

BAB II TINJAUAN PUSTAKA

2.1 Distribusi Multinomial

Distribusi multinomial merupakan suatu distribusi yang sering digunakan dalam analisis data kategori. Peubah respon dengan lebih dari dua kategori akan mengikuti distribusi multinomial. Misalnya terdapat j kategori pada peubah respon. Peluangnya dinotasikan $\{\pi_1, \pi_2, \dots, \pi_j\}$ dengan $\sum_j \pi_j = 1$. Untuk n sampel, peluang multinomial bahwa n_1 termasuk kategori 1, n_2 termasuk kategori 2, ..., n_j menjelaskan pada kategori j dengan $\sum_j n_j = n$ adalah :

$$P(n_1, n_2, \dots, n_j) = \left(\frac{n!}{n_1! n_2! \dots n_j!} \right) \pi_1^{n_1} \pi_2^{n_2} \dots \pi_j^{n_j}$$

Untuk kategori j , perhitungan n_j memiliki rata-rata $n\pi_j$ dan standar deviasi $\sqrt{[n\pi_j(1 - \pi_j)]}$ (Agresti, 1990).

2.2 Regresi Logistik Multinomial

Regresi logistik multinomial adalah regresi logistik yang menggunakan peubah respon dengan lebih dari dua kategori berskala nominal. Data berskala nominal merupakan data yang paling sederhana, dengan angka yang diberikan kepada objek mempunyai arti sebagai label dan tidak menunjukkan tingkatan apapun (Nazir, 2003).

Ciri-ciri data kategori berskala nominal yaitu hanya memiliki nama dan tidak memiliki urutan. Contoh data dengan peubah respon berskala nominal yaitu pemilihan jurusan pada sebuah Sekolah Menengah Atas (SMA) dengan IPA dikategorikan sebagai 0, IPS dikategorikan sebagai 1 dan Bahasa dikategorikan sebagai 2. Kasus lain misalnya faktor-faktor indikasi terjangkitnya suatu penyakit pada manusia usia lanjut dengan penyakit kanker dikategorikan sebagai 0, penyakit darah tinggi dikategorikan sebagai 1 serta penyakit tumor dikategorikan sebagai 2.

Aldrich dan Nelson (1984) menyatakan bahwa kata logit merupakan singkatan dari *Logistic Probability Unit* yaitu salah satu dari beberapa model yang mendasar dari fungsi peluang. Apabila terdapat j yang berarti banyaknya kategori pada peubah respon maka model logistik yang terbentuk sebanyak $j-1$. Hal ini dikarenakan salah

satu kategori akan dijadikan pembanding. Kategori mana saja tergantung dari tujuan yang ingin dicapai oleh peneliti. Namun menurut Hosmer dan Lemeshow (1989) kategori 0 pada peubah respon berskala nominal lebih baik digunakan sebagai pembanding. Menurut Agresti (1990), model umum regresi logistik multinomial untuk p banyaknya peubah prediktor yang dinyatakan dalam vektor $\mathbf{x}_i$ serta probabilitas kategori respon ke- j sebagai berikut :

$$\pi_j(\mathbf{x}_i) = P(y = j|\mathbf{x}_i) = \frac{\exp(g_j(\mathbf{x}_i))}{\sum_{k=0}^{j-1} \exp(g_k(\mathbf{x}_i))} \quad (2.1)$$

dengan $g_0(\mathbf{x}_i) = 0$, dengan $i=1, 2, \dots, p$ sebagai dasar permodelan untuk :

$$g_j(\mathbf{x}_i) = \beta_{j0} + \beta_{j1}\mathbf{x}_{i1} + \dots + \beta_{jp}\mathbf{x}_{ip} \quad (2.2)$$

$$\begin{aligned} g_j(\mathbf{x}_i) &= \ln \left(\frac{\pi_j(\mathbf{x}_i)}{\pi_0(\mathbf{x}_i)} \right) \\ &= \ln \left(\frac{\left(\frac{\exp(g_j(\mathbf{x}_i))}{\sum_{k=0}^{j-1} \exp(g_k(\mathbf{x}_i))} \right)}{\left(\frac{1}{\sum_{k=0}^{j-1} \exp(g_k(\mathbf{x}_i))} \right)} \right) \\ &= \ln \left(\frac{\exp(g_j(\mathbf{x}_i))}{\frac{1 + \exp(g_1(\mathbf{x}_i)) + \dots + \exp(g_{j-1}(\mathbf{x}_i))}{1}} \right) \\ &= \ln \left(\frac{\exp(g_j(\mathbf{x}_i))}{1 + \exp(g_1(\mathbf{x}_i)) + \dots + \exp(g_{j-1}(\mathbf{x}_i))} \right) \\ &= \ln \left(\exp(g_j(\mathbf{x}_i)) \right) \\ &= g_j(\mathbf{x}_i) \\ &= \beta_{j0} + \beta_{j1}\mathbf{x}_{i1} + \dots + \beta_{jp}\mathbf{x}_{ip} \\ &= \beta_{j0} + \sum_{k=0}^p \beta_{jk}\mathbf{x}_{ip} \end{aligned}$$

Pada regresi logistik dengan peubah respon 3 kategori ($j=0,1,2$) akan terbentuk 2 fungsi logit. Misal $j=0$ sebagai kategori pembanding untuk membentuk 2 fungsi logit, yaitu sebagai pembanding untuk $j=1$ dan $j=2$ maka :

$$g_0(x_i) = 0$$

Dua fungsi logit tersebut yaitu :

$$\begin{aligned}
 g_1(x_i) &= \ln \left(\frac{\pi_1(x_i)}{\pi_0(x_i)} \right) \\
 &= \ln \left[\frac{P(y = 1 | x_i)}{P(y = 0 | x_i)} \right] \\
 &= \ln \left(\frac{\left(\frac{\exp(g_1(x_i))}{\sum_{k=0}^{j-1} \exp(g_k(x_i))} \right)}{\left(\frac{1}{\sum_{k=0}^{j-1} \exp(g_k(x_i))} \right)} \right) \\
 &= \ln \left(\frac{\exp(g_1(x_i))}{\frac{1 + \exp(g_1(x_i)) + \exp(g_2(x_i))}{1}} \right) \\
 &= \ln \left(\exp(g_1(x_i)) \right) \\
 &= g_1(x_i) \\
 &= \beta_{10} + \beta_{11}x_{i1} + \beta_{12}x_{i2} + \dots + \beta_{1p}x_{ip} \\
 &= \beta_{10} + \sum_{k=0}^p \beta_{1k}x_{ik} \\
 &= \mathbf{X}'\boldsymbol{\beta}_1
 \end{aligned}$$

