# BAB III

# MODEL REGRESI DENGAN VARIABEL *DEPENDENT* ORDINAL YANG DIBENTUK OLEH SUATU VARIABEL KONTINU YANG TIDAK DIKETAHUI NILAINYA

Pada bab III ini akan dibahas mengenai model regresi dimana data variabel *dependent* yang diketahui merupakan data ordinal yang dibentuk oleh suatu variabel kontinu yang tidak diketahui nilainya.

Umumnya, nilai dari variabel penjelas dan variabel *dependent* pada model regresi didapatkan dari pengukuran. Apabila variabel *dependent* berupa variabel kategorik, maka model regresi logistik dapat diterapkan. Data yang dibutuhkan untuk melakukan analisis logistik adalah data dari variabel penjelas dan data dari variabel *dependent* yang bersifat kategorik. Jika data variabel *dependent* yang diketahui merupakan data dari variabel ordinal yang dibentuk dari suatu variabel kontinu yang nilainya tidak diketahui dan variabel kontinu tersebut diperhitungkan sebagai variabel *dependent* dalam model, maka analisis regresi logistik tidak dapat digunakan.

Karena itu dalam tulisan ini akan dicari suatu model regresi antara variabel penjelas dengan variabel *dependent* kontinu dimana data yang dimiliki adalah data dari variabel penjelas dan data ordinal yang dibentuk dari variabel kontinu tersebut. Sedangkan nilai dari variabel *dependent* kontinu tidak diketahui.

### 3.1 Model

Misalkan $Y$ merupakan variabel *dependent* yang berupa variabel kategorik bersifat ordinal dengan $M$ kategori yaitu misalkan $C_1, C_2, ..., C_M$ dan berasal dari suatu variabel kontinu $Z$ yang nilainya tidak diketahui, dimana :

$$
\begin{aligned}
& y \in C_1 \quad \textit{jika} \quad \gamma_0 < Z \le \gamma_1 \\
& y \in C_2 \quad \textit{jika} \quad \gamma_1 < Z \le \gamma_2 \\
& y \in C_3 \quad \textit{jika} \quad \gamma_2 < Z \le \gamma_3 \\
& \qquad \vdots \\
& y \in C_M \quad \textit{jika} \quad \gamma_{M-1} < Z < \gamma_M
\end{aligned}
$$

$\gamma_0, \gamma_1, ..., \gamma_M$ adalah $M+1$ bilangan real yang memenuhi

$\gamma_0 < \gamma_1 < \gamma_2 < \gamma_3 < ... < \gamma_M$ dengan $\gamma_0 = 0$ dan $\gamma_M = \infty$.

Misalkan hubungan antara variabel $Z$ sebagai variabel *dependent* dan $X$ ( $X_1, X_2, ..., X_p$ ) sebagai variabel penjelas dapat dituliskan dalam model sebagai berikut :

$$\mathbf{Z} = \mathbf{X\beta} + \mathbf{e} \;;\; \text{dimana } \mathbf{e} \sim N\left(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I}\right)$$

$$\mathbf{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \qquad \mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}.$$

Tujuan analisis ini adalah untuk mencari taksiran $\boldsymbol{\beta}$. Misal $Y_j$ adalah nilai dari $Y$ ( variabel kategori ordinal yang dibentuk dari variabel $Z$ ) untuk individu ke-*j*.

Definisikan :

$$Y_{jk} = \begin{cases} 1, Y_j \in C_k \\ 0, \text{yang lainnya} \end{cases} \quad ; \quad j = 1,2,...,n \quad k = 1,2,...,M$$

Nilai $Y_{jk}$ untuk seluruh pengamatan dapat dituliskan dalam bentuk matriks sebagai berikut :

$$\mathbf{Y} = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1M} \\ Y_{21} & Y_{22} & \cdots & Y_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nM} \end{bmatrix}$$

Karena :

1. $\mathbf{Z = X\boldsymbol{\beta} + e}$ ; dimana $\mathbf{e} \sim N\left(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I}\right)$.

2. $Y_j \in C_k$ *apabila* $\gamma_{k-1} < Z_j \leq \gamma_k$ ; $j = 1,2,...,n$

Fungsi probabilitas dari Y dapat ditulis sebagai berikut :

$$\Pr(Y_{jk}=1)=\Pr(Y_j \in C_k)$$

$$\begin{aligned}
&= \Pr(\gamma_{k-1} < Z_j \le \gamma_k) \\
&= \Pr(\gamma_{k-1} < \mathbf{x\beta + e} \le \gamma_k) \\
&= \Pr(\gamma_{k-1} < \sum \beta_i X_{ij} + e_j \le \gamma_k) \\
&= \Pr(\gamma_{k-1} - \sum \beta_i X_{ij} < e_j \le \gamma_k - \sum \beta_i X_{ij}) \\
&= \Pr(\frac{\gamma_{k-1} - \sum \beta_i X_{ij}}{\sigma} < \frac{e_j}{\sigma} \le \frac{\gamma_k - \sum \beta_i X_{ij}}{\sigma})
\end{aligned}$$

dimana $X_{0j}=1$, dengan $\sigma$ merupakan deviasi standar dari *error* $e$.

