

BAB II

LANDASAN TEORI

Pada bab ini akan dibahas mengenai regresi logistik, kesamaan sifatnya dengan regresi berganda, bagaimana cara menduga parameter dan kesesuaian model regresi logistik biner. Sebagai awalan, akan dijelaskan mengenai distribusi binomial dan model linear umum (*General Linear Models*). Uji *Pearson* serta *rasio odds* juga akan dibahas pada bab ini.

Apabila nilai-nilai dugaan peubah respon relatif sesuai dengan nilai-nilai amatan maka model dugaan dapat dikatakan layak tetapi jika perbedaan antara nilai dugaan dengan nilai yang diamati sangat besar maka model dugaan perlu untuk ditinjau kembali.

2.1 Distribusi Binomial

Jika peubah acak bernoulli diamati pada n percobaan dimana antara setiap percobaan bersifat saling bebas dan peluang munculnya kejadian sukses pada setiap percobaan bersifat konstan yaitu p , maka banyaknya kejadian sukses X diantara n percobaan akan berdistribusi binomial dengan parameter n dan p serta dapat dinotasikan dengan $X \sim \text{Binomial}(n, p)$. Fungsi kepadatan peluang untuk X adalah :

$$f(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & ; \quad x = 0, 1, \dots, n \\ 0 & ; \quad \text{selainnya} \end{cases} \quad (2.1)$$

Fungsi pembangkit momen untuk X adalah :

$$\begin{aligned}
 M_X(t) &= E(e^{tx}) \\
 &= \sum_{x=0}^n e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\
 &= \sum_{x=0}^n \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\
 &= (pe^t + q)^n
 \end{aligned} \tag{2.2}$$

Turunan pertama fungsi pembangkit momen dari Persamaan (2.2) adalah

$$\begin{aligned}
 M'_X(t) &= \frac{d}{dt} M_X(t) \\
 &= \frac{d}{dt} (pe^t + q)^n \\
 &= n(pe^t + q)^{n-1} pe^t
 \end{aligned} \tag{2.3}$$

Subtitusikan $t = 0$ dari Persamaan (2.3), maka diperoleh nilai harapan dari peubah acak X tersebut, yaitu

$$\begin{aligned}
 E(X) &= \frac{d}{dt} M_X(t) |_{t=0} \\
 &= n(pe^0 + q)^{n-1} pe^0 \\
 &= n(p + q)^{n-1} p \\
 &= np
 \end{aligned} \tag{2.4}$$

Turunan kedua fungsi pembangkit momen dari Persamaan (2.2) adalah

$$\begin{aligned}
 \frac{d^2}{dt^2} M_X(t) &= \frac{d}{dt} n(pe^t + q)^{n-1} pe^t \\
 &= n(n-1)(pe^t + q)^{n-2} (pe^t)^2 + n(pe^t + q)^{n-1} pe^t
 \end{aligned} \tag{2.5}$$

saat $t = 0$, maka dari Persamaan (2.5) akan diperoleh

$$\begin{aligned}
 E(X^2) &= \frac{d^2}{dt^2} M_X(t) \\
 &= n(n-1)(pe^0 + q)^{n-2}(pe^0)^2 + n(pe^0 + q)^{n-1}pe^0 \\
 &= n(n-1)(p+q)^{n-2}(p)^2 + n(p+q)^{n-1}p \\
 &= n(n-1)(p)^2 + np
 \end{aligned} \tag{2.6}$$

dari persamaan (2.6) diperoleh ragam dari X , yaitu

$$\begin{aligned}
 Var(X) &= E(X^2) - [E(X)]^2 \\
 &= n(n-1)(p)^2 + np - [np]^2 \\
 &= np - np^2 \\
 &= np(1-p)
 \end{aligned} \tag{2.7}$$

2.2 Analisis Regresi Linier Berganda

Analisis regresi linier berganda dipergunakan untuk mengetahui pengaruh antara satu peubah respon dengan banyak atau lebih dari satu peubah penjelas. Menurut Suyono (2015:26) untuk pengujian hipotesis dan untuk membuat interval konfidensi parameter galat-galat acak diasumsikan sebagai berikut :

1. Nilai rata-rata dari galat adalah nol

Dalam analisis regresi linier jumlah residual sama dengan 0 karena

$$\begin{aligned}
\sum_{i=1}^n e_i &= \sum_{i=1}^n (Y_i - b_0 - b_1 X_i) \\
&= \sum_{i=1}^n Y_i - nb_0 - b_1 \sum_{i=1}^n X_i \\
&= \sum_{i=1}^n Y_i - n(\bar{Y} - b_1 \bar{X}) - b_1 \sum_{i=1}^n X_i \\
&= \sum_{i=1}^n Y_i - \sum_{i=1}^n Y_i + b_1 \sum_{i=1}^n X_i - b_1 \sum_{i=1}^n X_i \\
&= 0
\end{aligned}$$

maka rata-rata residual juga sama dengan 0. Jadi, seolah-olah asumsi galat acak sama dengan 0 selalu terpenuhi. Namun perlu diperhatikan sebaran dari residual apakah memiliki pola dan pencilan atau tidak

2. Variansi dari galat adalah konstan (homoskedastik)

Variansi dari galat diasumsikan homogen karena jika tidak akan menyulitkan dalam mengukur standar deviasi yang benar dari prediksi galat (interval konfidensi terlalu lebar atau terlalu sempit) serta berimplikasi pada sulitnya memperoleh kesimpulan yang benar dalam uji hipotesis karena ada keterkaitan antara interval konfidensi dan uji hipotesis.

