BAB III

ESTIMASI PARAMETER PADA MODEL REGRESI LOGISTIK 2-LEVEL

Model hirarki 2-level merupakan model statistik yang digunakan untuk menganalisis data yang bersarang, atau data yang mempunyai struktur hirarki 2-level. Sebagai contoh data 2-level, murid bersarang pada sekolah, dimana murid merupakan data level-1 dan sekolah merupakan data level-2. Dengan menganalisis data yang bersifat hirarki 2-level menggunakan model regresi 2-level dapat diketahui informasi mengenai variasi pada level-2 yang tidak dapat diketahui jika analisis yang digunakan adalah model regresi biasa (tanpa memperhatikan struktur hirarki pada data). Jika variabel respon yang diamatinya berupa variabel respon biner (misal kejadian sukses dan gagal) atau data binomial, maka model yang dibutuhkan adalah model regresi logistik 2-level. Sebelum membahas mengenai model regresi logistik 2-level, akan dibahas mengenai estimasi parameter dalam model regresi 2-level dengan metode IGLS.

3.1 Estimasi Parameter Model Regresi 2-Level

Pada bab sebelumnya telah dibahas mengenai karakteristik data hirarki dan bentuk model regresi 2-level. Pada data hirarki, kemiripan karakteristik unit-unit pada level-1 dalam unit level-2 yang sama menyebabkan data hirarki tidak bersifat independen, sehingga metode *Ordinary Least Square* (OLS) kurang tepat untuk digunakan. Dalam subbab ini akan dibahas mengenai cara estimasi parameter-parameter dalam model regresi 2-level yang dapat mentolerir karakteristik data hirarki tersebut, yaitu dengan menggunakan metode *Iterative Generalized Least Square* (IGLS). Metode IGLS akan berguna dalam estimasi parameter dalam model regresi logistik 2-level yang akan dibahas di subbab selanjutnya.

Penaksiran dengan metode IGLS dilakukan dengan menaksir parameter-parameter tetap (*fixed parameter*) terlebih dahulu dengan diketahui suatu matriks varians-kovarians $\boldsymbol{V}$ menggunakan *Generalized Least Square* (GLS), selanjutnya hasil taksiran yang diperoleh digunakan untuk menaksir parameter random dalam model menggunakan GLS. Prosedur estimasi *fixed parameter* dan *random parameter* dilakukan berulang-ulang secara bergantian sampai mendapatkan taksiran yang konvergen. Untuk memudahkan penerapan prosedur estimasi, maka persamaan model regresi 2-level dengan *random intercept* dituliskan dalam bentuk matriks sebagai berikut :

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E} \tag{3.1}$$

dimana

- $\boldsymbol{Y}$ merupakan vektor berukuran *nx1* yang berisi observasi-observasi, dituliskan :

$$\mathbf{Y} = \{y_{ij}\} = \begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n_1 1} \\ y_{12} \\ y_{22} \\ \vdots \\ y_{n_2 2} \\ \vdots \\ y_{1m} \\ \vdots \\ y_{n_m m} \end{bmatrix},$$

y_{ij} = respon untuk unit ke-*i* dalam unit level-2 ke-*j*

j = 1, 2, … , *m* menyatakan unit-unit level-2

i = 1, 2, … , n_j menyatakan unit-unit level-1 yang bersarang dalam tiap unit level-2.

Total observasi dinyatakan oleh *n*, dengan $n = \sum_{j=1}^{m} n_j$

- ***X*** merupakan *design matrix* ukuran *nx(P+1)* berisi konstanta satu dan observasi untuk *P*-variabel penjelas, *m* unit level-2, dan *n* jumlah total observasi, ditulis sebagai berikut :

$$\mathbf{X} = \begin{bmatrix} 1 & x_{111} & \cdots & x_{P11} \\ 1 & x_{121} & & x_{P21} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n_m m} & \cdots & x_{Pn_m m} \end{bmatrix}$$

$x_{Pn_m m}$ = observasi variabel penjelas ke-*P* untuk unit level-1 ke-n_m dalam unit level-2 ke-*m*

- **β** adalah vektor ukuran *(P+1)x1* yang berisi *fixed parameter,*

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_P \end{bmatrix}$$

- ***E*** menyatakan matriks yang berisi penjumlahan residual level-1 dan level-2, dimana $\mathbf{E} = \{e_{ij}\}$, $e_{ij} = \varepsilon_{ij} + u_{0j}$. σ_ε^2 menyatakan variansi *error* level-1 dan σ_{u0}^2 menyatakan variansi *error* level-2.

Parameter-parameter yang akan ditaksir pada model regresi 2-level seperti yang telah dituliskan ulang dalam persamaan (3.1) adalah *fixed* parameter β_p dengan *p* = 1,2,..,*P* dan *random* parameter, σ_{u0}^2.

