PEMODELAN DENGAN REGRESI LOGISTIK

1. Data Biner

Data biner merupakan data yang hanya memiliki dua kemungkinan hasil. Secara umum, kedua hasil dilambangkan dengan $Y = 1$ (sukses) dan $Y = 0$ (gagal) dengan peluang masing-masing sebesar p dan $q = 1 - p$. Dalam beberapa kondisi, jika fokus perhatian bukan terhadap respon dari satu objek melainkan pada suatu kelompok objek yang memiliki kondisi yang sama, maka dalam hal ini data biasanya dicatat dalam bentuk y keberhasilan dalam kelompok berukuran n atau dalam bentuk peluang keberhasilan dalam kelompok tersebut. Data yang demikian dinamakan data biner berkelompok atau disebut juga sebagai data binomial.

Data binomial menyebar menurut distribusi Binomial dengan distribusi peluang sebagai berikut :

$$B(y; n, p) = \binom{n}{y} p^y (1-p)^{n-y} \qquad\qquad y = 0,1,2,\cdots,n$$

dengan nilai tengah dan ragam sebesar $E(Y) = np$ dan $Var(Y) = np(1-p)$. Sedangkan data biner yang tidak berkelompok menyebar menurut sebaran Bernoulli, yaitu bentuk khusus sebaran Binomial untuk $n = 1$, sehingga sebaran peluangnya adalah :

$$B(y; 1, p) = p^y (1-p)^{1-y} \qquad\qquad y = 0,1$$

dengan nilai tengah dan ragam masing-masing sebesar $E(Y) = p$ dan $Var(Y) = p(1-p)$.

2. Regresi Logistik Biner

Misalkan ingin dimodelkan hubungan antara beberapa variabel prediktor dengan suatu variabel respon biner yaitu variabel biner yang hanya mempunyai dua nilai kemungkinan yang biasanya dinyatakan dengan 0 (gagal) dan 1 (sukses).

Dengan analisis regresi biasa hubungan tersebut dapat dinyatakan dalam model [5]:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \qquad \begin{cases} i = 1,2,\dots,n \\ Y_i = 0,1 \end{cases} \tag{2.1}$$

karena Y_i hanya memiliki dua kemungkinan nilai akibatnya ε_i juga memiliki kondisi yang sama, dimana :

$$\varepsilon_i = \begin{cases} 1 - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \cdots - \beta_k X_{ki} & ; Y_i = 1 \\ -\beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \cdots - \beta_k X_{ki} & ; Y_i = 0. \end{cases}$$

Dalam keadaan seperti ini ε_i tidak dapat lagi diasumsikan berdistribusi normal, melainkan berdistribusi menurut distribusi Bernoulli yang memiliki distribusi peluang Y_i sebagai berikut :

Tabel 2.1 Distribusi Peluang Y_i

Y_i	Peluang ($P(Y_i = y_i)$)
0	$q_i = 1 - p_i$
1	p_i

dengan nilai tengah $E(Y_i)$, yaitu :

$$\begin{aligned} E(Y_i) &= 0 \cdot (q_i) + 1 \cdot (p_i) \\ &= p_i. \end{aligned} \tag{2.2}$$

dari Persamaan (2.1) diperoleh nilai tengah $E(Y_i)$, yaitu :

$$\begin{aligned} E(Y_i) &= E(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i) \\ &= E(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}) + E(\varepsilon_i) \end{aligned}$$

dari asumsi diketahui $E(\varepsilon_i) = 0$, maka :

$$E(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} \qquad (2.3)$$

Dengan mensubsitusi Persamaan (2.2) ke dalam Persamaan (2.3) diperoleh :

$$p_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} \qquad ; 0 \leq p_i \leq 1 \qquad (2.4)$$

Persamaan (2.4) disebut model peluang liniar (*linear probability model*). Model ini menunjukkan bahwa $E(Y_i)$ dapat dinyatakan sebagai peluang sukses p_i, sehingga diperoleh distribusi peluang $\varepsilon_i(\varepsilon_i = Y_i - E(Y_i))$, sebagai berikut :

Tabel 2.2 Distribusi Peluang ε_i

Y_i	ε_i	Peluang $(P(\varepsilon_i = e_i))$
0	$-p_i$	$q_i = 1 - p_i$
1	$q_i = 1 - p_i$	p_i

