

## BAB IV

### CONTOH APLIKASI MODEL REGRESI ORDINAL DUA LEVEL

Pada bab ini akan dibahas mengenai penerapan model regresi ordinal dua level untuk data *tercluster* dan untuk data longitudinal. Analisis dari data di mana siswa bersarang di dalam kelas digunakan untuk mengilustrasikan aplikasi model regresi ordinal dua level untuk data ordinal *tercluster*. Dalam menjelaskan penerapan model regresi ordinal dua level untuk data ordinal longitudinal, akan digunakan data psikiatrik di mana pasien diklasifikasikan pada beberapa tingkat keparahan penyakit terhadap beberapa titik waktu (*timepoints*). Struktur data dua level yang digunakan untuk data longitudinal adalah perulangan observasi bersarang dalam individu (pasien)

#### 4.1 CONTOH APLIKASI UNTUK DATA TERCLUSTER

*Television School and Family Smoking Prevention and Cessation Project (TVSFP)* merupakan studi yang dirancang untuk menguji efek dari *school-based social resistance curriculum (CC)*, *television-based program (TV)* dan interaksi keduanya terhadap pencegahan dan penghentian penggunaan rokok . Untuk ilustrasi dari penerapan model regresi ordinal dua

level untuk data ordinal tercluster, digunakan subset dari data TVSFP tersebut (Hedeker dan Gibbons, 1994).

#### 4.1.1 Data

Studi difokuskan pada siswa dari 28 sekolah di Los Angeles, di mana pada tiap sekolah dilakukan pengacakan untuk menerima satu dari empat kondisi studi berikut:

- (a) *social resistance classroom curriculum* (CC)
- (b) *media (television) intervention* (TV)
- (c) *social resistance classroom curriculum dan media (television) intervention* (CC\*TV)
- (d) tidak ada *treatment*.

Respon dalam studi ini merupakan skor pengetahuan siswa mengenai efek dari penggunaan rokok dan kesehatan (skor THKS). Skor THKS dari siswa didefinisikan sebagai total jawaban yang benar dari pertanyaan mengenai pengetahuan tentang efek penggunaan rokok.

Sampel dalam studi ini merupakan siswa tingkat tujuh yang telah mengikuti *pre-test* dan *post-intervention test* mengenai pengetahuan tentang efek dari penggunaan rokok. Untuk itu, terdapat 1600 siswa dari 135 kelas dan 28 sekolah yang memenuhi kriteria tersebut. Data yang dihasilkan merupakan data yang *unbalanced* dengan rentang data antara 1 sampai 13 kelas per sekolah, dan antara 2 sampai 28 siswa tiap kelas. Sebagian data

TVSFP dapat dilihat di lampiran 6. Data selengkapnya dapat dilihat di alamat <http://tigger.uic.edu/%7Ehedeker/ml.html> pada TVSFP2B. DAT. Untuk analisis data tersebut, digunakan struktur data dua level, dengan siswa bersarang dalam kelas, sedangkan kemungkinan bahwa kelas bersarang dalam sekolah diabaikan. Dalam contoh ini, siswa merupakan unit level 1 dan kelas merupakan unit level 2.

#### 4.1.2 Variabel-variabel dalam Penelitian

Variabel-variabel yang akan digunakan dalam aplikasi ini adalah:

1) Variabel respon (*dependent*)

$y$  = skor *post*-THKS, merupakan variabel laten kontinu yang tidak diketahui nilai pastinya, akan tetapi yang diketahui adalah kategori-kategorinya, yaitu

Kategori 1 : pengetahuan siswa tentang efek penggunaan rokok sangat rendah.

Kategori 2 : pengetahuan siswa tentang efek penggunaan rokok rendah.

Kategori 3 : pengetahuan siswa tentang efek penggunaan rokok sedang.

Kategori 4 : pengetahuan siswa tentang efek penggunaan rokok tinggi.

Hubungan antara variabel respon laten kontinu  $y$  dengan variabel respon ordinal ( $Y$ ) di atas didefinisikan sebagai berikut:

$$Y = 1, \text{ jika } -\infty < y \leq 0$$

$$Y = 2, \text{ jika } 0 < y \leq \gamma_2$$

$$Y = 3, \text{ jika } \gamma_2 < y \leq \gamma_3$$

$$Y = 4, \text{ jika } \gamma_3 < y \leq \infty$$

2) Variabel penjelas (*independent*).

