ANALISIS PEUBAH RESPON BINER

Kismiantini
Jurusan Pendidikan Matematika
FMIPA Universitas Negeri Yogyakarta

Abstrak

Pada regresi linier klasik, peubah respon diasumsikan merupakan peubah yang bersifat kontinu. Bila peubah respon tidak lagi kontinu melainkan berupa kategori (biner, cacahan) maka model regresi linier klasik tidak dapat digunakan. Permasalahan tersebut dapat diatasi dengan model linier terampat. Model linier terampat yang digunakan dalam menganalisis peubah berskala biner adalah model logit, model probit dan model *complementary log-log*. Pada tulisan ini akan dikaji penggunaan ketiga model tersebut dalam menganalisis peubah respon biner. Bila nilai galat baku Pearson semakin kecil maka semakin baik pula model yang digunakan.

Kata kunci : peubah respon biner, model logit, model probit, model *complementary log-log*

PENDAHULUAN

Latar Belakang

Suatu metode statistika yang dapat digunakan untuk mengetahui hubungan antara peubah respon dan beberapa peubah penjelas adalah regresi linear. Dalam regresi linear klasik, peubah respon harus bersifat kuantitatif dengan skala pengukuran minimal adalah interval dan peubah penjelas adalah *fixed*. Peubah respon juga diasumsikan berdistribusi normal dan mempunyai ragam yang homogen.

Bila peubah respon bukan lagi peubah kuantitatif melainkan berupa peubah kategorik yang hanya terdiri dari beberapa nilai maka regresi linear klasik tidak dapat digunakan. Adapun model regresi yang sering digunakan untuk menganalisis peubah respon berskala biner adalah model logit, model probit, dan model *complementary log-log*.

Model logit, model probit dan model *complementary log-log* termasuk dalam model linear terampat (*Generalized Linear Models*/GLM). GLM merupakan suatu metode yang dapat digunakan untuk memecahkan masalah peubah respon tidak lagi kontinu melainkan kategorik (misalnya biner), dengan menggunakan fungsi penghubung (*link function*) tertentu sehingga diperoleh suatu model yang mampu menganalisa hubungan antara peubah respon kategorik dengan satu atau beberapa peubah penjelas. Fungsi penghubung dari model logit, model probit, dan model *complementary log-log* masing-masing adalah logit, normit/probit, dan gompit/*complementary log-log*.

TINJAUAN PUSTAKA

Regresi Linear Klasik

Pada regresi linear klasik, model regresi dapat dituliskan sebagai $Y = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p + \varepsilon$ dengan Y merupakan peubah respon yang bersifat kontinu, X adalah peubah penjelas yang *fixed*, β_i $(i = 0,1,2,..., p)$ merupakan parameter dan ε adalah galat yang menyebar normal dengan nilai tengah 0 dan ragam σ^2.

Asumsi yang harus dipenuhi bila akan melakukan uji hipotesis pada regresi linear klasik adalah galat menyebar normal dengan nilai tengah 0 dan ragam σ^2 (ragam galat homogen) serta galat menyebar secara acak.

Generalized Linear Models (GLM) dan Keluarga Sebaran Eksponensial

Bila Y peubah respon tidak lagi mengikuti sebaran normal namun binomial atau Poisson (asalkan masih dalam keluarga eksponensial) dan ragam Y merupakan fungsi dari nilai tengahnya sehingga dapat dipastikan bahwa ragam tidak homogen maka digunakanlah suatu model yang disebut model linear terampat (*Generalized Linear Models* / GLM). GLM merupakan pengembangan dari model linear klasik dimana peubah respon Y merupakan suatu komponen yang bebas dengan nilai tengah μ.

Ada tiga komponen utama dalam GLM (McCullagh dan Nelder, 1989):

1. Komponen acak, yaitu komponen dari Y yang bebas dan fungsi sebaran peluang Y termasuk dalam keluarga sebaran eksponensial dengan $E(Y) = \mu$.
2. Komponen sistematik, yaitu $x_1, x_2, \ldots, x_p$ yang menghasilkan penduga linear η dimana $\eta = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p$.

3. Fungsi penghubung (*link function*) $g(.)$, yang menggambarkan hubungan antara penduga linear η dengan nilai tengah μ. Hubungan ini dapat ditulis dengan $\eta = g(\mu)$.

