Regresi Logistik Hierarki Untuk Pemodelan Data Hasil Penilaian Mata Kuliah Statistika Deskriptif Mahasiswa Fakultas Z Universitas Islam Bandung Tahun 2016/2017

[1]Caramina Marini, [2] Nusar Hajarisman, [3]Abdul Kudus
[1,2]Prodi Statistika, FakultasMatematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Jl. Tamansari No.1 Bandung 40116
email: [1]caraminamarini@gmail.com,[2]Nusarhajarisman@yahoo.com

Abstrak. Model regresi logistik hirarki dua level merupakan analisis multilevel yang digunakan untuk menganalisis data yang mempunyai struktur hirarki dua level dengan data respon biner (bernilai 0 atau 1). Data hirarki adalah data dengan unit-unit observasi yang bersarang pada unit yang lebih tinggi. Dalam skripsi ini, akan membahas penggunaan regresi logistik hirarki pada data hasil penilaian mata kuliah statistika deskriptif mahasiswa fakultas *Z* Universitas Islam Bandung tahun 2016/2017. Penaksiran parameter dilakukan menggunakan metode *Maksimum likelihood*. Kemudian model dapat di interpretasikan dengan menggunakan *Odds Ratio*. Perhitungan model regresi logistik hierarki menjelaskan bahwa latar belakang pendidikan S1 dosen statistika yang mengajar pada mata kuliah statistika deskriptif mempengaruhi kelulusan mahasiswa Fakultas *Z* Universitas Islam Bandung.

Kata Kunci: Regresi Logistik Hirarki Dua Level, Maksimum Likelihood, Odds Ratio.

A. Pendahuluan

Di dalam statistika, suatu teknik analisis yang digunakan untuk menganalisis hubungan antara satu variabel respon dengan satu atau lebih variabel prediktor adalah analisis regresi. Analisis Regresi ini telah banyak diterapkan dalam banyak bidang, termasuk bidang teknik, ekonomi, managemen, biologi, dan penelitian sosial. Apabila kita dihadapkan pada suatu masalah pendugaan atau peramalan nilai suatu variabel, katakanlah y, berdasarkan satu variabel lain, x. kemudian diambil suatu sampel acak berukuran n dari populasi [(x_i,y_i),untuk i=1,2,…,n]. Data yang diperoleh diplotkan untuk menghasilkan apa yang disebut diagram pencar. Apabila titik-titik dalam diagram pencar tersebut mengikuti garis lurus, maka hal ini menunjukkan bahwa kedua variabel tersebut saling berhubungan secara linear. Bila hubungan linear ada, maka kita berusaha menyatakan secara matematik dengan persamaan garis lurus yang disebut garis regresi linear. Persamaan tersebut dinyatakan oleh:

$$E(y|x) = \beta_0 + \beta_1 x \qquad (1)$$

Dimana β_0 menyatakan intersep atau perpotongan dengan sumbu tegak, dan β_1 adalah kemiringan (slope) atau gradiennya (Hajarisman ,2016)

B. Landasan Teori

Model Regresi Logistik Biasa

Regresi logistik merupakan model regresi yang digunakan bila variabel responnya bersifat kualitatif, (Hosmer dan Lemeshow, 1989). Model regresi logistik sederhana yaitu model regresi logistik untuk satu variabel prediktor X dengan variabel respon Y yang bersifat biner. Nilai variabel $Y = 1$ menyatakan adanya suatu karakteristik dan $Y = 0$ menyatakan tidak adanya suatu karakteristik. Menurut Hosmer dan Lemeshow (1989) model regresi logistik yang dipengaruhi oleh p variabel prediktor dapat dinyatakan sebagai nilai harapan dari Y bersyaratkan x

$$E(Y|x) = \frac{e^{(\beta_0 \sum_{k=1}^{p} \beta_k x_k)}}{1 + e^{(\beta_0 \sum_{k=1}^{p} \beta_k x_k)}} \qquad (2)$$

Atau bisa dinyatakan sebagai

$$\pi(x) = \frac{e^{(\beta_0 \sum_{k=1}^{p} \beta_k x_k)}}{1 + e^{(\beta_0 \sum_{k=1}^{p} \beta_k x_k)}} \qquad (3)$$

