

## MODEL REGRESI LOGISTIK POLITOMI DALAM APLIKASI BIOSTATISTIKA

Johan Harlan

Pusat Studi Informatika Komputer  
Universitas Gunadarma  
Jl. Margonda Raya No. 100 Depok

### Abstrak

Model regresi logistik politomi digunakan untuk menjelaskan hubungan prediktor-hasil-akhir, dengan hasil-akhir merupakan logit prediktor dan hasil-akhir bersifat kategorik politomi, baik nominal ataupun ordinal. Dengan proses pengestimasiannya yang bahkan lebih rumit daripada model regresi logistik sederhana, penggunaan model ini praktis baru berkembang dalam satu dasawarsa terakhir, dengan tersedianya program komputer statistik yang lebih canggih serta perangkat keras komputer dengan kapasitas memori yang lebih besar dan kecepatan prosesor yang lebih tinggi.

Kata kunci: Biostatistika, Epidemiologi, model regresi logistik, hasil-akhir politomi

### Pendahuluan

Istilah 'logistik' pertama kali digunakan bagi fungsi matematika berbentuk huruf 'S' oleh R. Pearl dan L.J. Reed pada tahun 1920, untuk mendeskripsikan 'kurva sigmoid' yang diajukan Verhulst satu abad sebelumnya, yaitu kurva yang menunjukkan pertumbuhan populasi dalam perjalanan waktu. Dalam bentuk sederhana (regresi logistik sederhana), persamaannya dapat dinyatakan sebagai:

$$Y_i = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)}$$

atau bentuk transformasi logit-nya:

$$\ln \frac{Y_i}{1 - Y_i} = \beta_0 + \beta_1 X_i$$

Untuk bentuk gandanya (regresi logistik ganda), suku  $\beta_0 + \beta_1 X_i$ , baik sebagai eksponen  $e$  pada rumus (1) ataupun ruas kanan rumus (2) diganti dengan  $\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi}$ .

Model regresi logistik cepat menjadi populer dalam aplikasi Biostatistika dan Epidemiologi, karena variabel dependennya yang bersifat dikotomi dan estimasinya yang rentang nilainya berkisar antara 0 sampai dengan 1 (sesuai untuk

deskripsi probabilitas), dianggap sangat tepat untuk merepresentasikan berbagai hasil-akhir dalam aplikasi Biostatistika dan Epidemiologi, misalnya mati-hidup, sembuh-tidak sembuh, jatuh sakit-tetap sehat, dan sebagainya. Model regresi logistik ganda juga dapat digunakan untuk menghasilkan perkiraan rasio imbalan yang valid. Sejak tahun 1990-an, aplikasi model regresi logistik dalam Biostatistika dan Epidemiologi praktis telah menggantikan penggunaan analisis fungsi diskriminan, yang sempat populer selama beberapa dekade sebelumnya.

Model regresi logistik juga dapat digunakan pada hasil-akhir yang bersifat politomi, walaupun aplikasi politomi ini pada awalnya kurang dikenal, bahkan di Indonesia sampai saat ini hampir tidak ada yang menggunakannya. Dikenal dua bentuk aplikasi politomi pada model regresi logistik, yaitu untuk variabel dependen berskala nominal dan berskala ordinal. Program statistik komputer yang dapat melakukan analisis regresi logistik politomi pada saat ini antara lain SPSS 14.0 dan Stata 9.0.

### Himpunan Kombinasi Dikotomi Bagi Hasil-akhir Politomi

Berbeda dengan model regresi linear, mengingat sifat matematis fungsi regresi logistik yang hasil-akhirnya bersifat dikotomi, variabel dependen yang berskala politomi tidaklah dapat dimasukkan begitu saja ke dalam satu fungsi regresi logistik. Untuk itu dilakukan modifikasi dengan mengubah hasil-akhir menjadi suatu himpunan dikotomi, dengan menggunakan satu fungsi logit bagi tiap dikotomi. Persyaratan lain yang harus dipenuhi ialah bahwa *likelihood* bagi masing-masing dikotomi tersebut dapat digabungkan dan bernilai sama dengan *likelihood* variabel dependen berskala politomi semula (Fox, 1984).