Dalam model linear klasik, komponen (1) menyebar normal dan komponen (3) merupakan fungsi identitas. Sedangkan dalam GLM, komponen (1) mungkin berasal dari salah satu anggota keluarga sebaran eksponensial lainnya dan komponen (3) merupakan fungsi monoton lainnya. Dengan demikian GLM dapat dimodelkan dengan $g(E(Y_i|x_i)) = g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} = \eta(x_i)$, sedangkan ragam Y_i merupakan fungsi dari nilai tengah respon yaitu $Var(Y_i) = \phi\, Var(\mu_i)$.

Dengan mengasumsikan Y memiliki fungsi sebaran peluang keluarga eksponensial, sehingga dapat dimodelkan sebagai berikut :

$$f(y|\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right) \qquad (1)$$

dengan θ dan ϕ adalah parameter dan $a(\phi)$, $b(\theta)$ dan $c(y, \phi)$ adalah suatu fungsi yang diketahui. Jika ϕ diketahui, model diatas adalah keluarga eksponensial dengan parameter kanonik θ. Selanjutnya hasil θ disebut sebagai fungsi penghubung pada GLM.

Fungsi log-likelihood adalah log $f(y|\theta,\phi)$, ditulis sebagai $l(\theta,\phi|y)$, merupakan suatu fungsi dari θ dan ϕ dengan y diketahui. Nilai harapan dan ragam Y dapat ditentukan dengan mengevaluasi fungsi turunan dari $l(\theta,\phi|y)$ yang bersesuaian yaitu :

$$E(Y) = \mu = b'(\theta) \text{ dan } Var(Y) = \sigma^2 = b''(\theta)a(\phi)$$

dengan $b'(\theta)$ dan $b''(\theta)$ adalah turunan pertama dan kedua dari $b(\theta)$.

Sebagai contoh, peubah respon Y yang memiliki sebaran normal dengan fungsi kepekatan peluang :

$$f(y|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}\right) \qquad (2)$$

Fungsi kepekatan peluang dari sebaran ini dapat ditulis juga sebagai berikut :

$$f\left(y\middle|\mu,\sigma^2\right)=\exp\left(\frac{y\mu-\frac{1}{2}\mu^2}{\sigma^2}-\frac{y^2}{2\sigma^2}-\frac{1}{2}\ln\left(2\pi\sigma^2\right)\right) \qquad (3)$$

Jadi hasil identitas θ adalah μ dan ϕ adalah σ^2, dengan $a(\phi)=\phi=\sigma^2$. Diperoleh $b(\theta)=\frac{1}{2}\theta^2$ dengan $\theta=\mu$, maka nilai tengah dan ragam dari Y adalah

$$E(Y)=b'(\theta)=\theta=\mu$$

$$Var(Y)=b''(\theta)a(\phi)=1.\sigma^2=\sigma^2$$

Pada penguraian fungsi kepekatan peluang sebaran normal diperoleh juga fungsi penghubung pada model linear klasik yaitu μ.

Peubah Respon Biner

Peubah respon Y biasa dinotasikan dengan Y = 1 dan Y = 0. Nilai harapan dari suatu peubah respon biner Y adalah

$$E(Y)=0\times P(Y=0)+1\times P(Y=1)=P(Y=1)$$

Jika digunakan model regresi linear klasik untuk peubah respon biner diatas maka modelnya adalah :

$$Y = \boldsymbol{x}'\beta + \varepsilon \qquad (4)$$

Model ini bukanlah model yang layak karena nilai harapan dari $E\left(Y\middle|x\right)= \boldsymbol{x}'\beta$ bisa saja bernilai diluar [0, 1] sehingga tidak lagi merepresentasikan suatu model peluang. Selain modelnya tidak layak, terdapat permasalahan ketidakhomogenan ragam dengan ragamnya tergantung oleh x_i yaitu $Var\left(Y\middle|x\right)= \boldsymbol{x}'\beta\,(1 - \boldsymbol{x}'\beta)$. Sehingga untuk mengetahui hubungan antara sebuah peubah respon biner dengan beberapa peubah penjelas, regresi linear klasik tidak dapat digunakan.

Model Logit

Peubah respon Y yang berskala biner dapat dituliskan Y = 1 dan Y = 0, sehingga mengikuti sebaran Bernoulli dengan fungsi peluang :

$$f\left(y_i|\pi_i\right)=\pi_i^{\,y_i}\left(1-\pi_i\right)^{1-y_i}=\left(1-\pi_i\right)\exp\left[y_i\ln\left(\frac{\pi_i}{1-\pi_i}\right)\right] \qquad (5)$$

dengan $y_i = 0{,}1$ dan π_i adalah peluang kejadian ke-*i* bernilai Y = 1, Sebaran ini merupakan keluarga sebaran eksponensial. $\ln\left[\pi_i/\left(1-\pi_i\right)\right]$ merupakan log odds pada Y = 1 yang disebut sebagai logit dari π_i.