Transformasi logit diterapkan pada model regresi logistik,

$$\text{Logit}\left(\pi(x)\right) = g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \sum\nolimits_{k=1}^{p} \beta_k x_k \qquad (4)$$

Model Regresi Logistik Hirarki

Dalam model linier (umum), diasumsikan bahwa pengamatan tersebut independen satu sama lain. Namun, ada banyak situasi dimana jenis independensi itu tidak berlaku. Salah satu jenis situasi yang melanggar asas independensi ini adalah atribut tingkat kluster yaitu nilai pengamatan masuk ke dalam kelompok-kelompok dan masing-masing kluster memiliki sifatnya sendiri. Model Hierarchical (juga disebut multi-level dan, dalam beberapa kasus disebut efek campuran), yang dirancang untuk menangani ketergantungan bersama antara pengamatan. Model spesifik dalam skala logistik yang digunakan dalam skripsi ini adalah:

$$Logit\left(\pi_{ij}\right) = \beta_{0j} + \sum_{k=1}^{p} \beta_k x_{kij} \qquad (5)$$

Setelah disederhanakan maka dapat ditulis menjadi

$$Logit\left(\pi_{ij}\right) = \beta_{0j} + \sum_{k=1}^{p} \beta_k x_{kij} \text{ dimana } \beta_{0j} = \beta_0 + \sum_{l=1}^{r} \gamma_l x_{l_j} + u_{0j} \qquad (6)$$

Persamaan untuk kedua level dapat dengan mudah dikombinasikan untuk membentuk suatu persamaan,

$$logit\left(\pi_{ij}\right) = \beta_0 + u_{0j} + \sum_{k=1}^{p} \beta_k x_{kij} + \sum_{l=1}^{r} \gamma_l x_{l_j} \qquad (7)$$

Penduga Parameter Model Hirarki

Metode penduga parameter yang dapat digunakan pada pemodelan regresi dua level adalah metode *Maximum Likelihood* (*ML*). Secara umum fungsi *Likelihood* adalah sebagai berikut (dalam notasi matriks):

$$L(\beta,\theta) = \prod_{j=1}^{J} (2\pi)^{\frac{-n_j}{2}} \left| \Sigma_{b_j} \right|^{-\frac{1}{2}} exp - 0.5\Sigma\left(y_j - X_j\beta\right)' \Sigma_{b_j}{}^{-1} \left(y_j - X_j\beta\right) \qquad (8)$$

Atau dapat juga ditulis sebagai:

$$ln_{ML}(\beta,\theta) = -0.5 \ln 2\pi - 0.5\, ln\left|\Sigma_{b_j}\right| - 0.5 \sum_{j=1}^{J} \left(y_j - X_j\beta_j\right)' \Sigma_{b_j}{}^{-1} \left(y_j - X_j\beta\right) \qquad (9)$$

Pengujian Hipotesis

Hipotesis untuk Level 1 dengan $k = 1,2,\cdots,p$ adalah :

$H_0: \beta_k = 0$; variabel bebas bukan merupakan penjelas yang signifikan terhadap variabel terikat.

$H_1: \beta_k \neq 0$; variabel bebas merupakan penjelas yang signifikan terhadap variabel terikat.

Hipotesis untuk Level 2 dengan $l = 1,2,\cdots,r$ adalah:

$H_0 : \gamma_l = 0$; variabel bebas bukan merupakan penjelas yang signifikan terhadap variabel terikat.

$H_1 : \gamma_l \neq 0$; variabel bebas merupakan penjelas yang signifikan terhadap variabel terikat.

statistik uji yang digunakan adalah statistik *Wald* sebagai berikut:

$$t = \frac{\hat{\beta}_k}{\sqrt{V(\hat{\beta}_k)}} \quad \text{dan} \quad t = \frac{\hat{\gamma}_l}{\sqrt{V(\hat{\gamma}_l)}} \tag{10}$$

Interpretasi Model

Pada regresi logistik, dalam menginterpretasikan parameter digunakan O*dds ratio* (ψ), dapat ditunjukkan pada Tabel 2.1 berikut ini:

