BAB II
TINJAUAN PUSTAKA

2.1 Model Regresi Linier Sederhana

Pandang model regresi linier sederhana :

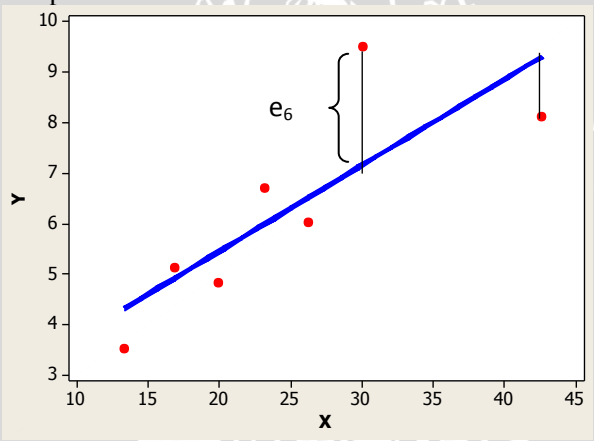
$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2 \ldots, n \tag{2.1}$$

Nilai harapan : $E\ y_i\ x_i\ = \beta_0 + \beta_1 x_i$

Ragam : $V\ y_i\ x_i\ = V(\beta_0 + \beta_1 x_i + \varepsilon_i) = \sigma^2$

Dengan demikian respon (y_i) adalah fungsi linier dari x meskipun ragam respon tidak bergantung pada nilai x. Karena galat tidak saling berkorelasi $cov\ \varepsilon_i, \varepsilon_j\ = 0$ maka respon juga tidak saling berkorelasi.

Secara geometrik, titik-titik data, model dan galat digambarkan pada Gambar 2.1:



Gambar 2.1. Garis Regresi

Titik-titik merah adalah nilai respon pada nilai x tertentu. Garis inilah yang akan diduga melalui pendugaan β_0 dan β_1, sehingga terbentuk persamaan $\hat{y}_i = \beta_0 + \beta_1 x_i$. Jarak vertikal yang menghubungkan titik respon dengan garis regresi y dinamai galat (Searle, 1971).

2.1.1 Pendugaan untuk Parameter β_0 dan β_1

Metode kuadrat terkecil digunakan untuk menduga β_0 dan β_1 dengan meminimumkan jumlah kuadrat selisih antara respon (y_i) dengan penduganya (y_i). Sebelum melakukan pendugaan parameter

β_0 dan β_1 persamaan (2.1), ada asumsi yang harus dipenuhi agar didapatkan penduga yang baik.

$$\varepsilon_i \sim \text{NIID}\ 0, \sigma^2$$
$$\text{E}\ \varepsilon_i\ = 0$$
$$\text{V}\ \varepsilon_i\ = \sigma^2$$
$$cov\ \varepsilon_i, \varepsilon_j\ = 0$$
$$\begin{aligned} E\ y_i\ &= E\ \beta_0 + \beta_1 x_i + \varepsilon_i \\ &= E\ \beta_0 + \beta_1 x_i\ + E\ \varepsilon_i \\ &= E\ \beta_0 + \beta_1 x_i\ + 0 \\ &= E\ \beta_0 + \beta_1 x_i \\ y_i &= \beta_0 + \beta_1 x_i \end{aligned}$$
$$\varepsilon_i = y_i - y_i$$
$$\varepsilon_i = y_i - \beta_0 + \beta_1 x_i$$
$$\varepsilon_i^2 = y_i - \beta_0 - \beta_1 x_i\ ^2$$

Prosedur metode kuadrat terkecil adalah sebagai berikut :

i. Menghitung jumlah kuadrat galat $\sum_{i=1}^{n} \varepsilon_i^2$

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i \right)^2$$

ii. Meminimumkan jumlah kuadrat galat dan menurunkan S secara parsial terhadap β_0 dan β_1 kemudian menyamakan dengan nol (0) yang menghasilkan 2 persamaan normal

$$\frac{\partial S}{\partial \beta_0} = \sum_{i=1}^{n} 2 \left(y_i - \beta_0 - \beta_1 x_i \right) \left(-1 \right)$$

$$\frac{\partial S}{\partial \beta_0} = 0$$

$$-2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) = 0$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{\beta}_0 - \sum_{i=1}^{n} \hat{\beta}_1 x_i = 0$$

$$\sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_i = 0$$

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \qquad (2.2)$$

$$\frac{\partial S}{\partial \beta_1} = \sum_{i=1}^{n} 2\left(y_i - \beta_0 - \beta_1 x_i\right)(-x_i)$$

$$\frac{\partial S}{\partial \beta_1} = 0$$

$$\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)(-x_i) = 0$$

$$\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} \hat{\beta}_0 x_i - \sum_{i=1}^{n} \hat{\beta}_1 x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \qquad (2.3)$$

iii. Menghitung β_0 dan β_1,

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i\right) = \bar{y} - \hat{\beta}_1 \bar{x}$$

kemudian disubstitusikan ke persamaan 2.3,

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

$$\left(\bar{y} - \hat{\beta}_1 \bar{x}\right) \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

$$\bar{y} \sum_{i=1}^{n} x_i - \hat{\beta}_1 \bar{x} \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

$$\hat{\beta}_1\left(\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

2.2 Model Regresi Linier Berganda

Pandang model regresi linier berganda :

