BAB II
TINJAUAN PUSTAKA

2.1 Data Longitudinal

Data longitudinal diperoleh dari hasil pengamatan pada beberapa unit *cross-sectional* selama beberapa periode waktu tertentu (Twisk, 2003). Data *cross-sectional* berasal dari pengamatan yang dilakukan pada individu berbeda pada waktu tertentu.

Menurut Everitt (2002), data longitudinal dapat dibedakan menjadi:

1. Data longitudinal seimbang, hasil pengamatan n subjek pada titik waktu sama.

Tabel 2.1. Struktur Data Longitudinal Seimbang

i	t	Y_{it}	X_{it}
1	1	Y_{11}	X_{11}
	2	Y_{12}	X_{12}
	⋮	⋮	⋮
	T	Y_{1T}	X_{1T}
⋮	⋮	⋮	⋮
n	1	Y_{n1}	X_{n1}
	2	Y_{n2}	X_{n2}
	⋮	⋮	⋮
	T	Y_{nT}	X_{nT}

unit *cross-sectional* ke-*i* = 1, 2, ..., n; n= ukuran contoh
unit waktu ke-t = 1, 2, ..., T; T = banyaknya waktu pengamatan
Y_{it}: respon unit *cross-sectional* ke-i pada unit waktu ke-t
X_{it}: prediktor unit *cross-sectional* ke-i pada unit waktu ke-t

2. Data longitudinal tidak seimbang dihasilkan dari pengamatan terhadap n subjek pada titik waktu berbeda.

Tabel 2.2. Struktur Data Longitudinal Tidak Seimbang

i	t	Y_{it}	X_{it}
1	1	Y_{11}	X_{11}
	2	Y_{12}	X_{12}
	⋮	⋮	⋮
	T_1	Y_{1T1}	X_{1T1}

2	1	Y_{21}	X_{21}
	2	Y_{22}	X_{22}
	⋮	⋮	⋮
	T_2	Y_{2T2}	X_{2T2}
⋮	⋮	⋮	⋮
n	1	Y_{n1}	X_{n1}
	2	Y_{n2}	X_{n2}
	⋮	⋮	⋮
	T_n	Y_{nTn}	X_{nTn}

unit waktu ke-t, t= 1, 2, ..., T_i; T_i = banyaknya waktu pengamatan pada unit *cross-sectional* ke -i

Y_{it}: respon unit *cross-sectional* ke-i pada unit waktu ke-t

X_{it}: prediktor unit *cross-sectional* ke-i pada unit waktu ke-t

2.2 *Generalized Linear Models* (GLM)

GLM pertama kali diperkenalkan oleh Nelder dan Wedderburn pada tahun 1972, merupakan pengembangan dari model linier klasik, yang mampu mengatasi masalah ketaknormalan respon. Peubah respon diasumsikan memiliki sebaran yang termasuk dalam keluarga eksponensial (Kleinbaum dan Klein, 2010).

Agresti (2002) menyajikan tiga komponen utama dalam GLM, yakni:

1. Komponen acak, yakni peubah respon Y yang saling bebas.
2. Komponen tetap yang disebut prediktor linier dalam bentuk:
$$\eta_i = \sum_{j=1}^{p} \beta_j x_{ij} \ , i = 1, \dots, n \qquad (2.1)$$
di mana x_{ij} adalah nilai peubah prediktor subjek ke-i pada peubah prediktor ke-j dan β_j parameter peubah prediktor ke-j
$$\eta_1 = \sum_{j=1}^{p} \beta_j x_{1j} = \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13} + \cdots + \beta_p x_{1p}$$
$$\eta_2 = \sum_{j=1}^{p} \beta_j x_{2j} = \beta_1 x_{21} + \beta_2 x_{22} + \beta_3 x_{23} + \cdots + \beta_p x_{2p}$$
$$\vdots$$
$$\eta_n = \sum_{j=1}^{p} \beta_j x_{nj} = \beta_1 x_{n1} + \beta_2 x_{n2} + \beta_3 x_{n3} + \cdots + \beta_p x_{np}$$
3. Fungsi penghubung $g(\mu_i)$ menghubungkan suatu fungsi nilai rata-rata komponen acak dengan komponen tetap η_i yaitu $\eta_i = g(\mu_i)$.