- Variabel penjelas pada level 1:

*Pre-THKS* = skor siswa pada saat *pretest* mengenai pengetahuan tentang rokok dan kesehatan. Variabel *pre-THKS* merupakan variabel penjelas kontinu.

- Variabel penjelas pada level 2:

*CC* = *social resistance classroom curriculum*. Variabel *CC* merupakan variabel kategorik dengan dua buah kategori, yaitu:

$$CC = \begin{cases} 1, & \text{jika ya} \\ 0, & \text{jika tidak} \end{cases}$$

*TV* = *media (television) intervention*. Variabel *TV* merupakan variabel kategorik dengan dua buah kategori, yaitu:

$$TV = \begin{cases} 1, & \text{jika ya} \\ 0, & \text{jika tidak} \end{cases}$$

*CC\*TV* = interaksi antara variabel *CC* dan *TV*. Variabel *CC\*TV* merupakan variabel kategorik dengan dua buah kategori, yaitu:

$$CC * TV = \begin{cases} 1, & \text{jika ya} \\ 0, & \text{jika tidak} \end{cases}$$

Model regresi ordinal dua-level dengan satu efek *random* (*random intercept*) digunakan untuk menganalisis data tersebut. Misalkan notasi  $i = 1, 2, \dots, 135$  menyatakan unit-unit level-2 (kelas) dan  $k = 1, 2, \dots, n_i$  menyatakan unit-unit level-1 (siswa) yang bersarang dalam tiap unit level-2. Banyaknya siswa dalam setiap kelas tidak sama.

Misalkan  $y_{ik}$  menyatakan variabel respon laten kontinu nilai post-THKS siswa ke- $k$  dalam kelas ke- $i$ . Model regresi ordinal dua level untuk studi di atas dapat ditulis sebagai berikut.

- Untuk model level-1:

$$y_{ik} = b_{0i} + \alpha_{(1)1} \text{pre-THKS}_{ik} + \varepsilon_{ik}$$

- Untuk model level-2:

$$b_{0i} = \mu_0 + \alpha_{(2)01} \text{CC}_i + \alpha_{(2)02} \text{TV}_i + \alpha_{(2)03} \text{CC}_i * \text{TV}_i + \delta_{0i}$$

Kedua model di atas dapat digabung menjadi model

$$y_{ik} = \mu_0 + \alpha_{(2)01} \text{CC}_i + \alpha_{(2)02} \text{TV}_i + \alpha_{(2)03} \text{CC}_i * \text{TV}_i + \alpha_{(1)1} \text{pre-THKS}_{ik} + \delta_{0i} + \varepsilon_{ik}$$

Asumsi:

- $\varepsilon_{ik}$  berdistribusi  $N(0,1)$
- $\delta_{0i}$  berdistribusi  $N(0, \sigma_\delta^2)$
- variabel *random*  $\varepsilon_{ik}$  dan  $\delta_{0i}$  saling bebas

### 4.1.3 Tujuan

Mencari taksiran parameter-parameter dalam model regresi ordinal dua-level tersebut. Metode estimasi yang digunakan dalam penaksiran parameter-parameter dalam model ini adalah metode MMLE yang telah dijelaskan di bab 3.

### 4.1.4 Pengolahan Data

Untuk mengolah data dalam penelitian ini, digunakan software MIXOR 2.0 for windows. Dengan menggunakan MIXOR, diperoleh taksiran parameter-parameter dalam model di atas sebagai berikut:

**Tabel 1**

Taksiran Parameter dalam model regresi ordinal dua level untuk data TVSFP dengan satu efek *random* (*random intercept*)

Variabel	Parameter	Nilai Taksiran
Intercept	$\hat{\mu}_0$	0.06220
Pre-THKS	$\hat{\alpha}_{(1)1}$	0.24247
CC	$\hat{\alpha}_{(2)01}$	0.50945
TV	$\hat{\alpha}_{(2)02}$	0.12363
CC*TV	$\hat{\alpha}_{(2)03}$	-0.19378
<i>Threshold 2</i>	$\hat{\gamma}_2$	0.76063
<i>Threshold 3</i>	$\hat{\gamma}_3$	1.48749
Efek random (kelas) Standar deviasi	$\hat{\sigma}_\delta$	0.26163