3. Tidak terjadi autokorelasi pada galat

Asumsi galat acak tidak berkorelasi dapat dicek atau diuji dengan menggunakan residual. Dalam prakteknya kita masih dapat berharap asumsi ini pada taraf signifikansi tertentu.

4. Galat berdistribusi normal

Jika asumsi galat berdistribusi normal terpenuhi, maka kita dapat mengestimasi parameter dengan menggunakan metode maksimum likelihood juga memudahkan dalam menguji hipotesis dan membentuk interval konfidensi.

Persamaan umum dari analisis regresi linier berganda sebagai berikut :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i \quad (2.8)$$

dimana Y =Variabel respon ; $X_1, X_2, \dots, X_k$ = Variabel penjelas ; β_0 = konstanta atau Intersep ; $\beta_1, \beta_2, \dots, \beta_k$ =Koefisien dengan ε_i sisa(*error*) untuk pengamatan ke- i yang diasumsikan berdistribusi normal yang saling bebas dan identik dengan rata-rata 0 (nol) dan variansi σ^2 .

Persamaan (2.8) dapat ditulis sebagai berikut

$$Y = X\beta + \varepsilon \quad (2.9)$$

dengan notasi matriks

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}; X = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1k} \\ 1 & X_{21} & \cdots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{nk} \end{pmatrix}; \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}; \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

dengan Y adalah vektor variabel respon $n \times 1$. X adalah matriks variabel penjelas berukuran $n \times (k+1)$ yang merupakan matriks non singular. β adalah vektor parameter berukuran $k \times 1$ dan ε adalah vektor error berukuran $n \times 1$. Dalam pengoperasian matriks, Suatu matriks dapat dibalik jika dan hanya jika matriks tersebut merupakan matriks persegi (matriks berukuran $n \times n$) dan matriks tersebut merupakan matriks non singular (determinan $\neq 0$)

Model regresi linier berganda dapat diperoleh dengan melakukan estimasi terhadap parameter-parameternya menggunakan metode kuadrat terkecil (MKT) (ordinary least square/OLS) dan metode kemungkinan maksimum (MKM) (maximum likelihood estimation/MLE)

Metode MKT ini bertujuan meminimumkan jumlah kuadrat *error*. Jika *error* diasumsikan saling bebas dan mempunyai mean 0 serta variansi σ^2 , maka

penduga yang dihasilkan dari metode MKT merupakan penduga yang tak bias, linier dan terbaik / *BLUE (Best Linier Unbiased Estimator)*. Berdasarkan Persamaan (2.9) dapat diperoleh penaksir (estimator) OLS untuk β adalah sebagai berikut (Kutner, et.al., 2004:223)

1. Bentuk Jumlah Kuadrat galat/*error* (JKG) sebagai fungsi dari β

$$\begin{aligned} JKG &= \varepsilon^T \varepsilon \\ &= (Y - X\beta)^T (Y - X\beta) \\ &= (Y^T - \beta^T X^T)(Y - X\beta) \end{aligned}$$

2. Diferensialkan JKG terhadap β dan setarakan dengan 0

$$\begin{aligned} \frac{\partial(JKG)}{\partial\beta} &= -X^T(Y\beta - X\beta) - (Y^T - \beta^T X^T)X \\ 0 &= -X^T Y + X^T X\beta - Y^T X + \beta^T X^T X \\ X^T X\beta + \beta^T X^T X &= X^T Y + Y^T X \\ X^T X\beta + (\beta^T X^T X)^T &= X^T Y + (Y^T X)^T \\ 2X^T X\beta &= 2X^T Y \\ X^T X\beta &= X^T Y \\ \hat{\beta} &= (X^T X)^{-1} X^T Y \end{aligned}$$

Jadi, Penduga untuk β adalah

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (2.10)$$

Metode kemungkinan maksimum (MKM) menunjukkan estimator yang sama terhadap Persamaan (2.8) sebagaimana yang telah ditunjukkan oleh persamaan (2.10). Fungsi kemungkinan maksimum untuk n pengamatan $Y_1, Y_2, \dots, Y_n$ dengan varians σ^2 maka fungsi kemungkinan maksimum terdiri dari 3 para-

meter β_0, β_1 , dan σ^2 :

$$\begin{aligned} L(\beta_0, \beta_1, \sigma^2) &= \prod_{i=1}^n \frac{1}{(2\pi\sigma^2)^{1/2}} \exp \left[-\frac{1}{2\sigma^2} (Y_i - \beta_0 - \beta_1 X_i)^2 \right] \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \right] \end{aligned} \quad (2.11)$$

Persamaan (2.11) saat di generalisasikan ke persamaan analisis regresi linier berganda maka akan menjadi :

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} - \dots - \beta_k X_{ik})^2 \right] \quad (2.12)$$

Memaksimalkan fungsi (2.12) terhadap $\beta_0, \beta_1, \dots, \beta_k$, maka akan ditemui Persamaan (2.10). Estimator (2.10) ialah estimator bagi metode kuadrat terkecil maupun metode kemungkinan maksimum. Estimator tersebut memiliki kriteria yaitu : Varians minimumnya tak bias, konsisten, dan cukup.