Karena diasumsikan $\mathbf{e}$ berdistribusi multivariat normal, dapat diperoleh :

$$\Pr\{Y_{jk}=1\}=\Pr\{Y_j \in C_k\}=\Phi\left[\frac{\gamma_k-\sum\beta_i X_{ij}}{\sigma}\right]-\Phi\left[\frac{\gamma_{k-1}-\sum\beta_i X_{ij}}{\sigma}\right]$$

dengan $\Phi$ adalah *cdf* dari distribusi normal standar.

Tanpa kehilangan perumuman, diasumsikan $\sigma=1$ maka dapat dituliskan :

$$\Pr\{Y_{jk}=1\}=\Phi\left[\gamma_k-\sum\beta_i X_{ij}\right]-\Phi\left[\gamma_{k-1}-\sum\beta_i X_{ij}\right]$$

Misalkan Q menyatakan jumlah parameter yang akan ditaksir.

$$\begin{aligned}
Q &= (M+1)+(p+1) \\
&= (M+p+2)
\end{aligned}$$

Karena telah ditentukan $\gamma_0=0$ dan $\gamma_M=\infty$, maka banyak parameter yang akan ditaksir menjadi :

$$\begin{aligned}
Q &= (M+p+2)-2 \\
&= M+p
\end{aligned}$$

Nilai-nilai $\left(\gamma_1,...,\gamma_{M-1},\beta_0,\beta_1,...,\beta_p\right)$ akan ditaksir dengan menggunakan penaksiran *maximum likelihood*.

### 3.2 Penaksiran Parameter pada Model

Sebagaimana telah disebutkan sebelumnya, pada sub bab ini akan dibahas mengenai penaksiran parameter $\left(\gamma_1,...,\gamma_{M-1},\beta_0,\beta_1,...,\beta_p\right)$ dengan menggunakan penaksiran *maximum likelihood*.

Untuk mempermudah notasi, sebut :

$$Z_{j,k} = \gamma_k - \sum_{i=0}^{p} \beta_i X_{i,j}$$

$$\Phi_{j,k} = \Phi\left(Z_{j,k}\right)$$

Dari $\Pr\left\{Y_{jk} = 1\right\} = \Phi\left[\gamma_k - \sum \beta_i X_{ij}\right] - \Phi\left[\gamma_{k-1} - \sum \beta_i X_{ij}\right]$ dapat ditulis sebagai berikut :

$$\Pr\left\{Y_{jk} = 1\right\} = \Phi_{j,k} - \Phi_{j,k-1}$$

Fungsi *likelihood* L dari *Y* adalah sebagai berikut :

$$\begin{aligned} L &= L\left(\beta_0,\beta_1,...,\beta_P,\gamma_1,...,\gamma_{M-1}\right) \\ &= \prod_{j=1}^{n}\prod_{k=1}^{M} \Pr(Y_{jk} = 1) \qquad (3.2.1) \\ &= \prod_{j=1}^{n}\prod_{k=1}^{M} \left(\Phi_{j,k} - \Phi_{j,k-1}\right)^{Y_{j,k}} \end{aligned}$$

Penaksir untuk parameter-parameter tersebut diperoleh dengan cara memaksimumkan fungsi *likelihood*. Akan tetapi, penaksir yang sama dapat juga diperoleh dengan memaksimumkan fungsi log *likelihood*, maka dari persamaan 3.2.1 diperoleh :

$$\begin{aligned} L^* &= \log L \\ &= \sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k} \log\left(\Phi_{j,k} - \Phi_{j,k-1}\right) \end{aligned} \qquad (3.2.2)$$

Fungsi log *likelihood* L* merupakan fungsi dari $\left(\gamma_1,...,\gamma_{M-1},\beta_0,\beta_1,...,\beta_p\right)$ dan akan dicari nilai taksiran dari $\left(\gamma_1,...,\gamma_{M-1},\beta_0,\beta_1,...,\beta_p\right)$ yang memaksimumkan L*.