Selanjutnya, berdasarkan hasil yang ditampilkan di tabel 1, maka diperoleh taksiran model regresi ordinal dua level dengan metode MMLE sebagai berikut:

$$\hat{y}_{ik} = 0.06220 + 0.50945 (CC)_i + 0.12363 (TV)_i - 0.19378 (CC*TV)_i + 0.24247 (pre-THKS)_{ik}$$

Dari output MIXOR juga diperoleh nilai dari *Intraclass Correlation Coefficient* (ICC) yaitu sebesar 0.064. ICC ini mengukur proporsi dari variansi antar kelas terhadap variansi total, dan dapat diinterpretasikan bahwa sebesar 6,4% dari total variansi dalam model adalah variansi antar kelas. Nilai ICC ini cukup besar (berada dalam rentang 5%-12%), sehingga mengindikasikan bahwa model regresi ordinal dua level dengan satu efek *random* perlu digunakan untuk menganalisis data TVSFP tersebut.

## 4.2 CONTOH APLIKASI UNTUK DATA LONGITUDINAL

### 4.2.1 Data

Untuk aplikasi dari model regresi ordinal dua level untuk data longitudinal, akan digunakan data psikiatrik yang dikumpulkan dari *National Institute of Mental Health (NIMH) Schizophrenia Collaborative Study* mengenai hubungan antara *treatment* pengobatan terhadap perubahan dalam tingkat keparahan penyakit *Schizophrenia*. Lebih lanjut, data yang

akan dianalisis dalam penelitian ini adalah Item 79 dari Skala *Inpatient Multidimensional Psychiatric* (IMPS79) yaitu tingkat keparahan penyakit (*severity of illness*).

Dalam studi ini, pasien secara acak diberikan satu dari dua *treatment* pengobatan: *placebo* atau obat anti-*psychotic*. Sampel dalam studi ini sebanyak 437 pasien klinik psikiatri di Chicago yang diukur pada tujuh titik waktu, yaitu minggu 0 (*baseline*), minggu ke-1, minggu ke-2, sampai minggu ke-6. Berdasarkan penelitian sebelumnya yang dilakukan oleh Gibbons, Waternaux, Hedeker, dan Davis (1988), untuk melinierkan hubungan antara skor IMPS79 terhadap waktu, dipilih transformasi akar kuadrat dari waktu (minggu). Sebagian dari data mengenai penelitian ini dapat dilihat di lampiran 8. Untuk data lengkapnya, dapat dilihat di situs <http://tigger.uic.edu/%7Ehedeker/ml.html> pada SCHIZX1. DAT.

Dari data tersebut, pada kolom ordinal IMPS79 score terdapat angka -9. Angka ini menandakan bahwa data dari pasien pada titik waktu tersebut tidak ada. Data dari titik waktu tersebut tidak akan digunakan dalam analisis, akan tetapi data pasien tersebut untuk titik waktu yang lain di mana tidak ada data yang hilang, akan digunakan dalam analisis. Jadi, banyaknya perulangan observasi per pasien bergantung pada jumlah titik waktu di mana tidak ada data yang hilang pada pasien tersebut. Secara keseluruhan terdapat total 1603 perulangan observasi dari 437 pasien yang akan digunakan dalam analisis dengan rentang perulangan observasi antara 2 sampai 5 observasi per pasien. Untuk menganalisis data tersebut, struktur



data dua level digunakan, di mana perulangan observasi bersarang dalam individu (pasien). Pasien merupakan unit level-2 dan perulangan observasi merupakan unit level 1.

#### 4.2.2 Variabel-variabel dalam Penelitian

Variabel-variabel yang akan digunakan dalam aplikasi ini adalah:

1) Variabel respon (*dependent*)

$y$  = skor IMPS79 score, merupakan variabel laten kontinu yang tidak diketahui nilai pastinya, akan tetapi yang diketahui adalah kategori-kategorinya. Dalam penelitian ini, tingkat keparahan penyakit *schizophrenia* (skor IMPS79) dibagi ke dalam empat kategori, yaitu:

Kategori 1 : Normal, tidak menunjukkan gejala penyakit.