Tabel 2.1 Nilai Model Regresi Logistik untuk *X* Biner

		Variabel Bebas	
		X_1	X_2
Variabel Terikat	y_1	$\pi(1) = \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1+e^{\beta_0}}$
	y_2	$1-\pi(1) = \frac{1}{1+e^{\beta_0+\beta_1}}$	$1-\pi(0) = \frac{1}{1+e^{\beta_0}}$

dengan menggunakan model regresi logistik sesuai tabel diatas maka *Odds Ratio* menjadi:

$$\psi = \frac{\left(\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right)\left(\frac{1}{1+e^{\beta_0}}\right)}{\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)\left(\frac{1}{1+e^{\beta_0+\beta_1}}\right)} = e^{\beta_1} \tag{11}$$

Deskripsi data

Data yang diperoleh adalah data hasil penilaian mata kuliah statistika deskriptif Fakultas *Z* Universitas Islam Bandung dengan total jumlah sebanyak 277 mahasiswa. Dengan $j = 1, 2, dan$ 3 menyatakan banyaknya dosen mata kuliah statistika deskriptif dan $i = 1, 2, \dots, n_j$ menyatakan banyaknya mahasiswa yang mengikuti perkuliahan dosen ke-*j*.

Jumlah total observasi mahasiswa yang bersarang pada dosen mata kuliah statistika deskriptif adalah:

$$N = \sum_{j=1}^{3} n_j \tag{12}$$

Dengan model :

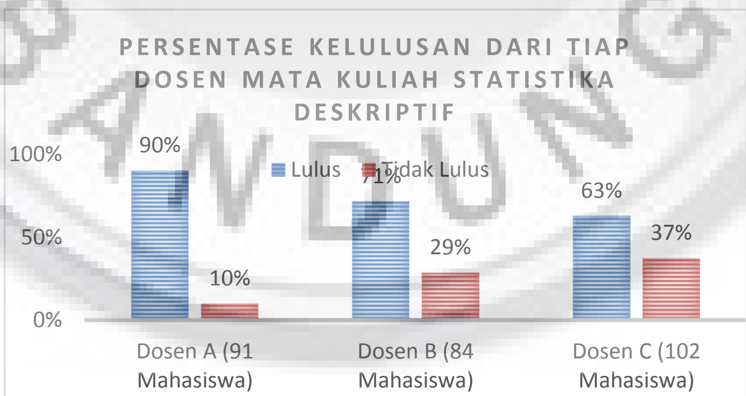
$$\log\left[\frac{\pi_{ij}}{1-\pi_{ij}}\right] = \beta_0 + u_{0j} + \beta_1 IPK_{ij} + \beta_2 \text{PEND}_j \tag{13}$$

Presentase Nilai Mutu pada mahasiswa Fakultas *Z* dapat di lihat pada Tabel 4.1.

Tabel 4.1 Presentase Nilai Mutu Mahasiswa Fakultas *Z*

Nilai Mutu	Frekuensi	Presentase
A	96	35%
B	60	22%
C	50	18%
D	40	14%
E	17	6%
T	6	2%
-	8	3%
Total	277	100%

Dari Tabel 4.1 terlihat bahwa dari total 277 mahasiswa fakultas *Z* yang mengikuti mata kuliah statistika deskriptif terdapat 35% mahasiswa yang mendapatkan nilai A, 22% atau setara dengan 60 mahasiswa mendapatkan nilai B, 18% mahasiswa mendapatkan nilai C, 14% mahasiswa mendapatkan nilai D, 6% mahasiswa mendapatkan nilai E, 6 mahasiswa mendapatkan nilai T dan 8 mahasiswa atau setara dengan 3% dari total mahasiswa yang tidak mengikuti pelajaran dan tidak mendapatkan Nilai Mutu. Dari 277 mahasiswa, sebanyak 91 merupakan mahasiswa Dosen 1, sebanyak 84 mahasiswa merupakan mahasiswa Dosen 2, dan sebanyak 102 mahasiswa merupakan mahasiswa Dosen 3 yang memiliki presentase tingkat kelulusan yang berbeda pada setiap dosen. Presentase tingkat kelulusan yang berbeda tersebut disajikan dalam Tabel 4.1 dan diagram batang pada Gambar 4.1.



