	0.072
	X4 (umur)	0.121	0.065	15.308	0.065

(\*) nilai signifikan pada taraf 0.05

Pada hasil tabel 4.3 diketahui bahwa nilai G yaitu sebesar 76.586 dan nilai  $\chi^2_{4(0.05)}$  sebesar 9.4877 maka dapat disimpulkan dengan taraf nyata sebesar 0.05 bahwa tolak  $H_0$  karena  $76.586 > 9.4877$ . Maka pengujian parameter secara simultan untuk data keinginan memiliki anak untuk Provinsi Aceh dapat disimpulkan bahwa ada salah satu peubah prediktor yang berpengaruh terhadap peubah respon. Sehingga dapat dilakukan uji secara parsial dengan nilai p-value pada statistika uji wald menunjukkan peubah prediktor yang berpengaruh terhadap peubah respon pada data keinginan memiliki anak yaitu anak lahir hidup dengan nilai p-value untuk logit-1 sebesar 0.0025 dan untuk logit-2 sebesar 0.0002.

**Hasil Pendugaan Parameter Terbaik**

**Tabel 4.** Hasil Pendugaan Parameter Terbaik

Peubah	$\hat{\beta}_{kj}$	
	$\hat{\beta}_{2j}$	$\hat{\beta}_{3j}$
Intersep	-9.2739	-4.0999
ALH	1.3918	1.0541

Pada tabel diatas disajikan hasil pendugaan parameter model terbaik, maka didapatkan dua model regresi logistik multinomial sebagai berikut:

1. Fungsi logit ke-1  
 $\hat{g}_1(x) = -9.2739 + 1.3918x_1$
2. Fungsi logit ke-2  
 $\hat{g}_2(x) = -4.0999 + 1.0541x_1$

Adapun peluang model regresi logistik multinomial terbaik untuk setiap kategori sebagai berikut:

1. Ingin menambah anak

$$\hat{\pi}_1(x) = \frac{1}{1 + \exp(-9.2739 + 1.3918x_1) + \exp(-4.0999 + 1.0541x_1)}$$

2. Belum menentukan

$$\hat{\pi}_2(x) = \frac{\exp(-9.2739 + 1.3918x_1)}{1 + \exp(-9.2739 + 1.3918x_1) + \exp(-4.0999 + 1.0541x_1)}$$

3. Tidak ingin menambah anak

$$\hat{\pi}_3(x) = \frac{\exp(-4.0999 + 1.0541x_1)}{1 + \exp(-9.2739 + 1.3918x_1) + \exp(-4.0999 + 1.0541x_1)}$$

**Pengujian Kelayakan Model**

**Tabel 5.** Hasil Kelayakan Model

Statistika uji $X^2_{pearson}$	p-value
27.288	0.018

Berdasarkan hasil pada pengujian kelayakan model pada tabel 4.5 menunjukkan nilai statistik uji *Chi-Kuadrat Pearson* sebesar 27.288 dan nilai p-value sebesar 0.018. Maka  $H_0$  ditolak sebab  $0.018 < 0.05$ . Artinya, dalam taraf nyata 0.05 dapat disimpulkan bahwa model belum layak bagi data keinginan memiliki anak untuk wilayah Provinsi Aceh.

**Interpretasi Model**

**Tabel 6.** Hasil Odds Ratio

Peubah( $X_{kj}$ )	$\beta_{kj}$	
	K=2	K=3

	$\beta_{2j}$	OR	$\beta_{3j}$	OR
Intersep	-9.2739	9.384E-05	-4.0999	0.0167
X1 (Anak Lahir Hidup)	1.3918	4.022	1.0541	2.869

Hasil dari odds ratio yang disajikan pada tabel 4.6 maka dapat disimpulkan sebagai berikut:

1. Logit-1

Nilai *odds ratio* peubah prediktor anak lahir hidup sebesar 4.022, artinya setiap wanita usia subur (WUS) yang menyatakan ingin anak dengan bertambahnya anak lahir hidup 1 orang maka terdapat 4.022 kali wanita usia subur yang (WUS) menyatakan belum menentukan ingin anak lagi.

2. Logit-2

Nilai *odds ratio* peubah prediktor anak lahir hidup sebesar 2.869, artinya setiap wanita usia subur (WUS) yang menyatakan ingin anak dengan bertambahnya anak lahir hidup 1 orang maka terdapat 2.869 kali wanita usia subur yang (WUS) menyatakan belum tidak ingin anak lagi.