$$\eta_1 = g(\mu_1) \rightarrow \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13} + \cdots + \beta_p x_{1p} = g(\mu_1)$$
$$\eta_2 = g(\mu_2) \rightarrow \beta_1 x_{21} + \beta_2 x_{22} + \beta_3 x_{23} + \cdots + \beta_p x_{2p} = g(\mu_2)$$
$$\vdots$$
$$\eta_n = g(\mu_n) \rightarrow \beta_1 x_{n1} + \beta_2 x_{n2} + \beta_3 x_{n3} + \cdots + \beta_p x_{np} = g(\mu_n)$$

2.3 Regresi Logistik

Hosmer dan Lemeshow (2002) menjelaskan bahwa regresi logistik merupakan salah satu metode regresi yang digunakan untuk mencari hubungan antara peubah respon dengan satu atau lebih peubah prediktor kontinyu maupun kategorik. Banyaknya kategori peubah respon dalam model logistik dibedakan menjadi dikotom dan politom.

Agresti (2002) mendefinisikan data biner sebagai hasil satu tindakan sukses (1) atau gagal (0) yang mengikuti sebaran Bernoulli. Untuk n tindakan maka banyaknya sukses (y) mengikuti sebaran Binomial dengan fungsi peluang:

$$P(Y = y) = \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y} \text{ ; y= 0, 1,..., n} \qquad (2.2)$$

n: banyaknya tindakan
p: peluang sukses

Subjek tidak hanya dikategorikan menjadi dua, bahkan lebih. Fahrmeir dan Gerhard (1994) menjelaskan bahwa subjek yang diklasifikasikan ke dalam lebih dari dua kategori, menghasilkan peubah respon politom. Peubah respon $Y_1, Y_2, ..., Y_k$ mengikuti sebaran multinomial dengan fungsi peluang:

$$P(Y_k = y_k) = P(Y_1 = y_1, \ldots, Y_q = y_q) = \frac{n!}{y_1! \ldots y_q!} (p_1)^{y_1} \ldots (p_q)^q \qquad (2.3)$$

n :banyak tindakan; $n = \sum_{k=1}^{q} y_k$
p_k : peluang sukses pada kategori ke-k , $\sum P_k = 1$.

Model umum persamaan regresi logistik dengan p peubah prediktor yakni:

$$\log[P(Y_k = 1|x_i)] = l \log\left[\frac{\pi_k}{1-\pi_k}\right] = \alpha + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}$$

$$\frac{\pi_k}{1-\pi_k} = \exp(\alpha + \beta_1 X_{i1} + \cdots + \beta_p X_{ip})$$

$$\pi_k = \exp(\alpha + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}) - \pi_j \left(\exp(\alpha + \beta_1 X_{i1} + \cdots + \beta_p X_{ip})\right)$$

$$\pi_k + \pi_k \left(\exp(\alpha + \beta_1 X_{i1} + \cdots + \beta_p X_{ip})\right) = \exp(\alpha + \beta_1 X_{i1} + \cdots + \beta_p X_{ip})$$

$$\pi_k(1+\exp(\alpha+\beta_1 X_{i1}+\cdots+\beta_p X_{ip}) = \exp(\alpha+\beta_1 X_{i1}+\cdots+\beta_p X_{ip})$$

$$[P(Y_k=1|x_i)] = \pi_k = \frac{\exp(\alpha+\beta_1 X_{i1}+\cdots+\beta_p X_{ip})}{1+\exp(\alpha+\beta_1 X_{i1}+\cdots+\beta_p X_{ip})} \quad (2.4)$$

α: intersep model logit

β: parameter koefisien peubah prediktor ($\beta_1, \beta_2, \ldots, \beta_p$)

$x_{i1},\ldots, x_{ip}$: peubah prediktor

p: banyaknya peubah prediktor (j= 1, 2, ...,p)

2.4 Regresi Logistik Ordinal

Fahrmeir dan Tutz (1994) menyarankan model regresi logistik ordinal digunakan untuk peubah respon Y dengan k≥2 berskala ordinal dan peubah prediktor X berupa kontinyu maupun kategorik.