Beberapa contoh himpunan dikotomi yang dapat digunakan untuk *hasil-akhir* berskala politomi antara lain yaitu:

1. Variabel hasil-akhir memiliki tiga kategori:
  - a. Model logit I: 

1	2
---	---
  - b. Model logit II: 

1	3
---	---

Atau:

  - a. Model logit I: 

1	2	3
---	---	---
  - b. Model logit II: 

2	3
---	---
2. Variabel hasil-akhir memiliki empat kategori:
  - a. Model logit I: 

1	2	3	4
---	---	---	---
  - b. Model logit II: 

1	2
---	---
  - c. Model logit III: 

3	4
---	---
3. Variabel hasil-akhir memiliki lima kategori:
  - a. Model logit I: 

1	2	3
	4	5
  - b. Model logit II: 

2	3	4
		5

c. Model logit III: 

3	4	5
---	---	---

d. Model logit IV: 

4	5
---	---

Tampak bahwa untuk tiap jumlah kategori hasil-akhir tertentu dapat digunakan lebih daripada satu himpunan (kombinasi) dikotomi. Kombinasi dikotomi yang dipilih sangat tergantung kepada pengetahuan biologis mengenai data yang dianalisis, selain skala data yang bersangkutan (nominal atau ordinal).

Pada contoh pertama untuk hasil-akhir dengan tiga kategori di atas, didapatkan satu kategori referensi (kategori 1), yang dapat digunakan untuk data nominal maupun ordinal, sedangkan contoh kedua untuk hasil-akhir dengan tiga kategori maupun hasil-akhir dengan lima kategori di atas umumnya digunakan untuk data ordinal, yaitu jika ada  $k$  kategori hasil-akhir, maka digunakan  $(k - 1)$  dikotomi dengan  $(k - 1)$  kategori secara berurutan masing-masing satu kali menjadi kelompok referensi. Contoh untuk hasil-akhir dengan empat kategori di atas umumnya hanya digunakan untuk data nominal, dan pasangan kategori yang diperbandingkan dipilih sesuai dengan tujuan penelitian.

Selanjutnya hanya akan dibahas satu contoh untuk hasil-akhir dengan tiga kategori, yaitu bentuk pertama yang diperlihatkan di atas.

### Hasil-akhir Dengan Tiga Kategori dan Satu Kategori Referensi

Misalkan dimiliki data variabel independen  $X_{1i}, X_{2i}, \dots, X_{ji}, \dots, X_{pi}$ ;  $j = 1, 2, \dots, p$  menyatakan jumlah variabel independen;  $i = 1, 2, \dots, n$  menyatakan jumlah subjek (ukuran sampel) dengan variabel dependen  $Y_i$  yang memiliki tiga kategori, masing-masing dengan nilai 0, 1, dan 2. Untuk menyederhanakan notasi, maka subskrip  $i$  selanjutnya akan dihilangkan dari



penulisan. Misalkan pula  $\mathbf{X}$  menyatakan vektor variabel independen dan  $\beta$  menyatakan vektor koefisien regresi, maka digunakan digunakan dua model logit: sebuah untuk  $Y = 1$  vs  $Y = 0$  dan sebuah lagi untuk  $Y = 2$  vs  $Y = 0$  (logit perbandingan  $Y = 2$  vs  $Y = 1$  dapat diperoleh sebagai selisih logit  $Y = 2$  vs  $Y = 0$  dan logit  $Y = 1$  vs  $Y = 0$ ).