Dari Tabel 2.2 dapat ditentukan :

$$E(\varepsilon_i) = \sum_{i=0}^{1} e_i P(\varepsilon_i = e_i)$$

$$= -p_i P(\varepsilon_i = -p_i) + 1 - p_i P(\varepsilon_i = 1 - p_i)$$

$$= (-p_i)(1 - p_i) + (1 - p_i)p_i$$

$$= 0$$

dan

$$E({\varepsilon_i}^2) = \sum_{i=0}^{1} {e_i}^2 P(\varepsilon_i = e_i)$$

$$= (-p_i)^2 P(\varepsilon_i = -p_i) + (1 - p_i)^2 P(\varepsilon_i = 1 - p_i)$$

$$= (-p_i)^2 (1 - p_i) + (1 - p_i)^2 p_i$$

$$= {p_i}^2 q_i + {q_i}^2 p_i$$

$$= p_i q_i (p_i + (1 - p_i))$$

$$= p_i q_i$$

dengan nilai ragam galat :

$$Var(\varepsilon_i) = E(\varepsilon_i{}^2) - [E(\varepsilon_i)]^2$$

$$= p_i q_i - 0$$

$$= p_i q_i$$

karena $Var(\varepsilon_i)$ memiliki nilai yang bergantung pada peluang sukses p_i, maka akibatnya asumsi kehomogenan ragam galat ε_i tidak terpenuhi.

Dari uraian diatas dapat diketahui bahwa analisis regresi biasa tidak dapat digunakan untuk memodelkan hubungan antara variabel respon biner dengan beberapa variabel prediktor. Salah satu pendekatan yang dapat digunakan untuk mengatasi masalah ini adalah analisis regresi logistik.

Tetapi sebelum melakukan analisis ini perlu dilakukan transformasi yang dapat menjamin nilai peluang sukses p_i akan selalu berada dalam selang [0,1]. Hal ini dikarenakan dalam menduga koefisien regresi $\widehat{\boldsymbol{\beta}}$ pada model penduga dari Persamaan (2.4), yaitu :

$$\widehat{p_i} = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki} \qquad (2.5)$$

Penduga koefisien regresi $\widehat{\boldsymbol{\beta}}$ tidak memiliki batasan nilai. Kombinasi linier dari nilai $\widehat{\boldsymbol{\beta}}$ dapat berada pada selang $(-\infty, \infty)$, sehingga tidak ada jaminan bahwa dugaan nilai peluang tersebut akan berada pada selang [0,1]. Oleh karena itu perlu dilakukan transformasi dimana salah satu bentuknya adalah transformasi logit.

3. Model Regresi Logistik Biner

Model regresi logistik adalah model regresi variabel respon biner yang melibatkan transformasi logit. Model regresi logistik diperoleh dari fungsi logistik dengan definisi sebagai berikut :

Definisi 3.1 [5]. Fungsi kepekatan peluang (fkp) bagi variabel random X yang terdistribusi logistik adalah :

$$f(x) = \frac{\exp\{(x-\mu)/\tau\}}{\tau[1+\exp\{(x-\mu)/\tau\}]^2} \qquad ; -\infty \leq x \leq \infty \, ; \, \tau > 0$$

dengan nilai tengah μ dan ragam $\sigma^2 = \pi^2\tau^2/3$

Langkah awal pembentukan model ini adalah dengan melakukan transformasi terhadap variabel respon pada Persamaan (2.4) yang bernilai [0,1] ke variabel respon yang memilki nilai $(-\infty, \infty)$. Kemudian membentuk model yang baru berdasarkan nilai yang telah ditransformasi tersebut. Diketahui bahwa :

$$p_i = P(X \leq x_i) = \int_{-\infty}^{x_i} f(u)\, du.$$

Berdasarkan fkp dari distribusi logistik diperoleh :

$$p_i = \int_{-\infty}^{x_i} \frac{\exp\{(u-\mu)/\tau\}}{\tau[1+\exp\{(u-\mu)/\tau\}]^2}\, du.$$