Definisikan $N_{j,k} = \dfrac{1}{\sqrt{2\pi}}\exp\left(-Z_{j,k}^2/2\right)$ *untuk* $j = 1,2,...,n$ dan didefinisikan pula $\delta_{i,j}$ Kroneker Delta, sebagai berikut :

$$\delta_{i,j} = \begin{cases} 1 & jika\ \ i = j \\ 0 & jika\ \ i \neq j \end{cases}$$

Maka didapatkan persamaan berikut :

$$\begin{aligned} \frac{\partial}{\partial \beta_u}\Phi_{j,k} &= \frac{\partial \Phi_{j,k}}{\partial Z_{j,k}} \cdot \frac{\partial Z_{j,k}}{\partial \beta_u} \\ &= \left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{Z_{j,k}{}^2}{2}\right)\right] \cdot -X_{u,j} \\ &= N_{j,k} \cdot -X_{u,j} \\ &= -N_{j,k}X_{u,j}\ , u = 0,1,2,...,P \end{aligned} \qquad (3.2.3)$$

$$\frac{\partial}{\partial \gamma_u}\Phi_{j,k} = \frac{\partial \Phi_{j,k}}{\partial Z_{j,k}} \cdot \frac{\partial Z_{j,k}}{\partial \gamma_u}$$

$$= \left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{Z_{j,k}{}^2}{2}\right)\right] \cdot 1 \quad , \quad ketika\ u = k \qquad (3.2.4)$$

$$= N_{j,k}\delta_{u,k} \ , u = 1, 2, 3, ..., M-1$$

dan

$$\frac{\partial}{\partial \beta_u}N_{j,k} = \frac{\partial N_{j,k}}{\partial Z_{j,k}} \cdot \frac{\partial Z_{j,k}}{\partial \beta_u}$$

$$= \left\{-\frac{2Z_{j,k}}{2} \cdot \left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{Z_{j,k}{}^2}{2}\right)\right]\right\} \cdot -X_{u,j} \qquad (3.2.5)$$

$$= \left\{Z_{j,k}N_{j,k}\right\}X_{u,j}$$

$$= Z_{j,k}N_{j,k}X_{u,j} \quad , u = 0, 1, 2, ..., P$$

$$\frac{\partial}{\partial \gamma_u}N_{j,k} = \frac{\partial N_{j,k}}{\partial Z_{j,k}} \cdot \frac{\partial Z_{j,k}}{\partial \gamma_u}$$

$$= \left\{-\frac{2Z_{j,k}}{2} \cdot \left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{Z_{j,k}{}^2}{2}\right)\right]\right\} \bullet 1 \quad , \quad ketika\ u = k$$

$$= \left\{-Z_{j,k}N_{j,k}\right\}\delta_{u,k}$$

$$= -Z_{j,k}N_{j,k}\delta_{u,k} \quad , u = 1, 2, 3, ..., M-1$$

(3.2.6)

Persamaan-persamaan 3.2.3 – 3.2.6 dapat digunakan untuk menyelesaikan turunan parsial dari persamaan 3.2.2 terhadap masing-

masing parameter. Metode untuk memaksimumkan fungsi *L** adalah dengan menggunakan turunan pertama. Ada sebanyak Q parameter yang akan ditaksir, maka harus diselesaikan Q persamaan berikut secara simultan :

$$
\begin{aligned}
\frac{\partial L^*}{\partial \beta_u} &= \frac{\partial L^*}{\partial \Phi_{j,k}} \cdot \frac{\partial \Phi_{j,k}}{\partial \beta_u} \\
&= \left[\sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k} \frac{1}{\Phi_{j,k}-\Phi_{j,k-1}}\right] \cdot \left[-N_{j,k}X_{u,j} - \left(-N_{j,k-1}X_{u,j}\right)\right] \\
&= \left[\sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k} \frac{1}{\Phi_{j,k}-\Phi_{j,k-1}}\right] \cdot \left[\left(N_{j,k-1}-N_{j,k}\right)X_{u,j}\right] \\
&= \sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k} \left[\frac{N_{j,k-1}-N_{j,k}}{\Phi_{j,k}-\Phi_{j,k-1}}\right] X_{u,j} \qquad , u = 0,1,2,\ldots,p
\end{aligned}
\tag{3.2.7}
$$

dan

$$
\begin{aligned}
\frac{\partial L^*}{\partial \gamma_u} &= \frac{\partial L^*}{\partial \Phi_{j,k}} \cdot \frac{\partial \Phi_{j,k}}{\partial \gamma_u} \\
&= \left[\sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k} \frac{1}{\Phi_{j,k}-\Phi_{j,k-1}}\right] \cdot \left[N_{j,k}\delta_{u,k} - N_{j,k-1}\delta_{u,k-1}\right] \\
&= \left[\sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k} \frac{1}{\Phi_{j,k}-\Phi_{j,k-1}}\right] \cdot \left[N_{j,k}\delta_{k,u} - N_{j,k-1}\delta_{k-1,u}\right] \\
&= \sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k} \left[\frac{N_{j,k}\delta_{k,u} - N_{j,k-1}\delta_{k-1,u}}{\Phi_{j,k}-\Phi_{j,k-1}}\right] , u = 1,2,3,\ldots,M-1
\end{aligned}
\tag{3.2.8}
$$

Taksiran *maximum likelihood* dari Q parameter tersebut diperoleh dengan menyelesaikan $\frac{\partial L^*}{\partial \beta_u} = 0$ dan $\frac{\partial L^*}{\partial \gamma_u} = 0$.