Kedua model logit tersebut adalah:

$$g_1(x) = \ln \frac{P(Y = 1|X)}{P(Y = 0|X)}$$

$$= \beta_{10} + \beta_{11}X_1 + \beta_{12}X_2 + \dots + \beta_{1p}X_p$$

$$= (\mathbf{1}, \mathbf{X}')\beta_1$$

$$g_2(x) = \ln \frac{P(Y = 2|X)}{P(Y = 0|X)}$$

$$= \beta_{20} + \beta_{21}X_1 + \beta_{22}X_2 + \dots + \beta_{2p}X_p$$

$$= (\mathbf{1}, \mathbf{X}')\beta_2$$

Probabilitas bersyarat bagi ketiga kategori hasil-akhir dengan syarat nilai-nilai variabel independen sesuai dengan vektor  $\mathbf{X}$  adalah:

$$P(Y = 0|X) = \frac{1}{1 + \exp[g_1(x)] + \exp[g_2(x)]} \quad (4.a)$$

$$P(Y = 1|X) = \frac{\exp[g_1(x)]}{1 + \exp[g_1(x)] + \exp[g_2(x)]} \quad (4.b)$$

$$P(Y = 2|X) = \frac{\exp[g_2(x)]}{1 + \exp[g_1(x)] + \exp[g_2(x)]} \quad (4.c)$$

Pada model regresi logistik dikotomi sederhana untuk data Epidemiologi dengan faktor pajanan (*exposure; E*) yang dikotomi -  $E = 1$  (terpajan) dan  $E = 0$  (tidak terpajan) - sebagai variabel independen dan faktor penyakit (*disease; D*) yang juga dikotomi -  $D = 0$  (tidak sakit) dan  $D = 1$  (sakit) - sebagai variabel dependen, maka struktur datanya dapat disajikan dalam bentuk satu tabel  $2 \times 2$  (lihat tabel 1).

Tabel 1.  
Struktur data Epidemiologi bagi pajanan dikotomi dan penyakit dikotomi

	$E$	$\bar{E}$	
$D$	$a$	$b$	$n_1$
$\bar{D}$	$c$	$d$	$n_0$
	$m_1$	$m_0$	$n$

Maka pada model regresi logistik politomi sederhana dimiliki satu variabel independen, yaitu faktor pajanan dengan dua kategori  $E = E_1 = 1$  (terpajan) dan  $\bar{E} = E_0 = 0$  (tidak terpajan), sedangkan faktor dependennya adalah penyakit dengan tiga kategori  $D_0 = 0$  (tidak sakit),  $D_1 = 1$  (sakit ringan), dan  $D_2 = 2$  (sakit berat), struktur datanya dapat disajikan dalam bentuk dua tabel  $2 \times 2$  (satu tabel

untuk  $D_1$  vs  $D_0$  dan satu tabel untuk  $D_2$  vs  $D_0$ ), ataupun diringkaskan menjadi satu tabel  $3 \times 2$  (lihat Tabel 2). Perhatikan bahwa dalam perhitungan maupun pengolahan data dengan komputer, nilai variabel dependen yang digunakan selalu adalah 1 vs 0, baik pada model logit 1 maupun model logit 2 (nilai variabel dependen sama dengan 2 hanya digunakan untuk memperjelas pembahasan).

Tabel 2.  
Struktur data Epidemiologi bagi pajanan dikotomi dan penyakit dengan tiga kategori

	$E_0$	$E_1$	
$D_0$	$a$	$b$	$n_0$
$D_1$	$c$	$d$	$n_1$
$D_2$	$e$	$f$	$n_2$
	$m_1$	$m_0$	$n$

Di sini  $g_1(x) = \beta_{10} + \beta_{11}X$ , sehingga: Untuk  $X = 0$ :  $g_2(x) = \beta_{20}$   
 Untuk  $X = 0$ :  $g_1(x) = \beta_{10}$  Untuk  $X = 1$ :  $g_2(x) = \beta_{20} + \beta_{21}$   
 $X = 1$ :  $g_1(x) = \beta_{10} + \beta_{11}$  Karena dengan metode *maximum likelihood* diupayakan agar nilai koefisien regresi logistik adalah sedemikian hingga:  
 Demikian pula  $g_2(x) = \beta_{20} + \beta_{21}X$ , sehingga:

$$P(Y = 0|X = 0) = a/m_1; P(Y = 1|X = 0) = c/m_1; \text{ dan } P(Y = 2|X = 0) = e/m_1,$$