Secara umum model respon pada regresi adalah

$Y_i = E\left(Y|x_i\right)+\varepsilon_i$ dengan ε_i merupakan komponen acak. (6)

Dalam model regresi biner $E\left(Y|x_i\right)=\pi_i$ dengan $Var\left(Y|x_i\right)=\pi_i\left(1-\pi_i\right)$, sehingga model responnya menjadi $Y_i=\pi_i+\varepsilon_i$. Asumsi yang mendasari model-model regresi biner adalah peubah respon biner Y_i merupakan peubah yang saling bebas antara satu dengan lainnya.

Galat ε_i menghasilkan dua nilai yaitu $\varepsilon_i = 1-\pi_i$ jika Y_i = 1 dengan peluang π_i dan $\varepsilon_i = -\pi_i$ jika Y_i = 0 dengan peluang 1 - π_i, sehingga ε_i menyebar dengan nilai tengah dan ragam sebagai berikut:

$$E\left(\varepsilon_i\right)=\left(1-\pi_i\right)\pi_i+\left(-\pi_i\right)\left(1-\pi_i\right)=0$$

$$Var\left(\varepsilon_i\right)=\left(1-\pi_i\right)^2\pi_i+\left(-\pi_i\right)^2\left(1-\pi_i\right)=\pi_i\left(1-\pi_i\right)$$

Dapat terlihat bahwa galat dari model-model biner menyebar dengan ragam yang tidak homogen.

Pada model regresi biner, Y_i mengikuti sebaran Bernoulli dengan parameter π_i. Sehingga $E(Y_i)=\pi_i$ maka $\pi_i=\beta_0+\beta_1 x_{i1}+\ldots+\beta_p x_{ip}$, π_i bersifat kontinu dengan $0<\pi_i<1$. Adanya masukkan peubah penjelas X_i menyebabkan π_i > 1 atau π_i < 0 sehingga model $\pi_i=\beta_0+\beta_1 x_{i1}+\ldots+\beta_p x_{ip}$ tidak baik untuk melakukan pendugaan $\hat{\pi}_i$.

Bagaimana menghubungkan π_i dengan $\eta_i=\beta_0+\beta_1 x_{i1}+\ldots+\beta_p x_{ip}$? Membuat fungsi π_i yang dapat bernilai dari -∞ sampai ∞.

$$-\infty < g\left(\pi_i\right) < \infty$$

$$\Rightarrow -\infty < \ln\left(\frac{\pi_i}{1-\pi_i}\right) < \infty$$

Fungsi π_i diperoleh dari penguraian fungsi peluang sebaran Bernoulli dalam bentuk fungsi sebaran keluarga eksponensial yaitu

$f(y_i|\pi_i) = \pi_i^{\,y_i}(1-\pi_i)^{1-y_i} = \exp\left[y_i \ln\left(\frac{\pi_i}{1-\pi_i}\right) + \ln(1-\pi_i)\right]$ dengan θ adalah $\ln\left(\frac{\pi_i}{1-\pi_i}\right)$ dan ϕ adalah 1, dengan $a(\phi) = \phi$ =1. Sehingga $\ln[\pi_i/(1-\pi_i)]$ merupakan fungsi penghubung yang disebut fungsi penghubung logit. Sehingga diperoleh model yaitu

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} \tag{7}$$

Model inilah yang dikenal dengan sebutan model logit, karena mempunyai fungsi penghubung logit.

Model umum regresi logistik dengan *p* peubah penjelas yaitu

$$\pi(x) = \frac{\exp(g(x))}{1+\exp(g(x))} \tag{8}$$

dengan melakukan transformasi logit diperoleh

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] \tag{9}$$

dengan $g(x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$, $g(x)$ merupakan penduga logit yang berperan sebagai fungsi linear dari peubah penjelas. Model regresi logistik merupakan transformasi logit dari peluang kejadian pengamatan ke-i yaitu π_i, sebagai fungsi linear dari peubah penjelas dalam vektor $\boldsymbol{x}_i$. Model regresi logistik menggunakan logit sebagai fungsi penghubung.