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} + \varepsilon_i \quad , \quad i = 1, 2, \ldots, n$$

$$y_i = \sum_{q=0}^{p} \beta_q x_{iq} + \varepsilon_i$$

di mana :

n = banyaknya pengamatan

p = banyaknya peubah prediktor (q = 0, 1, …, p)

Pendugaan parameter β_q sama halnya dengan model regresi linier sederhana yaitu menggunakan metode kuadrat terkecil.

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip}$$

$$\varepsilon_i = y_i - \hat{y}_i$$

$$\varepsilon_i = y_i - (\beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip})$$

$$\varepsilon_i^2 = (y_i - \beta_0 - \beta_1 x_{i1} - ... - \beta_p x_{ip})^2$$

Prosedur pendugaan koefisien regresi berganda :

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - ... - \beta_p x_{ip})^2$$

kemudian diturunkan terhadap β_0, β_1, …, β_p, untuk mendapatkan persamaan normal,

$$\frac{\partial S}{\partial \hat{\beta}_0} = 0, \ \frac{\partial S}{\partial \hat{\beta}_1} = 0, \ \ldots \ , \ \frac{\partial S}{\partial \hat{\beta}_p} = 0,$$

sehingga terbentuk (p + 1) persamaan :

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} + \ldots + \hat{\beta}_p \sum_{i=1}^{n} x_{ip} = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^{n} x_{i1}x_{i2} + \ldots + \hat{\beta}_p \sum_{i=1}^{n} x_{i1}x_{ip} = \sum_{i=1}^{n} x_{i1}y_i$$

$$\vdots$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{ip} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}x_{ip} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2}x_{ip} + \ldots + \hat{\beta}_p \sum_{i=1}^{n} x_{ip}^2 = \sum_{i=1}^{n} x_{ip}y_i$$

Dalam bentuk matriks :

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_{i1} & \cdots & \sum_{i=1}^{n} X_{ip} \\ \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i1}^2 & \cdots & \sum_{i=1}^{n} X_{i1}X_{1p} \\ \vdots & \vdots & \ddots & \\ \sum_{i=1}^{n} X_{ip} & \sum_{i=1}^{n} X_{i1}X_{1p} & & \sum_{i=1}^{n} X_{ip}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_{i1}Y_i \\ \vdots \\ \sum_{i=1}^{n} X_{ip}Y_i \end{bmatrix}$$

$$A_{(p+1)\times(p+1)} \quad \hat{\beta}_{(p+1)\times 1} = \boldsymbol{g}_{(p+1)\times 1}$$

Penduga koefisien regresi, $\boldsymbol{\beta} = A^{-1}\boldsymbol{g}$

2.3 Model Linier Umum

McCullagh dan Nelder (1983) menyatakan bahwa model linier umum atau *Generalized Linier Models* (GLM) pada regresi logistik terdiri atas tiga komponen, yaitu :

1. Komponen Acak atau fungsi sebaran peubah acak f(y) yang termasuk dalam keluarga eksponensial untuk suatu peubah acak yang tergantung pada parameter nilai tengah μ atau parameter lainnya.

2. Komponen Sistematik atau prediktor η yang mencakup p peubah prediktor $x_1, x_2, \dots, x_p$ dengan bentuk $\eta\ x\ = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$
3. Fungsi Penghubung menggambarkan hubungan antara komponen acak dan komponen sistematik dinyatakan dalam bentuk $\eta = g\ \mu_i$. Pandang $\mu_i = E\ Y_i$, μ_i dihubungkan oleh $\eta_i = g\ \mu_i$, dengan model fungsi penghubung antara μ_i dan p peubah prediktor adalah $g\ \mu_i\ = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$

Peubah respon k kategori berskala ordinal, dengan mempertimbangkan Y ϵ {1,2,…,k} maka peubah respon multikategori menyebar menurut sebaran multinomial dengan fungsi penghubung logit.