Pandang contoh acak dari sebaran bersama (Y, X), di mana Y merupakan peubah respon berskala ordinal dan $\mathbf{X'= (X_{i1}, X_{i2},..., X_{ip})}$ adalah vektor peubah prediktor. Peubah respon ordinal Y dengan k kategori dapat dipandang sebagai k titik pada peubah acak kontinyu tidak teramati (**U**). Bila Y= k dan α_j dalam selang $-\infty<\alpha_k<+\infty$, maka dapat dikatakan bahwa $\alpha_{k-1}<\mathbf{U}<\alpha_k$ untuk k= 1, 2, ..., q, dengan mengasumsikan bahwa peubah U ditentukan oleh peubah prediktor dengan bentuk linier:

$$\boldsymbol{U = X'\beta + \varepsilon} \quad (2.5)$$

di mana $\mathbf{\beta'= (\beta_1, \beta_2, ..., \beta_p)}$ adalah parameter dan ε peubah acak. Dari persamaan (2.5) diperoleh model peluang kumulatif peubah respon Y berikut:

$$P(Y \le k|X) = F(\boldsymbol{\alpha_k + X'\beta})$$
$$= P(Y=1|X) + P(Y=2|X) + \cdots P(Y=k|X) \quad (2.6)$$

Pandang fungsi sebaran logistik:

$$F(x) = \frac{1}{1+\exp(-x)} \quad (2.7)$$

dan model peluang kumulatif, yakni:

$$P(Y \le k|X) = F(\boldsymbol{\alpha_k + X'\beta})$$
$$= \frac{1}{1+\exp-(\boldsymbol{\alpha_k + X'\beta})}$$
$$= \frac{1}{1+\exp-(\boldsymbol{\alpha_k + X'\beta})} \cdot \frac{\exp(\boldsymbol{\alpha_k + X'\beta})}{\exp(\boldsymbol{\alpha_k + X'\beta})}$$

$$= \frac{\exp(\boldsymbol{\alpha_k}+\boldsymbol{X'\beta})}{1+\exp(\boldsymbol{\alpha_k}+\boldsymbol{X'\beta})} \quad (2.8)$$

k = 1, 2, ..., q

α_k : intersep

β : parameter peubah prediktor (β_1,β_2, ..., β_p)

Dari persamaan (2.8) diperoleh persamaan:

$$P(Y > k|X) = 1 - P(Y \leq k|X)$$
$$= 1 - \frac{\exp(\boldsymbol{\alpha_k} + \boldsymbol{X'\beta})}{1 + \exp(\boldsymbol{\alpha_k} + \boldsymbol{X'\beta})}$$
$$= \frac{1}{1+\exp(\boldsymbol{\alpha_k}+\boldsymbol{X'\beta})} \quad (2.9)$$

Nisbah (rasio) dua peluang menghasilkan

$$\frac{P(Y \leq k|X)}{P(Y > k|X)} = \frac{P(Y \leq k|X)}{1 - P(Y \leq k|X)}$$
$$= \frac{\left(\frac{\exp(\boldsymbol{\alpha_k} + \boldsymbol{X'\beta})}{1 + \exp(\boldsymbol{\alpha_k} + \boldsymbol{X'\beta})}\right)}{\left(\frac{1}{1 + \exp(\boldsymbol{\alpha_k} + \boldsymbol{X'\beta})}\right)}$$
$$= \exp(\boldsymbol{\alpha_k} + \boldsymbol{X'\beta}) \quad (2.10)$$

Untuk memperoleh model linier maka persamaan (2.10) dilogaritmakan menjadi

$$logit[(P(Y \leq k|X)] = ln\left(\frac{P(Y \leq k|X)}{1 - P(Y \leq k|X)}\right)$$
$$= \ln(\exp(\boldsymbol{\alpha_k} + \boldsymbol{X'\beta}))$$
$$g(\mu) = (\boldsymbol{\alpha_k} + \boldsymbol{X'\beta}) \quad (2.11)$$

2.4.1 *Proportional Odds*

Menurut Agresti (2002), model kumulatif logit yang digunakan secara bersamaan berdasarkan persamaan (2.11), yakni:

$$logit\lfloor P(Y \leq k|x_i)\rfloor = \boldsymbol{\alpha_k} + \boldsymbol{\beta x_i},\ i = 1, \ldots, n \quad (2.12)$$

Tiap kumulatif logit memiliki intersep tersendiri, α_k akan meningkat di q. Model persamaan (2.12) memiliki efek β sama untuk tiap logit karena β tidak dipengaruhi oleh i.

Persamaan kumulatif odds adalah

$$log\left\{\frac{P(Y\leq k|X)}{P(Y>k|X)}\right\} \quad k = 1,2,\dots,q \tag{2.13}$$

Bentuk persamaan (2.13) ditentukan oleh bentuk linier dari peubah prediktor. Jika terdapat dua populasi yang dimisalkan x_1 dan x_2 maka rasio dari kumulatif odds dua populasi yakni:

$$log\left\{\frac{P(Y\leq k|x_1)/P(Y>k|x_1}{P(Y\leq k|x_2)/P(Y>k|x_2)}\right\} = exp\{(x_1 - x_2)\beta\} \tag{2.14}$$

Persamaan (2.14) disebut kumulatif rasio odds, yang tidak bergantung pada kategori, sehingga kumulatif rasio odds dua populasi sama untuk semua kumulatif odds (Fahrmeir dan Tutz, 1994).