$$\begin{aligned}
 g_2(x_i) &= \ln \left(\frac{\pi_2(x_i)}{\pi_0(x_i)} \right) \\
 &= \ln \left[\frac{P(y = 2 | x_i)}{P(y = 0 | x_i)} \right] \\
 &= \ln \left(\frac{\left(\frac{\exp(g_2(x_i))}{\sum_{k=0}^{j-1} \exp(g_k(x_i))} \right)}{\left(\frac{\exp(g_0(x_i))}{\sum_{k=0}^{j-1} \exp(g_k(x_i))} \right)} \right)
 \end{aligned}$$

$$\begin{aligned}
&= \ln \left(\frac{\frac{\exp(g_2(\mathbf{x}_i))}{1 + \exp(g_1(\mathbf{x}_i)) + \exp(g_2(\mathbf{x}_i))}}{\frac{1}{1 + \exp(g_1(\mathbf{x}_i)) + \exp(g_2(\mathbf{x}_i))}} \right) \\
&= \ln \left(\exp(g_2(\mathbf{x}_i)) \right) \\
&= g_2(\mathbf{x}_i) \\
&= \beta_{20} + \beta_{21}\mathbf{x}_{i_1} + \beta_{22}\mathbf{x}_{i_2} + \dots + \beta_{2p}\mathbf{x}_{i_p} \\
&= \beta_{20} + \sum_{k=0}^{j-1} \beta_{2k}\mathbf{x}_{i_k} \\
&= \mathbf{X}'\boldsymbol{\beta}_2
\end{aligned}$$

Jika $\pi_j(\mathbf{x}_i) = P(y = j|\mathbf{x}_i)$; $j = 0,1,2$; didapatkan model peluang bersyarat terhadap setiap kategori peubah respon, yaitu :

$$\begin{aligned}
\pi_j(\mathbf{x}_i) &= P(y = j|\mathbf{x}_i) = \frac{\exp g_j(\mathbf{x}_i)}{\sum_{k=0}^{j-1} \exp g_k(\mathbf{x}_i)} \\
\pi_0(\mathbf{x}_i) &= P(y = 0|\mathbf{x}_i) = \frac{1}{1 + \exp(g_1(\mathbf{x}_i)) + \exp(g_2(\mathbf{x}_i))} \\
&= \frac{1}{1 + \exp(\beta_{10} + \sum_{k=1}^p \beta_{1k}\mathbf{x}_{i_k}) + \exp(\beta_{20} + \sum_{k=1}^p \beta_{2k}\mathbf{x}_{i_k})}
\end{aligned}$$

$$\begin{aligned}
\pi_1(\mathbf{x}_i) &= P(y = 1|\mathbf{x}_i) = \frac{\exp(g_1(\mathbf{x}_i))}{1 + \exp(g_1(\mathbf{x}_i)) + \exp(g_2(\mathbf{x}_i))} \\
&= \frac{\exp(\beta_{10} + \sum_{k=1}^p \beta_{1k}\mathbf{x}_{i_k})}{1 + \exp(\beta_{10} + \sum_{k=1}^p \beta_{1k}\mathbf{x}_{i_k}) + \exp(\beta_{20} + \sum_{k=1}^p \beta_{2k}\mathbf{x}_{i_k})}
\end{aligned}$$

dan

$$\begin{aligned}
\pi_2(\mathbf{x}_i) &= P(y = 2|\mathbf{x}_i) = \frac{\exp(g_2(\mathbf{x}_i))}{1 + \exp(g_1(\mathbf{x}_i)) + \exp(g_2(\mathbf{x}_i))} \\
&= \frac{\exp(\beta_{20} + \sum_{k=1}^p \beta_{2k}\mathbf{x}_{i_k})}{1 + \exp(\beta_{10} + \sum_{k=1}^p \beta_{1k}\mathbf{x}_{i_k}) + \exp(\beta_{20} + \sum_{k=1}^p \beta_{2k}\mathbf{x}_{i_k})}
\end{aligned}$$

Sehingga didapatkan bentuk umum model regresi logistik dengan j kategori respon, yaitu :

$$P(y = j|x_i) = \frac{\exp(g_j(x_i))}{\sum_{k=0}^{j-1} \exp(g_k(x_i))} \quad (2.3)$$

2.3 Pendugaan Parameter Regresi Logistik Multinomial

Metode pendugaan digunakan untuk mendapatkan nilai pendugaan terbaik bagi parameter. Pendugaan nilai β dapat dilakukan dengan banyak metode, pada regresi logistik digunakan metode kemungkinan maksimum (*Maximum Likelihood Estimation* atau MLE) untuk menduga parameter. Pada persamaan regresi logistik biner nilai ε (galat) mengikuti distribusi binomial, sedangkan pada regresi logistik multinomial nilai ε mengikuti distribusi multinomial. Kurniawan (2008) menjelaskan bahwa galat atau *error* dalam metode numerik adalah selisih yang ditimbulkan antara nilai sebenarnya dengan nilai yang dihasilkan dengan metode numerik. Menurut Hosmer dan Lemeshow (1989), MLE mempunyai banyak kelebihan dibandingkan dengan metode yang lain, yaitu dapat digunakan untuk model non linier seperti pada model regresi logistik, dan hasil pendugaannya mendekati nilai parameter yang diduga. Pada dasarnya MLE memberikan nilai pendugaan parameter model dengan memaksimalkan fungsi *likelihood*.

Parameter model diduga dari vektor $\hat{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$. Banyaknya parameter yang akan diduga adalah $p+1$. Dimisalkan terdapat peubah respon berskala nominal dengan tiga kategori, maka didapatkan fungsi *likelihood* untuk n pengamatan yang saling bebas adalah sebagai berikut :

$$L(\beta) = \ln \left(\prod_{i=1}^n [\pi_0(x_i)^{y_{0i}} \pi_1(x_i)^{y_{1i}} \pi_2(x_i)^{y_{2i}}] \right) \quad (2.4)$$

dengan :

$$\pi_0(x_i)^{y_{0i}} = \left(\frac{1}{1 + \exp(g_1(x_i)) + \exp(g_2(x_i))} \right)^{y_{0i}},$$

$$\pi_1(\mathbf{x}_i)^{y_{1i}} = \left(\frac{\exp(g_1(\mathbf{x}_i))}{1 + \exp(g_1(\mathbf{x}_i)) + \exp(g_2(\mathbf{x}_i))} \right)^{y_{1i}},$$

$$\pi_2(\mathbf{x}_i)^{y_{2i}} = \left(\frac{\exp(g_2(\mathbf{x}_i))}{1 + \exp(g_1(\mathbf{x}_i)) + \exp(g_2(\mathbf{x}_i))} \right)^{y_{2i}}$$

sehingga :