2.4 Regresi Logistik Ordinal

Menurut Fahrmeir dan Gerhard (1994) jika peubah respon mengandung lebih dari dua k kategori berskala ordinal, maka digunakan Regresi Logistik Ordinal yang didasarkan pada peluang kumulatif. Fox dan Andersen (2004) menyarankan k kategori peubah respon dapat dipandang sebagai k kelas peubah kontinyu tak teramati. Bentuk hubungan antara peubah respon teramati (*observable respons variable*) dengan peubah respon tak teramati (*unobservable respons variable*) dijelaskan sebagai berikut:

$$y_i = \begin{cases} 1, \ untuk\ u_i \le \theta_1 \\ 2, \ untuk\ \theta_1 < u_i \le \theta_2 \\ \vdots \\ k-1, \ untuk\ \theta_{k-2} x < u_i \le \theta_{k-1} \\ k, \ untuk\ \theta_{k-1} < u_i \end{cases}$$

berdasarkan asumsi:

$$u_i = \ \beta_1 x_{i1} + \cdots + \beta_p x_{ip}\ \ + \varepsilon_i$$

di mana:

u_i	= peubah respon tak teramati ke-i
y_i	= peubah respon teramati ke-i
$x_{i1}, \dots, x_{ip}$	= nilai ke-i peubah prediktor ke-p
ε_i	= galat ke-i
$\beta_q = \beta_1, \dots, \beta_p$	= koefisien peubah prediktor ke-q
$\theta_1, \dots, \theta_{k-1}$	= batas peubah respon tak teramati
n	= banyaknya pengamatan

p = banyaknya peubah prediktor
k = banyaknya kategori peubah respon (j = 1, 2, …, k-1)

Peubah respon ordinal juga dapat dipandang sebagai kelas-kelas dari peubah respon kontinyu tak teramati yang mempunyai batasan nilai tidak diketahui.

Peluang kumulatif peubah respon adalah:

$$\begin{aligned} P\ y \le j\ &= P\ u_i \le \theta_j \\ &= P\ \ \beta_1 x_{i1} + \cdots + \beta_p x_{ip}\ + \varepsilon_i \le \theta_j \\ &= P\ \varepsilon_i \le \theta_j + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} \\ &= F\ \varepsilon_i \end{aligned} \tag{2.4}$$

Galat diasumsikan menyebar logistik dengan fungsi kumulatif :

$$\begin{aligned} F\ \varepsilon_i\ &= \frac{1}{1 + exp\ -\varepsilon_i} \\ &= \frac{exp\ \varepsilon_i}{1 + exp\ \varepsilon_i} \end{aligned} \tag{2.5}$$

maka peluang kumulatif peubah respon persamaan (2.4) dapat ditulis:

$$\begin{aligned} P\ y \le j\ &= \pi_j\ X \\ &= \frac{exp\ \theta_j + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}}{1 + exp\ \theta_j + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}} \end{aligned}$$

$\pi_j\ X$ = model regresi logistik kategori ke-j

Model logit untuk peluang kumulatif:

$$\begin{aligned} logit\ P\ \boldsymbol{y} \le j\ &= log\ \frac{P\ \boldsymbol{y} \le j}{1 - P\ \boldsymbol{y} \le j} \\ &= log\ \frac{\dfrac{exp\ \boldsymbol{\theta}_j + \boldsymbol{X\beta}}{1 + exp\ \boldsymbol{\theta}_j + \boldsymbol{X\beta}}}{1 - \dfrac{exp\ \boldsymbol{\theta}_j + \boldsymbol{X\beta}}{1 + exp\ \boldsymbol{\theta}_j + \boldsymbol{X\beta}}} \end{aligned}$$

$$g\ \boldsymbol{x}\ = \boldsymbol{\theta}_j + \boldsymbol{X\beta} \tag{2.6}$$

$g\ \boldsymbol{x}_{\ n\times 1}$ = $logit\ P\ \boldsymbol{y} \le \boldsymbol{j}$
$\boldsymbol{y}_{\ nx1}$ = vektor peubah respon
$\boldsymbol{X}_{\ nxp}$ = matriks peubah prediktor
$\boldsymbol{\beta}_{\ px1}$ = vektor koefisien model logit
$\boldsymbol{\theta}_j$ = intersep model logit ke-j

Regresi logistik k kategori peubah respon ordinal sesuai persamaan (2.6) disebut Regresi Logistik Ordinal (Berrington *et al.*).

Model logit respon ordinal juga disebut *Cummulative Logit Model* karena didasarkan pada peluang kumulatif peubah respon untuk setiap kategori. Regresi logistik atau *Proportional Odds Model* (POM) dengan koefisien peubah prediktor (β) untuk model logit ke-j tidak tergantung pada kategori peubah respon atau sama untuk model logit ke-j.

Peubah respon kategori ke-j berdasarkan persamaan (2.5) dengan peluang kumulatif $P\ \boldsymbol{y} \le j\ = \boldsymbol{\pi_1}\ \boldsymbol{X}\ + \cdots + \boldsymbol{\pi}_j\ \boldsymbol{X}$, maka $\boldsymbol{\pi}_j\ \boldsymbol{X}$ adalah peluang terjadi suatu kejadian berdasarkan kategori ke-j peubah respon **y**:

$$P\ \boldsymbol{y} = j\ = \boldsymbol{\pi}_j\ \boldsymbol{X}$$
$$= P\ \boldsymbol{y} \le j\ - P\ \boldsymbol{y} \le j-1$$
$$= \frac{exp\ \boldsymbol{\theta}_j + \boldsymbol{X\beta}}{1 + exp\ \boldsymbol{\theta}_j + \boldsymbol{X\beta}} - \frac{exp\ \boldsymbol{\theta}_{j-1} + \boldsymbol{X\beta}}{1 + exp\ \boldsymbol{\theta}_{j-1} + \boldsymbol{X\beta}}$$