Misalkan $y = 1 + \exp\{(u-\mu)/\tau\}$ dan $\frac{dy}{du} = \exp\left(\frac{(u-\mu)}{\tau}\right) \cdot \frac{1}{\tau}$, maka :

$$p_i = \int_1^{1+\exp\{(x_i-\mu)/\tau\}} \frac{\exp\{(u-\mu)/\tau\}}{\tau(y)^2} \cdot \frac{\tau.\, dy}{\exp\{(u-\mu)/\tau\}}$$

$$= \int_1^{1+\exp\{(x_i-\mu)/\tau\}} \frac{1}{y^2}\, dy$$

$$= \left[-\frac{1}{y}\right]_1^{1+\exp\{(x_i-\mu)/\tau\}}$$

$$= -\frac{1}{1+\exp\{(x_i-\mu)/\tau\}} + 1$$

$$= \frac{\exp\{(x_i-\mu)/\tau\}}{1+\exp\{(x_i-\mu)/\tau\}}$$

Misalkan $\beta_0 = {}^{-\mu}/_{\tau}$ dan $\beta_1 = {}^{1}/_{\tau}$, maka :

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1+\exp(\beta_0 + \beta_1 x_i)} \tag{3.1}$$

Persamaan (3.1) adalah fungsi logistik dengan satu variabel prediktor. Sedangkan untuk dua atau lebih variabel prediktor $x_1, x_2, \dots, x_k$ maka Persamaan (3.1) dapat diperluas menjadi:

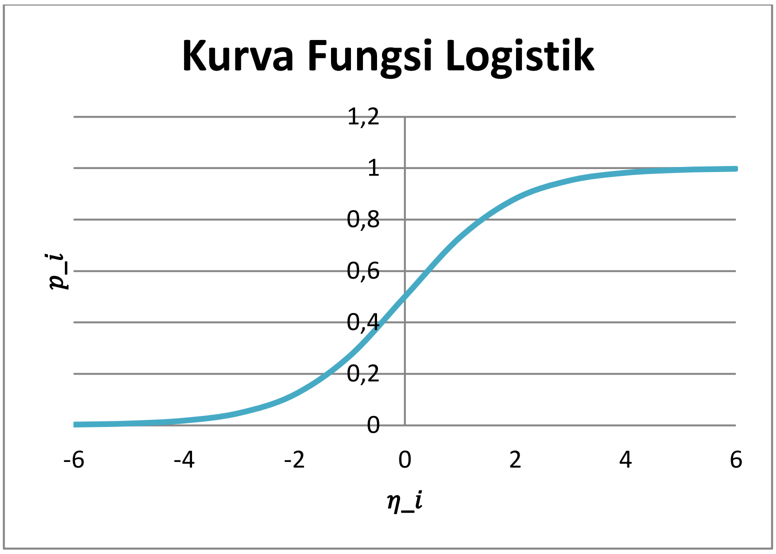
$$p_i = \frac{\exp\ (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki})}{1+\exp\ (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki})}$$

$$= \frac{1}{1+\exp[-(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki})]} \tag{3.2}$$

Jika $\eta_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$ maka Persamaan (3.2) dapat ditulis menjadi :

$$p_i = \frac{1}{1+e^{-\eta_i}} \tag{3.3}$$

Persamaan (3.3) adalah bentuk umum dari model logistik atau disebut juga sebagai fungsi logit. Jika $\eta_i \to -\infty$ maka $p_i \to 0$ dan jika $\eta_i \to \infty$ maka $p_i \to 1$. Hubungan ini berbentuk kurva seperti pada Gambar 3.1. Dengan demikian, dapat dijamin bahwa nilai p_i akan selalu berada pada selang [0,1] .



Gambar 3.1 Diagram Garis Kurva Fungsi Logistik.

.

Kemudian untuk membentuk model regresi logistik maka dilakukan transformasi terlebih dahulu terhadap Persamaan (2.4) untuk mempertahankan struktur linear dari model. Transformasi yang dilakukan adalah transformasi logit, dengan didefinisikan sebagai berikut :

$$\text{Logit }(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) \tag{3.4}$$

Dari Persamaan (3.3) diperoleh :

$$(1 + e^{-\eta_i})p_i = 1$$

$$e^{-\eta_i} = \frac{1}{p_i} - 1$$

$$\frac{1}{e^{\eta_i}} = \frac{1-p_i}{p_i}$$

$$e^{\eta_i} = \frac{p_i}{1-p_i}$$

sehingga diperoleh

$$\eta_i = \ln\left(\frac{p_i}{1-p_i}\right) \tag{3.5}$$

berdasarkan Persamaan (3.4) diperoleh

$$\eta_i = \text{Logit}\ (p_i)$$

karena $\eta_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$, maka diperoleh model linier sebagai berikut :

$$\text{Logit}\ (p_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}\ ;\ -\infty \leq \text{Logit}(p_i) \leq \infty \qquad (3.6)$$

Persamaan (3.6) merupakan hasil transformasi logit dan disebut sebagai model regresi logistik.