$$\begin{aligned} L(\boldsymbol{\beta}) &= \sum_{i=1}^n \left\{ y_{0i} \ln 1 + y_{1i} \exp(g_1(\mathbf{x}_i)) + y_{2i} \exp(g_2(\mathbf{x}_i)) \right. \\ &\quad \left. - \ln \left(1 + \exp(g_1(\mathbf{x}_i)) + y_{0i} \exp(g_2(\mathbf{x}_i)) \right)^{y_{0i}} \right. \\ &\quad \left. - \ln \left(1 + \exp(g_1(\mathbf{x}_i)) + y_{1i} \exp(g_2(\mathbf{x}_i)) \right)^{y_{1i}} \right. \\ &\quad \left. - \ln \left(1 + \exp(g_1(\mathbf{x}_i)) + y_{2i} \exp(g_2(\mathbf{x}_i)) \right)^{y_{2i}} \right\} \\ &= \sum_{i=1}^n \left\{ y_{0i} \ln 1 + y_{1i} \exp(g_1(\mathbf{x}_i)) + y_{2i} \exp(g_2(\mathbf{x}_i)) \right. \\ &\quad \left. - \ln \left(\left(1 + \exp(g_1(\mathbf{x}_i)) + \exp(g_2(\mathbf{x}_i)) \right)^{\sum_{j=0}^2 y_{ji}} \right) \right\} \end{aligned}$$

Karena $\ln 1 = 0$ dan $\sum_{j=0}^2 y_{ji} = 1$ maka

$$L(\boldsymbol{\beta}) = \sum_{i=1}^n y_{1i} g_1(\mathbf{x}_i) + y_{2i} g_2(\mathbf{x}_i) - \ln \left(1 + \exp(g_1(\mathbf{x}_i)) + \exp(g_2(\mathbf{x}_i)) \right) \quad (2.5)$$

$$\begin{aligned} L(\boldsymbol{\beta}) &= \sum_{i=1}^n y_{1i} \left(\beta_{10} + \sum_{k=1}^p \beta_{1k} \mathbf{x}_{ik} \right) + y_{2i} \left(\beta_{20} + \sum_{k=1}^p \beta_{2k} \mathbf{x}_{ik} \right) \\ &\quad - \ln \left\{ 1 + \exp \left(\beta_{10} + \sum_{k=1}^p \beta_{1k} \mathbf{x}_{ik} \right) \right. \\ &\quad \left. + \exp \left(\beta_{20} + \sum_{k=1}^p \beta_{2k} \mathbf{x}_{ik} \right) \right\} \end{aligned}$$

Langkah selanjutnya, agar nilai fungsi $L(\boldsymbol{\beta})$ mencapai maksimum maka dilakukan penurunan terhadap β_{10} lalu disamadengankan nol

$$\begin{aligned}
 \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_{10}} &= \frac{\partial}{\partial \beta_{10}} \left[\sum_{i=1}^n y_{1i} \beta_{10} + \left(y_{1i} \sum_{k=1}^p \beta_{1k} \mathbf{x}_{ik} \right) \right. \\
 &\quad \left. + y_{2i} \left(\beta_{20} + \sum_{k=1}^p \beta_{2k} \mathbf{x}_{ik} \right) \right. \\
 &\quad \left. - \ln \left\{ 1 + \exp \left(\beta_{10} + \sum_{k=1}^p \beta_{1k} \mathbf{x}_{ik} \right) \right. \right. \\
 &\quad \left. \left. + \exp \left(\beta_{20} + \sum_{k=1}^p \beta_{2k} \mathbf{x}_{ik} \right) \right\} \right] = 0 \\
 &= \sum_{i=1}^n y_{1i} - \frac{\exp(\beta_{10} + \sum_{k=1}^p \beta_{1k} \mathbf{x}_{ik})}{1 + \exp(\beta_{10} + \sum_{k=1}^p \beta_{1k} \mathbf{x}_{ik}) + \exp(\beta_{20} + \sum_{k=1}^p \beta_{2k} \mathbf{x}_{ik})} = 0 \\
 &= \sum_{i=1}^n y_{1i} - \pi_1(\mathbf{x}_i) = 0 \tag{2.6}
 \end{aligned}$$

Sedangkan pendugaan parameter β_{11} yang merupakan *slope*/koefisien peubah x_{i1} dari $(g_1(\mathbf{x}_i))$ adalah sebagai berikut :

$$\begin{aligned}
 \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_{11}} &= \frac{\partial}{\partial \beta_{11}} \left[\sum_{i=1}^n y_{1i} \beta_{10} \left(y_{1i} \beta_{10} \mathbf{x}_{i1} + y_{1i} \sum_{k=1}^p \beta_{1k} \mathbf{x}_{ik} \right) \right. \\
 &\quad \left. + y_{2i} \left(\beta_{20} + \sum_{k=1}^p \beta_{2k} \mathbf{x}_{ik} \right) \right. \\
 &\quad \left. - \ln \left\{ 1 + \exp \left(\beta_{10} + \beta_{11} x_{i1} + \sum_{k=1}^p \beta_{1k} \mathbf{x}_{ik} \right) \right. \right. \\
 &\quad \left. \left. + \exp \left(\beta_{20} + \sum_{k=1}^p \beta_{2k} \mathbf{x}_{ik} \right) \right\} \right] = 0
 \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n y_{1i} x_{i1} \\
&\quad - \frac{\exp(\beta_{10} + \beta_{11}x_{i1} + \sum_{k=2}^p \beta_{1k}x_{ik})x_{i1}}{1 + \exp(\beta_{10} + \sum_{k=1}^p \beta_{1k}x_{ik}) + \exp(\beta_{20} + \sum_{k=1}^p \beta_{2k}x_{ik})} \\
&= \sum_{i=1}^n y_{1i} x_{i1} - (\pi_1(x_i))x_{i1} = 0 \\
&= \sum_{i=1}^n x_{i1} (y_{1i} - \pi_1(x_i)) = 0 \tag{2.7}
\end{aligned}$$

Bentuk secara umum turunan parsial pertama fungsi log *likelihood* terhadap parameter yang akan diduga β_{jk} , adalah sebagai berikut :

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_{jk}} = \sum_{i=1}^n x_{ik}(y_{ji} - \pi_{ji}) \tag{2.8}$$

Sedangkan turunan parsial ke-2 adalah sebagai berikut :

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_{jk} \partial \beta_{j'k'}} = - \sum_{i=1}^n x_{ik'} x_{ik} \pi_{ji} (1 - \pi_{ji}) \tag{2.9}$$

dan

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_{jk} \partial \beta_{j'k'}} = \sum_{i=1}^n x_{ik'} x_{ik} \pi_{ji} \pi_{j'i} \tag{2.10}$$