Kategori 2 : penyakit *schizophrenia* yang diderita pasien tidak terlalu parah.

Kategori 3 : penyakit *schizophrenia* yang diderita pasien parah

Kategori 4 : penyakit *schizophrenia* yang diderita pasien sangat parah

Hubungan antara variabel laten kontinu ( $y$ ) dengan variabel respon ordinal ( $Y$ ) berdasarkan konsep *threshold* didefinisikan sebagai berikut:

$$Y = 1, \text{ jika } -\infty < y \leq 0$$

$$Y = 2, \text{ jika } 0 < y \leq \gamma_2$$

$$Y = 3, \text{ jika } \gamma_2 < y \leq \gamma_3$$

$$Y = 4, \text{ jika } \gamma_3 < y \leq \infty$$

2) Variabel penjelas (*independent*).

- Variabel penjelas pada level 1:

*Time* = merupakan minggu di mana data dari pasien dinilai, dan tidak ada data yang hilang. Seperti dijelaskan sebelumnya, untuk melinierkan hubungan antara skor IMPS79 terhadap waktu, maka variabel yang digunakan adalah akar kuadrat dari minggu (*SqrtWeek*). Variabel waktu ini mempunyai efek yang berbeda-beda untuk setiap pasien.

- Variabel penjelas pada level 2:

*Drug* = merupakan treatment pengobatan yang diberikan kepada pasien. Variabel *Drug* merupakan variabel kategorik dengan dua buah kategori, yaitu:

$$Drug = \begin{cases} 1, & \text{jika pasien diberikan obat anti-} \textit{psychotic} \\ 0, & \text{jika pasien diberikan } \textit{placebo} \end{cases}$$

Model regresi ordinal dua level dengan dua efek *random* (*random intercept* dan *slope*) digunakan untuk menganalisis data tersebut. Misalkan notasi  $i = 1, 2, \dots, 437$  menyatakan unit-unit level-2 (pasien) dan  $k = 1, 2, \dots, n_i$  menyatakan unit-unit level-1 (perulangan observasi) yang bersarang dalam tiap unit level-2. Banyaknya perulangan observasi untuk setiap pasien tidak sama.

Misalkan  $y_{ik}$  menyatakan variabel respon laten kontinu skor IMPS79 pasien ke- $i$  pada perulangan observasi ke- $k$  yang bersarang didalamnya. Model regresi ordinal dua level untuk penelitian di atas dapat dituliskan sebagai berikut:

- Untuk model level-1:

$$y_{ik} = b_{0i} + b_{1i} \text{Time}_{ik} + \varepsilon_{ik}$$

- Untuk model level-2:

$$b_{0i} = \mu_0 + \alpha_{(2)01} (\text{Drug})_i + \delta_{0i}$$

$$b_{1i} = \mu_1 + \alpha_{(2)11} (\text{Drug})_i + \delta_{1i}$$

Kedua model di atas dapat digabung menjadi model:

$$\begin{aligned} y_{ik} &= \mu_0 + \alpha_{(2)01} (\text{Drug})_i + \delta_{0i} + [\mu_1 + \alpha_{(2)11} (\text{Drug})_i + \delta_{1i}] (\text{Time})_{ik} + \varepsilon_{ik} \\ &= \mu_0 + \alpha_{(2)01} (\text{Drug})_i + \mu_1 (\text{Time})_{ik} + \alpha_{(2)11} (\text{Drug})_i \times (\text{Time})_{ik} + \delta_{1i} (\text{Time})_{ik} + \delta_{0i} + \varepsilon_{ik} \end{aligned}$$

Asumsi:

- $\varepsilon_{ik}$  berdistribusi  $N(0,1)$
- $(\delta_{0i}, \delta_{1i})$  berdistribusi *multivariate* normal  $N_2 \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\delta_0}^2 & \sigma_{\delta_0 \delta_1} \\ \sigma_{\delta_1 \delta_0} & \sigma_{\delta_1}^2 \end{bmatrix} \right)$
- variabel random  $\varepsilon_{ik}$  dan  $(\delta_{0i}, \delta_{1i})$  saling bebas untuk setiap  $i$  dan  $k$ .  $i = 1, \dots, 437$ ,  $k = 1, \dots, n_i$ .