2.4.1 Pendugaan Parameter

Metode pendugaan parameter pada Regresi Logistik Ordinal menggunakan metode *Maximum Likelihood Estimation* (MLE). Jika peubah respon $\boldsymbol{y} \sim Multinomial\ \boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_k; n; \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \dots, \boldsymbol{\pi}_k$ maka fungsi peluang :

$$P\ Y_j = y_j\ = P\ Y_1 = y_1, \dots, Y_k = y_k\ = \frac{n!}{y_1! \dots y_k!}\ p_1\ ^{y_1} \dots (p_k)^{y_k}$$

p_j = peluang sukses kategori ke-j

Fungsi *likelihood* untuk pengamatan y_i :

$$L\ \boldsymbol{\theta}, \boldsymbol{\beta}\ = \prod_{i=1}^{n} \boldsymbol{\pi}_1(\boldsymbol{X})^{\ y_{j1}}\ \boldsymbol{\pi}_2(\boldsymbol{X})^{\ y_{j2}} \dots\ \boldsymbol{\pi}_j(\boldsymbol{X})^{\ y_{ji}} \qquad (2.7)$$

Fungsi *log likelihood*:

$$\ell\ \boldsymbol{\theta},\boldsymbol{\beta}\ = \sum_{i=1}^{n} y_{1i}log(\boldsymbol{\pi_1}(\boldsymbol{X})) + y_{2i}log(\boldsymbol{\pi_2}(\boldsymbol{X})) + \cdots + y_{ki}log(\boldsymbol{\pi_j}(\boldsymbol{X}))$$

$$= \sum_{\boldsymbol{i=1}}^{n} y_{1i}log\ \frac{exp\ \boldsymbol{\theta}_1 + \boldsymbol{X\beta}}{1 + exp\ \boldsymbol{\theta}_1 + \boldsymbol{X\beta}} + y_{2i}log\ \frac{exp\ \boldsymbol{\theta}_2 + \boldsymbol{X\beta}\ - exp\ \boldsymbol{\theta}_1 + \boldsymbol{X\beta}}{exp\ \boldsymbol{\theta}_2 + \boldsymbol{X\beta}\ \ (1 + exp\ \boldsymbol{\theta}_j + \boldsymbol{X\beta}\ )} + \cdots + y_{ki}log\ \ 1 - \frac{exp\ \boldsymbol{\theta}_{k-1} + \boldsymbol{X\beta}}{1 + exp\ \boldsymbol{\theta}_{k-1} + \boldsymbol{X\beta}}$$

Nilai $\boldsymbol{\theta}$ dan $\boldsymbol{\beta}$ diperoleh dari turunan pertama fungsi *log likelihood*:

$$\frac{\partial \ell}{\partial \theta_1} = \sum_{i=1}^{n} y_{1i}\ \ 1 - \frac{exp\ \boldsymbol{\theta}_1 + \boldsymbol{X\beta}}{1 + exp\ \boldsymbol{\theta}_1 + \boldsymbol{X\beta}} + y_{2i}\ \ - \frac{exp(\boldsymbol{\theta}_1)}{exp\ \boldsymbol{\theta}_2 - \boldsymbol{\theta}_1} - \frac{exp\ \boldsymbol{\theta}_1 + \boldsymbol{X\beta}}{1 + exp\ \boldsymbol{\theta}_1 + \boldsymbol{X\beta}} = 0$$

$$\frac{\partial \ell}{\partial \theta_{k-1}} = \sum_{i=1}^{n} y_{k-1} - y_{k-2}\ \ \frac{exp(\boldsymbol{\theta}_{k-1})}{exp\ \boldsymbol{\theta}_{k-1} - \boldsymbol{\theta}_{k-2}} - \frac{exp\ \boldsymbol{\theta}_{k-1} + \boldsymbol{X\beta}}{1 + exp\ \boldsymbol{\theta}_{k-1} + \boldsymbol{X\beta}} - (n - y_{k-1})\ \frac{exp\ \boldsymbol{\theta}_{k-1} + \boldsymbol{X\beta}}{1 + exp\ \boldsymbol{\theta}_{k-1} + \boldsymbol{X\beta}} = 0$$