4. Pendugaan Regresi Logistik Biner

Pendugaan parameter dalam regresi logistik dilakukan dengan cara metode maksimum *likelihood*. Metode tersebut menduga koefisien $\boldsymbol{\beta}$ dengan cara memaksimumkan fungsi *likelihood* dan mensyaratkan bahwa data harus mengikuti suatu distribusi tertentu. Pada regresi logistik biner, setiap pengamatan mengikuti distribusi Bernoulli sehingga dapat ditentukan fungsi *likelihood*nya [5].

Jika x_i dan y_i adalah pasangan variabel prediktor dan respon pada pengamatan ke-i dan diasumsikan bahwa setiap pasangan pengamatan saling bebas dengan pasangan pengamatan lainnya, $i = 1,2, \dots, n$ maka fungsi peluang untuk setiap pasangan adalah sebagai berikut

$$f(x_i) = p_i^{\,y_i}(1 - p_i)^{\,i - y_i} \qquad ; y_i = 0,1 \qquad (4.1)$$

dengan

$$p_i = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki})]}$$

Setiap pasangan pengamatan diasumsikan bebas sehingga fungsi *likelihood*nya merupakan gabungan dari fungsi distribusi masing-masing pasangan yaitu sebagai berikut:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} p_i^{y_i}(1 - p_i)^{\,i-y_i} \qquad (4.2)$$

Fungsi *likelihood* tersebut lebih mudah dimaksimumkan dalam bentuk log $L(\boldsymbol{\beta})$ dan dinyatakan dengan $L(\boldsymbol{\beta})$

$$L(\boldsymbol{\beta}) = \ln l(\boldsymbol{\beta})$$

$$= \ln \left[\prod_{i=1}^{n} p_i^{y_i}(1 - p_i)^{\,1-y_i} \right]$$

$$= \ln[\ p_1^{y_1} \cdots p_n^{y_n}\ (1 - p_1)^{1-y_1} \cdots (1 - p_n)^{\,1-y_n}\ ]$$

berdasarkan sifat logaritma natural persamaan diatas dapat dibentuk seperti berikut:

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n} [\ln\ p_i^{y_i} + \ln\ (1 - p_i)^{1-y_i}]$$

$$= \sum_{i=1}^{n} [\ y_i \ln p_i + (1 -\ y_i)\ \ln(1 - p_i)\ ]$$

$$= \sum_{i=1}^{n} \left[\ y_i \ln\left(\frac{p_i}{1 - p_i}\right) + \ln(1 - p_i)\right]$$

dari Persamaan di atas diperoleh :

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[\ y_i\, \eta_i + \ln\ \left(1 - \frac{1}{1 +\ \ e^{-\eta_i}}\right)\right]$$

$$= \sum_{i=1}^{n} \left[\ y_i\, \eta_i + \ln\ \left(\frac{e^{-\eta_i}}{1 +\ \ e^{-\eta_i}}\right)\right]$$

$$= \sum_{i=1}^{n} [\ y_i\, \eta_i + \ln\ (1 + e^{\eta_i})^{-1}]$$

sehingga diperoleh :

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n} [\ y_i\, \eta_i - \ln\ (1 + e^{\eta_i})]. \qquad (4.3)$$

dengan $\eta_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$. Sehingga turunan pertamanya adalah :

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_0} = \frac{\partial}{\partial \boldsymbol{\beta}_0} \sum_{i=1}^{n} [\, y_i \, \eta_i - \ln \, (1 + e^{\eta_i})]$$

$$= \sum_{i=1}^{n} \left[y_i - \frac{e^{\eta_i}}{1 + e^{\eta_i}} \right]$$

berdasarkan Persamaan (3.4) diperoleh :

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_0} = \sum_{i=1}^{n} [\, y_i - p_i \,] \tag{4.4}$$

Selanjutnya turunan pertaman dari $L(\boldsymbol{\beta})$ terhadap β_1, yaitu :

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1} = \frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} [\, y_i \, \eta_i - \ln \, (1 + e^{\eta_i})]$$

$$= \sum_{i=1}^{n} \left[y_i x_{1i} - \frac{x_{1i} \, e^{\eta_i}}{1 + e^{\eta_i}} \right]$$

berdasarkan Persamaan di atas diperoleh :