$j = 0, 1, 2$; $j' = 0, 1, 2$; k dan $k' = 1, 2, \dots, p$ dengan $x_{i0} = 1$. π_{ji} merupakan penyederhanaan notasi untuk $\pi_j(x_i)$. Turunan kedua dari fungsi log *likelihood* merupakan elemen dari matriks Hessian. Akan tetapi, elemen-elemen tersebut bernilai negatif sehingga dapat didefinisikan matriks Hessian yang terbentuk adalah $-\mathbf{H}(\boldsymbol{\beta})$ dan disebut sebagai matriks Informasi, $\mathbf{I}(\boldsymbol{\beta})$ (Hosmer dan Lemeshow, 1989). Matriks tersebut berukuran $(p+1) \times (p+1)$. Matriks informasi yang diinverskan merupakan dugaan dari matriks kovarian dan dinyatakan dengan persamaan : $cov(\boldsymbol{\beta}) = (\mathbf{I}(\boldsymbol{\beta}))^{-1} = -(\mathbf{H}(\boldsymbol{\beta}))^{-1}$ atau $cov(\boldsymbol{\beta}) = \{\mathbf{X}' \text{Diag}[\pi_i(1 - \pi_i)]\mathbf{X}\}^{-1}$, $\mathbf{X}$ merupakan matriks berukuran $n \times (p+1)$ yang elemen-elemennya adalah data

dari peubah prediktor dan $Diag[\pi_i(1 - \pi_i)]$ adalah elemen diagonal dari matriks $\mathbf{V}$. Matriks $\mathbf{X}$ dan matriks $\mathbf{V}$ didefinisikan sebagai berikut:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad \text{dan}$$

$$\mathbf{V} = \begin{bmatrix} \pi_1(1 - \pi_1) & 0 & \dots & 0 \\ 0 & \pi_2(1 - \pi_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \pi_n(1 - \pi_n) \end{bmatrix}$$

Sebagaimana diketahui bahwa pendugaan parameter melalui metode MLE adalah melakukan turunan parsial fungsi *likelihood* terhadap parameter yang akan diduga. Akan tetapi, turunan parsial pertama dari fungsi log *likelihood* terhadap parameter yang akan diduga merupakan fungsi nonlinear (tidak dapat diturunkan secara parsial). Oleh karena itu, diperlukan metode iterasi *Newton Raphson*.

Persamaan-persamaan yang dipergunakan dalam metode iterasi *Newton Raphson* adalah sebagai berikut :

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \boldsymbol{\beta}^{(t)} - [\mathbf{H}^{(t)}]^{-1} \mathbf{g}^{(t)} \quad (2.11)$$

dengan nilai $\mathbf{H}^{(t)}$ dan $\mathbf{g}^{(t)}$ berupa matriks :

$$\mathbf{H}^{(t)} = \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_{jk} \partial \beta_{j'k'}} = \sum_{i=1}^n x_{ik} x_{ik'} \pi_{ji} \pi_{j'i} = \mathbf{X}' \mathbf{V}^{(t)} \mathbf{X} \quad (2.12)$$

$$\mathbf{g}^{(t)} = \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_{jk}} = \sum_{i=1}^n x_{ik} (y_{ji} - \pi_{ji}) = \mathbf{X}' (\mathbf{y} - \boldsymbol{\pi}^{(t)}) \quad (2.13)$$

Persamaan (2.11) dipergunakan sebagai panduan dalam iterasi *Newton Raphson*. $\mathbf{H}^{(t)}$ merupakan matriks turunan parsial ke-2 dari fungsi log *likelihood* terhadap parameter yang akan diduga. Sedangkan $\mathbf{g}^{(t)}$ adalah turunan parsial pertama dari fungsi log *likelihood* terhadap parameter yang akan diduga. Dan $[\mathbf{H}^{(t)}]^{-1}$ merupakan matriks kovarian (Agresti,1990). Adapun langkah-langkah iterasi *Newton Raphson* sebagai berikut:

1. Menentukan nilai awal dugaan parameter yaitu $\boldsymbol{\beta}^{(t)}$. Menurut Gerald dan Wheatley (1996) nilai $\boldsymbol{\beta}^{(t)}$ berdekatan dengan selisih antara nilai penduga dan dugaan yang diharapkan, hal ini

disebabkan agar proses iterasi tidak terlalu banyak. Sehingga untuk mempermudah perhitungan digunakan 0 sebagai nilai dugaan awal iterasi.

2. Membentuk keseluruhan data dalam bentuk matriks.

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \text{ dan } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

3. Menentukan model regresi logistik pada persamaan (2.2) yaitu :

$$g_j(x_i) = \beta_{j0} + \beta_{j1}x_{i1} + \dots + \beta_{jp}x_{ip}$$

4. Menentukan matriks $\mathbf{V}$

$$\mathbf{V}^{(0)} = \begin{bmatrix} \pi_i^{(t)}(1 - \pi_i^{(t)}) & 0 & \dots & 0 \\ 0 & \pi_2^{(t)}(1 - \pi_2^{(t)}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \pi_n^{(t)}(1 - \pi_n^{(t)}) \end{bmatrix}$$

5. Mensubstitusi hasil dari langkah pada persamaan (2.11) dengan

$$\mathbf{H}^{(t)} = \mathbf{X}'\mathbf{V}^{(t)}\mathbf{X}$$

$$= \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \pi_i^{(t)}(1 - \pi_i^{(t)}) & 0 & \dots & 0 \\ 0 & \pi_2^{(t)}(1 - \pi_2^{(t)}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \pi_n^{(t)}(1 - \pi_n^{(t)}) \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

dan

$$\mathbf{g}^{(t)} = \mathbf{X}'(\mathbf{y} - \boldsymbol{\pi}^{(t)})$$

$$\boldsymbol{\pi}^{(t)} = \begin{bmatrix} \pi_1^{(t)} \\ \pi_2^{(t)} \\ \vdots \\ \pi_n^{(t)} \end{bmatrix}$$

6. Diperoleh $\hat{\boldsymbol{\beta}}^{(t+1)}$

7. Menentukan selisih $(\hat{\beta}^{(t+1)} - \beta^{(t)})$, jika hasil selisihnya memenuhi kondisi konvergen yaitu bernilai $(1 \times 10^{-7}) \approx 0,0000001$ maka $\hat{\beta}^{(t+1)}$ merupakan hasil pendugaannya.

Bila kondisi konvergen belum terpenuhi maka iterasi dilanjutkan terus dengan cara menjadikan nilai $\hat{\beta}^{(t+1)}$ sebagai $\beta^{(t)}$ untuk iterasi selanjutnya, diulangi mulai langkah ke-3 sampai ke-7 hingga diperoleh nilai β yang konvergen.