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{n} y_1\ \ \boldsymbol{X} - \frac{\boldsymbol{x}_j exp\ \boldsymbol{\theta}_1 + \boldsymbol{X\beta}}{1 + exp\ \boldsymbol{\theta}_1 + \boldsymbol{X\beta}} + \ y_2 - y_1\ \ \ \boldsymbol{X}_j - \frac{\boldsymbol{x}_j exp\ \boldsymbol{\theta}_1 + \boldsymbol{X\beta}}{1 + exp\ \boldsymbol{\theta}_1 + \boldsymbol{X\beta}} - \frac{\boldsymbol{x}_j exp\ \boldsymbol{\theta}_1 + \boldsymbol{X\beta}}{1 + exp\ \boldsymbol{\theta}_1 + \boldsymbol{X\beta}} + \cdots + (1 - y_{k-1})\ - \frac{\boldsymbol{x}_j exp\ \boldsymbol{\theta}_{k-1} + \boldsymbol{X\beta}}{1 + exp\ \boldsymbol{\theta}_{k-1} + \boldsymbol{X\beta}} = 0$$

(Hyun, 2004)

Fungsi *log likelihood* tidak bersifat linier sehingga diperlukan metode iterasi *Newton-Rhapson* untuk menduga nilai θ dan β agar menghasilkan penduga parameter $\gamma = \theta, \beta$ yang konvergen pada nilai tertentu:

$$\boldsymbol{\gamma}^{(t+1)} = \boldsymbol{\gamma}^{(t)} - \boldsymbol{H}^{(t)\ -1}\boldsymbol{g}^{(t)}$$

$$\boldsymbol{H}^{\ t\ \ -1}_{p\times p} = \frac{\partial^2 \ell\ \gamma}{\partial \gamma^2} = \boldsymbol{X'VX}$$

$$\boldsymbol{g}^{(t)} = \frac{\partial \ell\ \gamma}{\partial \gamma} = \boldsymbol{X}\ \boldsymbol{Y} - \boldsymbol{\pi}$$

$\boldsymbol{H}^{(t)\ -1}$ adalah matriks ragam peragam.

Melalui teknik iterasi *Newton-Rhapson* $\boldsymbol{g}^{(t)}$ dan $\boldsymbol{H}^{(t)}$ digunakan untuk menduga $\boldsymbol{\gamma}^{(t)}$ pada iterasi ke-t (t = 0, 1, 2, …) hingga mencapai kondisi konvergen dengan syarat:

$$\gamma^{(t)} - \gamma^{(t-1)} < \varepsilon$$

2.4.2 Pengujian terhadap Parameter

Pengujian terhadap parameter merupakan suatu pemeriksaan apakah peubah-peubah prediktor dalam model mempunyai pengaruh nyata terhadap peubah respon dan dapat dilakukan secara parsial maupun serentak.

1. Pengujian parameter secara parsial

   Pengujian koefisien regresi secara parsial dilakukan untuk memeriksa peranan koefisien regresi setiap peubah prediktor secara individu dalam model berlandaskan hipotesis:

   1. $H_0 : \beta_q = 0$ vs $H_1 : \beta_q \neq 0$
   2. $H_0 : \theta_j = 0$ vs $H_1 : \theta_j \neq 0$

   Jika H_0 benar, statistik uji Wald adalah :

$$W = \frac{\beta_q}{se\ _{\beta_q}} \sim Z \quad \text{dan} \quad W = \frac{\theta_j}{se\ _{\theta_j}} \sim Z \qquad (2.8)$$

   di mana :

   β_q = penduga bagi β_q

   $se\ _{\beta_q}$ = salah baku β_q

   θ_j = penduga bagi θ_j

se_{θ_j} = salah baku θ_j

(Agresti, 1990)

H_0 ditolak jika P(|Z| > W) < α sehingga peubah prediktor mempengaruhi peubah respon.

Nilai se_{β_q} ditentukan dari akar diagonal utama matriks ragam peragam.

$$Var\ \beta_q = diag\ \boldsymbol{X'VX}^{-1}$$

$$se_{\beta_q} = \overline{Var\ \beta_q}$$

$$V_{n\times n} = \begin{bmatrix} \hat{\pi}_{11}(1-\hat{\pi}_{11}) & 0 & \cdots & 0 \\ 0 & \hat{\pi}_{12}(1-\hat{\pi}_{12}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\pi}_{1n}(1-\hat{\pi}_{1n}) \end{bmatrix}$$

Matriks **V** berukuran n×n. Matriks **X** berukuran n×p merupakan matriks peubah prediktor yang mempunyai unsur:

$$X_{n\times(p+1)} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{bmatrix}$$

2. Pengujian parameter secara simultan

   Pengujian secara simultan dilakukan untuk mengetahui pengaruh peubah prediktor secara bersama-sama, berlandaskan hipotesis :

   $H_0: \beta_q = 0$ lawan

   H_1: paling sedikit ada satu q di mana $\beta_q \neq 0$

   Jika H_0 benar,

   $G = -2\ L_0 - L_p \sim \chi^2_{(v)}$

   di mana :

   L_0 = nilai log likelihood model regresi logistik tanpa peubah prediktor

   L_p = nilai log likelihood model regresi logistik dengan peubah prediktor

   (Kleinbaum dan Klein, 2010)

Hipotesis nol ditolak jika $P\ \chi^2_{(v)} > G\ < \alpha$. Hal ini mengindikasikan bahwa paling sedikit ada satu β_q yang tidak sama dengan nol. Walaupun uji Wald cukup memadai untuk contoh berukuran besar, uji Likelihood lebih kuat dan lebih dapat diandalkan (Agresti, 1990).