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1} = \sum_{i=1}^{n} x_{1i} [\, y_i - p_i \,] \tag{4.5}$$

dengan melakukan hal yang sama pada turunan pertama dari $L(\boldsymbol{\beta})$ terhadap $(\beta_0, \beta_1, \cdots, \beta_k)$, maka diperoleh :

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_k} = \frac{\partial}{\partial \beta_k} \sum_{i=1}^{n} [\, y_i \, \eta_i - \ln \, (1 + e^{\eta_i})]$$

$$= \sum_{i=1}^{n} \left[y_i x_{ki} - \frac{x_{ki}\, e^{\eta_i}}{1 + e^{\eta_i}} \right]$$

berdasarkan Persamaan (3.4) diperoleh :

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_k} = \sum_{i=1}^{n} x_{ki} [\, y_i - p_i \,] \qquad (4.6)$$

dalam bentuk matriks diperoleh :

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{bmatrix} . \begin{bmatrix} y_1 - p_1 \\ y_2 - p_2 \\ \vdots \\ y_n - p_n \end{bmatrix} = \mathbf{X}^T (\mathbf{Y} - \mathbf{P_i})$$

Selanjutnya akan dicari turunan keduanya yaitu :

$$\frac{\partial^2 L(\boldsymbol{\beta})}{(\partial \beta_0)^2} = \frac{\partial^2}{(\partial \beta_0)^2} \sum_{i=1}^{n} [\, y_i\, \eta_i - \ln\ (1 + e^{\eta_i})]$$

$$= \frac{\partial}{\partial \beta_0} \sum_{i=1}^{n} [\, y_i - p_i \,]$$

berdasarkan Persamaan (3.4) diperoleh :

$$\frac{\partial^2 L(\boldsymbol{\beta})}{(\partial \beta_0)^2} = \frac{\partial}{\partial \beta_0} \sum_{i=1}^{n} \left[y_i - \frac{e^{\eta_i}}{1 + e^{\eta_i}} \right] \qquad (4.7)$$

sehingga diperoleh :

$$\frac{\partial^2 L(\boldsymbol{\beta})}{(\partial \beta_0)^2} = -\sum_{i=1}^{n} \left[\frac{e^{\eta_i}(1 + e^{\eta_i}) - e^{\eta_i}{}^2}{(1 + e^{\eta_i})^2} \right]$$

$$= -\sum_{i=1}^{n}\left[\frac{e^{\eta_i}}{(1+e^{\eta_i})} - \left(\frac{e^{\eta_i}}{(1+e^{\eta_i})}\right)^2\right]$$

berdasarkan Persamaan (3.4) diperoleh :

$$\frac{\partial^2 L(\boldsymbol{\beta})}{(\partial\beta_0)^2} = -\sum_{i=1}^{n}[\, p_i(1-p_i)] \tag{4.8}$$

Selanjutnya turunan dari $L(\boldsymbol{\beta})$ terhadap $\beta_0\beta_k$, yaitu :

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial\beta_0\beta_k} = \frac{\partial}{\partial\beta_k}\sum_{i=1}^{n}\left[\, y_i - \frac{e^{\eta_i}}{1+e^{\eta_i}}\right]$$

$$= -\sum_{i=1}^{n}\left[\frac{x_{ki}\;\, e^{\eta_i}(1+e^{\eta_i}) - x_{ki}\;\, {e^{\eta_i}}^2}{(1+e^{\eta_i})^2}\right]$$

berdasarkan Persamaan (3.4) diperoleh :

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial\beta_0\beta_k} = -\sum_{i=1}^{n} x_{ki}[\, p_i(1-p_i)] \tag{4.9}$$

Misal turunan parsial pertama dari $L(\boldsymbol{\beta})$ terhadap $\beta_j, j \leq k$ adalah

$$\frac{\partial L(\boldsymbol{\beta})}{\partial\beta_j} = \frac{\partial}{\partial\beta_j}\sum_{i=1}^{n}[\, y_i\,\eta_i - \ln\;(1+e^{\eta_i})]$$

$$= \sum_{i=1}^{n}\left[\, y_i x_{ji} - \frac{x_{ji}\, e^{\eta_i}}{1+e^{\eta_i}}\right]$$

berdasarkan Persamaan (3.4) diperoleh :