2.3 Model Linear Umum

Model linear umum (*Generalized Linear Model*) ialah pengembangan dari model linear klasik yang variabel responnya saling bebas. Perbedaan antara model linier klasik dengan model linier umum, yaitu pada model linier klasik komponen acaknya menyebar normal dan fungsi penghubungnya merupakan fungsi identitas sehingga $\eta_i = \mu_i$, sedangkan pada model linear umum komponen acaknya mungkin berasal dari salah satu anggota sebaran keluarga eksponensial dan fungsi penghubungnya merupakan fungsi g yang bersifat monoton dan dapat diturunkan (*differentiable*). Model linear umum diasumsikan mengikuti distribusi keluarga eksponensial yang merupakan distribusi dengan sifat yang lebih umum. Model linear umum dapat dituliskan sebagai berikut :

$$g(\mu_i) = \eta_i \quad (2.13)$$

dengan nilai dari

$$\eta_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi} \quad (2.14)$$

Menurut (Gill J, 2001:31) Model Linear umum memiliki tiga komponen, yaitu :

1. Komponen acak, yaitu komponen X_i , $i = 1, 2, \dots, n$ merupakan variabel acak yang distribusinya masuk ke dalam distribusi keluarga eksponensial dengan rata-rata μ_i , misal $\mu_i = E(Y_i)$, $i=1,2,\dots,N$.
2. Komponen sistematis yang menunjukkan fungsi linier dari variabel penjelas. Misal X_{ij} menyatakan nilai penduga j dimana ($j=1,2,\dots,p$) untuk subjek i .

$$\eta_i = \sum_j X_{ij} \beta_j; i=1,2,\dots,N$$

atau secara umum dapat dituliskan dengan

$$\eta_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi}$$

β adalah sebuah vektor dari parameter model dan X merupakan model matriks yang mencakup nilai-nilai variabel penjelas untuk amatan sebanyak N . $X_{ij} = 1$ untuk koefisien *intercept* pada model untuk tiap- i .

3. Fungsi penghubung, yaitu suatu fungsi g yang menghubungkan komponen acak dan komponen sistematis sehingga $E(Y_i) = \mu_i = g^{-1}(\eta)$.

2.4 Analisis Regresi Logistik Biner

Analisis regresi logistik adalah analisis yang digunakan untuk menganalisis hubungan antara peubah respon (Y) yang bersifat kategorik atau numerik. Apabila untuk setiap objek amatan ke- i , $i = 1, 2, \dots, n$ kategori dari peubah respon hanya memiliki dua kemungkinan nilai (misalnya $Y_i = 1$ untuk kejadian

sukses dan $Y_i = 0$ untuk kejadian gagal) dan Y_i diasumsikan berdistribusi binomial dengan parameter n_i dan p_i maka analisis regresi logistik tersebut dinamakan juga sebagai analisis regresi logistik biner. Jika $Pr(Y_i = 1) = p_i$ dan $Pr(Y_i = 0) = 1 - p_i$ maka mean untuk Y_i adalah $n_i p_i$.

Pada analisis regresi linier dimana Y_i bertipe kontinu, hubungan antara mean dari peubah respon Y_i dengan sejumlah peubah penjelas yaitu $X = (X_{1i}, X_{2i}, \dots, X_{ki})$ diasumsikan bersifat linier, atau dapat dituliskan sebagai berikut

$$E(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} \quad (2.15)$$

Pada analisis regresi logistik biner, pemodelan mengenai bagaimana peluang kejadian sukses p_i dapat dijelaskan oleh sejumlah peubah penjelas, relatif lebih bermakna dibandingkan pemodelan antara nilai-nilai Y_i dengan peubah penjelas seperti halnya pada analisis regresi linier. Namun demikian, karena nilai p_i adalah peluang terjadinya kejadian sukses yang nilainya terbatas pada interval $(0,1)$, padahal nilai dugaan mungkin berada diluar interval tersebut. Oleh karena itu dilakukan transformasi terhadap p_i ke dalam suatu fungsi yang memiliki nilai tak terbatas $(-\infty, +\infty)$. Transformasi logit terhadap p_i didefinisikan sebagai

$$\text{logit}(p_i) = \log \left(\frac{p_i}{1 - p_i} \right)$$

Model regresi logistik biner yang menunjukkan hubungan linier antara nilai-nilai dari transformasi logit dengan peubah penjelasnya diberikan oleh :

$$\log \left(\frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} \quad (2.16)$$

Berdasarkan Persamaan (2.16), maka diperoleh p_i sebagai berikut

$$p_i = \frac{\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})}{1 + \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})} \quad (2.17)$$

Pada Persamaan (2.17), parameter $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ merupakan parameter-parameter regresi dan $X_{1i}, X_{2i}, \dots, X_{ki}$ adalah peubah-peubah penjelas, sekaligus menyatakan bahwa hubungan antara p_i dengan peubah bebas tidak bersifat linear.