maka diperoleh:

$$\begin{aligned} \exp \beta_{10} &= c/a & \text{atau: } \beta_{10} &= \ln \left( \frac{c/a}{\text{Var } \beta_{11}} \right) & \text{[5.a]} \\ \exp (\beta_{10} + \beta_{11}) &= d/b & \text{atau: } \beta_{10} + \beta_{11} &= \ln \left( \frac{d/b}{\text{Var } \beta_{11}} \right) & \text{[5.a]} \\ \beta_{11} &= \ln (d/b) - \ln (c/a) & & & = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} & \text{[5.b]} \\ &= \ln (ad/bc) = \ln OR_{(10)} & & & & \end{aligned}$$

Dengan cara yang sama diperoleh:

$$\begin{aligned} \beta_{20} &= \ln (e/a) & & = \frac{1}{110} + \frac{1}{7} + \frac{1}{42} + \frac{1}{10} = 0.2768 & \text{[6.a]} \\ \beta_{21} &= \ln (af/be) = \ln OR_{(20)} & & SE \beta_{11} = 0.2758^{(6.b)} = 0.5251 & \\ & & & \beta_{21} &= \ln OR_{(20)} & \\ & & & Var \beta_{21} &= \text{Var} \left[ \ln OR_{(20)} \right] & \\ & & & & = \frac{1}{a} + \frac{1}{b} + \frac{1}{e} + \frac{1}{f} & \end{aligned}$$

Contoh 1:

Misalkan dimiliki data hipotetis yang menggambarkan distribusi pajanan dan penyakit pada tabel 3x2 berikut (tabel 3: pajanan bersifat dikotomi; penyakit dengan tiga kategori). Nilai rasio imbalanced dapat dihitung secara langsung dan disajikan pada kolom keempat.

Kesalahan standar bagi koefisien regresi logistik juga dapat dihitung dari data tabel dengan menggunakan metode Woolf:

$$\begin{aligned} &= \frac{1}{110} + \frac{1}{7} + \frac{1}{31} + \frac{1}{6} = 0.3509 \\ SE \beta_{21} &= 0.3509^{1/2} = 0.5923 \end{aligned}$$



Tabel 3.  
Contoh data hipotetis hubungan antara pajanan dikotomi dengan penyakit yang memiliki tiga kategori

	$E = 0$	$E = 1$		OR
$D = 0$	110	7	117	1.00
$D = 1$	42	10	52	3.74
$D = 2$	31	6	37	3.04
	183	23	206	

Dengan program komputer MULTLR bagi data di atas diperoleh keluaran seperti terlihat pada Tabel 4.

Tabel 4.  
Contoh keluaran komputer pada analisis regresi logistik politomi

a. Perkiraan koefisien regresi logistik

Model logit	Variabel	Perkiraan koefisien	$SE$	Skor Z	Nilai $p$
1	EXP *)	1.3195	0.5251	2.5127	0.0120
	Konstante	-0.9628			
2	EXP	1.1123	0.5923	1.8779	0.0604
	Konstante	-1.2665			

\*) Variabel pajanan (= exposure)

b. Perkiraan rasio imbangan

Model Logit	Variabel	$\hat{OR}$	Interval konfidensi 95%	
			Batas bawah	Batas atas
1	EXP	3.741	1.337	10.472
2	EXP	3.041	0.953	9.712

Tampak bahwa hasil perhitungan perkiraan rasio imbangan yang diperoleh melalui metode stratifikasi pada tabel tepat sama nilainya dengan perkiraan rasio imbangan yang diperoleh melalui analisis regresi logistik sederhana. Selain itu perkiraan kesalahan standar rasio imbangan yang diperoleh dengan metode Woolf dari tabel juga tepat sama dengan perkiraan kesalahan standar yang diperoleh melalui analisis regresi logistik sederhana. Walaupun demikian, semakin banyak jumlah variabel independen, akan semakin rumit pula perhitungan perkiraan rasio imbangan berikut perkiraan kesalahan standar-nya dengan metode stratifikasi

yang menggunakan tabel, sedangkan dengan model logit, semua perhitungan ini dapat diselesaikan secara cepat dengan menggunakan program komputer.