2.4 Nilai Odds Ratio

Odds ratio (OR) merupakan perbandingan peluang suatu kejadian sukses dengan peluang kejadian gagal. Dalam regresi logistik *odds ratio* dapat digunakan untuk mempermudah interpretasi model yang dihasilkan. Menurut Agresti (2000), *Odds ratio* pada kategori $y = j$ merupakan perbandingan antara $x = a$ dan $x = b$, dengan $P(y = j)$ adalah peluang sukses dan $P(y = 0)$ adalah peluang gagal. Untuk menjelaskan *Odds ratio* pada 3 peubah respon multinomial dapat dilihat pada Tabel 2.1.

Tabel 2.1 Nilai Peluang *Odds ratio* Model Regresi Logistik

	$y = 0$	$y = 1$	$y = 2$
$x = 0$	$P(y = 0 x_i) = \frac{1}{1 + e^{\beta_{10}} + e^{\beta_{20}}}$	$P(y = 1 x_i) = \frac{e^{\beta_{10}}}{1 + e^{\beta_{10}} + e^{\beta_{20}}}$	$P(y = 2 x_i) = \frac{e^{\beta_{20}}}{1 + e^{\beta_{10}} + e^{\beta_{20}}}$
$x = 1$	$P(y = 0 x_i) = \frac{1}{1 + e^{\beta_{10} + \beta_{11}} + e^{\beta_{20} + \beta_{21}}}$	$P(y = 1 x_i) = \frac{e^{\beta_{10} + \beta_{11}}}{1 + e^{\beta_{10} + \beta_{11}} + e^{\beta_{20} + \beta_{21}}}$	$P(y = 2 x_i) = \frac{e^{\beta_{20} + \beta_{21}}}{1 + e^{\beta_{10} + \beta_{11}} + e^{\beta_{20} + \beta_{21}}}$

Dari Tabel 2.1 didapatkan *Odds ratio* sebagai berikut :

$$OR(a, b) = \frac{P(y=j|x=a)/P(y=0|x=a)}{P(y=j|x=b)/P(y=0|x=b)}, j = 1, 2 \quad (2.14)$$

Menurut Kleinbaum dan Klein (2010) pada regresi logistik dengan tiga kategori peubah respon akan terbentuk dua *Odds ratio*, yang pertama membandingkan peluang antara peubah respon kategori 1 ($y = 1$) dan yang kedua membandingkan peluang antara peubah respon kategori ke 2 ($y = 2$) dengan peubah respon kategori pembanding ($y = 0$).

Nilai *Odds ratio* digunakan untuk menunjukkan kecenderungan hubungan suatu peubah prediktor terhadap peubah respon. Misal nilai *Odds ratio* sama dengan 1 berarti bahwa $x = a$ mempunyai resiko yang sama dengan $x = b$ untuk menghasilkan $y = j$. Bila $1 < Odds$

$Ratio < \infty$ berarti $x = a$ memiliki resiko lebih tinggi *Odds ratio* kali dibandingkan $x = b$ untuk menghasilkan $y = j$ dan sebaliknya untuk $0 < Odds\ ratio < 1$, maka $x = a$ memiliki resiko lebih rendah *Odds ratio* kali dibandingkan $x = b$ untuk menghasilkan $y = j$ (Hosmer dan Lemeshow, 2000).

Sebagai contoh misalnya peneliti ingin memprediksi kelainan refraksi berdasarkan usia pasien (x). Kelainan refraksi terdiri dari miopi=0, hipermetropi=1 dan astigmatisma=2. Model regresi logistik yang terbentuk untuk persamaan 1 (logit 1) yakni $P(y = 1) = -0,523 + 0,183x_{11}$. Ketika nilai $x_{11} = 1$ untuk pasien berusia <50 tahun maka nilai *Odds ratio* yang dihasilkan adalah :

$$\begin{aligned} & \frac{P(y = j|x = a)/P(y = 0|x = a)}{P(y = j|x = b)/P(y = 0|x = b)} \\ &= \frac{\exp(-0,523 + 0,183(1) + 0,117(0))}{\exp(-0,523 + 0,183(0) + 0,117(0))} \\ &= \exp(0,183) = 1,201 \end{aligned}$$

Hal ini berarti bahwa orang yang berusia diatas 50 tahun memiliki resiko menderita kelainan refraksi hipermetropi 1,201 kali lebih tinggi dibanding kelainan mata yang lainnya.

2.5 Regresi Probit

Model regresi probit merupakan metode analisis yang digunakan untuk menggambarkan hubungan antara peubah prediktor dan peubah respon lebih dari 1 kategori. Regresi probit yang merupakan kependekan dari *Probability Unit* berdasarkan fungsi sebaran peluang normal kumulatif baku yang dikenal juga sebagai model Normit singkatan dari *Normal Probability Unit* (Aldrich dan Nelson, 1984). Hampir sama dengan model logit, probit adalah sebuah metode analisis untuk menduga kemungkinan sebuah peristiwa dengan peubah respon nominal, ordinal serta biner (Vasisht, 2000).

Borooah (2002) menjelaskan bahwa bentuk model probit adalah :

$$\pi_j(x_i) = \Phi(\beta_{j0} + \beta_{j1}x_{i1} + \beta_{j2}x_{i2} + \dots + \beta_{jp}x_{ip}) \quad (2.15)$$

dengan :

$\pi_j(x_i)$ = Peluang peubah respon kategori ke-j pada pengamatan ke-i

Φ	= Fungsi sebaran kumulatif normal baku
$x_{i1}, x_{i2}, \dots, x_{ip}$	= Nilai peubah prediktor ke- p pada pengamatan ke- i
$\beta_{j0}, \beta_{j1}, \dots, \beta_{jp}$	= Koefisien regresi parsial untuk setiap peubah
j	= Peubah respon ke- j
p	= Peubah prediktor ke- p

Transformasi dari model probit atau fungsi normit adalah :

$g_j(\mathbf{x}_i)$	$= \Phi^{-1}(\pi_j(\mathbf{x}_i))$
	$= \Phi^{-1}(\Phi(\beta_{j0} + \beta_{j1}x_{i1} + \beta_{j2}x_{i2} + \dots + \beta_{jp}x_{ip}))$
	$= \beta_{j0} + \beta_{j1}x_{i1} + \beta_{j2}x_{i2} + \dots + \beta_{jp}x_{ip}$ (2.16)

Jika diasumsikan bahwa peubah y^* adalah fungsi linier dari peubah prediktor dengan bentuk :

$$y^* = \mathbf{X}'\boldsymbol{\beta} + e \quad (2.17)$$

Peubah y adalah peubah respon teramati dengan memberikan *threshold* (batas-batas) $\gamma_1 < \gamma_2 < \dots < \gamma_{j-1}$ pada y^* . Maka peubah respon y akan membentuk j kategori, secara rinci adalah sebagai berikut :

$$\begin{aligned} y = 0 & \quad \text{jika } y^* \leq \gamma_0, \\ y = 2 & \quad \text{jika } \gamma_0 < y^* \leq \gamma_2, \\ & \vdots \\ y = m & \quad \text{jika } \gamma_{m-1} < y^* \leq \gamma_m, \\ & \vdots \\ y = j & \quad \text{jika } y^* \leq \gamma_{j-1} \end{aligned}$$

Dimisalkan peubah respon y terdapat 2 kategori yang mengikuti distribusi multinomial. $Y_i = j$ sedemikian sehingga $\gamma_1 < y^* \leq \gamma_2$, dengan $j = 1, \dots, m$ dengan $\gamma_1 = -\infty$ dan $\gamma_m = +\infty$.