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{n} x_{ji}[\, y_i - p_i \,] \tag{4.10}$$

maka turunan parsial kedua terhadap β_u, $u \leq k$ adalah

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_u \beta_j} = \frac{\partial}{\partial \beta_u} \sum_{i=1}^{n} \left[y_i x_{ji} - \frac{x_{ji}\ e^{\eta_i}}{1 + e^{\eta_i}} \right]$$

$$= -\sum_{i=1}^{n} \left[\frac{x_{ui} x_{ji}\ e^{\eta_i}(1 + e^{\eta_i}) - x_{ui} x_{ji}\ e^{\eta_i}{}^2}{(1 + e^{\eta_i})^2} \right]$$

berdasarkan Persamaan (3.4) diperoleh :

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_u \boldsymbol{\beta}_j} = -\sum_{i=1}^{n} x_{ui} x_{ji}[\, p_i(1 - p_i)] \quad \text{untuk } u, j = 1,2, \dots, k \tag{4.11}$$

dengan melakukan hal sama pada turunan kedua dari $L(\boldsymbol{\beta})$ terhadap β_j, maka diperoleh :

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\left(\partial \beta_j\right)^2} = \frac{\partial}{\partial \beta_j} \sum_{i=1}^{n} x_{ji}[\, y_i - p_i \,]$$

$$= \frac{\partial}{\partial \beta_u} \sum_{i=1}^{n} \left[y_i x_{ji} - \frac{x_{ji}\ e^{\eta_i}}{1 + e^{\eta_i}} \right]$$

$$= -\sum_{i=1}^{n} \left[\frac{x_{ji} x_{ji}\ e^{\eta_i}(1 + e^{\eta_i}) - x_{ji} x_{ji}\ e^{\eta_i}{}^2}{(1 + e^{\eta_i})^2} \right]$$

berdasarkan Persamaan (3.5) diperoleh :

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\left(\partial \boldsymbol{\beta}_j\right)^2} = -\sum_{i=1}^{n} x_{ji}{}^2 [\, p_i(1-p_i)] \tag{4.12}$$

Jika dinyatakan dalam bentuk matriks adalah sebagai berikut :

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{bmatrix} \cdot \begin{bmatrix} p_1(1-p_i) & 0 & \cdots & 0 \\ 0 & p_2(1-p_i) & \cdots & 0 \\ \vdots \quad \vdots & & \ddots & \vdots \\ 0 \quad 0 & & \cdots & p_n(1-p_i) \end{bmatrix} \cdot$$

$$\begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix} = \mathbf{X}^T\mathbf{V}\mathbf{X}$$

Pendugaan parameter $\boldsymbol{\beta}$ dengan metode iterasi Newton-Raphson :

a. Dipilih taksiran awal untuk $\boldsymbol{\beta}$ misal $\hat{\beta} = 0$.

b. Dihitung $\mathbf{X}^T(\mathbf{Y} - \mathbf{P_i})$ dan $\mathbf{X}^T\mathbf{V}\mathbf{X}$ selanjutnya dihitung invers dari $\mathbf{X}^T\mathbf{V}\mathbf{X}$.

c. Pada setiap $i+1$ dihitung taksiran baru yaitu $\hat{\beta}_{i+1} = \hat{\beta}_i + \{\mathbf{X}^T\mathbf{V}\mathbf{X}\}^{-1}\{\mathbf{X}^T(\mathbf{Y} - \mathbf{P_i})\}$.

d. Iterasi berakhir jika diperoleh $\hat{\beta}_{i+1} \cong \hat{\beta}_i$.

5. Interpretasi Koefisien Regresi Logistik

Misalkan diketahui model regresi logistik dengan k variabel prediktor sebagai berikut :

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} \qquad ; \; i = 1,2,\ldots,n$$

Persamaan ini juga dapat ditulis menjadi :

$$\frac{p_i}{1-p_i} = \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}\ ) \tag{5.1}$$