2.5 Multikolinieritas

Menurut Myers *et al*., analisis regresi selain bertujuan untuk mengetahui hubungan antara peubah prediktor dengan peubah respon juga digunakan untuk menentukan peubah-peubah penyusun model dan untuk prediksi. Pada analisis regresi berganda, terdapat asumsi non-multikolinieritas. Multikolinieritas adalah hubungan linier antar peubah predictor yang akan menimbulkan masalah dalam penentuan peubah-peubah penyusun model.

Multikolinieritas dapat terjadi apabila matriks ($\mathbf{X'X}$) tidak memiliki kebalikan karena bersifat singular, sehingga hasil pendugaan tidak akan unik (Toutenburg, 2002).

Hosmer dan Lemeshow (2000) menyatakan bahwa analisis regresi logistik berganda sensitif terhadap multikolinieritas. Multikolinieritas timbul ketika model yang dispesifikasikan tidak tepat dan banyaknya peubah prediktor lebih sedikit dibandingkan ukuran contoh. Salah satu cara mengatasi multikolinieritas menggunakan *Partial Least Square Regression*.

Pengujian multikolinieritas dilakukan dengan *Variance Inflation Factor* (VIF):

$$\mathrm{VIF_q} = \frac{1}{1-\mathrm{R_q^2}} \qquad (2.9)$$

$\mathrm{R_q^2}$ adalah koefisien determinasi ganda jika X_q diregresikan dengan p-1 peubah prediktor lain. Asumsi non-multikolinieritas tidak terpenuhi jika $\mathrm{VIF_q} > 10$.

Multikolinieritas dalam suatu data akan menghasilkan penduga parameter yang kurang baik, galat besar dan ragam galat besar. Galat besar akan memperkecil statistik uji-t dan memperlebar selang kepercayaan bagi β_q (Hamilton, 1992).

Pada regresi logistik, pendugaan parameter dengan metode maksimum likelihood tidak dapat dilakukan karena matriks ragam peragam $\boldsymbol{X'VX}^{-1}$ bersifat non singular jika matriks ragam-peragam

terdapat korelasi maka $[\boldsymbol{X'VX}] = 0$. Multikolinieritas berpengaruh terhadap galat penduga koefisien regresi β yang bernilai semakin besar. Hal ini dapat dilihat dari matriks $\boldsymbol{X'VX}^{-1}$ yang dibutuhkan dalam menentukan salah baku bagi penduga koefisien regresi logistik melalui persamaan :

$$se_{\beta_q} = \sqrt{Var\ \beta_q}$$

Multikolinieritas terjadi apabila $\boldsymbol{X'VX}$ bersifat singular atau mendekati singular $\boldsymbol{X'VX} \rightarrow 0$, sehingga unsur-unsur matriks $\boldsymbol{X'VX}^{-1}$ akan semakin besar atau mendekati tak hingga. Hal ini menyebabkan se_{β_q} semakin besar pula. Dengan demikian multikolinieritas mengakibatkan presisi penduga koefisien regresi semakin rendah. Semakin besar se_{β_q} maka statistik W semakin kecil. Jadi, multikolinieritas akan menurunkan kekuatan uji Wald.

2.6 *Partial Least Square Regression* (PLSR)

PLSR merupakan model yang menghubungkan peubah respon dengan peubah prediktor numerik maupun kategorik, di mana banyak peubah prediktor lebih dari banyaknya pengamatan. Model ini juga dikembangkan di bidang biologi sebagai metode reduksi peubah untuk membentuk peubah baru yang disebut komponen (Bastien *et al*., 2004).

Konsep dasar PLSR adalah menguraikan peubah respon dan peubah prediktor menurut persamaan:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E}$$

$$\mathbf{y} = \mathbf{Tc}' + \mathbf{f} \qquad (2.10)$$

$\mathbf{X}_{(nxp)}$ = matriks peubah prediktor
$\mathbf{y}_{(nx1)}$ = vektor peubah respon
$\mathbf{T}_{(nxm)}$ = matriks komponen PLSR
$\mathbf{P}_{(pxm)}$ = matriks koefisien komponen PLSR
$\mathbf{c}_{(1xm)}$ = vektor koefisien PLSR
$\mathbf{E}_{(nxp)}$ = matriks residual **X**
$\mathbf{f}_{(nx1)}$ = vektor residual **y**
m = banyaknya komponen PLSR

n = banyaknya pengamatan
p = banyaknya peubah prediktor

T komponen PLSR adalah kombinasi linier peubah prediktor, berdasarkan persamaan:

$$\mathbf{T} = \mathbf{XW}$$

$\mathbf{W}_{(pxm)}$ adalah matriks pembobot dan setiap komponen $\boldsymbol{t}_h$ dapat dituliskan:

$$\begin{aligned} \mathbf{t}_1 &= \mathbf{w}_{11}\mathbf{x}_1 + \mathbf{w}_{21}\mathbf{x}_2 + \cdots + \mathbf{w}_{p1}\mathbf{x}_p \\ \ldots &= \quad \ldots \\ \mathbf{t}_m &= \mathbf{w}_{1m}\mathbf{x}_1 + \mathbf{w}_{2m}\mathbf{x}_2 + \cdots + \mathbf{w}_{pm}\mathbf{x}_p \end{aligned} \qquad (2.11)$$

dengan h = 1, 2, …, m dan skor komponen diperoleh dari persamaan (2.11) (Bouleisteix dan Strimmert, 2005).

Bastien *et al*. (2004) mengemukakan bahwa koefisien $\mathbf{c}_h$ pada persamaan (2.10) diduga dengan regresi berganda $\boldsymbol{t}_h$ terhadap **y** jika dikembalikan dalam bentuk peubah prediktor asal maka penduga bagi peubah respon dapat diperoleh dengan:

$$\mathbf{y} = \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \cdots + \beta_p\mathbf{x}_p$$

Menurut Ding dan Gentleman (2004), beberapa penelitian tidak selalu melibatkan peubah respon numerik, sehingga dikembangkan pula PLSR untuk *Generalized Linear Models* (GLM) untuk peubah respon kategori. Pada modifikasi PLSR dengan GLM untuk peubah respon kategorik disebut dengan *Partial Least Square Generalized Linear Regression* (PLS-GLR).

Secara umum model PLS-GLR (fungsi penghubung g $\boldsymbol{\mu}$) dapat dituliskan:

$$g\ \boldsymbol{\mu} = c_1\mathbf{t}_1 + c_2\mathbf{t}_2 + \cdots + c_m\mathbf{t}_m \qquad (2.12)$$

PLS-GLR untuk regresi logistik disebut *Logistics Partial Least Square* (LPLS). Bastien, Vinzi dan Tenenhaus (2002) memperkenalkan PLS *Ordinal Logistics Regression* (PLSOLR) sebagai penggabungan antara *Partial Least Square Regression* dan *Ordinal Logistics Regression* ketika peubah respon berskala ordinal, sehingga g $\boldsymbol{\mu}$ (2.12) merupakan logit P $\mathbf{y} \leq \mathbf{j}$ pada persamaan (2.6). Pendugaan parameter model ini dilakukan dengan MLE (*Maximum Likelihood Estimation*).

2.6.1 Pembentukan Komponen

Menurut Bastien *et al*. (2004), tiap-tiap tahap algoritma pada pembentukan model PLS *Ordinal Logistics Regression* dilakukan dengan OLR.

Pembentukan $\mathbf{t}_h$ komponen PLSOLR untuk h = 1, 2, …, m dapat dijelaskan sebagai berikut:

$$t_h = \mathbf{w}_{1h}\mathbf{x}_1 + \mathbf{w}_{2h}\mathbf{x}_2 + \cdots + \mathbf{w}_{ph}\mathbf{x}_p \quad (2.13)$$

dan

$$\mathbf{w}_h = \frac{\mathbf{a}_h}{\mathbf{a}_h} \quad (2.14)$$

vektor koefisien $\mathbf{a}_h$ dengan elemen $\mathbf{a}_{qh}$ merupakan koefisien regresi bagi peubah prediktor dari:

$$g\ \boldsymbol{\mu} = \boldsymbol{\theta}_j + c_1\mathbf{t}_1 + \cdots + c_m\mathbf{t}_m + \mathbf{a}_{qh}\mathbf{x}_{qh} + \text{galat} \quad (2.15)$$

dan untuk tiap-tiap $\mathbf{x}_{qh}$ diperoleh dengan persamaan

$$\mathbf{x}_{q(h)} = \mathbf{x}_{q(h-1)} - \mathbf{p}_{q1}\mathbf{t}_1 - \cdots - \mathbf{p}_{qm}\mathbf{t}_m \quad (2.16)$$

di mana :

$\mathbf{t}_h$ = vektor komponen PLSOLR ke-h berukuran (n x 1)

$\mathbf{w}_h$ = vektor pembobot $\mathbf{x}_q$ pada komponen $\mathbf{k}_h$ berukuran (p x 1)

$\boldsymbol{\alpha}_j$ = vektor intersep regresi logistik ke-j

$c_1, \ldots, c_m$ = koefisien regresi logistik untuk $\mathbf{t}_1, \ldots, \mathbf{t}_m$

$\mathbf{a}_{qh}$ = koefisien regresi logistik untuk tiap-tiap $\mathbf{x}_{qh}$

$\mathbf{x}_{qh}$ = vektor peubah prediktor ke-q berukuran (n x 1)

$\mathbf{p}_{q1}, \ldots, \mathbf{p}_{qm}$ = koefisien komponen $\mathbf{t}_h$

Untuk komponen $\mathbf{t}_h$, $\mathbf{x}_{qh} = \mathbf{x}_{q(0)} = \mathbf{x}_q$ dengan $\mathbf{x}_q$ adalah vektor peubah prediktor awal atau sebagai vektor inisialisasi.