#### D. Kesimpulan

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

Berdasarkan hasil tahapan analisis pada data keinginan memiliki anak untuk Provinsi Aceh, maka dapat disimpulkan bahwa data yang bersumber dari SDKI 2012 untuk Provinsi Aceh terdapat pemisahan kurang sempurna sehingga menggunakan metode PLME. Faktor yang mempengaruhi keinginan yaitu untuk peubah prediktor anak lahir hidup. Dan terdapat model terbaik untuk regresi multinomial sebagai berikut:

1. Fungsi logit ke-1

$$\hat{g}_1(x) = -9.2739 + 1.3918x_1$$

2. Fungsi logit ke-2

$$\hat{g}_2(x) = -4.0999 + 1.0541x_1$$

Dan Adapun peluang model regresi logistik multinomial terbaik untuk setiap kategori yaitu:

1. Ingin menambah anak

$$\hat{\pi}_1(x) = \frac{1}{1 + \exp(-9.2739 + 1.3918x_1) + \exp(-4.0999 + 1.0541x_1)} = 0.954257$$

2. Belum menentukan

$$\hat{\pi}_2(x) = \frac{\exp(-9.2739 + 1.3918x_1)}{1 + \exp(-9.2739 + 1.3918x_1) + \exp(-4.0999 + 1.0541x_1)} = 0.00036$$

3. Tidak ingin menambah anak

$$\hat{\pi}_3(x) = \frac{\exp(-4.0999 + 1.0541x_1)}{1 + \exp(-9.2739 + 1.3918x_1) + \exp(-4.0999 + 1.0541x_1)} = 0.045383$$

#### Acknowledge

Terimakasih kepada Allah SWT atas rahmat dan hidayah-Nya, penulis dapat menyelesaikan penelitian ini dengan lancar, terimakasih kepada bapak Suliadi, Ph.D selaku dosen wali, bapak Dr. Yayat Karyana, M.Si selaku dosen pembimbing, ibu Marizsa Herlina, M.Sc dan bapak Dr. Suwanda selaku dosen pembahas penelitian ini. Terimakasih kepada seluruh dosen statistika dan staff fakultas matematika dan ilmu pengetahuan alam. Serta teman-teman prodi statistika universitas islam bandung khususnya angkatan 2018.

#### Daftar Pustaka

- [1] Agresti, A. (2006). An Introduction to Categorical Data Analysis: Second Edition. In *An Introduction to Categorical Data Analysis: Second Edition*. <https://doi.org/10.1002/0470114754>
- [2] Apriyanti, O. (n.d.). *HUBUNGAN TINGKAT PENDIDIKAN DAN NILAI ANAK*

*DENGAN FERTILITAS PASANGAN PERKAWINAN USIA MUDA 1*).  
<http://www.novapdf.com/>

- [3] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (n.d.). *Applied logistik regression*.
- [4] Jaya, B. (2021). Estimasi Parameter Regresi Logistik Multinomial Menggunakan Maksimum Likelihood. *Repositori Universitas Sumatera Utara*, 45. <https://repositori.usu.ac.id/bitstream/handle/123456789/30765/160803079.pdf?sequence=1&isAllowed=y#:~:text=Regresi logistik multinomial atau disebut,satu atau lebih variabel bebas.>
- [5] Sholihin, M., Hadi, A. F., Anggraeni, D., Estimation, M. L., Estimation, M. L., Regresi, M., Biner, L., & Penalized, S. (2014). *Implementasi Metode Penalized Maximum Likelihood Estimation Pada Model Regresi Logistik Biner 1 Pendahuluan*. November, 215–222.
- [6] Cook, S. J., Niehaus, J., & Zuhlke, S. (2018). A warning on separation in multinomial logistik models. *Research and Politics*, 5(2), 0–4. <https://doi.org/10.1177/2053168018769510>
- [7] Davis, K., & Blake, J. (1956). Social Structure and Fertility: An Analytic Framework. In *Source: Economic Development and Cultural Change* (Vol. 4, Issue 3). The University of Chicago Press.
- [8] Efendi, A., & Ramadhan, H. W. (2018). Parameter estimation of multinomial logistik regression model using least absolute shrinkage and selection operator (LASSO). *AIP Conference Proceedings, 2021*(October 2018). <https://doi.org/10.1063/1.5062766>
- [9] Fagerland, M. W., & Hosmer, D. W. (2012). A generalized Hosmer-Lemeshow goodness-of-fit test for multinomial logistik regression models. *Stata Journal*, 12(3), 447–453. <https://doi.org/10.1177/1536867x1201200307>
- [10] Greene, W. H. (2002). The Multinomial Logit Model. *Econometric Analysis*, 720–723.
- [11] Horowitz, J. L., & Nesheim, L. (2021). Using penalized likelihood to select parameters in a random coefficients multinomial logit model. *Journal of Econometrics*, 222(1), 44–55. <https://doi.org/10.1016/j.jeconom.2019.11.008>
- [12] Fauziah, Ghina. (2021). *Estimasi Pseudo Poisson Maximum Likelihood untuk Mengatasi Masalah dalam Model Log-Linear pada Kasus Kusta di Jawa Barat Tahun 2018*. *Jurnal Riset Statistika*, 1(1), 57-62.