Kategori ke-1: $y = 1$:

$$\begin{aligned} P(y = 1) &= P(\gamma_1 < y^* \leq \gamma_2) \\ &= P(-\infty < y^* \leq \gamma_2) \\ &= P(y^* \leq \gamma_2) \\ &= P(\mathbf{X}'\boldsymbol{\beta} + e \leq \gamma_2) \\ &= P(e \leq \gamma_2 - \mathbf{X}'\boldsymbol{\beta}) \\ &= \Phi(\gamma_2 - \mathbf{X}'\boldsymbol{\beta}) \end{aligned}$$

Kategori ke-2: = 2 :

$$\begin{aligned}
 P(y = 2) &= P(\gamma_1 < y^* \leq \gamma_2) \\
 &= P(\gamma_1 < \mathbf{X}'\boldsymbol{\beta} + e \leq \gamma_2) \\
 &= P(\gamma_1 - \mathbf{X}'\boldsymbol{\beta} < e \leq \gamma_2 - \mathbf{X}'\boldsymbol{\beta}) \\
 &= \Phi(\gamma_2 - \mathbf{X}'\boldsymbol{\beta}) - \Phi(\gamma_1 - \mathbf{X}'\boldsymbol{\beta}) \\
 &\vdots
 \end{aligned}$$

Kategori ke-j: = j :

$$\begin{aligned}
 P(y = j) &= \Phi(\gamma_j - \mathbf{X}'\boldsymbol{\beta}) - \Phi(\gamma_{j-1} - \mathbf{X}'\boldsymbol{\beta}) \\
 &= 1 - \Phi(\gamma_{j-1} - \mathbf{X}'\boldsymbol{\beta}) \quad (2.18)
 \end{aligned}$$

Greene (2008), menyatakan bahwa metode pendugaan parameter model regresi probit untuk peubah respon multinomial adalah metode estimasi kemungkinan maksimum (*Maximum Likelihood Estimation* atau MLE), dalam hal ini sama dengan model logistik yang telah dibahas sebelumnya.

$$L(\boldsymbol{\beta}) = \sum_{i=1}^n \sum_{j=1}^m Y_{ij} \ln(\Phi_{ij} - \Phi_{ij-1})$$

dengan $\Phi_{ij} = \Phi(\gamma_j - \mathbf{X}'\boldsymbol{\beta})$ dan $\Phi_{ij-1} = \Phi(\gamma_{j-1} - \mathbf{X}'\boldsymbol{\beta})$ (Jackman, 2003)

Pendugaan parameter model probit juga menggunakan iterasi Newton-Raphson seperti pada regresi logistik, yaitu :

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \boldsymbol{\beta}^{(t)} - [\mathbf{H}^{(t)}]^{-1} \mathbf{g}^{(t)} \quad (2.19)$$

Dengan nilai $\mathbf{H}^{(t)}$ dan $\mathbf{g}^{(t)}$ berupa matriks :

$$\mathbf{H}^{(t)} = \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_{jk} \partial \beta_{j'k'}} = \mathbf{X}' \mathbf{V}^{(t)} \mathbf{X}$$

$$\mathbf{g}^{(t)} = \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_{jk}} = \mathbf{X}'(\mathbf{y} - \boldsymbol{\pi}^{(t)})$$

Langkah-langkah iterasi *Newton Raphson* :

1. Menentukan nilai awal dugaan parameter yaitu $\boldsymbol{\beta}^{(t)}$. Menurut Gerald dan Wheatley (1996) nilai $\boldsymbol{\beta}^{(t)}$ digunakan nilai yang berdekatan dengan selisih antara nilai penduga dan dugaan yang diharapkan, hal ini disebabkan agar proses iterasi tidak terlalu banyak. Sehingga untuk mempermudah perhitungan digunakan 0 sebagai nilai dugaan awal iterasi.

2. Membentuk keseluruhan data dalam bentuk matriks.

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \text{ dan } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

3. Menentukan model regresi probit pada persamaan (2.17) yaitu :

$$g_j(\mathbf{x}_i) = \beta_{j0} + \beta_{j1}x_{i1} + \beta_{j2}x_{i2} + \dots + \beta_{jp}x_{ip}$$

4. Menentukan matriks $\mathbf{V}$

$$\mathbf{V}^{(t)} = \begin{bmatrix} \pi_i^{(t)}(1 - \pi_i^{(t)}) & 0 & \dots & 0 \\ 0 & \pi_2^{(t)}(1 - \pi_2^{(t)}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \pi_n^{(t)}(1 - \pi_n^{(t)}) \end{bmatrix}$$

5. Mensubstitusi hasil dari langkah pada persamaan (2.20) dengan

$$\mathbf{H}^{(t)} = \mathbf{X}'\mathbf{V}^{(t)}\mathbf{X} \quad (2.21)$$

dan

$$\mathbf{g}^{(t)} = \mathbf{X}'(\mathbf{y} - \boldsymbol{\pi}^{(t)}) \quad (2.22)$$

$$\boldsymbol{\pi}^{(t)} = \begin{bmatrix} \pi_1^{(t)} \\ \pi_2^{(t)} \\ \vdots \\ \pi_n^{(t)} \end{bmatrix}$$

6. Diperoleh $\hat{\boldsymbol{\beta}}^{(t+1)}$

7. Menentukan selisih $(\hat{\boldsymbol{\beta}}^{(t+1)} - \boldsymbol{\beta}^{(t)})$, jika hasil selisihnya memenuhi kondisi konvergen yaitu bernilai $(1 \times 10^{-7}) \approx 0,0000001$ maka $\hat{\boldsymbol{\beta}}^{(t+1)}$ merupakan hasil dugaannya.

Bila kondisi konvergen belum terpenuhi maka iterasi dilanjutkan terus dengan cara menjadikan nilai $\hat{\boldsymbol{\beta}}^{(t+1)}$ sebagai $\boldsymbol{\beta}^{(t)}$ untuk iterasi

selanjutnya, diulangi mulai langkah ke-3 sampai ke-7 hingga diperoleh nilai β yang konvergen.