2.4.2 Pendugaan Parameter Model Regresi Logistik Ordinal

Salah satu metode pendugaan parameter yang dapat dipergunakan pada regresi logistik ordinal adalah *Maximum Likelihood Estimator* (MLE). Jika nilai X tertentu menghasilkan lebih dari satu pengamatan Y dan digolongkan menjadi k kategori, maka n_k adalah banyaknya pengamatan pada kategori ke-k menyebar y~Multinomial (y_1, y_2, ..., y_q, π_1, π_2,... π_q) maka *likelihood* untuk pengamatan y adalah:

$$\begin{aligned}\ell(\alpha,\beta) &= \prod(P_k)^{y_k} \\ &= \prod_{i=1}^{n}\left(\pi_1(X)\right)^{y_{i1}}\left(\pi_2(X)\right)^{y_{i2}}\dots\left(\pi_q(X)\right)^{y_{iq}} \\ &= \prod_{i=1}^{n}\left(\frac{e^{\alpha_1+X'\beta}}{1+e^{\alpha_1+X'\beta}}\right)^{y_{i1}}\left(\frac{e^{\alpha_2+X'\beta}}{1+e^{\alpha_2+X'\beta}}\right)^{y_{i2}}\dots \\ &\qquad \left(\frac{e^{\alpha_{k-1}+X'\beta}}{1+e^{\alpha_{k-1}+X'\beta}}\right)^{y_{iq}}\end{aligned} \tag{2.15}$$

y_k : vektor peubah untuk k kategori
n : total pengataman (y_{i1}, y_{i2},...,y_{iq})
$\pi_k(X)$: vektor peluang untuk y_k

Untuk mendapatkan nilai maksimum dari fungsi *likelihood* maka kedua ruas harus dilogaritmakan. Fungsi log-*likelihood* untuk persamaan (2.16)

$$\begin{aligned}L(\alpha,\beta) &= Log\left[\prod_{i=1}^{n}\left(\pi_1(X)\right)^{y_{i1}}\left(\pi_2(X)\right)^{y_{i2}}\dots\left(\pi_q(X)\right)^{y_{iq}}\right] \\ &= \sum_{i=1}^{n}[y_{i1}\log\left(\pi_1(X)\right) + y_{i2}\log\left(\pi_2(X)\right) + \cdots \\ &\qquad + y_{iq}\log\left(\pi_q(X)\right)]\end{aligned}$$

$$= \sum_{i=1}^{n} \left[y_{i1} log \left(\frac{e^{\alpha_1 + X'\beta}}{1+e^{\alpha_1 + X'\beta}} \right) + y_{i2} log \left(\frac{e^{\alpha_2 + X'\beta}}{1+e^{\alpha_2 + X'\beta}} \right) + \cdots + y_{iq} log \left(\frac{e^{\alpha_{k-1} + X'\beta}}{1+e^{\alpha_{k-1} + X'\beta}} \right) \right] \quad (2.16)$$

Nilai $\boldsymbol{\alpha}=[\alpha_1,\ldots,\alpha_{k-1}]$ dan $\boldsymbol{\beta}=[\beta_1,\beta_2,\ldots,\beta_p]$ didapatkan dari turunan parsial pertama masing – masing fungsi yang disamakan nol.

Hasil turunan pertama terhadap α_1 adalah sebagai berikut:

$$\frac{\partial L}{\partial \alpha_1} = \sum_{i=1}^{n} \left\{ y_{i1} \left(1 - \frac{e^{\boldsymbol{\alpha_1 + X'\beta}}}{1+e^{\boldsymbol{\alpha_1 + X'\beta}}} \right) + y_{i2} \left(- \frac{e^{\boldsymbol{\alpha_1}}}{e^{\boldsymbol{\alpha_2 - \alpha_1}}} - \frac{e^{\boldsymbol{\alpha_1 + X'\beta}}}{1+e^{\boldsymbol{\alpha_1 + X'\beta}}} \right) \right\}$$