Gambar 4.1 Persentase Kelulusan Mahasiswa Berdasarkan Dosen Mata Kuliah Statistika Deskriptif

Gambar 4.1 menunjukkan bahwa presentase tingkat kelulusan mahasiswa Dosen A adalah 90% dari jumlah mahasiswa Dosen A atau sebanyak 82 mahasiswa. Presentase kelulusan mahasiswa Dosen B adalah 71% dari jumlah mahasiswa Dosen B atau sebanyak 60 mahasiswa. Dan presentase kelulusan mahasiswa Dosen C adalah 63% dari jumlah mahasiswa Dosen C atau sebanyak 64 mahasiswa. Banyaknya siswa yang lulus mata kuliah statistika dengan dosen berlatar belakang statistika dan bukan statistika dapat dilihat pada Tabel 4.2.

Tabel 4.2 Perbandingan Kelulusan Mata Kuliah Statistika Deskriptif Dengan Latar Belakang Pendidikan S1 Dosen.

		Latar belakang pedidikan S1 dosen	
		statistika	*Non*-statistika
Y	Lulus	142	64
	Tidak Lulus	33	38
	Total	175	102

Dari Tabel 4.2 bisa disimpulkan bahwa lebih banyak mahasiswa dengan dosen berlatar belakang pendidikan S1 statistika yang lulus mata kuliah statistika deskriptif dibandingkan dengan dosen berlatar belakang pendidikan S1 bukan statistika. Dengan menggunakan rumus dari *Odds Ratio* data diatas bisa diinterpretasikan sebagai berikut:

Kecenderungan (*odd*) lulus dari mahasiswa dengan dosen yang berlatar belakang S1 statistika adalah $\left(\frac{142}{175} \div \frac{33}{175}\right)$ = 4.3 kali dari mahasiswa yang tidak lulus dengan dosen berlatar belakang S1 statistika. Sedangkan kecenderungan lulus dari mahasiswa dengan dosen yang bukan berlatar belakang S1 statistika adalah $\left(\frac{64}{102} \div \frac{38}{102}\right)$ = 1.6 kali dari mahasiswa yang tidak lulus dengan dosen berlatar belakang S1 statistika. Dari uraian diatas maka dapat disimpulkan *Odds Ratio* antara dosen berlatar belakang S1 statistika dengan dosen yang bukan berlatar belakang statistika adalah 2.5, artinya tingkat kelulusan dosen berlatar belakang statistika adalah 2.5 kali tingkat kelulusan dosen yang berlatar belakang bukan statistika.

Hasil Perhitungan Estimasi Parameter

Penaksiran parameter model regresi logistik hirarki diperoleh dari hasil perhitungan PROC NLMIXED dengan menggunakan program SAS 9.4. Hasil pemograman akan disajikan pada Tabel 4.3

Tabel 4.3 Hasil Perhitungan Estimasi Model Regresi Logistik Hirarki

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	95% Confidence Limits	
b0	-10.0019	1.5516	2	-6.45	0.0232	-16.6778	-3.3260
b_IPK	3.9698	0.5552	2	7.15	0.0190	1.5809	6.3587

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	95% Confidence Limits	
b_Latar	2.3809	0.7063	2	3.37	0.0779	-0.6582	5.4200
s2u	0.1732	0.2954	2	0.59	0.6171	-1.0980	1.4443

Dengan Hipotesis sebagai berikut:

H_0 : $\beta_k = 0$, dengan k = 1 dan 2; Parameter k yaitu IPK dan Latar belakang pendidikan S1 dosen tidak berpengaruh secara signifikan terhadap Kecenderungan Lulus Mata Kuliah Statistika Deskriptif(Y)

H_1 : $\beta_k \neq 0$, dengan k = 1 dan 2 ; Parameter k yaitu IPK dan Latar belakang pendidikan S1 dosen berpengaruh secara signifikan terhadap Kecenderungan Lulus Mata Kuliah Statistika Deskriptif(Y).

Dengan menggunakan $\alpha = 10\%$ dilihat dari nilai Pr > |t| maka Tolak H_0 artinya parameter yang diuji, β_1 dan β_2, signifikan pada tingkat signifikansi α atau berpengaruh secara signifikan terhadap Kecenderungan Lulus Mata Kuliah Statistika Deskriptif(Y).