### Statistik Log Likelihood Pada Model Logit Politomi

Salah satu statistik terpenting yang lazim digunakan pada analisis regresi logistik adalah log *likelihood* (dan  $-2 \log$  *likelihood* yang berdistribusi khi-kuadrat), antara lain untuk digunakan dalam uji rasio *likelihood* untuk menilai signifikansi.

Perhatikan kembali struktur data pada tabel 2. *Likelihood* untuk model logit1 adalah:

$$L_1 = [P(Y=0|X=0)]^a [P(Y=0|X=1)]^b [P(Y=1|X=0)]^c [P(Y=1|X=1)]^d$$

$$= [a/m_1]^a [b/m_0]^b [c/m_1]^c [d/m_0]^d \quad (7.a)$$

sedangkan *likelihood* untuk model logit 2 adalah:

$$L_2 = [P(Y=0|X=0)]^a [P(Y=0|X=1)]^b [P(Y=2|X=0)]^e [P(Y=2|X=1)]^f$$

$$= [a/m_1]^a [b/m_0]^b [e/m_1]^e [f/m_0]^f \quad (7.b)$$

*Likelihood* untuk keseluruhan data (gabungan kedua model logit) adalah:

$$L = [P(Y=0|X=0)]^a [P(Y=0|X=1)]^b [P(Y=1|X=0)]^c$$

$$[P(Y=1|X=1)]^d [P(Y=2|X=0)]^e [P(Y=2|X=1)]^f$$

$$= [a/m_1]^a [b/m_0]^b [c/m_1]^c [d/m_0]^d [e/m_1]^e [f/m_0]^f \quad (7)$$

Contoh 2:

Perhatikan kembali data pada contoh 1. *Likelihood* untuk berbagai model yang digunakan dapat dihitung secara langsung

dengan menggunakan Persamaan 7, 7.a, dan 7.b di atas, maupun diperoleh dari keluaran program komputer.

Tabel 5.  
Contoh *likelihood* untuk beberapa model logit

Model logit	Log <i>likelihood</i>	-2 log <i>likelihood</i>
Dasar *)	-201.2995	402.5990
1	-134.4626	268.9251
2	-127.4213	254.8425
Gabungan **)	-197.5660	395.1320

\*) Model tanpa variabel dependen

\*\* ) Model dengan variabel dependen memiliki tiga kategori

Log *likelihood* untuk model dasar tanpa variabel independen (hanya memuat konstante) dapat dihitung sebagai berikut (lihat Tabel 2):

$$\ln L_0 = \ln \left[ (n_0/n)^{n_0} (n_1/n)^{n_1} (n_2/n)^{n_2} \right]$$

$$= \sum_i n_i \ln (n_i/n)$$

Untuk data pada contoh 1 (Tabel 3), *likelihood* untuk model dasar ini adalah:

$$\ln L_0 = 117 \ln(117/206) + 52 \ln(52/206) + 37 \ln(37/206)$$

$$= -201.2995$$

Perhatikan bahwa *likelihood* untuk model logit 1 dan 2 tak dapat dibandingkan signifikansinya dengan *likelihood* model

dasar, karena model dasar menggunakan seluruh data ( $Y=0$ ,  $Y=1$ , dan  $Y=2$ ), sedangkan model 1 dan 2 masing-masing hanya menggunakan sebagian data ( $Y=0$  dan  $Y=1$  untuk model 1;  $Y=0$  dan  $Y=2$  untuk model 2). Signifikansi pajanan pada model 1 dan 2 menurut hasil uji Wald dapat dilihat pada tabel 4.a ( $p = 0.012$  untuk variabel *EXP* pada model 1 dan  $p = 0.060$  pada model 2), walaupun uji Wald demikian kurang dianjurkan penggunaannya dalam analisis regresi logistik.