Disamakan dengan nol:

$$\frac{\partial L}{\partial \alpha_1} = \sum_{i=1}^{n} \left\{ y_{i1} \left(1 - \frac{e^{\boldsymbol{\alpha_1 + X'\beta}}}{1+e^{\boldsymbol{\alpha_1 + X'\beta}}} \right) + y_{i2} \left(- \frac{e^{\boldsymbol{\alpha_1}}}{e^{\boldsymbol{\alpha_2 - \alpha_1}}} - \frac{e^{\boldsymbol{\alpha_1 + X'\beta}}}{1+e^{\boldsymbol{\alpha_1 + X'\beta}}} \right) \right\} = 0 \quad (2.17)$$

Turunan pertama terhadap α_{k-1} adalah:

$$\frac{\partial L}{\partial \alpha_{k-1}} = \sum_{i=1}^{n} \begin{Bmatrix} (y_{k-1} - y_{k-2}) \left(- \frac{e^{\boldsymbol{\alpha_{k-1}}}}{e^{\boldsymbol{\alpha_{k-1} - \alpha_{j-2}}}} - \frac{e^{\boldsymbol{\alpha_{k-1} + X'\beta}}}{1+e^{\boldsymbol{\alpha k + X'\beta}}} \right) - \\ (n - y_{k-1}) \left(\frac{e^{\boldsymbol{\alpha_{k-1} + X'\beta}}}{1+e^{\boldsymbol{\alpha_{k-1} + X'\beta}}} \right) \end{Bmatrix}$$

Disamakan dengan nol:

$$\frac{\partial L}{\partial \alpha_{k-1}} = \sum_{i=1}^{n} \left\{ (y_{k-1} - y_{k-2}) \left(- \frac{e^{\boldsymbol{\alpha_{k-1}}}}{e^{\boldsymbol{\alpha_{k-1} - \alpha_{j-2}}}} - \frac{e^{\boldsymbol{\alpha_{k-1} + X'\beta}}}{1+e^{\boldsymbol{\alpha_{k-1} + X'\beta}}} \right) - (n - y_{k-1}) \left(\frac{e^{\alpha_{k-1} + X'\beta}}{1+e^{\alpha_{k-1} + X'\beta}} \right) \right\} = 0 \quad (2.18)$$

Hasil turunan pertama β adalah sebagai berikut:

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n} \left\{ y_1 \left(X' - \frac{X' e^{\alpha_1 + X'\beta}}{1+e^{\alpha_1 + X'\beta}} \right) + (y_2 - y_1) \left(X' - \frac{X' e^{\alpha_1 + X'\beta}}{1+e^{\alpha_1 + X'\beta}} \right) + \cdots + (1 - y_{k-1}) \left(X' - \frac{X' e^{\alpha_{k-1} + X'\beta}}{1+e^{\alpha_{k-1} + X'\beta}} \right) \right\}$$

Disamakan dengan nol:

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n} \left\{ y_1 \left(X' - \frac{X' e^{\alpha_1 + X'\beta}}{1+e^{\alpha_1 + X'\beta}} \right) + (y_2 - y_1) \left(X' - \frac{X' e^{\alpha_1 + X'\beta}}{1+e^{\alpha_1 + X'\beta}} \right) + \cdots + (1 - y_{k-1}) \left(X' - \frac{X' e^{\alpha_{k-1} + X'\beta}}{1+e^{\alpha_{k-1} + X'\beta}} \right) \right\} = 0 \quad (2.19)$$

Agresti (2002) menjelaskan bahwa dengan metode iterasi *Newton-Raphson* akan didapat penduga parameter secara berulang – ulang hingga konvergen pada nilai tertentu. Metode iterasi *Newton-Raphson* dapat menyelesaikan persamaan fungsi *likelihood*, untuk menentukan nilai $\hat{\alpha}$ dan $\hat{\beta}$ dengan rumus iterasi:

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - (\mathbf{H}^{(t)})^{-1}\mathbf{g}^{(t)} \quad (2.20)$$

sehingga $H^{(t)}$ dan $g^{(t)}$ berupa matriks

$$\boldsymbol{H}^{(t)} = \frac{\partial^2 \ell(\theta)}{\partial \theta_i} = \boldsymbol{X'V^{-1}X} \quad (2.21)$$

$$\boldsymbol{g}^{(t)} = \frac{\partial \ell(\theta)}{\partial \theta_i} = \boldsymbol{X'(Y-\pi)} \quad (2.22)$$