Langkah pertama dalam menaksir parameter dengan metode IGLS adalah menaksir *fixed parameter* **β***,* untuk suatu matriks varians-kovarians ***V*** yang diketahui, dengan menggunakan *Generalized Least Square* (GLS) :

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y} \tag{3.2}$$

Sebagai taksiran inisial, matriks varians-kovarians yang digunakan pada persamaan (3.2) adalah $\mathbf{V} = \sigma_\varepsilon^2\mathbf{I}$ (diasumsikan $\sigma_{u0}^2 = 0$), dimana ***I*** adalah matriks identitas ukuran *nxn*. Dengan diketahui $\mathbf{V} = \sigma_\varepsilon^2\mathbf{I}$, artinya pada taksiran inisial nilai taksiran diperoleh seperti pada *Ordinary Least Square*,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

Setelah taksiran dari $\boldsymbol{\beta}$ diketahui, hitung nilai-nilai taksiran untuk $\mathbf{Y}$, yaitu $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$. Sehingga dapat diperoleh nilai residual yang dinyatakan dalam bentuk :

$$\tilde{\mathbf{Y}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

Bentuk *cross product* matriks $\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T$:

$$\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T = \left[\begin{bmatrix} \tilde{y}_{11} \\ \tilde{y}_{21} \\ \vdots \\ \tilde{y}_{n_1 1} \\ \tilde{y}_{12} \\ \tilde{y}_{22} \\ \vdots \\ \tilde{y}_{n_2 2} \\ \vdots \\ \tilde{y}_{1m} \\ \vdots \\ \tilde{y}_{n_m m} \end{bmatrix}\right] \begin{bmatrix} \tilde{y}_{11} & \tilde{y}_{21} & \cdots & \tilde{y}_{n_1 1} & \tilde{y}_{12} & \tilde{y}_{22} & \cdots & \tilde{y}_{n_2 2} & \cdots & \tilde{y}_{1m} & \cdots & \tilde{y}_{n_m m} \end{bmatrix}$$

$$
= \begin{bmatrix}
\tilde{y}_{11}^2 & \tilde{y}_{11}\tilde{y}_{21} & \cdots & \tilde{y}_{11}\tilde{y}_{n_1 1} & \tilde{y}_{11}\tilde{y}_{12} & \tilde{y}_{11}\tilde{y}_{22} & \cdots & \tilde{y}_{11}\tilde{y}_{n_2 2} & \cdots & \tilde{y}_{11}\tilde{y}_{1m} & \cdots & \tilde{y}_{11}\tilde{y}_{n_m m} \\
\tilde{y}_{21}\tilde{y}_{11} & \tilde{y}_{21}^2 & \cdots & \tilde{y}_{21}\tilde{y}_{n_1 1} & \tilde{y}_{21}\tilde{y}_{12} & \tilde{y}_{21}\tilde{y}_{22} & \cdots & \tilde{y}_{21}\tilde{y}_{n_2 2} & \cdots & \tilde{y}_{21}\tilde{y}_{1m} & \cdots & \tilde{y}_{21}\tilde{y}_{n_m m} \\
\vdots & \vdots & \ddots & \vdots & & & & \vdots & & \vdots & & \vdots \\
\tilde{y}_{n_1 1}\tilde{y}_{11} & \tilde{y}_{n_1 1}\tilde{y}_{21} & \cdots & \tilde{y}_{n_1 1}^2 & \tilde{y}_{n_1 1}\tilde{y}_{12} & \tilde{y}_{n_1 1}\tilde{y}_{22} & \cdots & \tilde{y}_{n_1 1}\tilde{y}_{n_2 2} & \cdots & \tilde{y}_{n_1 1}\tilde{y}_{1m} & \cdots & \tilde{y}_{n_1 1}\tilde{y}_{n_m m} \\
\tilde{y}_{12}\tilde{y}_{11} & \tilde{y}_{12}\tilde{y}_{21} & \cdots & \tilde{y}_{12}\tilde{y}_{n_1 1} & \tilde{y}_{12}^2 & \tilde{y}_{12}\tilde{y}_{22} & \cdots & \tilde{y}_{12}\tilde{y}_{n_2 2} & \cdots & \tilde{y}_{12}\tilde{y}_{1m} & \cdots & \tilde{y}_{12}\tilde{y}_{n_m m} \\
\tilde{y}_{22}\tilde{y}_{11} & \tilde{y}_{22}\tilde{y}_{21} & \cdots & \tilde{y}_{22}\tilde{y}_{n_1 1} & \tilde{y}_{22}\tilde{y}_{12} & \tilde{y}_{22}^2 & \cdots & \tilde{y}_{22}\tilde{y}_{n_2 2} & \cdots & \tilde{y}_{22}\tilde{y}_{1m} & \cdots & \tilde{y}_{22}\tilde{y}_{n_m m} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots & & \vdots \\
\tilde{y}_{n_2 2}\tilde{y}_{11} & \tilde{y}_{n_2 2}\tilde{y}_{21} & \cdots & \tilde{y}_{n_2 2}\tilde{y}_{n_1 1} & \tilde{y}_{n_2 2}\tilde{y}_{12} & \tilde{y}_{n_2 2}\tilde{y}_{22} & \cdots & \tilde{y}_{n_2 2}^2 & \cdots & \tilde{y}_{n_2 2}\tilde{y}_{1m} & \cdots & \tilde{y}_{n_2 2}\tilde{y}_{n_m m} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \ddots & \vdots & & \vdots \\
\tilde{y}_{1m}\tilde{y}_{11} & \tilde{y}_{1m}\tilde{y}_{21} & \cdots & \tilde{y}_{1m}\tilde{y}_{n_1 1} & \tilde{y}_{1m}\tilde{y}_{12} & \tilde{y}_{1m}\tilde{y}_{22} & \cdots & \tilde{y}_{1m}\tilde{y}_{n_2 2} & \cdots & \tilde{y}_{1m}^2 & \cdots & \tilde{y}_{1m}\tilde{y}_{n_m m} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & & \ddots & \\
\tilde{y}_{n_m m}\tilde{y}_{11} & \tilde{y}_{n_m m}\tilde{y}_{21} & \cdots & \tilde{y}_{n_m m}\tilde{y}_{n_1 1} & \tilde{y}_{n_m m}\tilde{y}_{12} & \tilde{y}_{n_m m}\tilde{y}_{22} & \cdots & \tilde{y}_{n_m m}\tilde{y}_{n_2 2} & \cdots & \tilde{y}_{n_m m}\tilde{y}_{1m} & \cdots & \tilde{y}_{n_m m}^2
\end{bmatrix}
$$