Penghitungan komponen $\mathbf{t}_h$ dilakukan sampai didapatkan matriks **X** menjadi matriks nol (**0**). Jika telah terbentuk komponen $\mathbf{t}_h$, maka dilakukan regresi logistik $\mathbf{t}_h$ terhadap **y**. Sesuai persamaan (2.12), maka:

$$\begin{aligned}
&\text{logit } P\ \mathbf{y} \le 1 &&= \boldsymbol{\theta}_1 + c_1\mathbf{t}_1 + c_2\mathbf{t}_2 + \cdots + c_m\mathbf{t}_m \\
&\text{logit } P\ \mathbf{y} \le 2 &&= \boldsymbol{\theta}_2 + c_1\mathbf{t}_1 + c_2\mathbf{t}_2 + \cdots + c_m\mathbf{t}_m \\
&\quad\vdots \\
&\text{logit } P\ \mathbf{y} \le k-1 &&= \boldsymbol{\theta}_{k-1} + c_1\mathbf{t}_1 + c_2\mathbf{t}_2 + \cdots + c_m\mathbf{t}_m
\end{aligned} \quad (2.17)$$

Model logit pada persamaan (2.17) disebut dengan *Partial Least Square Ordinal Logistics Regression* (Bastien *et al*., 2004).

2.6.2 Transformasi Model

PLSR adalah model yang menghubungkan peubah prediktor dengan peubah respon dengan tujuan awal ingin mengetahui bentuk hubungan peubah prediktor **X** dengan peubah respon **y**, maka berdasarkan model yang sudah terbentuk dari persamaan (2.17) dilakukan transformasi model dalam bentuk peubah prediktor asal atau peubah prediktor **X**. Transformasi dilakukan dengan mensubstitusikan persamaan (2.12) ke persamaan (2.17), dengan model hasil transformasi dapat dituliskan:

$$\begin{aligned}
&\text{logit } P\ \mathbf{y} \le 1 \quad = \boldsymbol{\theta}_1 + \boldsymbol{\beta}_1\mathbf{x}_1 + \boldsymbol{\beta}_2\mathbf{x}_2 + \cdots + \boldsymbol{\beta}_p\mathbf{x}_p \\
&\text{logit } P\ \mathbf{y} \le 2 \quad = \boldsymbol{\theta}_2 + \boldsymbol{\beta}_1\mathbf{x}_1 + \boldsymbol{\beta}_2\mathbf{x}_2 + \cdots + \boldsymbol{\beta}_p\mathbf{x}_p \\
&\qquad\vdots \\
&\text{logit } P\ \mathbf{y} \le k-1 \quad = \boldsymbol{\theta}_{k-1} + \boldsymbol{\beta}_1\mathbf{x}_1 + \boldsymbol{\beta}_2\mathbf{x}_2 + \cdots + \boldsymbol{\beta}_p\mathbf{x}_p
\end{aligned} \tag{2.18}$$

PLS *Ordinal Logistics Regression* untuk **X** diperoleh dari:

$$\begin{aligned}
\boldsymbol{\beta}_1 &= \mathbf{c}_1\mathbf{w}_{11} + \mathbf{c}_2\mathbf{w}_{12} + \cdots + \mathbf{c}_m\mathbf{w}_{1m} \\
\boldsymbol{\beta}_2 &= \mathbf{c}_1\mathbf{w}_{21} + \mathbf{c}_2\mathbf{w}_{22} + \cdots + \mathbf{c}_m\mathbf{w}_{2m} \\
\ldots &= \quad \ldots \\
\boldsymbol{\beta}_p &= \mathbf{c}_1\mathbf{w}_{p1} + \mathbf{c}_2\mathbf{w}_{p2} + \cdots + \mathbf{c}_m\mathbf{w}_{pm}
\end{aligned}$$

2.7 Koefisien Determinasi (R^2)

Liu *et al*. (2006), menyatakan bahwa PLSOLR mampu meningkatkan keakuratan klasifikasi peubah respon. Koefisien determinasi dengan nilai *cross validation* R^2 digunakan untuk mengetahui seberapa baik klasifikasi yang dihasilkan dari model yang terbentuk. Nilai R^2 digunakan untuk mengetahui keefektifan PLSOLR dalam meningkatkan keakuratan klasifikasi ketika data memenuhi asumsi non-multikolinieritas:

$$R^2 = 1 - \frac{y_i - y_i^{\ 2}}{y_i - y^{\ 2}} \tag{2.19}$$

y_i = peubah respon ke-i
y_i = nilai penduga y_i
y = rata-rata peubah respon