2.6 Interpretasi probit

Menurut Widhiarso (2012) koefisien probit (β_{jp}) merupakan pengaruh perubahan satu unit peubah prediktor (x_{ip}) pada peluang normal kumulatif (z) dari peubah respon (y). Dalam hal ini peluang dari y bukan fungsi linier dari z , tetapi fungsi normal kumulatif z . Pengaruh dari perubahan satu unit x pada peluang y tergantung pada kategori peubah prediktor. Sehingga perlu dipilih salah satu kategori peubah prediktor untuk dijadikan titik acuan atau pembanding. Interpretasi koefisien model probit dilakukan dengan melihat tanda dari koefisien probit (β_{jp}). Jika β_{jp} yang diperoleh positif, maka kecenderungan peubah respon kategori 1 ($y=1$) lebih besar pada peubah prediktor kategori 1 ($x=1$) dibandingkan dengan peubah prediktor kategori 0 ($x=0$). Dalam hal ini, hanya peubah prediktor signifikan yang diinterpretasikan.

Menurut Fox (1989) jika peneliti ingin mengetahui faktor-faktor penyebab kecelakaan lalu lintas (x). Kecelakaan yang disebabkan karena kelalaian pengemudi=1 dan yang disebabkan karena pelanggaran lalu lintas=0. Model regresi probit yang terbentuk $\Phi(1,092 + 0,241x_{11})$. Ketika nilai $x_{11} = 1$ (kategori kelalaian pengemudi) dibandingkan dengan kategori 0 (kategori pelanggaran lalu lintas). Maka interpretasinya yakni kecenderungan kecelakaan lalu lintas yang disebabkan oleh kelalaian pengemudi beresiko lebih tinggi dibandingkan kecelakaan yang disebabkan oleh pelanggaran lalu lintas.

2.7 P-value

Menurut Kurniawan (2008) dalam konsep analisis statistika, peneliti harus memutuskan menerima atau menolak H_0 . Untuk itu diperlukan adanya kriteria uji. Kriteria uji yang paling sering digunakan yakni p -value. P -value adalah peluang terkecil sehingga nilai suatu uji statistik yang sedang diamati masih mempunyai arti. P -value lebih banyak digunakan daripada kriteria uji lain seperti tabel distribusi dan selang kepercayaan. Hal ini disebabkan karena P -value memberikan 2 informasi sekaligus, yaitu seberapa baik H_0 harus

ditolak serta peluang terjadinya kejadian pada H_o (dengan asumsi H_o dianggap benar).

P -value dihitung dari peluang perbandingan antara nilai statistik uji dengan statistik yang didapatkan dari tabel. Dimisalkan menghitung nilai P -value untuk uji t , $P\text{-value} = P(|t_{n-2}| \geq t_{hitung})$. Jika nilai P -value sebesar 0,021, hal ini berarti bahwa jika H_o dianggap benar maka kejadian pada H_o hanya akan terjadi sebanyak 21 kali dari 1000 kali percobaan yang sama. Oleh karena kecilnya peluang terjadinya kejadian pada H_o , maka dapat diambil keputusan untuk menolak pernyataan yang ada pada H_o . Sebagai gantinya pernyataan pada H_1 diterima.

P -value dapat juga diartikan sebagai besarnya peluang melakukan kesalahan pada saat memutuskan untuk menolak H_o . Pada umumnya P -value dibandingkan dengan taraf nyata α tertentu, biasanya 0,05 atau 5%. Taraf nyata α diartikan sebagai peluang melakukan kesalahan untuk menyimpulkan bahwa H_o salah, padahal H_o benar. Kesalahan semacam ini dikenal sebagai galat atau kesalahan jenis I. Misal α yang digunakan adalah 0,05, jika P -value sebesar 0,021 ($<0,05$) maka H_o ditolak. Hal ini disebabkan karena keputusan menolak H_o (menganggap H_o salah), kemungkinan kesalahan masih lebih kecil dari 0,05, dimana 0,05 merupakan batas maksimal dalam membuat keputusan.

2.8 Nonmultikolinieritas

Nonmultikolinearitas merupakan asumsi dalam regresi yang berarti antara peubah prediktor tidak terjadi hubungan mendekati sempurna atau hubungan sempurna. Jika dalam suatu model regresi terdapat hubungan antara dua atau lebih peubah prediktor, hal ini berarti terdapat hubungan linier yang tinggi di antara peubah-peubah prediktor dalam model. Penduga yang dihasilkan akan bersifat bias dan memiliki ragam mendekati tak hingga. Seperti pada regresi linier, model logistik dan model probit juga sensitif terhadap nonmultikolinearitas (Hosmer dan Lemeshow, 1989).

Apabila terdapat sebuah data dengan peubah prediktor $x_1, x_2, \dots, x_p$ bersifat kategorik maka untuk mengetahui ada tidaknya multikolinearitas menggunakan uji *Pearson*. Adapun uji *Pearson* yang digunakan yaitu :

$$\chi^2_{hitung} = \sum \sum \frac{(|n_{jp} - e_{jp}|)^2}{e_{jp}} \sim \chi^2_{\alpha} \quad (2.22)$$

dengan:

n_{jp}	= banyaknya pengamatan peubah respon kategori ke-j pada peubah prediktor ke-p
e_{jp}	= frekuensi harapan pengamatan peubah respon kategori ke-j pada peubah prediktor ke-p
$e_{ij} = \frac{n_j n_p}{n..}$	
$n..$	= jumlah pengamatan
n_j	= banyaknya pengamatan pada peubah respon ke-j
n_p	= banyaknya pengamatan pada peubah prediktor ke-p

Jika nilai $P[\chi^2_{hitung} < \chi^2_{\alpha}] > \alpha$ menunjukkan bahwa antar peubah prediktor terdapat nonmultikolinearitas.