$\mathbf{(H^{(t)})^{-1}}$ adalah matriks peragam. Matriks **X** berukuran (nxp+1) dengan p peubah prediktor $\mathbf{V^{-1}}$ adalah matriks dengan diagonal utama $\hat{\pi}_{ik}(1-\hat{\pi}_{ik})$, dengan k=1,2 dan i=1,...,n sehingga didapatkan iterasi *Newton Raphson* sebagai berikut:

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - (\boldsymbol{X'V^{-1}X})^{-1}(\boldsymbol{X'(Y-\pi)}) \quad (2.23)$$

Agresti (2002) menyarankan bahwa melalui iterasi *Newton Raphson* $\mathbf{H^{(t)}}$ dan $\mathbf{g^{(t)}}$ digunakan untuk menduga $\boldsymbol{\theta}^{(t)}$ pada iterasi ke-t=0, 1, 2,... Iterasi untuk pendugaan nilai θ_j terus dilakukan hingga mencapai kondisi konvergen berdasarkan indikator:

$$\left|\hat{\theta}_j^{(t)} - \hat{\theta}_j^{(t-1)}\right| < \varepsilon = 10^{-6}$$

2.4.3 Pengujian Parameter

Pengujian parameter secara serentak dan parsial dilakukan untuk mengetahui apakah peubah prediktor menentukan nilai peubah respon.

a. Pengujian secara parsial

Dilakukan untuk mengetahui pengaruh peubah prediktor terhadap peubah respon secara individu dengan uji Wald berdasarkan hipotesis:

H_0: $\beta_j = 0$ H_1: $\beta_j \neq 0$

Jika H_0 benar

$$\hat{\beta}_j \sim N(0, \sigma^2_{\hat{\beta}_j})$$

$$W = \frac{\hat{\beta}_j - 0}{\sigma_{\hat{\beta}_j}} \sim Z(0,1) \quad (2.24)$$

$\sigma_{\hat{\beta}_j}$ adalah salah baku penduga β_j. Nilai $\sigma_{\hat{\beta}_j}$ diperoleh dari akar nilai diagonal utama matriks ragam peragam $(\hat{\beta}_j) = \mathbf{[X'V^{-1}X]^{-1}}$

$$\boldsymbol{V}^{-1} = \begin{bmatrix} n\hat{\pi}_1\,(1-\hat{\pi}_1) & 0 & \dots & 0 \\ 0 & n\hat{\pi}_2\,(1-\hat{\pi}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & n\hat{\pi}_n\,(1-\hat{\pi}_n) \end{bmatrix}$$

(nxn)

$\sigma^2_{\hat{\beta}_j}$ = diagonal utama matriks $\mathbf{[X'V^{-1}X]^{-1}}$

H_0 ditolak jika P(|Z| >W) < α, yang menunjukkan bahwa peubah prediktor berpengaruh terhadap peubah respon (Hedeker dan Gibbons, 2006).

b. Pengujian secara serentak

Menurut Hosmer and Lemeshow(2002) pengujian dilakukan untuk menguji keberartian parameter secara serentak menggunakan uji G yang diperoleh dari pendekatan *likelihood ratio test* berdasarkan hipotesis:

H_0: β_j=0

H_1: Paling tidak ada satu j di mana $\beta_j \neq 0$

Jika H_0 benar

$$G = -2(L_0 - L_p) \sim \chi^2_{(p)} \qquad (2.25)$$

L_0 : log-likelihood model intersep $L_0 = \prod_{k=1}^{q} \left(\frac{n_{..k}}{n_{...}}\right)$

$n_{..k}$: frekuensi setiap kategori ke-*k*

L_1 : log-likelihood model dengan intersep dan semua peubah penjelas (model penuh)

$$L_1 = \prod_{k=1}^{q} \prod_{j=1}^{p} \prod_{l=1}^{m} (\hat{\pi}_k(x_i))^{n_{ljk}}$$

n_{ljk} = frekuensi setiap kategori ke-*l* pada peubah penjelas ke-*j* kategori respon ke-*k*

H_0 akan ditolak jika P[$\chi^2_{(p)}$ >G]<α yang berarti paling sedikit terdapat satu β_j yang tidak sama dengan nol.