Ruas kiri dari Persamaan (5.1) di atas merupakan perbandingan antara peluang berhasil p_i dengan peluang gagal $1 - p_i$ yang disebut *odds*. Sedangkan perbandingan nilai *odds* antara dua individu disebut *odds ratio* [15], yang dinotasikan :

$$\theta = \left[\frac{p_i(1)/(1 - p_i(1))}{p_i(0)/(\,1 - p_i(0))}\right]$$

$$= \frac{\exp(\beta_0 + \ \beta_1.(1) + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}\ )}{\exp(\beta_0 + \ \beta_1.(0) + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}\ )}$$

$$= \frac{\exp\big(\beta_0 + \ \beta_1.(1)\big)\exp(\beta_2 X_{2i} + \cdots + \beta_k X_{ki})}{\exp\big(\beta_0 + \ \beta_1.(0)\big)\exp(\beta_2 X_{2i} + \cdots + \beta_k X_{ki})}$$

$$= \exp(\beta_1)$$

Artinya, *odds* (resiko) terjadinya $Y = 1$ pada kategori $X = 1$ adalah sebesar $\exp(\beta_1)$ kali *odds* (resiko) terjadinya $Y = 1$ pada kategori $X = 0$.

Jika variabel prediktor merupakan variabel kategorik dengan lebih dari dua kategori (polikotomi), maka interpretasinya dilakukan dengan cara yang sama dengan interpretasi variabel dikotomi, hanya saja perlu dibentuk variabel boneka (dummy) terlebih dahulu.

Sementara untuk variabel prediktor kontinu, interpretasinya adalah setiap kenaikan nilai X sebesar satu satuan (unit) akan mengakibatkan perubahan nilai *odds* (resiko) terjadinya $Y = 1$ sebesar exp (β_1) kali.

6. Keakuratan Model

Setelah model peluang logit terbaik diperoleh, maka pengelompokkan suatu objek dapat dilakukan. Pengelompokkan suatu objek dilakukan berdasarkan pada nilai peluang logistiknya. Pengelompokkan objek tersebut dilakukan dengan kriteria sebagai berikut:

1. Alokasikan objek tersebut ke dalam kelompok 1 ($Y = 1$), jika $\hat{p}_i > \frac{1}{2}$
2. Alokasikan objek tersebut ke dalam kelompok 2 ($Y = 0$), jika $\hat{p}_i \leq \frac{1}{2}$

Setelah dilakukan pengelompokkan objek, selanjutnya diperlukan suatu ukuran yang akan mengukur seberapa akurat model yang terbentuk. Keakuratan model yang terbentuk ditentukan oleh suatu ukuran yang dinamakan *hit ratio*, yang didefinisikan sebagai berikut:

$$hit\ ratio = \frac{\text{banyak objek yang diklasifikas dengan benar}}{\text{total banyak pengamatan}} \times 100\% \qquad (6.1)$$

7. Keberartian Model dan Koefisien Regresi Logistik Biner

7.1 Statistik Uji G

Dalam regresi linier biasa untuk menguji hipotesis tentang pengaruh koefisien regresi $\boldsymbol{\beta}$ secara bersama digunakan uji F, sementara dalam regresi logistik digunakan uji perbandingan *likelihood* (*ratio test*) atau disebut juga statistik uji G. Hipotesis pada uji ini yaitu :

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$

$H_1 :$ ada $\beta_j \neq 0$; $j = 1,2, \dots, k$.

Rumus untuk statistik uji G adalah [15] :

$$G = -2 \ln L_0 - (-2 \ln L_k)$$
$$= -2 \ln \left(\frac{L_0}{L_k}\right)$$

dengan : L_k = likelihood model yang terdiri dari k peubah.

L_0 = likelihood model yang hanya terdiri atas β_0.

Statistik uji G mengikuti sebaran khi-kuadrat (χ^2) dengan derajat bebas k dengan aturan pengambilan keputusan tolak H_0 , yaitu :

$G > \chi^2_{k,\ db}$, maka H_0 ditolak.

$G \leq \chi^2_{k,\ db}$, maka H_0 tidak ditolak.

7.2 Statistik Uji Wald

Statistik uji Wald pada regresi logistik analog dengan uji t pada regresi linier biasa yang bertujuan untuk menguji pengaruh koefisien regresi $\boldsymbol{\beta}$ pada model secara terpisah, dengan hipotesis sebagai berikut :

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$; $j = 1,2, \dots , k$

dengan rumus umum untuk uji Wald adalah [11]:

$$W_j = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2 \quad ; \quad j = 1,2, \dots , k$$

dengan : $\hat{\beta}_j$ = nilai koefiesien regresi ke-j

$SE(\hat{\beta}_j)$ = nilai galat baku (*standard error*) koefisien regresi ke-j

Statistik uji ini berdistribusi khi-kuadrat (χ^2) dengan derajat bebas 1, dengan kesimpulan tolak H_0 jika :

$W_j > \chi^2_{\alpha,\ 1}$, maka H_0 ditolak.