2.4.1 Pendugaan Parameter Model Regresi Logistik

Biner

Metode kuadrat terkecil (MKT) tidak dapat digunakan untuk menduga parameter pada regresi logistik biner karena dapat mengakibatkan diperolehnya taksiran yang terlalu rendah (*underestimate*) atau taksiran yang terlalu tinggi (*overestimate*) yang akan mempengaruhi perhitungan interval kepercayaan. Pada regresi logistik biner, Fungsi likelihood dari peubah acak $Y_i \sim \text{Binomial}(n_i, p_i)$, diberikan oleh

$$L(p) = \prod_{i=1}^n \binom{n_i}{Y_i} p_i^{Y_i} (1 - p_i)^{n_i - Y_i} \quad (2.18)$$

Logaritma untuk fungsi likelihood pada Persamaan (2.18) diberikan oleh ;

$$\begin{aligned} \ln L(p) &= \sum_{i=1}^n \ln \left\{ \binom{n_i}{Y_i} p_i^{Y_i} (1 - p_i)^{n_i - Y_i} \right\} \\ &= \sum_{i=1}^n \left\{ Y_i \ln p_i + (n_i - Y_i) \ln(1 - p_i) + \ln \binom{n_i}{Y_i} \right\} \\ &= \sum_{i=1}^n \left\{ Y_i \ln \left(\frac{p_i}{1 - p_i} \right) + n_i \ln(1 - p_i) + \ln \binom{n_i}{Y_i} \right\} \end{aligned}$$

karena

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

maka diperoleh

$$\begin{aligned}
 \ln L(\beta) &= \sum_{i=1}^n \left\{ Y_i (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}) + \right. \\
 &\quad \left. n_i \ln \left(\frac{1}{1 + (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})} \right) + \ln \left(\begin{matrix} n_i \\ Y_i \end{matrix} \right) \right\} \\
 &= \sum_{i=1}^n \left\{ Y_i (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}) \right. \\
 &\quad \left. - n_i \ln (1 + \exp (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})) + \ln \left(\begin{matrix} n_i \\ Y_i \end{matrix} \right) \right\}
 \end{aligned}$$

Turunan pertama terhadap parameter-parameter regresi tersebut adalah

$$\begin{aligned}
 \frac{\partial \ln L(\beta)}{\partial \beta_0} &= \sum_{i=1}^n \left\{ Y_i - \frac{n_i \exp (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})}{1 + \exp (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})} \right\} \\
 &= \sum_{i=1}^n \{Y_i - n_i p_i\}
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial \ln L(\beta)}{\partial \beta_1} &= \sum_{i=1}^n \left\{ X_{1i} Y_i - \frac{X_{1i} n_i \exp (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})}{1 + \exp (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})} \right\} \\
 &= \sum_{i=1}^n \{X_{1i} Y_i - X_{1i} n_i p_i\} \\
 &= \sum_{i=1}^n X_{1i} \{Y_i - n_i p_i\}
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial \ln L(\beta)}{\partial \beta_k} &= \sum_{i=1}^n \left\{ X_{ki} Y_i - \frac{X_{ki} n_i \exp (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})}{1 + \exp (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})} \right\} \\
 &= \sum_{i=1}^n \{X_{ki} Y_i - X_{ki} n_i p_i\} \\
 &= \sum_{i=1}^n X_{ki} \{Y_i - n_i p_i\}
 \end{aligned}$$

Persamaan likelihood tersebut tidak linier pada parameter $(\beta_0, \beta_1, \beta_2, \dots, \beta_k)$, oleh karena itu diperlukan metode numerik untuk mencari dugaan $\hat{\beta}$ menggunakan metode Newton-Raphson. Penggunaan metode Newton-Raphson membutuhkan turunan parsial kedua dari fungsi $L(\beta)$.

Turunan pertama dari $\ln L(\beta)$ terhadap β_r , $r \leq k$ adalah :

$$\frac{\partial \ln L(\beta)}{\partial \beta_r} = \sum_{i=1}^n X_{ri} \{Y_i - n_i p_i\} \quad (2.19)$$

dimana $r=0,1,\dots,k$

Bentuk matriks yang menunjukkan turunan pertama $\ln L(\beta)$ yaitu

$$U' = \begin{bmatrix} 1 & X_{11}(Y_1 - n_1 p_1) & \cdots & X_{1n}(Y_n - n_n p_n) \\ 1 & X_{21}(Y_1 - n_1 p_1) & \cdots & X_{2n}(Y_n - n_n p_n) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1}(Y_1 - n_1 p_1) & \cdots & X_{nn}(Y_n - n_n p_n) \end{bmatrix}$$