Persamaan 3.2.7 dan 3.2.8 tidak linear terhadap parameter yang tidak diketahui yaitu $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$, sehingga tidak dapat diselesaikan secara analitis. Oleh karena itu akan digunakan penurunan dengan metode Newton Raphson.

Dalam metode Newton Raphson dibutuhkan matriks turunan kedua. Sehingga dari persamaan 3.2.7 dan 3.2.8 akan dicari turunan parsialnya terhadap $\boldsymbol{\beta}$ dan terhadap $\boldsymbol{\gamma}$, sehingga diperoleh turunan kedua sebagai berikut :

$$
\begin{aligned}
\frac{\partial^2 L^*}{\partial\beta_u\beta_v} &= \frac{\partial}{\partial\beta_u}\left(\frac{\partial L^*}{\partial\beta_v}\right)\\
&= \frac{\partial}{\partial\beta_u}\left(\sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k}\left[\frac{N_{j,k-1}-N_{j,k}}{\Phi_{j,k}-\Phi_{j,k-1}}\right]X_{v,j}\right)\\
&= \sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k}\left[\frac{\left(\left(\left(Z_{j,k-1}N_{j,k-1}X_{u,j}\right)-\left(Z_{j,k}N_{j,k}X_{u,j}\right)\right)\Phi_{j,k}-\Phi_{j,k-1}\right)-\left(\left(-N_{j,k}X_{u,j}-\left(-N_{j,k-1}X_{u,j}\right)\right)\left(N_{j,k-1}-N_{j,k}\right)\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}\right]X_{v,j}\\
&= \sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k}\left[\frac{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)X_{u,j}\left(\left(Z_{j,k-1}N_{j,k-1}\right)-\left(Z_{j,k}N_{j,k}\right)\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}-\frac{X_{u,j}\left(-N_{j,k}-(-N_{j,k-1})\right)\left(N_{j,k-1}-N_{j,k}\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}\right]X_{v,j}\\
&= \sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k}\left[\frac{X_{u,j}\left(\Phi_{j,k}-\Phi_{j,k-1}\right)\left(N_{j,k-1}Z_{j,k-1}-N_{j,k}Z_{j,k}\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}-\frac{X_{u,j}\left(N_{j,k-1}-N_{j,k}\right)\left(N_{j,k-1}-N_{j,k}\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}\right]X_{v,j}\\
&= \sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k}\left[\frac{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)\left(N_{j,k-1}Z_{j,k-1}-N_{j,k}Z_{j,k}\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}-\frac{\left(N_{j,k-1}-N_{j,k}\right)\left(N_{j,k-1}-N_{j,k}\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}\right]X_{u,j}X_{v,j}\\
&= \sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k}\left[\frac{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)\left(N_{j,k-1}Z_{j,k-1}-N_{j,k}Z_{j,k}\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}-\frac{\left(N_{j,k-1}-N_{j,k}\right)^2}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}\right]X_{u,j}X_{v,j}
\end{aligned}
$$

(3.2.9)

$$
\begin{aligned}
\frac{\partial^2 L^*}{\partial \beta_u \gamma_v} &= \frac{\partial^2 L^*}{\partial \gamma_v \beta_u} \\
&= \frac{\partial}{\partial \gamma_v}\left(\frac{\partial L^*}{\partial \beta_u}\right) \\
&= \frac{\partial}{\partial \gamma_v}\left(\sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k}\left[\frac{N_{j,k-1}-N_{j,k}}{\Phi_{j,k}-\Phi_{j,k-1}}\right]X_{u,j}\right) \\
&= \sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k}\left[\frac{\left(\left(-Z_{j,k-1}N_{j,k-1}\delta_{v,k-1}\right)-\left(-Z_{j,k}N_{j,k}\delta_{v,k}\right)\right)\left(\Phi_{j,k}-\Phi_{j,k-1}\right)-\left(\left(N_{j,k}\delta_{v,k}\right)-\left(N_{j,k-1}\delta_{v,k-1}\right)\right)\left(N_{j,k-1}-N_{j,k}\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}\right]X_{u,j} \\
&= \sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k}\left[\frac{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)\left(\left(Z_{j,k}N_{j,k}\delta_{v,k}\right)-\left(Z_{j,k-1}N_{j,k-1}\delta_{v,k-1}\right)\right)-\left(N_{j,k-1}-N_{j,k}\right)\left(\left(N_{j,k}\delta_{v,k}\right)-\left(N_{j,k-1}\delta_{v,k-1}\right)\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}\right]X_{u,j} \\
&= \sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k}\left[\frac{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)\left(\left(Z_{j,k}N_{j,k}\delta_{k,v}\right)-\left(Z_{j,k-1}N_{j,k-1}\delta_{k-1,v}\right)\right)-\left(N_{j,k-1}-N_{j,k}\right)\left(\left(N_{j,k}\delta_{k,v}\right)-\left(N_{j,k-1}\delta_{k-1,v}\right)\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}\right]X_{u,j} \\
&= \sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k}\left[\frac{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)\left(N_{j,k}Z_{j,k}\delta_{k,v}-N_{j,k-1}Z_{j,k-1}\delta_{k-1,v}\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}-\frac{\left(N_{j,k-1}-N_{j,k}\right)\left(N_{j,k}\delta_{k,v}-N_{j,k-1}\delta_{k-1,v}\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}\right]X_{u,j}
\end{aligned}
$$