### 4.2.3 Tujuan

Mencari taksiran parameter-parameter dalam model regresi ordinal dua level untuk data longitudinal tersebut. Metode estimasi yang digunakan dalam penaksiran parameter-parameter dalam model ini adalah metode MMLE yang telah dijelaskan di bab 3

### 4.2.4 Pengolahan Data

Untuk mengolah data dalam penelitian ini, digunakan software MIXOR 2.0 for windows. Dengan menggunakan MIXOR, diperoleh taksiran parameter-parameter dalam model di atas sebagai berikut:

**Tabel 2**

Taksiran Parameter dalam model regresi ordinal dua level untuk data longitudinal dengan dua efek *random* (*random intercept* dan *slope*)

Variabel	Parameter	Nilai Taksiran
Intercept	$\hat{\mu}_0$	4.10997
Time	$\hat{\mu}_1$	-0.50548
Drug	$\hat{\alpha}_{(2)01}$	0.03823
Drug x Time	$\hat{\alpha}_{(2)11}$	-0.95018
Threshold 2	$\hat{\gamma}_2$	2.18411
Threshold 3	$\hat{\gamma}_3$	3.65352
Random effect variance & covariance terms (Cholesky of var-covariance matrix)		

----- <i>Intercept (Cholesky)</i>	$\hat{\sigma}_{\delta_0}$	1.48627
<i>Covariance (Cholesky)</i>	$\sigma_{\delta_1\delta_0} / \sigma_{\delta_0}$	-0.31473
<i>Time (Cholesky)</i>	$\sqrt{\hat{\sigma}_{\delta_1}^2 - (\hat{\sigma}_{\delta_1\delta_0}^2 / \hat{\sigma}_{\delta_0}^2)}$	0.73025

Selanjutnya, berdasarkan hasil yang ditampilkan di tabel 2, maka diperoleh taksiran model regresi ordinal dua level dengan metode MMLE sebagai berikut:

$$\begin{aligned}\hat{y}_{ik} &= 4.10997 + 0.03823(\text{Drug})_i - [0.50548 + 0.95018(\text{Drug})_i](\text{Time})_{ik} \\ &= 4.10997 + 0.03823(\text{Drug})_i - 0.50548(\text{Time})_{ik} - 0.95018(\text{Drug})_i \times (\text{Time})_{ik}\end{aligned}$$

Selain itu, dari tabel 2 juga diperoleh taksiran faktor *Cholesky* dari matriks variansi kovariansi untuk efek *random*, yaitu:

$$\hat{\mathbf{T}} = \begin{pmatrix} 1.48627 & 0 \\ -0.31473 & 0.73025 \end{pmatrix}$$

Sehingga, taksiran matriks variansi kovariansi dari efek *random* adalah :

$$\begin{aligned}\hat{\Sigma}_{\delta} &= \hat{\mathbf{T}}\hat{\mathbf{T}}' \\ &= \begin{pmatrix} 1.48627 & 0 \\ -0.31473 & 0.73025 \end{pmatrix} \begin{pmatrix} 1.48627 & -0.31473 \\ 0 & 0.73025 \end{pmatrix} \\ &= \begin{pmatrix} 2.209 & -0.468 \\ -0.468 & 0.632 \end{pmatrix}\end{aligned}$$

Dari taksiran matriks variansi-kovariansi untuk efek *random* di atas terlihat bahwa komponen variansi untuk *intercept* tidak nol, sehingga diduga bahwa terdapat variasi di dalam rata-rata skor IMPS79 pasien yang tidak

dijelaskan oleh variabel penjelas dan interaksinya (terdapat variasi pada *intercept*). Komponen variansi untuk kemiringan (*slope*) dari variabel waktu (*Time*) juga cukup besar (0.0632), sehingga diduga bahwa variabel waktu mempunyai efek yang berbeda-beda pada setiap pasien. Selanjutnya, didapat juga kovariansi antara *intercept* dan *slope* sebesar -0.468. Hal ini mengindikasikan bahwa terdapat hubungan atau korelasi negatif antara *intercept* dan variabel waktu (*Time*). Akan tetapi, untuk lebih memastikan hal-hal tersebut, perlu dilakukan pengujian lebih lanjut (tidak dibahas dalam skripsi ini)