Turunan parsial kedua dari Persamaan (2.15) terhadap β_s , $s \leq k$ adalah :

$$\begin{aligned} \frac{\partial^2 \ln L(\beta)}{\partial \beta_r \partial \beta_s} &= - \sum_{i=1}^n \left\{ \frac{[X_{si} X_{ri} n_i \exp(\beta_0 + \dots + \beta_k X_{ki})][1 + \exp(\beta_0 + \dots + \beta_k X_{ki})]}{[1 + \exp(\beta_0 + \dots + \beta_k X_{ki})]^2} \right. \\ &\quad \left. - \frac{[X_{ri} n_i \exp(\beta_0 + \dots + \beta_k X_{ki})][X_{si} \exp(\beta_0 + \dots + \beta_k X_{ki})]}{[1 + \exp(\beta_0 + \dots + \beta_k X_{ki})]^2} \right\} \\ &= - \sum_{i=1}^n \left\{ \frac{X_{si} X_{ri} n_i [\exp(\beta_0 + \dots + \beta_k X_{ki})]}{1 + \exp(\beta_0 + \dots + \beta_k X_{ki})} \right. \\ &\quad \left. - \frac{X_{si} X_{ri} n_i [\exp(\beta_0 + \dots + \beta_k X_{ki})]^2}{[1 + \exp(\beta_0 + \dots + \beta_k X_{ki})]^2} \right\} \\ &= - \sum_{i=1}^n X_{si} X_{ri} n_i \left\{ \left[\frac{\exp(\beta_0 + \dots + \beta_k X_{ki})}{1 + \exp(\beta_0 + \dots + \beta_k X_{ki})} \right] \right. \\ &\quad \left. - \frac{[\exp(\beta_0 + \dots + \beta_k X_{ki})]^2}{[1 + \exp(\beta_0 + \dots + \beta_k X_{ki})]^2} \right\} \\ &= - \sum_{i=1}^n X_{si} X_{ri} n_i \left\{ \left[\frac{\exp(\beta_0 + \dots + \beta_k X_{ki})}{1 + \exp(\beta_0 + \dots + \beta_k X_{ki})} \right] \right. \\ &\quad \left. - \left[\frac{\exp(\beta_0 + \dots + \beta_k X_{ki})}{1 + \exp(\beta_0 + \dots + \beta_k X_{ki})} \right]^2 \right\} \end{aligned}$$

karena

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

maka

$$\begin{aligned}\frac{\partial^2 \ln L(\beta)}{\partial \beta_r \partial \beta_s} &= - \sum_{i=1}^n X_{si} X_{ri} n_i (p_i - p_i^2) \\ &= - \sum_{i=1}^n X_{si} X_{ri} n_i p_i (1 - p_i)\end{aligned}\quad (2.20)$$

untuk $r, s = 0, 1, \dots, k$

Bentuk matriks yang menunjukkan turunan parsial kedua dari logaritma fungsi likelihood terhadap β adalah

$$H = - \begin{bmatrix} 1 & \cdots & 1 \\ X_{11} & \cdots & X_{1n} \\ \vdots & \ddots & \vdots \\ X_{k1} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} n_1 p_1 (1 - p_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & n_n p_n (1 - p_n) \end{bmatrix} \begin{bmatrix} 1 & X_{11} & \cdots & X_{1k} \\ 1 & X_{21} & \cdots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{nk} \end{bmatrix}$$

atau

$$H = -X^T V X \quad (2.21)$$

dimana V adalah matriks diagonal dengan elemen pada diagonal utamanya yaitu $n_i p_i (1 - p_i)$ untuk $i = 1, 2, \dots, n$

Metode Newton Raphson adalah metode iterasi untuk menyelesaikan persamaan tak linier. Jika $U' = \left(\partial L(\beta) / \partial \beta_1, \partial L(\beta) / \partial \beta_2, \dots, \partial L(\beta) / \partial \beta_k \right)$ dan H ialah matriks *Hessian* yang memiliki entri-entri $h_{ab} = \left(\partial^2 L(\beta) / \partial \beta_r \partial \beta_s \right)$, maka menurut (Agresti, 2002:143) prosedur iterasi Newton Raphson untuk mencari nilai taksiran adalah sebagai berikut

1. Memilih taksiran awal $\widehat{\beta}^0$
2. Menentukan vektor gradien (U') yang beranggotakan turunan pertama dari fungsi likelihood terhadap β
3. membentuk matriks Hessian yang beranggotakan turunan kedua fungsi likelihood terhadap β

4. Pada setiap iterasi ke-(t+1), dengan t=0,1,.. dihitung dengan nilai dugaan baru yaitu $\hat{\beta}^{(t+1)}$ secara iteratif dengan menggunakan formula

$$\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} [H^{(t)}]^{-1} - U^{(t)}$$

5. Iterasi berlanjut hingga $\hat{\beta}^{(t+1)} \approx \hat{\beta}^{(t)}$, atau iterasi dihentikan jika selisih antara $\hat{\beta}^{(t+1)}$ dengan $\hat{\beta}^{(t)}$ sangat kecil.