2.5 *Generalized Estimation Equations* (GEE)

Liang dan Zeger (1986) memperkenalkan *Generalized Estimating Equation* (GEE) sebagai metode pengembangan dari GLM yang digunakan untuk menduga parameter model berdasarkan data yang mengandung autokorelasi dan tidak menyebar Normal. Persamaan GEE secara umum adalah:

$$g(\mu) = \boldsymbol{X'\beta} \quad (2.26)$$

Fungsi penghubung pada regresi logistik adalah log e sehingga model GEE dapat ditulis:

$$g(\mu) = logit[(P(Y \leq k|X) = \boldsymbol{\alpha_k} + \sum_{j=1}^{p} \boldsymbol{\beta_j} \mathbf{X_{ij}} \quad (2.27)$$

1.5.1 Pendugaan Parameter

Hedeker dan Gibbons (2006) menjelaskan bahwa penduga parameter β didapatkan dengan menyelesaikan fungsi *quasi score (S(β))* yang disebut sebagai Generalized Estimating Equations:

$$\boldsymbol{S(\beta)} = \sum_{i=1}^{n} \boldsymbol{D^T V^{-1}(Y_i - \mu_i)} = \mathbf{0} \quad (2.28)$$

$$\boldsymbol{D'} = \left(\frac{\partial \mu_i}{\partial \beta}\right)' = \begin{bmatrix} \frac{x_{i11}}{g'(\mu_{i1})} & \cdots & \frac{x_{it1}}{g'(\mu_{iT})} \\ \vdots & & \vdots \\ \frac{x_{i1p}}{g'(\mu_{i1})} & \cdots & \frac{x_{itp}}{g'(\mu_{iT})} \end{bmatrix}$$

V merupakan matriks ragam-peragam Y_i yakni:

$$\boldsymbol{V} = \boldsymbol{\phi A^{\frac{1}{2}} R(\alpha) A^{\frac{1}{2}}} \quad (2.29)$$

di mana **A** merupakan matriks berukuran nxn dengan ragam ($V(\mu_{ij})$) sebagai elemen diagonal ke-n dan $\boldsymbol{R(\alpha)}$ adalah matriks korelasi berukuran nxn (Hedeker dan Gibbons, 2006).

Burton *et al*. dalam Twisk (2003) menjelaskan pendugaan parameter GEE terdiri tiga tahap, yakni:

1. Pendugaan parameter β menggunakan GLM dengan metode *Maximum likelihood*.
2. Pembentukan matriks korelasi $\boldsymbol{R(\alpha)}$ berdasarkan nilai *Pearson residual* dan parameter dispersi yang dihasilkan pada tahap pertama.

   $$e_{ij} = \frac{y_{ij} - \bar{y}_{.j}}{s_j} \quad (2.30)$$

   di mana e_{ij} adalah *Pearson residual* dan y_{ij} adalah ragam pada waktu ke-t yang nilainya sama dengan $\bar{y}_{.j}$. Parameter dispersi diduga dengan persamaan

   $$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^{n} \sum_{j=1}^{p} e_{ij}^2 \quad (2.31)$$

di mana $\hat{\phi}$ merupakan parameter dispersi, n adalah ukuran contoh dan p banyaknya parameter regresi.

3. Iterasi pendugaan parameter.
   Untuk mendapatkan penduga parameter yang konvergen maka dipergunakan iterasi *Newton Raphson* hingga $\hat{\beta}$ konvergen.

Pada GEE, korelasi yang terjadi dibentuk dalam sebuah matriks korelasi $\boldsymbol{R(\alpha)}$ berukuran nxn, di mana struktur matriks korelasi tidak diketahui dan harus diduga. Kleinbaum dan Klein (2010) menjelaskan bahwa terdapat lima struktur matriks korelasi yakni:

1. *Independent*, diasumsikan bahwa tidak terdapat korelasi antar pengamatan. Pada data longitudinal, struktur korelasi ini tidak tepat digunakan karena ulangan pada setiap subjek sama akan menyebabkan korelasi.
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
2. *Exchangable*, diasumsikan bahwa korelasi antar pengamatan memiliki nilai yang hampir sama.
$$\begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$
3. *Autoregressive AR(1)*, diasumsikan bahwa semua korelasi antar pengamatan dipisahkan sejauh t = 1 bernilai ρ, pada t = 2 bernilai ρ^2, hingga t waktu pengamatan bernilai ρ^t.
$$\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$
4. *m-dependent*, diasumsikan bahwa korelasi antar pengamatan yang dipisahkan oleh t waktu akan bernilai sama, begitu juga pada t+1 waktu, dengan t = 1,2,..,m. Korelasi yang berjarak lebih dari m akan bernilai nol. Contoh: *2-dependent (m =2, n=4)*
$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & 0 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ 0 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