(3.2.10)

$$
\begin{aligned}
\frac{\partial^2 L^*}{\partial \gamma_u \gamma_v} &= \frac{\partial}{\partial \gamma_u}\left(\frac{\partial L^*}{\partial \gamma_v}\right) \\
&= \frac{\partial}{\partial \gamma_u}\left(\sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k}\left[\frac{N_{j,k}\delta_{k,v}-N_{j,k-1}\delta_{k-1,v}}{\Phi_{j,k}-\Phi_{j,k-1}}\right]\right) \\
&= \sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k}\left[\frac{\left(\left(-Z_{j,k}N_{j,k}\delta_{u,k}\delta_{k,v}\right)-\left(-Z_{j,k-1}N_{j,k-1}\delta_{u,k-1}\delta_{k-1,v}\right)\right)\left(\Phi_{j,k}-\Phi_{j,k-1}\right)-\left(\left(N_{j,k}\delta_{u,k}\right)-\left(N_{j,k}\delta_{u,k-1}\right)\right)\left(N_{j,k}\delta_{k,v}-N_{j,k-1}\delta_{k-1,v}\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}\right] \\
&= \sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k}\left[\frac{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)\left(\left(Z_{j,k-1}N_{j,k-1}\delta_{u,k-1}\delta_{k-1,v}\right)-\left(Z_{j,k}N_{j,k}\delta_{u,k}\delta_{k,v}\right)\right)-\left(N_{j,k}\delta_{u,k}-N_{j,k}\delta_{u,k-1}\right)\left(N_{j,k}\delta_{k,v}-N_{j,k-1}\delta_{k-1,v}\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}\right] \\
&= \sum_{j=1}^{n}\sum_{k=1}^{M} Y_{j,k}\left[\frac{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)\left(N_{j,k-1}Z_{j,k-1}\delta_{u,k-1}\delta_{v,k-1}-N_{j,k}Z_{j,k}\delta_{u,k}\delta_{v,k}\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}-\frac{\left(N_{j,k}\delta_{u,k}-N_{j,k-1}\delta_{u,k-1}\right)\left(N_{j,k}\delta_{v,k}-N_{j,k-1}\delta_{v,k-1}\right)}{\left(\Phi_{j,k}-\Phi_{j,k-1}\right)^2}\right]
\end{aligned}
$$

(3.2.11)

Untuk mendapatkan parameter-parameter $\left(\gamma_1,...,\gamma_{M-1},\beta_0,\beta_1,...,\beta_p\right)$ digunakan metode Newton-Raphson sebagai berikut :

1. Pilih taksiran awal dari $\boldsymbol{\theta}^{(0)}=\left(\boldsymbol{\beta}^{(0)},\boldsymbol{\gamma}^{(0)}\right)$.
2. Tentukan taksiran dari $\boldsymbol{\theta}$ pada iterasi ke-m+1 (m=0,1,2,3,...), yaitu $\hat{\boldsymbol{\theta}}^{(m+1)}$, secara iteratif menggunakan formula :

   $$\hat{\boldsymbol{\theta}}^{(m+1)}=\hat{\boldsymbol{\theta}}^{(m)}-\left(\mathbf{H}^{-1}\left(\hat{\boldsymbol{\theta}}^{(m)}\right)\mathbf{f}'\left(\hat{\boldsymbol{\theta}}^{(m)}\right)\right)$$

   dengan :

   $\hat{\boldsymbol{\theta}}^{(m)}$ : adalah taksiran dari $\boldsymbol{\theta}$ pada iterasi ke-m.

   $\mathbf{f}'\left(\hat{\boldsymbol{\theta}}^{(m)}\right)$ : adalah vektor turunan parsial pertama dari $L^*(\boldsymbol{\theta})$ dihitung pada $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(m)}$.

   $\mathbf{H}\left(\hat{\boldsymbol{\theta}}^{(m)}\right)$ : adalah matriks turunan parsial kedua dari $L^*(\boldsymbol{\theta})$ dihitung pada $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(m)}$.
3. Hentikan proses iterasi apabila $\hat{\boldsymbol{\theta}}^{(m+1)}\approx\hat{\boldsymbol{\theta}}^{(m)}$ (misalkan $\left\|\boldsymbol{\theta}^{(m+1)}-\hat{\boldsymbol{\theta}}^{(m)}\right\|<10^{-5}$). Kemudian ambil $\hat{\boldsymbol{\theta}}^{(m+1)}$ sebagai taksiran dari $\hat{\boldsymbol{\theta}}$.