Lakukan pemvektorisasian pada matriks $\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{\mathbf{T}}$:

$$
\mathbf{Y^*} = vec(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{\mathbf{T}}) = \begin{bmatrix} \tilde{y}_{11}^2 \\ \vdots \\ \tilde{y}_{n_m m}\tilde{y}_{11} \\ \tilde{y}_{11}\tilde{y}_{21} \\ \vdots \\ \tilde{y}_{n_m m}\tilde{y}_{21} \\ \vdots \\ \tilde{y}_{11}\tilde{y}_{n_m m} \\ \vdots \\ \tilde{y}_{n_m m}^2 \end{bmatrix} \tag{3.3}
$$

Operator *vec* merupakan operator yang membuat matriks ukuran *nxn* menjadi vektor ukuran *nnx1* dengan menyusun entri-entri matriks pada kolom (*s+1*) di bawah entri terakhir kolom ke-*s*, dengan *s* = 1,2,…,*n*.

Matriks varians-kovarians ***V*** ukuran *nxn* adalah matriks *block diagonal* yang dinyatakan dalam bentuk sebagai berikut :

$$\mathbf{V} = \begin{bmatrix} \mathbf{A_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A_2} & & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A_m} \end{bmatrix} \tag{3.4}$$

m = banyaknya unit level-2 yang diobservasi, dan $\mathbf{A_1}, \mathbf{A_2}, ..., \mathbf{A_m}$ adalah matriks varians-kovarians untuk masing-masing unit level-2, yang didefinisikan sebagai berikut :

$$\begin{aligned} \mathbf{A_1} &= \sigma_{u0}^2 \mathbf{J}_{(n_1)} + \sigma_\varepsilon^2 \mathbf{I}_{(n_1)} \\ \mathbf{A_2} &= \sigma_{u0}^2 \mathbf{J}_{(n_2)} + \sigma_\varepsilon^2 \mathbf{I}_{(n_2)} \\ &\vdots \\ \mathbf{A_m} &= \sigma_{u0}^2 \mathbf{J}_{(n_m)} + \sigma_\varepsilon^2 \mathbf{I}_{(n_m)} \end{aligned} \tag{3.5}$$

Jika dijabarkan, matriks varians-kovarians untuk unit level-2 ke-*j*, $\mathbf{A_j}$, dijabarkan sebagai berikut :

$$\mathbf{A_j} = \begin{bmatrix} \sigma_{u0}^2 + \sigma_\varepsilon^2 & \sigma_{u0}^2 & \cdots & \sigma_{u0}^2 \\ \sigma_{u0}^2 & \sigma_{u0}^2 + \sigma_\varepsilon^2 & & \sigma_{u0}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{u0}^2 & \sigma_{u0}^2 & \cdots & \sigma_{u0}^2 + \sigma_\varepsilon^2 \end{bmatrix}$$

$\mathbf{A_j}$ berukuran $n_j x n_j$, dengan $\mathbf{I}_{(n_j)}$ adalah matriks identitas berukuran $n_j x n_j$, dan $\mathbf{J}_{(n_j)}$ adalah matriks yang entri-entrinya berisi konstanta 1 ukuran $n_j x n_j$. Dari (3.4) dan (3.5), matriks varians-kovarians untuk *n* observasi, dimana $n = \sum_{j=1}^{m} n_j$, dinyatakan dalam bentuk :

$$\mathbf{V}=\begin{bmatrix} \sigma_{u0}^2\mathbf{J}_{(n_1)}+\sigma_\varepsilon^2\mathbf{I}_{(n_1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_{u0}^2\mathbf{J}_{(n_2)}+\sigma_\varepsilon^2\mathbf{I}_{(n_2)} & & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sigma_{u0}^2\mathbf{J}_{(n_m)}+\sigma_\varepsilon^2\mathbf{I}_{(n_m)} \end{bmatrix} \quad (3.6)$$

Lakukan pemvektorisasian pada matriks varians-kovarians ***V*** dengan menyusun entri-entri dari kolom ke-(*s+1*) di bawah entri terakhir dari kolom ke-*s* dari matriks ***V***, dengan *s* = 1,2,…,*n*. Vektorisasi matriks *V* dinyatakan dalam notasi ***V****, dengan ***V**** berukuran *nnx1* :

$$\mathbf{V}^* = vec(\mathbf{V}) = \begin{bmatrix} \sigma_{u0}^2+\sigma_\varepsilon^2 \\ \sigma_{u0}^2 \\ \vdots \\ \sigma_{u0}^2+\sigma_\varepsilon^2 \\ \sigma_{u0}^2 \\ \vdots \\ \sigma_{u0}^2+\sigma_\varepsilon^2 \end{bmatrix} \quad (3.7)$$