Model Probit

Model probit adalah model peluang yang menggunakan suatu fungsi transformasi untuk memetakan fungsi linier $x'\beta$ pada selang [0,1].

$$P(Y_i = 1|x_i) = F(x'\beta)$$

Fungsi transformasi yang sesuai adalah yang memenuhi kriteria:

$$F(-\infty) = 0,\ F(\infty) = 1,\ \partial F(z)/\partial z > 0.$$

Kriteria di atas dapat dipenuhi oleh sembarang fungsi sebaran kumulatif (*cumulative density function*/cdf). Dua fungsi yang sering digunakan adalah sebaran normal baku dan sebaran logistik. Fungsi transformasi dalam model probit adalah fungsi sebaran kumulatif dari sebaran normal baku, yaitu :

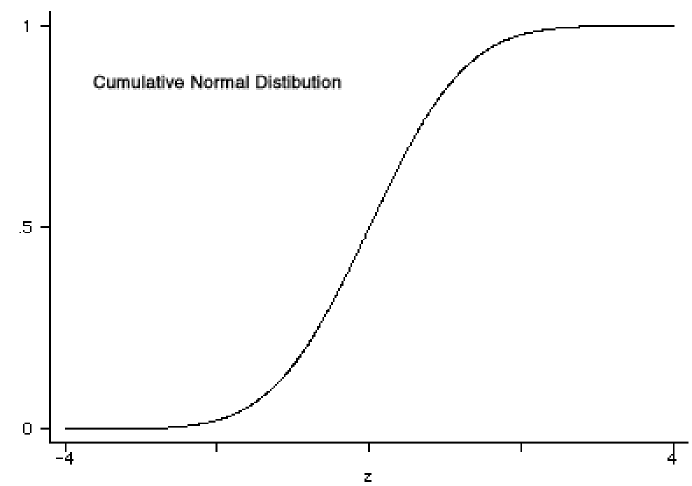
$$P(Y_i = 1|x_i) = \Phi(x_i'\beta) = \int_{-\infty}^{x_i'\beta} \phi(z)dz \qquad (10)$$

Di mana $\phi(.)$ adalah fungsi kepekatan peluang dan $\Phi(.)$ adalah fungsi sebaran kumulatif dari sebaran normal baku.

Dengan kata lain, untuk peubah respon biner :

$$F(g(x)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{g(x)} e^{-\frac{z^2}{2}} dz \qquad (11)$$

Gambar 1. Kurva Sebaran Normal Kumulatif



Secara umum model probit dapat dinyatakan sebagai berikut :

$$\pi_i = F(Z_i) = F(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots + + \beta_p x_{pi} + \varepsilon_i) \qquad (12)$$

Di mana F merupakan fungsi sebaran kumulatif dan X_{ji} adalah peubah bebas yang bersifat stokastik. Oleh karena model peluang probit berkaitan dengan fungsi sebaran kumulatif normal, maka dapat dituliskan model peluang probit sederhana sebagai berikut:

$$Z_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots + + \beta_p x_{pi} + \varepsilon_i \qquad (13)$$

Untuk memperoleh suatu dugaan dari nilai Z_i, maka dapat digunakan invers dari fungsi sebaran kumulatif normal baku, sehingga diperoleh :

$$Z_i = F^{-1}(\pi_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots + + \beta_p x_{pi} + \varepsilon_i \qquad (14)$$

Invers dari fungsi sebaran kumulatif normal baku ini dalam GLM tidak lain adalah fungsi penghubung *(link function).* Dengan demikian model probit diperoleh dengan menggunakan invers fungsi sebaran kumulatif normal baku sebagai *link function* pada GLM. Fungsi penghubung pada model probit disebut juga sebagai normit.

Peluang π_i yang dihasilkan dari suatu model probit dapat diinterpretasikan sebagai suatu dugaan dari peluang bersyarat bahwa suatu objek pengamatan atau kelompok akan mengalami suatu kejadian berdasarkan nilai tertentu dari X.

Model *Complementary log-log*

Transformasi *complementary log-log* adalah invers fungsi sebaran kumulatif $F^{-1}(\pi)$ dari sebaran Gompertz. Fungsi sebaran kumulatif Gompertz adalah

$$F(x,c,b) = 1 - \exp\left(\frac{-b(c^{x-1})}{\log c}\right),\ x \geq 0, b > 0, c > 0. \qquad (15)$$

Seperti model logit dan probit, transformasi *complementary log-log* memastikan bahwa peluang prediksi diantara interval [0,1]. Jika peluang sukses diekspresikan sebagai fungsi dari parameter yang sudah diketahui yaitu :

$$\pi_i = 1 - \exp\{-\exp(\Sigma_k\, \beta_k x_{ik})\} \qquad (16)$$

Maka model adalah linear dalam invers fungsi sebaran kumulatif Gompertz yaitu log dari log negative p_i atau log{-log(1-π_i)} yaitu

$$\log\{-\log(1-\pi_i)\} = \Sigma_k\, \beta_k x_{ik} \qquad (17)$$

Invers fungsi sebaran kumulatif Gompertz ini dalam GLM yang disebut sebagai fungsi penghubung. Sehingga model *complementary log-log* dengan fungsi penghubung yang sering disebut dengan gompit/*complementary log-log*.