Untuk mendapatkan hasil akhir dari proses iterasi ini, pada umumnya digunakan software matematika seperti Matlab atau yang lainnya.

## 3.3 Pengujian Hipotesis

Setelah diperoleh taksiran parameter-parameter dalam model, selanjutnya dilakukan pengujian hipotesis terhadap parameter dalam model. Pengujian terbagi ke dalam dua bagian yaitu pengujian kegunaan model dan pengujian terhadap setiap parameter dalam model.

### 3.3.1 Pengujian kegunaan model

Pengujian signikansi model bertujuan untuk menguji apakah ada hubungan linear antara variabel *dependent* dan variabel penjelas $X_1, X_2, ..., X_p$ dalam model. Pengujian dilakukan dengan menggunakan uji rasio likelihood.

Misalkan parameter $\left(\gamma_1, ..., \gamma_{M-1}, \beta_0, \beta_1, ..., \beta_p\right)$ berada dalam suatu ruang parameter $\Omega$ dan parameter $\left(\gamma_1, ..., \gamma_{M-1}, \beta_0, \beta_1, ..., \beta_p\right)$ berada dalam ruang parameter $\omega$ yang merupakan subset dari $\Omega$.

Misal $L(\hat{\omega})$ adalah nilai dari fungsi likelihood dibawah $H_0$ dan $L(\hat{\Omega})$ adalah nilai dari fungsi likelihood secara keseluruhan dimana $\hat{\omega}$ merupakan taksiran *maximum likelihood* untuk $\omega$ dan $\hat{\Omega}$ merupakan merupakan taksiran *maximum likelihood* untuk $\Omega$.

Hipotesis sebagai berikut :

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

$$H_1 : \exists \beta_i \neq 0 \, , \quad untuk\ i = 1,2,...,p$$

Jika $H_0$ ditolak, berarti paling tidak ada satu dari variabel penjelas $X_1, X_2, ..., X_p$ mempunyai kontribusi yang signifikan pada model. Atau dengan perkataan lain setidaknya ada satu variabel penjelas yang dapat digunakan untuk memprediksi variabel *dependent*.

Misal $r_k$ menyatakan jumlah seluruh observasi dengan $Y \in C_k$.

$$r_k = \sum_{j=1}^{n} Y_{j,k} \, ,$$

Dibawah hipotesis $H_0$ benar diperoleh :

$$\Pr(Y_{j,k} = 1) = \Pr(Y_j \in C_k) = \Phi\left(\gamma_k - \beta_0\right) - \Phi\left(\gamma_{k-1} - \beta_0\right) = \frac{r_k}{n} \;\; ; \; k = 1,2,...,M$$

dan fungsi likelihood sebagai berikut :

$$L\left(\hat{\omega}\right) = \prod_{k=1}^{M} \left(\frac{r_k}{n}\right)^{r_k} .$$

Statistik uji yang digunakan untuk pengujian di atas adalah statistik uji rasio likelihood :

$$\begin{aligned} G &= -2 \log \frac{L\left(\hat{\omega}\right)}{L\left(\hat{\Omega}\right)} \\ &= -2\left(\sum_{k=1}^{M} r_k \log r_k - n \log n - \log L\left(\hat{\Omega}\right)\right) \end{aligned}$$

dengan $G$ akan mendekati distribusi $\chi^2$ dan derajat bebasnya sebesar p.

Aturan keputusan : $H_0$ akan ditolak pada tingkat signifikansi $\alpha$ apabila

$G > \chi^2_{\alpha;p}$ .

Apabila keputusannya $H_0$ ditolak menyatakan bahwa pada tingkat signifikansi $\alpha$ ada paling sedikit satu parameter diantara $(\beta_1, \beta_2, ..., \beta_p)$ yang tidak sama dengan nol, yang artinya paling tidak ada satu dari variabel penjelas $X_1, X_2, ..., X_p$ mempunyai kontribusi yang signifikan pada model. Atau dengan perkataan lain setidaknya ada satu variabel penjelas yang dapat digunakan untuk memprediksi variabel *dependent* dalam model pada tingkat signifikansi $\alpha$ .

### 3.3.2 Pengujian terhadap masing-masing koefisien regresi ( $\beta$ )

Setelah dilakukan pengujian model, selanjutnya dilakukan pengujian untuk tiap-tiap parameter dalam model. Pengujian ini dilakukan dengan tujuan untuk menentukan variabel penjelas mana yang signifikan memberikan kontribusi pada model atau dengan perkataan lain memberikan kontribusi yang signifikan dalam memprediksi variabel *dependent*.