Diketahui nilai ekspekstasi dari $\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\mathsf{T}$ adalah ***V*** :

$$\begin{aligned} E((\mathbf{Y}-E(\mathbf{Y}))(\mathbf{Y}-E(\mathbf{Y}))^\mathsf{T}) &= E((\mathbf{Y}-\hat{\mathbf{Y}})(\mathbf{Y}-\hat{\mathbf{Y}})^\mathsf{T}) \\ &= E((\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})^\mathsf{T}) \\ &= E(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\mathsf{T}) = \mathbf{V} \end{aligned} \quad (3.8)$$

Dengan pengaturan sedemikian rupa, bisa dibentuk model linier berdasarkan (3.8) :

$$\begin{aligned} E(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\mathsf{T}) &= \mathbf{V} \\ E(vec(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\mathsf{T})) &= vec(\mathbf{V}) \\ E(\mathbf{Y}^*) &= \mathbf{V}^* \end{aligned}$$

Sehingga diperoleh model $\mathbf{Y}^* = \mathbf{V}^* + \mathbf{R}$, dengan **R** menyatakan residual, yang dijabarkan dalam bentuk berikut :

$$\mathbf{Y}^* = \mathbf{V}^* + \mathbf{R}$$

$$\begin{bmatrix} \tilde{y}_{11}^2 \\ \vdots \\ \tilde{y}_{n_m m}\tilde{y}_{11} \\ \tilde{y}_{11}\tilde{y}_{21} \\ \vdots \\ \tilde{y}_{n_m m}\tilde{y}_{21} \\ \vdots \\ \tilde{y}_{11}\tilde{y}_{n_m m} \\ \vdots \\ \tilde{y}_{n_m m}^2 \end{bmatrix} = \begin{bmatrix} \sigma_{u0}^2 + \sigma_\varepsilon^2 \\ \vdots \\ 0 \\ \sigma_{u0}^2 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ \sigma_{u0}^2 + \sigma_\varepsilon^2 \end{bmatrix} + \mathbf{R} = \begin{bmatrix} 1 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 1 \end{bmatrix} \sigma_{u0}^2 + \begin{bmatrix} 1 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 1 \end{bmatrix} \sigma_\varepsilon^2 + \mathbf{R} \qquad (3.9)$$

Pada model linier yang terbentuk dalam persamaan (3.9), ***Y**** dijadikan sebagai respon, σ_{u0}^2 dan σ_ε^2 menjadi koefisisen-koefisien model, dan vektor-vektor berisi konstanta 0 dan 1 yang bersesuaian dengan σ_{u0}^2 dan σ_ε^2 menjadi variabel-variabel penjelas. Sehingga pada (3.9), parameter-parameter yang akan ditaksir adalah σ_{u0}^2 dan σ_ε^2.

Jika vektor-vektor yang bersesuaian dengan σ_{u0}^2 dan σ_ε^2 dalam (3.9) dinotasikan sebagai $\mathbf{Z}_1^*$ dan $\mathbf{Z}_2^*$, kemudian dibentuk matriks $\mathbf{Z}^* = [\mathbf{Z}_1^* \quad \mathbf{Z}_2^*]$, dan parameter-parameter random yang akan ditaksir tergabung dalam vektor $\boldsymbol{\theta}$, dimana $\boldsymbol{\theta} = \begin{bmatrix} \sigma_{u0}^2 \\ \sigma_\varepsilon^2 \end{bmatrix}$, maka (3.9) dapat dimodelkan dalam persamaan :

$$E(\mathbf{Y}^*) = \mathbf{Z}^* \boldsymbol{\theta} \qquad (3.10)$$

Dengan membentuk model yang dinyatakan dalam persamaan (3.10), parameter-parameter random yang ingin diketahui (σ_{u0}^2 dan σ_ε^2) dapat ditaksir. Penaksiran parameter-parameter random dilakukan dengan metode yang sama seperti pada penaksiran parameter-parameter tetap β_p, p = 1,2,…,*P,* yaitu dengan menggunakan metode *Generalized least Square* (GLS) :

$$\boldsymbol{\theta} = (\mathbf{Z}^{*T}(\mathbf{V}^*)^{-1}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*T}(\mathbf{V}^*)^{-1}\mathbf{Y}^* \qquad (3.11)$$

dengan $\mathbf{V}^* = \mathbf{V} \otimes \mathbf{V}$, ***V**** berukuran *nnxnn*.

Setelah diperoleh taksiran dari parameter-parameter random, ulangi langkah pengestimasian *fixed parameter* dengan nilai matriks varians-kovarians yang baru, kemudian hasil penaksiran *fixed parameter* digunakan untuk menaksir *random parameter,* selanjutnya dilakukan penaksiran berulang-ulang secara bergantian antara *fixed parameter* dan *random parameter* sampai konvergen, yaitu nilai taksiran tidak lagi berfluktuasi pada iterasi-iterasi berikutnya.

3.2 Model Regresi Logistik 2-Level

Model regresi logistik 2-level dapat digunakan untuk menganalisis data yang mempunyai struktur hirarki 2-level dan respon biner. Model regresi logistik 2-level dengan efek random terjadi pada perpotongan dengan sumbu-*y* (*intercept*) saja disebut sebagai model regresi logistik 2-level dengan *random intercept*.

Asumsikan terdapat *m* unit level-2 yang diambil secara acak dari populasi unit level-2, dan diambil n_j unit level-1 secara acak dari tiap-tiap unit level-2 yang telah diperoleh sebelumnya, dengan

j = 1, 2, … , *m* menyatakan unit-unit level-2

i = 1, 2, … , n_j menyatakan unit-unit level-1 yang bersarang dalam tiap unit level-2.