Penduga untuk varians dan kovarians dari $\sum(\hat{\beta})$ diperoleh dari matrik varians-kovarians $\sum(\beta)$, dimana matriks varians kovarians tersebut merupakan invers matriks informasi $I(\beta)$. Matriks informasi $I(\beta)$ adalah negatif dari matriks turunan parsial kedua dari logaritma fungsi likelihood terhadap β (negatif matriks Hessian). Berdasarkan Persamaan (2.16), matriks informasi $I(\beta)$ diberikan oleh:

$$I(\beta) = -(-X^T V X) \quad (2.22)$$

$$= (X^T V X) \quad (2.23)$$

Dugaan untuk matriks $\sum(\hat{\beta})$, yaitu, $\widehat{\sum(\hat{\beta})}$ diperoleh dengan mengganti parameter-parameter yang tidak diketahui dengan penduga maksimum likelihood $\hat{\beta}_0, \dots, \hat{\beta}_k$. Elemen-elemen pada diagonal utama dari $\widehat{\sum(\hat{\beta})}$ adalah penduga dari varians $\hat{\beta}_j$ yang dinyatakan oleh $\widehat{\sigma^2(\beta)}$ sedangkan elemen-elemen lainnya merupakan penduga dari varians-kovarians antara $\hat{\beta}_j$ dan $\hat{\beta}_u$ yang dinyatakan oleh $\widehat{\sigma(\beta_j \beta_u)}$ dimana $j, u = i = 0, 1, \dots, k$. Penduga galat baku dari parameter $\hat{\beta}_i$ adalah

$$\widehat{SE(\beta_i)} = \sqrt{\widehat{\sigma^2(\beta_i)}} \quad (2.24)$$

Dugaan galat baku ini digunakan untuk menguji signifikansi dari suatu peubah bebas dalam menjalankan peluang sukses p_i .

2.4.2 Uji Signifikansi Parameter

Parameter yang telah diduga kemudian diuji signifikansinya baik secara simultan atau serempak dengan menggunakan uji perbandingan rasio kemungkinan atau *Likelihood Ratio Test (LRT)* dan uji parsial dengan menggunakan uji Wald.

1. Uji Serempak

Uji Serempak dilakukan untuk melihat pengaruh variabel penjelas secara bersama-sama terhadap variabel respon. Uji serempak melalui perbandingan rasio kemungkinan atau uji G memiliki hipotesis:

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \text{Tidak semua } \beta_i = 0$$

dimana

$i = 1, 2, \dots, k$ dengan k = banyaknya variabel bebas X dan

β_i = Parameter ke - i model regresi linier dengan statistik uji G

$$G = -2 \ln \left[\frac{L(\hat{p}_i)}{L(\tilde{p}_i)} \right] \sim \chi_k^2$$

dimana

$\hat{p}_i$: Fungsi likelihood model dugaan

$\tilde{p}_i$: Fungsi likelihood model penuh

Jika $G \geq \chi_k^2$ maka H_0 ditolak yang berarti peubah penjelas X berpengaruh nyata terhadap peubah respon Y .

2. Uji Parsial

Uji Parsial dilakukan untuk melihat peranan masing-masing variabel penjelas terhadap variabel respon. Uji signifikan parameter parsial dengan menggunakan Uji Wald. Hipotesis yang akan diuji :

$H_0 : \beta_i = 0$ (koefisien β_i tidak signifikan secara statistik)

$H_1 : \beta_i \neq 0$ (koefisien β_i signifikan secara statistik)

dimana

$i = 1, 2, \dots, k$ dengan $k =$ banyaknya variabel bebas X dan

$\beta_i =$ Parameter ke - i model regresi linier

Statistik Uji Wald :

$$W = \frac{(\hat{\beta}_i)}{se(\hat{\beta}_i)} \quad (2.25)$$

dengan :

$se(\hat{\beta}_i)$: Galat baku penduga β_i terdapat pada (2.24)

$\hat{\beta}_i$: koefisien dugaan peubah prediktor

$i : 0, 1, 2, \dots, n$ Jika ukuran contoh besar ($n \geq 30$) maka statistik uji *Wald* mengikuti sebaran normal dengan kriteria keputusan menerima H_0 jika $W \leq Z$ berarti β_i tidak signifikan dan dapat disimpulkan bahwa peubah penjelas β_i tidak berpengaruh terhadap peubah respon. Di bawah ini akan diberikan contoh mengenai uji signifikansi parameter dengan menggunakan uji Parsial Wald.

Contoh

Diberikan hasil dari fitting model regresi logistik mengenai data 100 responden penderita jantung koroner berdasarkan umur penderita. Variabel respon merupakan banyaknya penderita jantung koroner yang dipengaruhi oleh variabel penjelas umur penderita, dimana tingkatan umur

yang berbeda-beda dari 100 responden memperlihatkan penyakit jantung koroner menginfeksi seluruh tingkatan umur. (Hosmer dan Lemeshow, 2000 : 10

Tabel 2.1: Hasil Fitting Model Regresi Logistik Data 100 Responden Penderita Jantung Koroner Berdasarkan Umur Penderita

Variable	Coeff.	Std.Err.	z	$P > z $
AGE	0.111	0.0241	4.61	< 0.001
Constant	-5.309	1.1337	-4.68	< 0.001

Sumber: Hosmer dan Lemeshow, 2000

Uji Parsial Wald membandingkan estimasi nilai maksimum likelihood dengan parameter $\hat{\beta}_i$ dengan estimasi nilai standard error atau galat baku penduga parameter. Berdasarkan Tabel (2.1) dapat ditentukan hipotesis dari kasus tersebut yaitu sebagai berikut.