H adalah matriks berukuran QxQ dengan entri berupa turunan parsial kedua dari log-*likelihood* dan $H^{-1}$ inverse dari matriks H. Matriks variansi-kovariansi dapat ditaksir dari $-[H]^{-1}$ (bukti dapat dilihat dalam lampiran). Taksiran matriks variansi-kovariansi dinyatakan dalam bentuk $\hat{V}(\hat{\theta})$ (dapat dilihat dalam lampiran). Elemen diagonal utama taksiran matriks variansi-kovariansi $\hat{V}(\hat{\theta})$ menunjukkan taksiran variansi dari $\hat{\beta}_i$ yang dinyatakan dengan $Var\left(\hat{\beta}_i\right)$ dan juga taksiran variansi dari $\hat{\gamma}_i$ yang dinyatakan dengan $Var\left(\hat{\gamma}_i\right)$. Taksiran standar *error* dari $\hat{\beta}_i$ dinyatakan dengan $Std(\hat{\beta}_i) = \sqrt{Var\left(\hat{\beta}_i\right)}$ sedangkan taksiran standar *error* dari $\hat{\gamma}_i$ dinyatakan dengan $Std(\hat{\gamma}_i) = \sqrt{Var\left(\hat{\gamma}_i\right)}$. Nilai $s(\hat{\beta}_i)$ dan $s(\hat{\gamma}_i)$ akan digunakan dalam pengujian ini.

Hipotesis sebagai berikut :

$H_0 : \beta_i = 0$

Variabel $X_i$ tidak memberikan kontribusi yang signifikan dalam memprediksi variabel *dependent*.

$H_1 : \beta_i \neq 0$

Variabel $X_i$ memberikan kontribusi yang signifikan dalam memprediksi variabel *dependent*.

untuk $i = 1, 2, ..., p$.

Karena $\hat{V}(\hat{\theta}) = -\left[H^{-1}\right]$ adalah matriks variansi-kovariansi maka :

$$Std\left(\beta_i\right) = \sqrt{-\left[H^{-1}\right]_{i,i}}$$

dengan $-\left[H^{-1}\right]_{i,j}$ adalah entri ke i,j dari matriks varians-kovarians.

Dalam jumlah sampel yang besar :

$$z = \frac{\hat{\beta}_i - \beta_i}{\sqrt{-\left[H^{-1}\right]_{i,i}}} \rightarrow N\left(0,1\right).$$

Maka untuk pengujian di atas digunakan statistik uji :

$$\begin{aligned} z &= \frac{\hat{\beta}_i - 0}{\sqrt{-\left[H^{-1}\right]_{i,i}}} \\ &= \frac{\hat{\beta}_i}{\sqrt{-\left[H^{-1}\right]_{i,i}}} \end{aligned}$$

dengan $z$ akan berdistribusi $N\left(0,1\right)$.

Aturan keputusan : $H_0$ akan ditolak pada tingkat signifikansi $\alpha$ apabila $|z| > z_{\alpha/2}$.

Apabila keputusannya $H_0$ ditolak menyatakan bahwa pada tingkat signifikansi $\alpha$ parameter $\beta_i$ tidak sama dengan nol, yang artinya variabel penjelas $X_i$ memberikan kontribusi informasi yang signifikan untuk memprediksi nilai variabel *dependent*.

### 3.3.3 Pengujian terhadap $\gamma$

Pengujian ini bertujuan untuk melihat apakah taksiran untuk $\gamma_i$ yang diperoleh sesuai dengan suatu nilai $\gamma_{i,0}$ yang telah ditetapkan oleh peneliti.

Hipotesis sebagai berikut :

$H_0 : \gamma_i = \gamma_{i,0}$

$H_1 : \gamma_i \neq \gamma_{i,0}$

untuk $i = 2,...,M-1$.

Apabila $H_0$ ditolak menyatakan bahwa taksiran $\gamma_i$ yang diperoleh tidak sama atau berbeda dengan suatu nilai tertentu $\gamma_{i,0}$ yang ditetapkan oleh peneliti.

$\hat{\mathrm{V}}(\hat{\theta}) = -\left[H^{-1}\right]$ merupakan matriks variansi-kovariansi dengan :

$$Std(\gamma_i) = \sqrt{-\left[H^{-1}\right]_{i,i}}$$

dengan $-\left[H^{-1}\right]_{i,j}$ adalah entri ke i,i dari matriks varians-kovarians.

Dalam jumlah sampel yang besar :

$$z = \frac{\hat{\gamma}_i - \gamma_{i,0}}{\sqrt{-\left[H^{-1}\right]_{i,i}}} \rightarrow N(0,1).$$

Maka untuk pengujian di atas digunakan statistik uji :

$$z = \frac{\hat{\gamma}_i - \gamma_{i,0}}{\sqrt{-\left[H^{-1}\right]_{i,i}}}$$

dengan $z$ akan berdistribusi $N(0,1)$.