Dengan model yang diperoleh dari Tabel 4.3 perbedaan peluang dapat ditunjukkan dengan Tabel 4.4.

Tabel 4.4 Peluang Kelulusan Mahasiswa Fakultas *Z* pada Kuliah Statistika Deskriptif pada Berbagai Nilai IPK untuk Dosen Berlatar Belakang Statistika dan Non-Statistika.

IPK	Peluang Kelulusan	
	Dosen Statistika	Dosen Non-Statistika
1.0	0.002	0.02
1.5	0.02	0.16
2.0	0.11	0.58
2.5	0.48	0.90
3.0	0.87	0.98
3.5	0.98	0.99
4.0	0.99	1

Pada Tabel 4.4 terlihat perbedaan peluang antara IPK dengan Dosen Berlatar Belakang Statistika dan Non-Statistika. Mahasiswa dengan IPK 1.0 dengan latar belakang dosen statistika memiliki peluang 10x lipat lebih besar dibandingkan dengan mahasiswa dengan IPK 1.0 dengan latar belakang dosen non-statistika. Perbedaan tersebut membuktikan adanya peluang kelulusan yang lebih besar untuk mahasiswa Fakultas *Z* dengan Dosen berlatar belakang statistika dibandingkan dengan dosen berlatar belakang non-statistika.

C. Kesimpulan

1. Parameter IPK merupakan variabel penjelas yang signifikan terhadap perolehan peluang kelulusan mahasiswa Fakultas Z Universitas Islam Bandung, dengan kesimpulan bahwa kecenderungan (odd) lulus dari mahasiswa yang berIPK satu satuan lebih tinggi adalah 53 kalinya kecenderungan (odd) lulus dari mahasiswa yang berIPK satu satuan lebih rendah.
2. Parameter Pendidikan dosen berlatar belakang S1 statistika merupakan variabel penjelas yang signifikan terhadap perolehan peluang kelulusan mahasiswa Fakultas Z Universitas Islam Bandung dengan kesimpulan bahwa kecenderungan (odd) kelulusan mahasiswa dengan dosen yang mempunyai latar belakang pendidikan S1 statistika adalah 11 kali kecenderungan kelulusan mahasiswa dengan dosen yang tidak mempunyai latar belakang pendidikan S1 statistika.
3. Dengan model yang diperoleh dari Tabel 4.3 perbedaan peluang dapat ditunjukkan dengan Tabel 4.4. Pada Tabel 4.4 terlihat perbedaan peluang antara IPK dengan Dosen Berlatar Belakang Statistika dan Non-Statistika. Mahasiswa dengan IPK 1.0 dengan latar belakang dosen statistika memiliki peluang 10x lipat lebih besar dibandingkan dengan mahasiswa dengan IPK 1.0 dengan latar belakang dosen non-statistika. Perbedaan tersebut membuktikan adanya peluang kelulusan yang lebih besar untuk mahasiswa Fakultas Z dengan Dosen berlatar belakang statistika dibandingkan dengan dosen berlatar belakang non-statistika.

Daftar Pustaka

Cox, D. R., and Snell, E. J. (1989). *Analysis of Binary Data. London*: Chapman and Hall.

Dai Jian, Zhongmin Li dan David Rocke, *Hierarchical Logistic Regression Modeling with SAS GLIMMIX* (Online), (http://www.lexjansen.com/wuss/2006/Analytics/ANL-Dai.pdf, Diakses 1 Mei 2017).

Hajarisman, N. (2016), *Analisis Regresi Lanjut.* Bandung: Pusat Penerbitan Unisba.

Hosmer, D.W., and Lemeshow, S. (1989). *Applied Logistic Regression*. John Willey, dari New York.

Levy, Roger. (2012), *Probabilistic Models in the Study of Language* (Online), (https://id.scribd.com/document/200838920/Probabilistic-Models-in-the-Study-of-Language, diakses 1 Mei 2017).

McCullagh, P., and J.A. Nelder (1989*). Generalized Linear Models*. (Second Edition). New York: Chapman and Hall.

Wilson, J.R., dan K.A. Lorenz, *Modeling Binary Correlated Responses using SAS, SPSS and R,* ICSA Book Series in Statistics 9 (Online), (www.springer.com/fr/book/9783319238043, Diakses 1 Mei 2017)