Jumlah total observasi level-1 dalam unit-unit level-2 adalah :

$$n = \sum_{j=1}^{m} n_j$$

Secara umum bentuk representasi multilevel dari model regresi logistik 2-level dengan *random intercept* dituliskan sebagai berikut :

Model level-1 :

$$\operatorname{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_{0j} + \sum_{p=1}^{P} \beta_p x_{pij} \tag{3.12}$$

dengan

y_{ij} = variabel respon untuk unit ke-*i* pada level-1 dalam unit ke-*j* pada level-2

β_{0j} = *random intercept* untuk unit ke-*j* pada level-2

β_p = efek tetap (*fixed effects*) untuk variabel penjelas ke-*p*

x_{pij} = variabel penjelas ke-*p* di level-1 untuk unit ke-*i* pada level-1 dalam unit ke-*j* pada level-2

Model level-2 :

$$\beta_{0j} = \beta_0 + u_{0j} \tag{3.13}$$

dengan

β_0 = *fixed intercept*, merupakan rata-rata keseluruhan

u_{0j} = efek random (*error*) untuk unit ke-*j* pada level-2, diasumsikan berdistribusi $N(0, \sigma_{u_0}^2)$

Jika model level-1 dan level-2 digabungkan diperoleh :

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 x_{1ij} + \ldots + \beta_P x_{Pij} + u_{0j} = \mathbf{x}_{ij}^{'}\boldsymbol{\beta} + u_{0j} \tag{3.14}$$

dimana

$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right)$ merupakan prediktor linier yang terdiri dari *fixed part* dari model , $\mathbf{x}_{ij}^{'}\boldsymbol{\beta}$ dan *random part* , u_{0j}

$\boldsymbol{\beta}$ = vektor *fixed effect*, berukuran *(P+1)x1*, $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_P \end{bmatrix}$

$\mathbf{x}_{ij}^{'}$ = vektor berisi variabel penjelas ukuran *1x(P+1)*,

$\mathbf{x}_{ij}^{'} = [1 \quad x_{1ij} \quad x_{2ij} \quad \ldots \quad x_{Pij}]$

Seperti pada model regresi logistik biasa, bentuk persamaan model regresi logistik 2-level dengan *random intercept* (3.14) dapat juga ditulis sebagai :

$$\pi_{ij} = f(\mathbf{x}_{ij}^{'}\boldsymbol{\beta} + u_{oj}) = \frac{1}{1 + \exp\{-(\mathbf{x}_{ij}^{'}\boldsymbol{\beta} + u_{0j})\}} \quad (3.15)$$

Seperti yang telah dijelaskan pada bab 2, pada variabel respon biner, respon kini dapat ditulis sebagai jumlah dari probabilitas π_{ij} dengan residual level-1 ε_{ij} yang diasumsikan mempunyai distribusi kumulatif logistik :

$$y_{ij} = \pi_{ij} + \varepsilon_{ij} \quad (3.16)$$

dimana π_{ij} seperti ditunjukkan dalam persamaan (3.15), ε_{ij} merupakan residual model level-1 yang mempunyai mean nol dan variansi $\pi_{ij}(1-\pi_{ij})$, y_{ij} berdistribusi Bernoulli dengan probabilitas π_{ij} : $y_{ij} \sim Bernoulli(1, \pi_{ij})$.

3.3 Estimasi Parameter Model Regresi Logistik 2-Level

Dalam skripsi ini estimasi parameter-parameter dalam model regresi logistik 2-level dilakukan dengan metode *Penalized Quasi Likelihood* (PQL). Estimasi parameter dengan metode PQL dilakukan dengan melinierkan bagian yang non-linier dari model terlebih dahulu sehingga diperoleh bentuk linier dari model. Selanjutnya, parameter-parameter dalam model yang telah dilinierkan diestimasi menggunakan prosedur estimasi model linier 2-level. Hasil taksiran parameter yang diperoleh digunakan untuk mendapatkan bentuk model yang dilinierisasi, kemudian didapat bentuk model linier yang baru untuk dicari taksiran parameter-parameternya. Proses tersebut dilakukan secara berulang-ulang sampai konvergen.

Langkah pertama dari metode PQL yaitu melinierisasi bagian yang non-linier dari model (3.16), yaitu π_{ij}, dimana π_{ij} merupakan fungsi dari $\mathbf{x}_{ij}\boldsymbol{\beta}+u_{0j}$ seperti ditunjukkan pada persamaan (3.15). Maka melinierkan π_{ij} sama saja dengan melinierkan fungsi dari $\mathbf{x}_{ij}\boldsymbol{\beta}+u_{0j}$. Cara melinierisasi $f(\mathbf{x}_{ij}\boldsymbol{\beta}+u_{0j})$ yang digunakan adalah dengan menggunakan perluasan deret Taylor.