Signifikansi faktor pajanan (variabel *EXP*) pada model gabungan (dengan variabel dependen memiliki tiga kategori) dapat dinilai dengan uji rasio log *likelihood*. Statistik pengujinya adalah:

$$G = -2 \ln [L_0/L_1] \quad (9)$$



yang berdistribusi khi-kuadrat dengan derajat bebas satu.

Untuk data pada tabel 5, statistik pengujian ini adalah:

$$G = 402.5990 - 395.1320 = 7.467$$

### Variabel Independen Lebih Daripada Satu

Seperti halnya dengan model logit dikotomi, penggunaan model logit politomi ganda ditujukan untuk:

1. Menghasilkan prediksi hasil-akhir yang lebih tepat.
2. Menghasilkan perkiraan rasio imbalan yang lebih valid.

Untuk tujuan pertama, penerimaan suatu model logit terlebih dahulu harus melalui proses diagnostik regresi, yaitu menilai kecocokan/ ketidakcocokan variabel independen yang dimasukkan dalam model dengan proses uji statistik: uji rasio *likelihood*, yang sering sekali perlu dibantu dengan uji Wald (walaupun uji terakhir ini sebenarnya kurang dianjurkan). Pada tahap berikutnya, setelah ditentukan variabel independen mana saja yang dipertahankan dalam/dikeluarkan dari model, barulah dilakukan penilaian seberapa tepat model akhir tersebut memprediksi hasil-akhirnya, yang juga dinilai dengan proses uji statistik, yaitu uji kebaikan-suai.

Perhatikan bahwa berbeda dengan analisis regresi linear, pada analisis regresi logistik penggunaan model ganda tidaklah ditujukan untuk meningkatkan presisi, karena pada analisis regresi logistik, semakin banyak kovariabel yang dimasukkan dalam model, semakin besar variansi koefisien regresi logistik (Kleinbaum, 1982).

Penggunaan model ganda dalam bidang Epidemiologi lebih banyak dimanfaatkan untuk tujuan kedua di atas, yaitu memperoleh perkiraan rasio imbalan yang lebih valid, dengan kata lain mengurangi/memperkecil bias oleh

efek konfaunding. Untuk tujuan kedua ini, penilaian dimasukkan/dikeluarkannya suatu kovariabel dari model semata-mata dinilai secara substantif (kemaknaan biologis), dan tidak melalui proses uji statistik. Suatu kovariabel (akan dipertahankan dalam model jika pengeluarannya dari model mengakibatkan perubahan perkiraan rasio imbalan yang bermakna secara biologis, walaupun kovariabel tersebut mungkin saja tidak bermakna secara statistik. Penggunaan uji statistik pada tujuan kedua ini hanyalah untuk menilai ada tidaknya interaksi antar kovariabel (juga dilakukan dengan uji rasio *likelihood*).

Sebagai hasilnya pada penggunaan model logit untuk tujuan kedua ini akan dihasilkan model akhir yang mungkin saja kurang memenuhi persyaratan kecocokan maupun kebaikan-suai secara statistik, sehingga penilaian kedua persyaratan tersebut melalui uji statistik pun kurang dipentingkan pada keadaan tersebut.

#### Contoh 3:

Sebagai ilustrasi bagi model regresi logistik politomi ganda, diperlihatkan contoh analisis data yang dikutip dari Hosmer dan Lemeshow (1989). Jumlah subjek adalah 412 orang wanita dewasa dengan *hasil-akhir* pernah tidaknya mereka melakukan Pemeriksaan Mammografi (PM: 0 = Tidak pernah; 1 = Pernah dalam satu tahun terakhir; dan 2 = Pernah lebih daripada satu tahun sebelum saat penelitian). Variabel yang dinilai berikut kode dan penjelasan pengukurannya diperlihatkan pada tabel 6, sedangkan keluaran komputer hasil analisis data diperlihatkan pada Tabel 7.