Model logit, model probit dan model *complementary log-log* termasuk dalam model linier terampat (*Generalized Linear Models*/GLM), dengan berturut-turut mempunyai fungsi penghubung logit, probit, dan *complementary log-log*.

Secara umum, ada 3 fungsi penghubung yang dapat digunakan untuk menganalisa model respon biner. Fungsi ini adalah :

1) logit, yang merupakan invers fungsi sebaran kumulatif dari sebaran logistik.
2) normit (juga disebut probit), yang merupakan invers fungsi sebaran kumulatif dari sebaran normal baku.
3) gompit (juga disebut *complementary log-log*), yang merupakan invers fungsi sebaran kumulatif dari sebaran Gompertz.

Tabel 1. Fungsi Penghubung

Nama	Fungsi Penghubung	Sebaran	Rataan	Ragam
Logit	$\ln\{\pi_i/(1-\pi_i)\}$	Logistik	0	$\pi^2/3$
Normit (probit)	$\Phi^{-1}(\pi_i)$	Normal	0	1
Gompit (Complementary Log-log)	$\ln\{-\ln(1-\pi_i)\}$	Gompertz	$-\gamma$ (konstanta Euler)	$\pi^2/6$

Pendugaan Parameter

Pendugaan parameter β_i pada model respon biner dilakukan dengan metode penduga kemungkinan maksimum, karena asumsi kehomogenan ragam galat tidak dipenuhi. Jika antara amatan yang satu dengan amatan yang lain diasumsikan bebas, maka fungsi kemungkinan maksimumnya adalah :

$$l(\beta) = \prod_{i=1}^{n} f(y_i|\pi_i) \tag{18}$$

Parameter β_i diduga dengan memaksimumkan persamaan diatas. Untuk memudahkan perhitungan dilakukan pendekatan logaritma, sehingga fungsi log-kemungkinan sebagai berikut :

$$L(\beta) = \ln[l(\beta)] \tag{19}$$

Nilai dugaan β_i dapat diperoleh dengan memaksimumkan $\ln[l(\beta)]$ yaitu dengan membuat turunan pertama $\ln[l(\beta)]$ terhadap β_i dengan i = 0, 1, 2, …, p. Secara analitik penurunan ini sangatlah tidak mudah, oleh karena itu secara teknis pendugaan β_i diperoleh dari proses iterasi yaitu dengan menggunakan algoritma *Iteratively Reweighted Least Square* (*IRLS*) (McCullagh dan Nelder, 1989). Paket program Minitab dapat digunakan untuk menghitung nilai-nilai dugaan β_i.

Uji Taraf Nyata Parameter

Pengujian terhadap parameter model respon biner dilakukan untuk mengetahui peranan peubah bebas dalam model. Uji parameter yang digunakan adalah statistik :

1. Uji Wald (W)
2. Uji rasio kemungkinan (G)

Statistik Uji Wald digunakan untuk menguji parameter β_i (Hosmer dan Lemeshow, 1989). Rumus untuk Uji Wald berdasarkan hipotesis

H_0 : $\beta_i = 0$ lawan H_1 : $\beta_i \neq 0$ (i = 1, 2, …, p) adalah $W_i = \dfrac{\hat{\beta}_i}{\hat{SE}(\hat{\beta}_i)}$ dengan $\hat{\beta}_i$ merupakan penduga β_i dan $\hat{SE}(\hat{\beta})$ merupakan penduga galat baku dari $\hat{\beta}_i$. Statistik W mengikuti sebaran normal baku. Kriteria keputusan adalah H_0 ditolak jika $|W_{hitung}| > Z_{\frac{\alpha}{2}}$.