Dimisalkan $\mathbf{x}_{ij}\boldsymbol{\beta}+u_{0j}=H$, sehingga dapat ditulis $f(H)=\pi_{ij}$ dan turunan pertama dari $f(H)$ adalah $f'(H)=\pi_{ij}(1-\pi_{ij})$ (ditunjukkan dalam Lampiran 1). Maka perluasan deret Taylor sampai order pertama untuk fungsi $f(H)$ pada dinyatakan sebagai berikut :

$$f(H)\approx f(H_0)+(H-H_0)f'(H_0)$$

Dengan mensubstitusi H dengan $\mathbf{x}_{ij}\boldsymbol{\beta}+u_{0j}$, dan H_0 menyatakan suatu nilai, maka persamaan di atas menjadi :

$$\begin{aligned}
f(\mathbf{x}_{ij}\boldsymbol{\beta}+u_{0j}) &\approx f(H_0)+((\mathbf{x}_{ij}\boldsymbol{\beta}+u_{0j})-(\mathbf{x}_{ij}\boldsymbol{\beta}^{(0)}+u_{0j}^{(0)}))f'(H_0)\\
&= f(H_0)+((\mathbf{x}_{ij}\boldsymbol{\beta}-\mathbf{x}_{ij}\boldsymbol{\beta}^{(0)})+(u_{0j}-u_{0j}^{(0)}))f'(H_0)\\
&= f(H_0)+(\mathbf{x}_{ij}\boldsymbol{\beta}-\mathbf{x}_{ij}\boldsymbol{\beta}^{(0)})f'(H_0)+(u_{0j}-u_{0j}^{(0)})f'(H_0)\\
&= f(H_0)+\mathbf{x}_{ij}(\boldsymbol{\beta}-\boldsymbol{\beta}^{(0)})f'(H_0)+(u_{0j}-u_{0j}^{(0)})f'(H_0) \qquad (3.17)
\end{aligned}$$

Persamaan (3.17) merupakan perluasan deret Taylor order pertama di sekitar nilai H_0. Order dari perluasan deret Taylor mendasari order dari metode PQL yang digunakan. Metode PQL order kedua (PQL-2) melinierkan bagian non-linier dari model dengan menggunakan perluasan deret Taylor order kedua, sementara metode PQL order pertama (PQL-1) melinierkan bagian non-linier dari model menggunakan perluasan deret Taylor order pertama.

Pada skripsi ini pembahasan estimasi parameter untuk model regresi logistik 2-level dengan *random intercept* dikhususkan pada metode PQL order pertama, sehingga perluasan deret Taylor dilakukan pada nilai $\boldsymbol{\beta}$ dan u_{0j} dan deret Taylor hanya dilakukan sampai order pertama. Metode PQL merupakan metode yang bersifat iteratif, dimana pada setiap iterasi dilakukan pelinierisasian pada nilai $\boldsymbol{\beta}$ dan u_{0j} dengan nilai keduanya berbeda untuk setiap iterasi. Hal tersebut disebabkan nilai $\boldsymbol{\beta}$ dan u_{0j} pada iterasi ke-*t* merupakan nilai yang diperoleh dari hasil taksiran pada iterasi sebelumnya, sehingga nilai $\boldsymbol{\beta}$ dan u_{0j} tidak selalu sama pada tiap iterasi. Untuk selanjutnya pelinierisasian bagian yang non-linier dari model pada iterasi ke-*t* mengikuti ketentuan metode *Penalized Quasi Likelihood* order pertama dituliskan dalam bentuk :

$$\pi_{ij} = f(\mathbf{x}_{ij}\boldsymbol{\beta} + u_{0j}) \approx f(H_t) + \mathbf{x}_{ij}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)})f'(H_t) + (u_{0j} - u_{0j}^{(t)})f'(H_t) \qquad (3.18)$$

dengan

t = menyatakan langkah iterasi, *t* = 0,1,2,…(sampai iterasi berhenti).
t = 0 menyatakan nilai inisial (seperti pada persamaan (3.17))

$f(H_t)$ = merupakan nilai dari fungsi *f(H)* pada titik H_t saat iterasi ke-*t*,

dimana $f(H_t) = \pi_{ij}^{(t)}$

$\pi_{ij}^{(t)}$ = nilai π untuk unit level-1 ke-*i* pada unit level-2 ke-*j* saat iterasi ke-*t*

$\mathbf{x}_{ij}$ = vektor ukuran *1x(P+1)* berisi observasi dari variabel-variabel penjelas untuk unit ke-*i* pada level-1 dalam unit ke-*j* pada level-2,

$\mathbf{x}_{ij} = \begin{bmatrix} 1 & x_{1ij} & x_{2ij} & \cdots & x_{Pij} \end{bmatrix}$

$\boldsymbol{\beta}$ = adalah vektor ukuran *(P+1)x1* berisi *fixed parameter,* $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_P \end{bmatrix}$

$\boldsymbol{\beta}^{(t)}$ = adalah vektor ukuran *(P+1)x1* berisi nilai *fixed parameter* saat iterasi ke-*t* , $\boldsymbol{\beta}^{(t)} = \begin{bmatrix} \beta_0^{(t)} \\ \beta_1^{(t)} \\ \vdots \\ \beta_P^{(t)} \end{bmatrix}$

$f'(H_t)$ = merupakan nilai turunan pertama dari fungsi *f(H)* pada titik H_t saat iterasi ke-*t*, dimana $f'(H_t) = \pi_{ij}^{(t)}(1-\pi_{ij}^{(t)})$, $f'(H_t)$ disimbolkan sebagai $w_{ij}^{(t)}$.

u_{0j} = menyatakan efek random untuk unit ke-*j* pada level-2

$u_{0j}^{(t)}$ = menyatakan nilai efek random untuk unit ke-*j* pada level-2 saat iterasi ke-*t*