Model logit 1 adalah model yang memperbandingkan PM = 0 vs PM = 1, sedangkan model logit 2 adalah yang memperbandingkan PM = 0 vs PM = 2. Pada tabel 7 diperlihatkan hasil analisis untuk model yang memasukkan seluruh kovariabel secara lengkap. Tampak bahwa jika tujuan regresi logistik adalah untuk menghasilkan prediksi *hasil-akhir* yang

tepat, variabel yang potensial untuk *dummy*-nya menunjukkan nilai *p* yang dikeluarkan adalah DET, yang walaupun relatif cukup besar. hanya dengan uji Wald, kedua variabel

Tabel 6.  
Contoh kode dan penjelasan pengukuran pada analisis regresi logistik politomi ganda

Variabel	Kode	Keterangan
<b>Variabel Dependen:</b>		
- Pemeriksaan Mammografi	PM	0 = Tidak pernah; 1 = Pernah tahun lalu; 2 = Pernah lebih daripada satu tahun lalu
<b>Variabel Independen:</b>		
- 'Anda tidak memerlukan pemeriksaan Mammografi jika Anda tidak menderita gejala kanker payudara'	GEJ	1 = Sangat setuju; 2 = Setuju; 3 = Tidak setuju; 4 = Sangat tidak setuju
- Persepsi mengenai manfaat Mammografi	PERS	Skala 0 (sangat bermanfaat) s/d 20 (tidak bermanfaat)
- Ibu atau saudara perempuan dengan riwayat kanker payudara	RIW	0 = Tidak; 1 = Ya
- 'Pernahkah Anda diajarkan untuk melakukan pemeriksaan sendiri payudara (PSP)?'	PSP	0 = Tidak; 1 = Ya
- 'Besarkah kemungkinan pemeriksaan Mammografi untuk mendeteksi kanker payudara?'	DET	1 = Tidak mungkin; 2 = Cukup mungkin; 3 = Sangat mungkin

(Hosmer & Lemeshow, 1989)

Tabel 7.  
Contoh keluaran komputer pada analisis regresi logistik politomi ganda

Model logit	Variabel	Perkiraan koefisien	$S\hat{E}$	Skor Z	Nilai <i>p</i>
1	GEJ(1)	0.110	0.923	0.12	0.905
	GEJ(2)	1.925	0.778	2.48	0.013
	GEJ(3)	2.457	0.775	3.17	0.002
	PERS	-0.219	0.076	-2.91	0.004
	RIW	1.366	0.438	3.12	0.002
	PSP	1.292	0.530	2.44	0.015
	DET(1)	0.017	1.162	0.01	0.988
	DET(2)	0.904	1.127	0.80	0.422
	Konstante	-2.999			
2	GEJ(1)	-0.290	0.644	-0.45	0.652
	GEJ(2)	0.817	0.540	1.51	0.130
	GEJ(3)	1.130	0.548	2.06	0.039



PERS	-0.148	0.076	-1.94	0.051
RIW	1.065	0.459	2.32	0.020
PSP	1.052	0.515	2.04	0.041
DET(1)	-0.924	0.714	-1.29	0.196
DET(2)	-0.691	0.687	-1.01	0.315
Konstante	-0.986	1.111	-0.89	

Log likelihood = -346.95

(Hosmer & Lemeshow, 1989)

Sebaliknya jika tujuan regresi logistik adalah untuk menghasilkan perkiraan rasio imbangan yang valid, setelah dilakukan analisis statistik untuk menguji kemungkinan adanya interaksi (tiap kemungkinan interaksi derajat dua/interaksi antar dua kovariabel harus diperiksa sendiri-sendiri), maka penilaian kemungkinan untuk mengeluarkan kovariabel yang tidak terlibat interaksi dari model tidak dilakukan melalui proses uji statistik.

#### Rangkuman

Dalam aplikasi Biostatistika/Epidemiologi, analisis regresi logistik merupakan alternatif yang populer bagi analisis fungsi diskriminan (dan juga merupakan alternatif utama bagi model log linear, apabila variabel independen seluruhnya merupakan variabel kategorik). Dalam kenyataannya, analisis regresi logistik sejak tahun 1990-an telah hampir seluruhnya menggeser penggunaan analisis fungsi diskriminan dan model log linear.