Aturan keputusan : $H_0$ akan ditolak pada tingkat signifikansi $\alpha$ apabila $|z| > z_{\alpha/2}$.

Apabila keputusannya $H_0$ ditolak menyatakan bahwa pada tingkat signifikansi $\alpha$ parameter $\gamma_i$ tidak sama dengan suatu nilai tertentu $\gamma_{i,0}$ yang ditetapkan oleh peneliti. Atau dengan perkataan lain, taksiran $\gamma_i$ berbeda dengan nilai $\gamma_{i,0}$ yang ditetapkan oleh peneliti.

Untuk skripsi ini nilai $\gamma_{i,0}$ tidak diketahui sehingga tidak dilakukan pengujian terhadap $\gamma$.

### 3.4 Koefisien Determinasi

Selain melakukan pengujian seperti di atas, ada beberapa metode untuk mengukur kecocokan model. Pada umumnya untuk mengukur kecocokan dari suatu model digunakan plot residual atau koefisien determinasi $R^2$. Akan tetapi dalam skripsi ini plot residual tidak dapat dilakukan karena data dari variabel *dependent* kontinu tidak diketahui nilainya sehingga tidak dimungkinkan untuk membuat plot residual. Oleh karena itu dalam pembahasan berikut untuk mengukur kecocokan model hanya dengan menggunakan koefisien determinasi $R^2$. Selain koefisien determinasi $R^2$, akan dilihat pula misklasifikasi yang terjadi.

$R^2$ merupakan proporsi variasi dari variabel *dependent* yang dapat dijelaskan oleh variabel penjelas melalui model. Akan tetapi dalam pembahasan ini, variabel *dependent* kontinu tidak diketahui nilainya, sehingga tidak dimungkinkan untuk mengetahui variasi dari variabel *dependent* kontinu tersebut. Walaupun demikian koefisien determinasi $R^2$ dapat ditaksir dengan cara berikut :

Misalkan diasumsikan variabel *dependent* kontinu $Z$ memenuhi model regresi dengan taksiran parameter $\hat{\beta} = \left(\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p\right)$, $\hat{\beta}$ merupakan taksiran *maximum likelihood* dari $\beta$ yang diperoleh dari pembahasan sebelumnya.

Definisikan :

$$\hat{Z}_j = \sum_{i=0}^{p} \hat{\beta}_i X_{ij}$$

$$\hat{Z}_j = \hat{\beta}_0 + \hat{\beta}_1 X_{1j} + \hat{\beta}_2 X_{2j} + ... + \hat{\beta}_p X_{pj}$$

Didefinisikan $e_j = \left(Z_j - \hat{Z}_j\right)$.

Telah dibuktikan bahwa :

$\dfrac{\sum e_j^2}{n}$ konvergen secara probabilitas ke $\sigma^2$.

$$SSE = \sum e_j^2$$

Maka nilai $SSE$ akan mendekati $n\sigma^2$.

Karena telah diasumsikan $\sigma = 1$, maka $SSE = n$.

*Sum square* dari model dapat ditaksir sebagai berikut

$$S\hat{S}_{(Model)} = \sum_{i=1}^{n} \left(\hat{Z}_j - \bar{\hat{Z}}\right)^2 ; dengan\ \bar{\hat{Z}} = \sum_{i=1}^{n} \frac{\hat{Z}_i}{n}.$$

Taksiran *sum of square* total adalah sebagai berikut :

$$
\begin{aligned}
S\hat{S}_{(Total)} &= S\hat{S}_{(Model)} + SSE \\
&= \sum_{j=1}^{n}\left(\hat{Z}_j - \bar{\hat{Z}}\right)^2 + n
\end{aligned}
$$

Maka $R^2$ dapat ditaksir dengan :

$$
\begin{aligned}
\hat{R}^2 &= \frac{S\hat{S}_{(Total)} - SSE}{S\hat{S}_{(Total)}} \\
&= \frac{(S\hat{S}_{(Model)} + SSE) - SSE}{S\hat{S}_{(Total)}} \\
&= \frac{S\hat{S}_{(Model)}}{S\hat{S}_{(Total)}}
\end{aligned}
$$

$$
\hat{R}^2 = \frac{\sum_{j=1}^{n}\left(\hat{Z}_j - \bar{\hat{Z}}\right)^2}{\left[\sum_{j=1}^{n}\left(\hat{Z}_j - \bar{\hat{Z}}\right)^2\right] + n}
$$

$\hat{R}^2$ akan mengestimasi nilai dari $R^2$. Nilai $R^2$ selanjutnya dapat digunakan untuk menilai baik atau tidaknya model di mana apabila nilai $R^2$ semakin dekat dengan nilai 1 menyatakan model semakin baik.