5. *Unstructured*, diasumsikan bahwa korelasi antar pengamatan memiliki nilai yang berbeda. Struktur korelasi ini umum digunakan karena tidak ada satupun struktur korelasi yang secara logis dapat merepresentasikan data, jumlah ulangan kecil dan seimbang untuk tiap unit penelitian (Swan, 2006).

$$\begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{12} & 1 & \rho_{23} & \rho_{24} \\ \rho_{13} & \rho_{23} & 1 & \rho_{34} \\ \rho_{14} & \rho_{24} & \rho_{34} & 1 \end{bmatrix}$$

2.5.2 Pendugaan Matriks Ragam-Peragam Parameter

Hedeker dan Gibbons (2006) menjelaskan matriks ragam-peragam parameter cov (β) perlu diduga untuk memilih struktur korelasi dan pengujian signifikansi parameter. Penduga matriks cov (β) terdapat dua jenis yakni:

1. *Naive/model-based estimator*

   Digunakan apabila struktur korelasi diketahui secara pasti

$$\boldsymbol{V_m(\widehat{\beta}) = \left[D'\ V^{-1} D\ \right]^{-1}} \quad (2.32)$$

2. *Robust /empirical/sandwich estimator*

   Digunakan apabila struktur korelasi tidak diketahui secara pasti

$$\boldsymbol{V_e(\widehat{\beta}) = M_0^{-1} M_1 M_0^{-1}} \quad (2.33)$$

$$\boldsymbol{M_0 = D'\ V^{-1} D}$$

$$\boldsymbol{M_1 = \sum_{i=1}^{n} D'\ V^{-1} D\ \ (y_i - \widehat{\mu}_i)(y_i - \widehat{\mu}_i)' V^{-1} D}$$

2.5.3 Generalized Wald Test dan Wald Test

Pengujian parameter pada GEE dilakukan secara simultan maupun parsial. Pada pengujian signifikansi parameter secara simultan digunakan *Generalized Wald Test* didasarkan hipotesis

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

H_1: paling tidak ada satu j di mana $\beta_j \neq 0$

Jika H_0 benar , statistik uji

$$\boldsymbol{\beta' C' (C V_e(\widehat{\beta}) C')^{-1} C \beta} \ \sim {\chi^2}_q \quad (2.34)$$

di mana C adalah matriks berukuran p x q, p = banyaknya parameter, q = banyaknya parameter yang diuji dan $\boldsymbol{V_e(\hat{\beta})}$ matriks ragam–peragam *empirical estimator*. Setiap baris dalam matriks C mengandung nilai 1 untuk parameter yang akan diuji dan 0 untuk parameter lain. H_0 ditolak jika $P[\chi^2_{(q)} > \chi^2] < \alpha$ (Hedeker dan Gibbons, 2006).
Sedangkan statistik uji *Wald* digunakan untuk pengujian secara parsial (2.24).

2.5.4 Pemilihan Struktur Korelasi

Pemilihan struktur korelasi berguna untuk mendapatkan struktur korelasi yang sesuai dengan data. Swan (2006) memberikan petunjuk dalam pemilihan struktur korelasi, yakni:

1. Unstructured, digunakan bila jumlah ulangan kecil dan seimbang.
2. Autoregressive AR(m), digunakan bila ulangan tiap objek berdasarkan waktu, dengan menetukan orde m terlebih dahulu.
3. Exchangable, digunakan bila ulangan tiap objek tidak berdasarkan waktu.
4. Independent, digunakan bila jumlah objek pengamatan kecil.
5. Bila terdapat lebih dari satu kemungkinan, maka digunakan *Quasi-likelihood under the independence Information Criterion* (QIC) .
   Rumus QIC sebagai berikut:

$$QIC = -2Q(\beta) + 2(\boldsymbol{V_m}^{-1}(\hat{\boldsymbol{\beta}})\boldsymbol{V_e}(\hat{\boldsymbol{\beta}})) \quad (2.35)$$

dengan Q(β) adalah nilai Quasi-likelihood dengan mengabaikan struktur korelasi atau menggunakan struktur korelasi independent, $\boldsymbol{V_m}(\hat{\boldsymbol{\beta}})$ dan $\boldsymbol{V_e}(\hat{\boldsymbol{\beta}})$ adalah matriks ragam peragam penduga parameter *model-based estimator* dan *sandwich estimator*. Model dengan nilai QIC terkecil merupakan model dengan struktur korelasi terbaik.