β_i diduga dengan metode kemungkinan maksimum maka untuk menguji peranan peubah penjelas di dalam model secara bersama-sama digunakan uji rasio kemungkinan yaitu Uji G (Hosmer dan Lemeshow, 1989). Rumus untuk uji G berdasarkan hipotesis :
H_0 : $\beta_1 = \beta_2 = ... = \beta_p = 0$ lawan H_1 : paling sedikit ada satu $\beta_i \neq 0$ (i = 1, 2, …, p) adalah

$$G = -2\ln\left[\frac{L_0}{L_1}\right]$$

L_0 **:** likelihood tanpa peubah penjelas

L_1 **:** likelihood dengan peubah penjelas

Statistik G akan mengikuti sebaran χ^2 dengan derajat bebas p. Kriteria Keputusan yang diambil yaitu menolak H_0 jika $G_{\text{hitung}} > \chi^2_{\alpha(p)}$ (Hosmer dan Lemeshow, 1989).

Kriteria Pemilihan Model Terbaik

Kriteria yang digunakan untuk memilih model terbaik antara model logit, model probit, dan model *complementary log-log* adalah galat baku. Galat baku pada model-model ini berbeda dengan galat baku pada regresi linier klasik, karena model logit, probit dan *complementary log-log* merupakan bagian dari model linier terampat. Galat baku pada model linier terampat berupa galat baku Pearson dan galat baku *deviance* yang dirumuskan sebagai berikut (Agresti, 2002) :

$$r_{P_i} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{Var}(Y_i)}} \text{ dan } r_{D_i} = \sqrt{d_i} \times sign(y_i - \hat{\mu}_i) \quad (20)$$

dengan $d_i = 2\omega_i \left[y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i) \right]$ adalah *deviance*.

Intepretasi Koefisien

Interpretasi koefisien untuk model logit dapat dilakukan dengan melihat rasio oddsnya. Makna rasio odds sebesar ψ adalah bahwa untuk X = 1 memiliki kecenderungan Y = 1 sebesar ψ kali dibandingkan dengan untuk X = 0 (Hosmer dan Lemeshow, 1989). Jika suatu peubah penjelas mempunyai tanda koefisien positif, maka nilai rasio oddsnya akan lebih besar dari satu, sebaliknya jika tanda koefisiennya negatif maka nilai rasio oddsnya akan lebih kecil dari satu.

Interpretasi koefisien model probit dan *complementary log-log* tidak dapat dilakukan seperti pada model logit. Dalam prakteknya, interpretasi koefisien probit dan *complementary log-log* kadangkala dilakukan hanya dengan melihat tanda dari koefisien, yang berarti hanya melihat arah dari pengaruh peubah bebas terhadap peubah respon, apakah memberikan pengaruh positif atau negatif.

APLIKASI DENGAN DATA

Dua buah jenis bohlam lampu yaitu A dan B dicobakan pada 10 voltase yang berbeda. Untuk masing-masing level voltase dicobakan 50 bohlam A dan 50 bohlam B. Respon yang diamati adalah jumlah bohlam yang putus untuk masing-masing voltase yang dicobakan setelah 800 jam. (Data diambil dari data *lightbul* di Minitab versi 13). Berikut data yang diperoleh :

Tabel 2. Data Bohlam Lampu

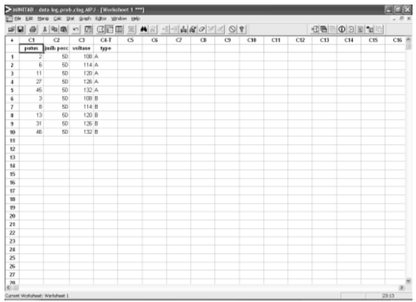
# bohlam putus	# percobaan	Level voltase	Jenis bohlam
2	50	108	A
6	50	114	A
11	50	120	A
27	50	126	A
45	50	132	A
3	50	108	B
8	50	114	B
13	50	120	B
31	50	126	B
46	50	132	B

Keterangan :

- Peubah respon adalah jumlah bohlam yang putus
- Peubah penjelas adalah level voltase dan jenis bohlam lampu

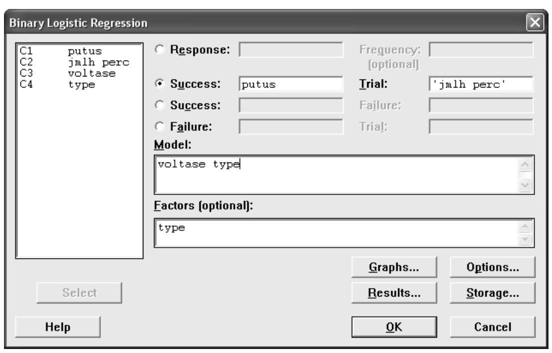
Analisis data diolah menggunakan *software* Minitab versi 13 dengan langkah-langkah sebagai berikut :