2.9 Uji Signifikansi Model

Pengujian signifikansi koefisien β dari model dapat dilakukan secara serentak (*overall*) maupun parsial. Hal ini dimaksudkan untuk menentukan apakah peubah-peubah prediktor dalam model mempunyai pengaruh nyata terhadap peubah respon. Pengujian signifikansi dilakukan sebagai berikut :

1. Pengujian signifikansi secara serentak (*overall*)

Pengujian ini dilakukan sebagai upaya untuk memeriksa pengaruh koefisien regresi peubah prediktor secara bersamaan. Pengujian secara serentak menggunakan metode *likelihood-ratiotest*. *Likelihood-ratiotest* adalah metode pengujian signifikansi model dengan membandingkan *likelihood* untuk model lengkap (L_1) dan *likelihood* untuk model yang semua parameternya sama dengan nol (L_0). Adapun statistik uji untuk *Likelihood-ratiotest* adalah sebagai berikut, (Hosmer dan Lemeshow, 1989) :

Hipotesis :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

H_1 : minimal ada satu $\beta_k \neq 0$; $k = 1, 2, \dots, p$ p = banyak peubah prediktor dalam model.

$$G^2 = -2 \ln \left(\frac{L_0}{L_1} \right) = -2 (\ln(L_0) - \ln(L_1)) \quad (2.23)$$

dengan :

L_0 = Nilai *likelihood* model regresi tanpa peubah predictor

$$L_0 = \left(\frac{n_{j1}}{n}\right)^{n_{1j}} \left(\frac{n_{j2}}{n}\right)^{n_{2j}} \dots \left(\frac{n_{jp}}{n}\right)^{n_{jp}}$$

n_{jp} = Banyaknya peubah respon kategori ke- j pada peubah prediktor kategori ke- p

L_1 = Nilai *likelihood* model regresi dengan peubah prediktor

$$L_1 = \prod_{i=1}^n (\pi_1(x))^{y_{j1}} (\pi_2(x))^{y_{j2}} \dots (\pi_j(x))^{y_{jp}}$$

y_{jp} = Banyaknya pengamatan peubah respon kategori ke- j pada peubah prediktor ke- p

Tolak H_0 bila $G^2 > \chi^2_{(p;\alpha)}$ dengan p adalah jumlah peubah prediktor dalam model atau p -value kurang dari α . Hal ini berarti peubah prediktor di dalam model secara serentak berpengaruh terhadap peubah respon.

2. Pengujian signifikansi secara parsial

Pengujian parameter yang digunakan adalah uji Wald. Pengujian ini dimaksudkan untuk melihat apakah suatu peubah prediktor layak masuk ke dalam model. Hipotesis yang digunakan adalah :

$H_0 : \beta_k = 0$; (tidak ada pengaruh antara peubah prediktor ke- p dengan peubah respon).

$H_1 : \beta_k \neq 0$; (ada pengaruh antara peubah prediktor ke- p dengan peubah respon).

Dengan: $k = 1, 2, \dots, p$; p adalah banyak peubah prediktor dalam model dengan statistik uji (Hosmer dan Lemeshow, 1989):

$$|W| = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \quad (2.24)$$

$$SE(\hat{\beta}_k) = \sqrt{var(\hat{\beta}_k)}$$

dengan :

$\hat{\beta}_k$	= penduga β_k
$SE(\hat{\beta}_k)$	= standart error dari penduga β_k

Statistik Wald mengikuti sebaran χ^2 , H_0 ditolak jika $W > \chi^2_\alpha$ atau p -value $< \alpha$, sehingga dapat disimpulkan bahwa peubah

prediktor secara parsial (berdiri sendiri) berpengaruh pada peubah respon.

2.10 Uji Kelayakan Model (*Goodness Of Fit*)

Menurut Oltman dan Lackritz (1991), uji khi-kuadrat digunakan untuk beberapa tujuan, yaitu :

1. Menguji ragam sebuah populasi
2. Menguji kesesuaian model (*Goodness of fit*)
3. Menguji kebebasan antara dua peubah atau lebih

Dalam hal ini, uji khi-kuadrat digunakan untuk pengujian signifikansi, yang bertujuan menguji kesesuaian model untuk mengetahui apakah model yang dihasilkan sudah sesuai dengan data. Statistik uji yang digunakan untuk melihat nilai *Goodness of fit* dalam analisis regresi logistik dan probit adalah Uji *Pearson*, dengan hipotesis :

H_0 : model sesuai dengan data

H_1 : model tidak sesuai dengan data

$$\text{Apabila } H_0 \text{ benar : } \sum_{i=0}^j \frac{(y_{ab} - n_a \pi_{ab})^2}{n_a \pi_{ab}} \sim \chi^2_{p(j-1)-h} \quad (2.25)$$

dengan :

y_{ab}	= Jumlah peubah respon pada kategori ke- b dan peubah prediktor pada kategori ke- a
π_{ab}	= Peluang peubah respon kategori ke- b dan peubah prediktor kategori ke- a
n_a	= Banyak pengamatan pada peubah prediktor kategori ke- a
p	= Banyak peubah prediktor
j	= Banyak kategori peubah respon
h	= Banyak parameter yang diduga dalam model

Jika nilai statistik uji pada persamaan (2.25) lebih dari dan sama dengan nilai kritis khi-kuadrat atau *p-value* kurang dari sama dengan 0,05 keputusan yang akan diambil adalah menerima H_0 atau berarti model yang digunakan telah sesuai dengan data (Hosmer dan Lemeshow, 1989).

2.11 Kriteria Model Terbaik

Penelitian ini membandingkan model regresi logistik dan probit dengan peubah respon multinomial berdasarkan nilai R^2 *Mc.Fadden*. R^2 *Mc.Fadden* adalah indikator model menggunakan nilai *likelihood-ratio* yang didasarkan pada nilai *likelihood* model penuh yang mengandung semua parameter (L_1) dengan model yang semua parameternya samadengan nol (L_0) :

$$\bar{R}_{MCF} = 1 - \frac{\ln L_1}{\ln L_0} \quad (2.26)$$

Model terbaik memiliki nilai R^2 *Mc.Fadden* terbesar. Semakin besar nilai R^2 *Mc.Fadden* maka semakin baik model menjelaskan data (Fujimoto, 2003).

2.12 Uji t Berpasangan

Uji t berpasangan merupakan metode untuk menguji perbedaan rata-rata antara dua kelompok data yang berpasangan. Sama seperti uji t tidak berpasangan, uji t berpasangan memiliki asumsi yang harus dipenuhi, yaitu :

1. Data berdistribusi normal.
2. Kedua kelompok data berpasangan (dependen)
3. Peubah yang dihubungkan berbentuk numerik atau kategorik (dengan hanya 2 kelompok).

Rumus yang digunakan, sebagai berikut :

$$t = \frac{d}{sd/\sqrt{n}} \sim t_{(n-1, \alpha/2)} \quad (2.27)$$

dengan hipotesis :

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

dengan :

d = Rata-rata deviasi (selisih sampel sebelum dan sampel sesudah)

sd = Standar deviasi dari d (selisih sampel sebelum dan sampel sesudah)

n = Banyaknya sampel

Jika nilai statistik uji pada persamaan (2.27) lebih dari sama dengan nilai t tabel atau *p-value* kurang dari sama dengan 0,05 maka keputusan yang akan diambil adalah tolak H_0 atau terdapat perbedaan yang nyata antar dua kelompok data (Yitnosumarto, 1990).

UNIVERSITAS BRAWIJAYA