$H_0 : \beta_i = 0$ (variabel umur tidak mempengaruhi terinfeksi jantung koroner)

$H_1 : \beta_i \neq 0$ (variabel umur mempengaruhi terinfeksi jantung koroner)

Maka dapat diperoleh perhitungan statistik Uji Wald yaitu

$$W = \frac{(\hat{\beta}_i)}{se(\hat{\beta}_i)} = \frac{0.111}{0.024} = 4.61$$

dari hasil perhitungan tersebut didapatkan bahwa nilai dari W sebesar 4.61 dan nilai tersebut lebih kecil dari nilai Z maka disimpulkan bahwa tolak H_0 sehingga variabel penjelas mempengaruhi variabel respon dimana variabel umur 100 responden mempengaruhi terinfeksi jantung koroner.

2.5 Kesesuaian Model Logistik Biner

Kesesuaian antara dugaan peubah respon terhadap nilai amatan merupakan hal yang menjadi perhatian selain pembentukan model dugaan terhadap data. Apabila nilai-nilai dugaan peubah respon relatif sesuai dengan nilai-nilai amatan maka model dugaan dapat dikatakan layak tetapi jika perbedaan antara nilai dugaan dengan nilai yang diamati sangat besar maka model dugaan perlu untuk ditinjau kembali. Pada analisis regresi logistik biner, pengujian kesesuaian model dugaan didasarkan pada statistik uji chi kuadrat *Pearson* atau statistik *Pearson*.

Statistik uji *Pearson* merupakan fungsi dari sisaan, yaitu selisih dari nilai aktual dengan nilai dugaan. Untuk suatu peubah bebas tertentu, nilai sisaan *Pearson* untuk amatan ke- i didefinisikan sebagai berikut :

$$r(y_i, \hat{p}_i) = \frac{(y_i - n_i \hat{p}_i)}{\sqrt{n_i \hat{p}_i (1 - \hat{p}_i)}} \quad (2.26)$$

sehingga chi kuadrat *Pearson* dapat dinyatakan sebagai berikut:

$$\chi^2 = \sum_{i=1}^n r(y_i, \hat{p}_i)^2 \quad (2.27)$$

Chi kuadrat *Pearson* mengikuti sebaran χ^2 dengan derajat bebas $(n-p)$, dengan p adalah banyaknya parameter dalam model yang diduga. Jika model regresi logistik yang digunakan terhadap data layak, maka nilai chi kuadrat *Pearson* akan mendekati nilai derajat bebasnya. Hal ini dapat dijelaskan kaena nilai harapan dari sebaran χ^2 sama dengan derajat bebasnya. Jika nilai chi kuadrat *Pearson* jauh lebih besar dari derajat bebasnya, maka asumsi dari keragaman binom tidak terpenuhi.

Akaike's Information Criterion (AIC) atau Kriteria Informasi Akaike ialah ukuran relatif kebaikan untuk mengepas dari model statistik dengan membandingkan antara akurasi dan kompleksitas model. Berdasarkan data set, beberapa

model kandidat diurutkan menurut nilai-nilai AIC mereka. Dapat diambil suatu kesimpulan, misalnya, dua model teratas adalah model yang tepat dan sisanya adalah model yang tidak tepat jadi AIC merupakan sarana perbandingan antar model. AIC mampu menunjukkan seberapa tepat model tersebut dengan data yang ada namun jika semua calon model tersebut tidak tepat, AIC tidak memberikan peringatan apapun. Rumus umum AIC, (Grasa,1989:71), yaitu :

$$AIC(k) = -2 \ln(\tilde{\theta}) + 2k \quad (2.28)$$

dimana k = Jumlah parameter dalam model statistik dan $\tilde{\theta}$ = nilai maksimal dari fungsi kemungkinan maksimum untuk model dugaan.

2.6 Rasio Odds

Rasio odds merupakan salah satu parameter yang sering menjadi perhatian dalam analisis data biner yang membandingkan dua kelompok pada respon biner. Menurut Hosmer dan Lemeshow (2000:50), rasio odds merupakan parameter pada analisis regresi logistik yang interpretasinya lebih mudah untuk dipahami maknanya dibandingkan koefisien regresi (β). Pada analisis regresi logistik dengan peubah bebas ($X_j, j=1,2,\dots,k$) terdiri dari dua kemungkinan nilai ($X_{ji}=1$ dan $X_{ji}=0$, untuk $i=1,2,\dots,n$) maka rasio odds didefinisikan sebagai ukuran seberapa besar kemungkinan munculnya kejadian sukses ($Y=1$) pada individu dengan nilai $X_{ji}=1$ dibandingkan dengan munculnya kejadian sukses tersebut pada individu yang mempunyai nilai $X_{ji}=0$. Odds dari kejadian sukses ($Y=1$) pada individu $X_{ji}=1$ adalah

$$\frac{P[Y = 1 | X_{ji} = 1]}{P[Y = 0 | X_{ji} = 1]} = \frac{p_{(1)}}{[1 - p_{(1)}]}$$

sedangkan odds kejadian sukses ($Y=1$) pada individu $X_{ji}=0$ didefinisikan sebagai

$$\frac{P[Y = 1 | X_{ji} = 0]}{P[Y = 0 | X_{ji} = 0]} = \frac{p_{(0)}}{[1 - p_{(0)}]}$$