1. Masukkan data pada **worksheet Minitab**.

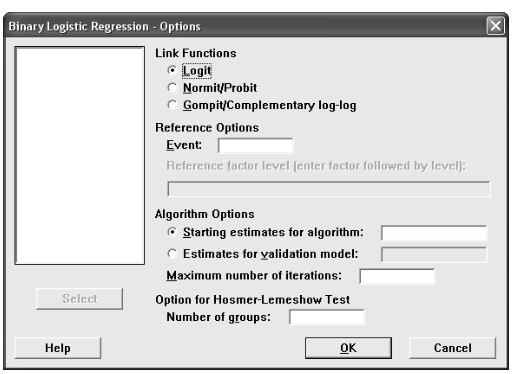


2. Langkah analisis dimulai dengan memilih menu ***Stat-Regression-Binary Logistic Regression***.

3. Masukkan **peubah respon** yaitu putus ke **Success**, lalu masukkan **jumlah percobaan** yaitu jmlh perc ke **Trial**, pada **Model** masukkan **peubah penjelas** yaitu voltase dan type. Karena peubah type merupakan peubah kategorik maka masukkan type ke **Factors(optional)**.



4. Pada **Options…** pilih **Link Functions Logit** lalu tekan **OK.**



5. Abaikan yang lain lalu tekan **OK.**
6. Ulangi langkah 2-5 dengan mengganti ***link function***.

Tabel 3. Ringkasan Output Minitab versi 13 untuk Ketiga Fungsi Penghubung

	Logit	**Probit**	**Complementary Log-log**
Model dugaan	$\hat{Y}$ = -27.075 + 0.21759 Volts + 0.2874TypeB	$\hat{Y}$ = -15.454 + 0.124375 Volts + 0.1660TypeB	$\hat{Y}$ = -20.839 + 0.16360 Volts + 0.1805TypeB
Uji Taraf Nyata Parameter Constant = 0 Volts = 0 Type B = 0	 Z=-11.68; P=0.000 Z=11.60; P=0.000 Z=1.20; P=0.232	 Z=-12.96; P=0.000 Z=12.83; P=0.000 Z=1.21; P=0.225	 Z=-12.97; P=0.000 Z=12.87; P=0.000 Z=1.13; P=0.260
Log-Likelihood	-216.723	-217.778	-213.836
Uji semua slope = 0	G = 232.543 P-Value = 0.000	G = 230.432 P-Value = 0.000	G = 238.316 P-Value = 0.000
Goodness-of-Fit-Tests Pearson Deviance Hosmer-Lemeshow	 χ^2=7.375 P-Value=0.391 χ^2=7.512 P-Value=0.378 χ^2=7.375 P-Value=0.497	 χ^2=9.506 P-Value=0.218 χ^2=9.623 P-Value=0.211 χ^2=9.506 P-Value=0.301	 χ^2=1.755 P-Value=0.972 χ^2=1.740 P-Value=0.973 χ^2=1.755 P-Value=0.988
Concordant	84.1%	84.1%	84.1%
Measures of Association Somers'D Goodman-Kruskal Gamma Kendall's Tau-a	 0.74 0.78 0.35	 0.74 0.78 0.35	 0.74 0.78 0.35

Pembandingan Model

Untuk memilih model terbaik diantara model logit, model probit, dan model *complementary log-log* akan digunakan tabel dibawah ini.

Tabel 4. Dugaan Parameter Model Logit, Probit, dan *Complementary log-log*

Peubah bebas	Model Logit			Model Probit			Model *Complementary log-log*		
	bl_i	Z	P-Value	bp_i	Z	P-Value	bc_i	Z	P-Value
Intersep	-27.075	-11.69	0.000	-15.454	-12.96	0.000	-20.839	-12.97	0.000
Volts	0.21759	11.60	0.000	0.124375	12.83	0.000	0.16360	12.87	0.000
TypeB	0.2874	1.20	0.232	0.1660	1.21	0.225	0.1805	1.13	0.260
Statistik-G = 232.543 P-Value = 0.000				Statistik-G = 230.432 P-Value = 0.000			Statistik-G = 238.316 P-Value = 0.000		
Rata-rata Galat Baku = 0.196				Rata-rata Galat Baku = 0.160			Rata-rata Galat Baku = 0.039		

Berdasarkan Tabel 4 dapat dilihat bahwa model dengan dua peubah penjelas yaitu volts dan typeB menghasilkan nilai statistik-G untuk model logit sebesar 232.543 dengan p-value=0.000, untuk model probit sebesar 230.432 dengan p-value=0.000, untuk model *complementary log-log* sebesar 238.316 dengan p-value=0.000. Bila dipilih $\alpha = 0.05$ maka hasil tersebut menyatakan bahwa secara bersama-sama peubah penjelas tersebut berperan nyata terhadap peubah respon. Hal ini menunjukkan bahwa level voltase dan jenis bohlam lampu berpengaruh terhadap jumlah bohlam lampu yang putus.