Setelah π_{ij} mempunyai bentuk linier dengan pendekatan deret Taylor seperti pada persamaan (3.18), substitusikan bentuk linier tersebut ke dalam persamaan (3.16) :

$$y_{ij} = f(H_t) + \mathbf{x}_{ij}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)})f'(H_t) + (u_{0j} - u_{0j}^{(t)})f'(H_t) + \varepsilon_{ij} \tag{3.19}$$

dengan

y_{ij} = Nilai respon untuk unit ke-*i* pada level-1 dalam unit ke-*j* pada level-2

ε_{ij} = Nilai residual untuk unit ke-*i* pada level-1 dalam unit ke-*j* pada level-2

Pada langkah awal (t = 0), persamaan (3.19) ditulis :

$$y_{ij} = f(H_0) + \mathbf{x}_{ij}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)})f'(H_0) + (u_{0j} - u_{0j}^{(0)})f'(H_0) + \varepsilon_{ij} \tag{3.20}$$

dengan diberikan suatu nilai awal $\pi_{ij}^{(0)}$, $\boldsymbol{\beta}^{(0)}$, dan $u_{0j}^{(0)}$.

Selanjutnya untuk memperoleh model dalam bentuk linier, bagi persamaan (3.20) dengan $f'(H_t) = \pi_{ij}^{(t)}(1 - \pi_{ij}^{(t)}) = w_{ij}^{(t)}$ sehingga diperoleh :

$$y_{ij} = \pi_{ij} + \varepsilon_{ij}$$

$$y_{ij} = f(H_t) + \mathbf{x}_{ij}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)})f'(H_t) + (u_{0j} - u_{0j}^{(t)})f'(H_t) + \varepsilon_{ij}$$

$$\frac{y_{ij}}{f'(H_t)} = \frac{f(H_t) + \mathbf{x}_{ij}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)})f'(H_t) + (u_{0j} - u_{0j}^{(t)})f'(H_t) + \varepsilon_{ij}}{f'(H_t)}$$

$$\frac{y_{ij}}{f'(H_t)} = \frac{f(H_t)}{f'(H_t)} + \mathbf{x}_{ij}\boldsymbol{\beta} - \mathbf{x}_{ij}\boldsymbol{\beta}^{(t)} + u_{0j} - u_{0j}^{(t)} + \frac{\varepsilon_{ij}}{f'(H_t)}$$

$$\mathbf{x}_{ij}\boldsymbol{\beta}^{(t)} + u_{0j}^{(t)} + \frac{y_{ij} - f(H_t)}{f'(H_t)} = \mathbf{x}_{ij}\boldsymbol{\beta} + u_{0j} + \frac{\varepsilon_{ij}}{f'(H_t)}$$

$$\mathbf{x}_{ij}\boldsymbol{\beta}^{(t)} + u_{0j}^{(t)} + \frac{y_{ij} - f(H_t)}{f'(H_t)} = \mathbf{x}_{ij}\boldsymbol{\beta} + u_{0j} + \frac{\varepsilon_{ij}}{f'(H_t)}$$

$$\mathbf{x}_{ij}\boldsymbol{\beta}^{(t)} + u_{0j}^{(t)} + \frac{y_{ij} - \pi_{ij}^{(t)}}{\pi_{ij}^{(t)}(1 - \pi_{ij}^{(t)})} = \mathbf{x}_{ij}\boldsymbol{\beta} + u_{0j} + \frac{\varepsilon_{ij}}{\pi_{ij}^{(t)}(1 - \pi_{ij}^{(t)})}$$

$$\mathbf{x}_{ij}\boldsymbol{\beta}^{(t)} + u_{0j}^{(t)} + \frac{y - \pi_{ij}^{(t)}}{w_{ij}^{(t)}} = \mathbf{x}_{ij}\boldsymbol{\beta} + u_{0j} + \frac{\varepsilon_{ij}}{w_{ij}^{(t)}} \tag{3.21}$$

Dari persamaan di atas, lakukan pengaturan sedemikian rupa sehingga diperoleh :

$$y_{ij}^{**(t)} = \mathbf{x}_{ij}\boldsymbol{\beta} + u_{0j} + \varepsilon_{ij}^{**} \tag{3.22}$$

Model (3.22) adalah model yang telah dalam bentuk linier, dimana berdasarkan persamaan (3.21) dan persamaan (3.22),

$$y_{ij}^{**(t)} = \mathbf{x}_{ij}\boldsymbol{\beta}^{(t)} + u_{0j}^{(t)} + \frac{(y_{ij} - \pi_{ij}^{(t)})}{w_{ij}^{(t)}} \tag{3.23}$$

dengan

t = menyatakan langkah iterasi, t = 0,1,2,…(sampai iterasi berhenti). t = 0 menyatakan nilai inisial

$y_{ij}^{**(t)}$ = merupakan nilai respon yang telah ditransformasi untuk unit ke-*i* pada level-1 dalam unit ke-*j* pada level-2 saat iterasi ke-*t*

$\mathbf{x}_{ij}$ = vektor ukuran *1x(P+1)* berisi observasi dari variabel-variabel penjelas untuk unit ke-*i* pada level-1 dalam unit ke-*j* pada level-2, $\mathbf{x}_{ij} = \begin{bmatrix} 1 & x_{1ij} & x_{2ij} & \cdots & x_{Pij} \end{bmatrix}$

$\boldsymbol{\beta}^{(t)}$ = adalah vektor ukuran *(P+1)x1* berisi nilai *fixed parameter* saat iterasi ke-*t*, $\boldsymbol{\beta}^{(t)} = \begin{bmatrix} \beta_0^{(t)} \\ \beta_1^{(t)} \\ \vdots \\ \beta_P^{(t)} \end{bmatrix}$