Analisis regresi logistik politomi relatif masih kurang dikenal, baik karena kurangnya kepustakaan yang membahasnya, maupun karena karakteristik penelitian Epidemiologi yang sangat menekankan hasil-akhir yang bersifat dikotomi. Walaupun demikian, dalam berbagai penelitian Kesehatan/Kedokteran, klasifikasi yang bersifat multikategori tidaklah sepenuhnya ditinggalkan, dan dalam keadaan demikian analisis regresi logistik politomi merupakan

metode statistika yang relevan bagi analisis probabilitas pada hasil-akhir yang multikategorik demikian.

Pada analisis regresi logistik politomi, prinsip penggunaan hasil-akhir yang bersifat dikotomi tidak ditinggalkan, dalam kenyataannya hasil-akhir yang bersifat politomi demikian diuraikan menjadi suatu himpunan kombinasi hasil-akhir yang bersifat dikotomi: Bagi data dengan hasil-akhir yang memiliki  $k$  kategori digunakan ( $k - 1$ ) dikotomi, yang berarti menggunakan ( $k - 1$ ) buah model logit. Pada hasil-akhir dengan  $k$  kategori demikian dapat dipilih berbagai himpunan ( $k - 1$ ) dikotomi, pemilihan himpunan dikotomi yang akan digunakan sangat ditentukan oleh teori biologis yang mendasari penelitian.

Seperti halnya pada analisis regresi logistik dikotomi, pada analisis regresi logistik politomi ini pun model ganda lebih dipentingkan daripada model sederhana. Pada penggunaan model ganda ini pun, harus dibedakan tujuan analisis regresi: Menghasilkan prediksi hasil-akhir yang lebih tepat atau menghasilkan perkiraan rasio imbangan yang lebih valid.

Apabila tujuan analisis regresi logistik adalah memprediksi hasil-akhir secara lebih tepat, maka semua persyaratan proses diagnostik regresi harus dipenuhi, termasuk proses untuk menilai kecocokan model regresi yang digunakan dengan uji rasio *likelihood* (dan uji Wald yang kurang dianjurkan pemakaiannya) serta analisis kebaikan-suai-nya. Sebaliknya, jika tujuan analisis regresi logistik adalah untuk mengestimasi rasio imbangan secara lebih

valid, penggunaan metode statistika terbatas pada uji interaksi antar kovariabel. Selanjutnya, penilaian (ada tidaknya) konfounding dilakukan berdasarkan pertimbangan biologis dan tidak didasarkan atas uji statistik.

Stanley, KE (eds). *Statistics in Medical Research: Methods and Issues, with Applications in Cancer Research*. New York: John Wiley & Sons, 1982.

#### DAFTAR PUSTAKA

- Afifi, AA, Clark, V. *Computer-Aided Multivariate Analysis*. New York: Van Nostrand Reinhold Company, 1984.
- Bishop, YMM, Fienberg, SE, Holland, PW. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Massachusetts: The Massachusetts Institute of Technology Press, 1984.
- Breslow, NE, Day, NE. *Statistical Methods in Cancer Research: Vol 1 – The Analysis of Case Control Studies*. Lyon: International Agency for Research On Cancer, 1980.
- Feinstein, AR. *Multivariable Analysis: An Introduction*. New Haven: Yale University Press, 1996.
- Fox, J. *Linear Statistical Models and Related Methods: With Applications to Social Research*. New York: John Wiley & Sons, 1984.
- Hosmer, DW, Lemeshow, S. *Applied Logistic Regression*. New York: John Wiley & Sons, 1989.
- Kleinbaum, DG. *Logistic Regression: A Self-Learning Text*. New York: Springer-Verlag, 1994.
- Kleinbaum, DG, Kupper, LL, Morgenstern, H. *Epidemiologic Research: Principles and Quantitative Methods*. New York: Van Nostrand reinhold Company, 1982.
- Pagano, M, Gauvreau, K. *Principles of Biostatistics*. Belmont, California: Wadsworth Publishing Company, 1993.
- Schoenfeld, DA. *Analysis of Categorical Data: Logistic Models*. In: Mike, V,