Uji Wald digunakan untuk mengetahui signifikansi parameter. Untuk model logit terlihat bahwa koefisien-koefisien volts dan typeB memiliki nilai Z berturut-turut adalah –11.69, 11.60, 1.20 dengan p-value adalah 0.000, 0.000, 0.232. Untuk model probit memiliki nilai Z berturut-turut adalah –12.96, 12.83, 1.21 dengan p-value adalah 0.000, 0.000, 0.225. Sedangkan untuk model *complementary log-log* memiliki nilai Z berturut-turut adalah –12.97, 12.87, 1.13 dengan p-value adalah 0.000, 0.000, 0.260. Bila dipilih $\alpha = 0.05$ maka hal ini menunjukkan bahwa level voltase dapat menerangkan jumlah bohlam lampu yang putus, sedangkan jenis bohlam lampu type B tidak dapat menerangkan jumlah bohlam lampu yang putus.

Rata-rata galat baku dari model logit adalah 0.196, model probit sebesar 0.160, dan model *complementary log-log* adalah 0.039. Dapat dilihat bahwa rata-rata galat baku dari model *complementary log-log* mempunyai nilai yang paling kecil diantara ketiga model tersebut. Dari kriteria ini dapat ditarik suatu kesimpulan bahwa model *complementary log-log* yang paling baik untuk menganalisa data tersebut.

Interpretasi Koefisien

Dugaan koefisien untuk model logit, probit, dan *complementary log-log* mempunyai tanda yang sama, sehingga pengaruh peubah penjelas pada peubah respon mempunyai arah yang sama. Besar koefisien ini berbeda antara model logit, model probit, dan model *complementary log-log*. Model logit mempunyai koefisien yang lebih besar jika dibandingkan koefisien pada model probit dan *complementary log-log*.

Penduga *odds ratio* untuk peubah kontinu volts sebesar 1.24 yang lebih besar dari 1 sehingga dapat diartikan bahwa semakin besar voltase akan semakin besar jumlah bohlam lampu yang putus, sedangkan odds ratio untuk peubah kategorik yaitu peubah type B adalah sebesar 1.33 yang berarti jumlah bohlam lampu yang putus dari type bohlam B 1.33 kali dibandingkan dengan type bohlam A.

Interpretasi koefisien pada model probit dan model *complementary log-log* dengan melihat tanda koefisiennya, yaitu koefisien untuk peubah volts dan typeB bertanda positif yang berarti peluang jumlah bohlam lampu yang putus akan semakin besar bila voltase ditingkatkan dan jumlah bohlam type B diperbanyak.

SIMPULAN

Beberapa simpulan yang dapat diambil dari tulisan ini adalah :

1. Model *complementary log-log* merupakan model yang terbaik diantara model logit dan model probit untuk data bohlam lampu karena memiliki nilai galat baku yang paling kecil.
2. Nilai dugaan koefisien baik dari model logit, model probit, maupun model *complementary log-log* mempunyai tanda yang sama.
3. Model logit mempunyai nilai dugaan koefisien yang paling besar dibandingkan dengan dugaan koefisien pada model probit dan model *complementary log-log*.
4. Interpretasi koefisien model logit lebih mudah dibandingkan interpretasi koefisien model probit dan model *complementary log-log* yaitu dengan melihat dugaan *odds ratio*.

DAFTAR PUSTAKA

Agresti, Allan. 2002. *Categorical Data Analysis*. New York: John Wiley & Sons.

Agresti, Allan. 1996. *An Introduction to Categorical Data Analysis*. New York: John Wiley & Sons.

Hosmer, D.W. dan Lemeshow, S. 1989. *Applied Logistic Regression*. New York: John Wiley & Sons.

McCullagh, P. dan Nelder, J. A. 1989. *Generalized Linear Models 2nd Edition*. London: Chapman & Hall.

Myers, R. H. 1989. *Classical and Modern Regression with Application*. Boston: PWS-KENT.