$u_{0j}^{(t)}$ = menyatakan nilai efek random untuk unit ke-*j* pada level-2 saat iterasi ke-*t*

y_{ij} = merupakan nilai observasi respon untuk unit ke-*i* pada level-1 dalam unit ke-*j* pada level-2

$\pi_{ij}^{(t)}$ = nilai π untuk unit ke-*i* pada level-1 dalam unit ke-*j* pada level-2 saat iterasi ke-*t*

$w_{ij}^{(t)} = f'(H_t) = \pi_{ij}^{(t)}(1-\pi_{ij}^{(t)})$.

dan

$$\varepsilon_{ij}^{**} = \frac{\varepsilon_{ij}}{w_{ij}^{(t)}} \tag{3.24}$$

Model dalam persamaan (3.22) merupakan model yang telah dilinierisasi dengan respon transformasi inisial adalah $y_{ij}^{**(t)}$ dan

$$\text{var}(\varepsilon_{ij}^{**}) = \text{var}\left(\frac{\varepsilon_{ij}}{w_{ij}^{(t)}}\right) \approx \frac{1}{w_{ij}^{(t)}}$$

Parameter-parameter dalam model yang telah dilinierisasi pada iterasi ke-*t* yang dinyatakan dalam persamaan (3.22) dapat diestimasi dengan menggunakan prosedur estimasi dari model multilevel linier, yaitu metode *Iterative Generalized Least Square* (IGLS) seperti dijelaskan dalam subbab 3.1. Prosedur estimasi parameter-parameter dalam model (3.22) menggunakan IGLS di tiap iterasi PQL serupa dengan yang dijelaskan pada subbab 3.1, dengan mengganti $\mathbf{Y} = \{y_{ij}\}$ menjadi $\mathbf{Y} = \{y_{ij}^{**(t)}\}$.

Setelah diperoleh taksiran parameter tetap dan random yang konvergen dengan IGLS, hitung nilai taksiran residual yang akan digunakan pada proses linierisasi di iterasi berikutnya dengan :

$u_{0j}^{(t)} = \dfrac{n_j \sigma_{uo}^2}{n_j \sigma_{uo}^2 + \sigma_\varepsilon^2} \bar{\bar{y}}$ untuk nilai residual level-2,

sehingga nilai residual level-1 nya adalah

$\varepsilon_{ij}^{(t)} = y_{ij}^{**} - u_{0j}^{(t)}$

dimana

$u_{0j}^{(t)}$ = Nilai efek random (residual level-2) untuk unit level-2 ke-*j* pada iterasi ke-*t*. Disini baru bisa dicari pada iterasi pertama dan seterusnya (*t* = 1,2,..)

n_j = Banyaknya unit level-1 dalam unit ke-*j* pada level-2

σ_{uo}^2 = Variansi efek random (residual level-2) yang diperoleh dari penaksiran IGLS

σ_ε^2 = Variansi *error* level-1

$\bar{\bar{y}}_j$ = Rata-rata dari $\tilde{y}_{ij}$ untuk setiap unit level-2 ke-*j* ($\bar{\bar{y}}_{ij} = \tilde{y}_{ij} / n_j$), dengan $\tilde{y}_{ij} = y_{ij} - \hat{y}_{ij}$ seperti pada prosedur IGLS.

Untuk lebih ringkasnya, prosedur estimasi parameter-parameter dalam model logistik 2-level dengan *random intercept* menggunakan metode *Penalized Quasi Likelihood* order pertama (PQL-1) adalah :

(1) Linierisasi bagian non-linier dari model, yaitu π_{ij}, dengan menggunakan perluasan deret Taylor order pertama dengan diberikan nilai taksiran π_{ij}, $\boldsymbol{\beta}$, dan u_{0j} yang diperoleh dari iterasi sebelumnya (khusus untuk langkah awal (t = 0), diberikan suatu nilai-nilai inisial). Pada bentuk model yang telah dilinierkan diperoleh nilai respon yang telah ditransformasi.

(2) Parameter-parameter pada model dalam bentuk linier yang diperoleh dari langkah (1) diestimasi dengan prosedur estimasi model multilevel linier, yaitu *Iterative Generalized Least Square* (IGLS).

(3) Hasil taksiran parameter random dalam model digunakan untuk mencari nilai taksiran u_{0j}.

(4) Ulangi langkah (1) dengan menggunakan hasil taksiran *fixed* dan *random parameter* dari langkah (2) serta hasil taksiran u_{0j} dari langkah (3).

(5) Langkah (1) sampai (4) dilakukan sampai nilai taksiran-taksiran parameter yang diperoleh konvergen.

Intra-class correlation

Seperti halnya pada model regresi 2-level pada bab sebelumnya, pada model logistik 2-level dengan *random intercept* juga dapat diketahui korelasi antara dua unit level-1 dalam unit level-2 yang sama, atau disebut "*intra-class correlation*", yang dinyatakan dengan :

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_\varepsilon^2} \tag{3.25}$$

dimana $\sigma_\varepsilon^2 = \pi^2 / 3$.

Semakin tinggi nilai ρ menunjukkan semakin miripnya dua unit level-1 dari unit level-2 yang sama, dibandingkan dengan dua unit level-1 yang diambil dari dua unit level-2 yang berbeda.