Rasio Odds didefinisikan sebagai rasio antara odds untuk $X_{ji}=1$ terhadap odds untuk $X_{ji}=0$, yaitu

$$\psi = \frac{p_{(1)}/[1 - p_{(1)}]}{p_{(0)}/[1 - p_{(0)}]} \quad (2.29)$$

Logaritma dari rasio odds disebut logit, yaitu

$$\begin{aligned} \ln(\psi) &= \ln \left[\frac{p_{(1)}/[1 - p_{(1)}]}{p_{(0)}/[1 - p_{(0)}]} \right] \\ &= \ln \left[\frac{p_{(1)}}{1 - p_{(1)}} \right] - \ln \left[\frac{p_{(0)}}{1 - p_{(0)}} \right] \\ &= g(1) - g(0) \end{aligned}$$

dimana $g(1)$ adalah logit untuk kejadian sukses ($Y=1$) pada individu $X_{ji}=1$, yakni

$$\begin{aligned} g(1) &= \ln \left[\frac{p_{(1)}}{1 - p_{(1)}} \right] \\ &= \beta_0 + \beta_1 X_{1i} + \dots + \beta_j \times 1 + \dots + \beta_k X_{ki} \end{aligned}$$

dan $g(0)$ adalah logit untuk kejadian sukses ($Y=1$) pada individu $X_{ji}=0$, yakni

$$\begin{aligned} g(0) &= \ln \left[\frac{p_{(0)}}{1 - p_{(0)}} \right] \\ &= \beta_0 + \beta_1 X_{1i} + \dots + \beta_j \times 0 + \dots + \beta_k X_{ki} \end{aligned}$$

Dengan demikian rasio odds untuk model regresi logistik dapat didefinisikan sebagai selisih logit, yaitu

$$\begin{aligned} \ln(\psi) &= g(1) - g(0) \\ &= (\beta_0 + \beta_1 X_{1i} + \dots + \beta_j \times 1 + \dots + \beta_k X_{ki}) \\ &\quad - (\beta_0 + \beta_1 X_{1i} + \dots + \beta_j \times 0 + \dots + \beta_k X_{ki}) \\ &= \beta_j \end{aligned} \quad (2.30)$$

Dari Persamaan (2.30) maka diperoleh rasio odds yaitu $\psi = \exp(\beta_j)$. Rasio odds tersebut diinterpretasikan sebagai peluang munculnya kejadian sukses ($Y=1$) pada individu dengan $X_{ji}=1$ adalah sebesar $\exp(\beta_j)$ kali dibandingkan dengan munculnya kejadian sukses tersebut pada individu yang mempunyai nilai $X_{ji}=0$.

Hosmer dan Lemeshow (2000:53) menyatakan bahwa log odds untuk model regresi logistik dengan nilai peubah bebas lebih dari satu didefinisikan sebagai selisih antara penduga logit untuk suatu peubah bebas X_{ji} bernilai a dengan penduga logit untuk peubah bebas X_{ji} bernilai b dimana peubah bebas lainnya diasumsikan konstan, yaitu

$$\begin{aligned}
 \ln \left[\hat{\psi}(a, b) \right] &= \hat{g}(X_{ji} = a) - \hat{g}(X_{ji} = b) \\
 &= \left(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_j \times a + \dots + \hat{\beta}_k X_{ki} \right) \\
 &\quad - \left(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_j \times b + \dots + \hat{\beta}_k X_{ki} \right) \\
 &= \hat{\beta}_j (a - b)
 \end{aligned} \tag{2.31}$$

Dengan demikian penduga rasio oddsnya yaitu

$$\hat{\psi}(a - b) = \exp \left[\hat{\beta}_j (a - b) \right]$$

Pada beberapa kasus penelitian, untuk model dugaan logit $\hat{g}(X) = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_j X_{ji} + \dots + \hat{\beta}_k X_{ki}$, dimana peubah bebas X_{ji} bertipe kontinu maka koefisien $\hat{\beta}_j$ menunjukkan perubahan fungsi logit untuk setiap kenaikan satu satuan unit peubah bebas atau dapat dituliskan $\hat{\beta}_j = \hat{g}(X + 1) - \hat{g}(X)$. Namun demikian, pada umumnya perubahan satu satuan unit tidak memberikan makna yang berarti sehingga untuk memberikan interpretasi yang lebih bermakna pada model regresi logistik dengan peubah bebas bertipe kontinu,

rasio odds dihitung untuk perubahan c satuan unit peubah bebas, yaitu

$$\begin{aligned}\widehat{\psi}(X+c, X) &= \widehat{\psi}(c) \\ &= \exp [\widehat{g}(X+c) - \widehat{g}(X)] \\ &= \exp \left(c\widehat{\beta}_j \right)\end{aligned}$$

Interpretasi rasio odds pada kasus ini adalah untuk setiap kenaikan c satuan unit peubah bebas X , maka kemungkinan suatu individu mempunyai peluang sukses adalah $\exp \left(c\widehat{\beta}_j \right)$ kali dibandingkan dengan individu tersebut akan berpeluang gagal untuk suatu peubah bebas bernilai tentu.