$W_j \leq \chi^2_{\alpha,\ 1}$, maka H_0 tidak ditolak.

7.3 $\mathbf{R^2}$ (*R Square*)

Ukuran kebaikan model pada regresi linier ditentukan dengan menggunakan R^2, dimana:

$$R^2 = \frac{JK_{total} - JK_{sisaan}}{JK_{total}} = \frac{JK_{regresi}}{JK_{total}}$$

Jumlah kuadrat (JK) sisaan pada regresi linier analog dengan devians pada regresi logistik [14]. Devians merupakan model log likelihood dari k variabel dikali-2, yaitu :

$$D = -2 \ln L_k$$

sedangkan JK total analog dengan $-2\ \ln L_0$, sehingga JK regresi pada regresi linier biasa dapat dikatakan analog dengan statistik uji G yaitu :

$$G = -2 \ln L_0 - (-2 \ln L_k)$$

$$= -2 \ln \left(\frac{L_0}{L_k}\right)$$

dengan demikian, R^2 pada regresi logistik yang analog dengan R^2 pada regresi linier biasa adalah :

$$R^2 = \frac{-2 \ln L_0 - (-2 \ln L_k)}{-2 \ln L_0}$$

$$= 1 - \ln \left(\frac{L_k}{L_0}\right).$$

Bentuk lain dari R^2 untuk model regresi logistik yaitu :

$$R^2 = 1 - \left\{\ln \left(\frac{L_0}{L_k}\right)\right\}^{2/n}$$

Sementara bentuk modifikasi dari R^2 diatas, seperti berikut :

$$\bar{R}^2 = \frac{R^2}{\max(R^2)}$$

dengan max $(R^2) = 1 - (\ln L_0)^{2/n}$ [6]. R^2 merupakan proporsi keragaman data yang dapat diterangkan oleh model. Jika R^2 semakin mendekati 1, maka semakin baik model yang diperoleh.

STUDI KASUS

Suatu penelitian dilakukan untuk mengetahui seberapa besar peluang seseorang dapat diterima bekerja di PT Makmur Jaya. Penelitian dilakukan dengan mengambil sampel sebesar 40 orang dan menggunakan analisis regresi logistik. Adapun variabel yang diteliti adalah sebagai berikut.

y (keputusan) = 1 jika diterima dan 0 jika tidak diterima

x1 = lama pendidikan terakhir (tahun)

x2 = lama pengalaman kerja (tahun)

x3 (jenis kelamin) = 1 jika pelamarnya laki-laki dan 0 jika pelamarnya perempuan

Pertanyaan:

1. Carilah Model logit dan model regresi logistik dari kasus tersebut.
2. Interpretasikan nilai estimasi β (dengan menggunakan nilai odd dari masing-masing variabel)
3. Berapa peluang diterimanya seorang wanita yang ingin melamar pekerjaan di PT Makmur Jaya, jika diketahui memiliki lama pendidikan 4 tahun dan pengalaman bekerja 1 tahun ?

Data 40 orang yang melamar pekerjaan

Pelamar	Education (X1)	Experience (X2)	SEX (X3)	HIRED (Y)
1	6	6	1	1
2	6	3	1	1
3	8	3	0	1
4	8	10	0	1
5	4	5	1	1
6	6	1	1	1
7	8	5	1	1
8	4	10	1	1
9	6	12	0	1
10	6	2	0	1
11	4	0	1	0

12	4	1	0	0
13	4	2	1	0
14	4	4	0	0
15	6	1	0	0
16	4	2	1	0
17	8	5	1	0
18	4	2	0	0
19	6	7	0	0
20	6	4	0	0
21	8	0	1	0
22	4	0	0	0
23	4	1	1	0
24	4	5	1	0
25	6	0	1	0
26	4	9	0	0
27	8	1	0	0
28	6	1	1	0
29	6	6	1	1
30	6	3	1	1
31	8	3	0	1
32	8	10	0	1
33	4	5	1	1
34	6	1	1	1
35	8	5	1	1
36	4	10	1	1
37	6	12	0	1
38	6	2	0	1
39	4	0	1	0
40	4	10	1	0