

## **APLIKASI REGRESI LOGISTIK ORDINAL MULTILEVEL UNTUK PEMODELAN DAN KLASIFIKASI HURUF MUTU MATA KULIAH METODE STATISTIKA**

*(The Application of Multilevel Ordinal Logistic Regression for Modeling and Classification The Final Grade of Statistical Methods Course at Faculty of Mathematics and Natural Science, Bogor Agricultural University)*

Indahwati<sup>1</sup>, Dian Kusumaningrum<sup>1</sup>, Iin Maena<sup>2</sup>

<sup>1</sup>Departemen Statistika, FMIPA-IPB

<sup>2</sup>Alumni S1 Departemen Statistika, FMIPA-IPB

E-mail : <sup>1</sup> [indah\\_stk@yahoo.com](mailto:indah_stk@yahoo.com)

### **Abstract**

*Statistical Methods (STK211) is an interdept course under coordination of Statistic Departement Faculty of Mathematics and Natural Science, Bogor Agricultural University (BAU). The final grade received by student who follow Statistical Methods is measurement in ordinal scale, that is A, B, C, D and E. In the 2008/2009 academic year there are 7 parallel classes in the Faculty of Mathematics and Natural Science, BAU. By considering the hierarchical structure contained in the score of student achievement data, the student (first level) is nested in a parallel class (second level), hence this study used multilevel ordinal logistic regression analysis to model the final score of Statistical Methods with the factors that influence it. Explanatory variables that significantly affect the final score of Statistical Methods are the GPA of TPB (student's first year of college) and gender, with the variability of the intercepts across parallel classes in the logit function as 1.184. Percentage classification accuracy obtained by using multilevel ordinal logistic regression model was 56.85%.*

**Keywords** : hierarchical, multilevel modeling, multilevel ordinal logistic regression, classification

### **PENDAHULUAN**

Metode Statistika (STK211) merupakan mata kuliah interdep yang diasuh oleh Departemen Statistika IPB. Pada tahun akademik 2008/2009, di Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) IPB terdapat 7 kelas paralel yang mengambil mata kuliah ini. Setiap kelas paralel umumnya terdiri dari mahasiswa satu departemen dan berada di bawah koordinasi Departemen Statistika. Dengan demikian dapat dikatakan bahwa mahasiswa tersarang pada kelas paralel. Struktur seperti ini disebut struktur hirarkhi (*hierarchical*).

Struktur hirarkhi mengindikasikan bahwa data yang dianalisis berasal dari beberapa level, dimana level yang lebih rendah tersarang pada level yang lebih tinggi. Pada struktur hirarkhi ini, individu-individu dalam kelompok yang sama cenderung mempunyai karakteristik yang mirip, sehingga antar amatan tidak saling bebas. Pelanggaran terhadap asumsi kebebasan akan membuat nilai dugaan galat baku koefisien regresi berbias ke bawah, sehingga dalam pengujian hipotesis akan cenderung menolak hipotesis nol dan menyimpulkan terdapat hubungan yang nyata antara peubah bebas dengan peubah responnya (Hox 2002). Oleh karena itu, untuk data dengan

struktur hirarkhi perlu dilakukan analisis menggunakan pemodelan multilevel untuk mengatasi hal tersebut. Selain itu adanya keperluan untuk menganalisis peubah-peubah yang berasal dari beberapa level secara simultan juga menjadi pertimbangan digunakannya pemodelan multilevel (Hox 2002).

Penelitian mengenai pemodelan multilevel terhadap nilai akhir Metode Statistika berupa data kontinu telah dilakukan oleh Widiyani (2009), dan pada data biner oleh Husniyati (2010). Pada penelitian ini akan diterapkan analisis regresi logistik ordinal multilevel terhadap peubah respon ordinal berupa huruf mutu mata kuliah Metode Statistika dengan lima kategori yaitu A, B, C, D dan E. Selain itu ingin diketahui tingkat ketepatan klasifikasi dari model yang diperoleh

### **TINJAUAN PUSTAKA**

#### **Model Regresi Multilevel**

Pemodelan multilevel merupakan suatu teknik statistika yang digunakan untuk menganalisis data dengan struktur hirarkhi. Pada model multilevel peubah respon diukur pada level kesatu, sedangkan peubah penjelas dapat didefinisikan pada setiap level. Bentuk sederhana dari model regresi multilevel adalah model regresi dua level. Secara

matematis, model regresi dua level dengan satu peubah bebas pada level kesatu dapat ditulis sebagai berikut (Hox 2002) :

$$Y_{ij} = \beta_{0j} + \beta_{1j} + \varepsilon_{ij} \quad \dots\dots\dots(1)$$

dengan  $i = 1, \dots, n_j$  adalah indeks individu pada level satu,  $j = 1, \dots, J$  adalah indeks kelompok pada level dua,  $Y_{ij}$  adalah nilai respon pada individu ke- $i$  dan kelompok ke- $j$ ,  $\beta_{0j}$  adalah intersep pada level dua ke- $j$ ,  $\beta_{1j}$  adalah kemiringan garis pada level dua ke- $j$ ,  $\varepsilon_{ij}$  adalah galat yang menyebar  $N(0, \sigma_\varepsilon^2)$ .

Pada persamaan (1), koefisien regresi  $\beta_0$  dan  $\beta_1$  memiliki indeks  $j$  untuk kelompok pada level dua, yang mengindikasikan bahwa koefisien regresi pada level dua dapat memiliki nilai yang berbeda. Jika terdapat satu peubah penjelas pada level dua, maka keragaman koefisien regresi tersebut dimodelkan melalui persamaan (2), yaitu :

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}Z_j + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}Z_j + u_{1j} \end{aligned} \quad \dots\dots\dots(2)$$

dengan  $Z_j$  adalah peubah penjelas pada level kedua,  $u_{0j}$  dan  $u_{1j}$  adalah galat pada level kedua. Diasumsikan  $u_{0j} \sim N(0, \sigma_{u_0}^2)$  dan  $u_{1j} \sim N(0, \sigma_{u_1}^2)$  serta  $u_{0j}, u_{1j}$  dan  $\varepsilon_{ij}$  saling bebas (Hox 2002).

Untuk data kategorik, model multilevel yang dapat diterapkan adalah model logistik multilevel, yaitu model logistik biner multilevel untuk respon biner dan model logistik ordinal multilevel untuk respon ordinal.

**Regresi Logistik Ordinal**

Regresi logistik ordinal digunakan untuk memodelkan hubungan antara peubah respon yang berskala ordinal dengan peubah-peubah penjelasnya. Jika diasumsikan terdapat peubah respon  $Y$  berskala ordinal dengan  $J$  kategori dan  $\mathbf{x}' = (x_1, x_2, \dots, x_p)$  adalah vektor peubah penjelas, maka peluang dari peubah respon kategori ke- $j$  pada peubah penjelas  $X$  tertentu dapat dinyatakan dengan  $P[Y=j | x] = \pi_j(x)$  dan peluang kumulatifnya adalah (Hosmer & Lemeshow 2000) :

$$P[Y \leq j | x] = \pi_1(x) + \dots + \pi_j(x)$$

Model logit kumulatif didefinisikan dengan:

$$\begin{aligned} L_j(x) &= \text{logit} (P[Y \leq j | x]) \\ &= \log \left( \frac{P[Y \leq j | x]}{1 - P[Y \leq j | x]} \right) \\ &= \log \left( \frac{\pi_1(x) + \dots + \pi_j(x)}{\pi_{j+1}(x) + \dots + \pi_J(x)} \right) \\ &= \alpha_j - \mathbf{x}'\boldsymbol{\beta} \end{aligned}$$

dimana  $j=1, \dots, J-1$  dan  $\alpha_1, \dots, \alpha_{J-1}$  adalah *threshold model* serta  $\boldsymbol{\beta}$  merupakan vektor koefisien regresi.

Metode pendugaan parameter yang dapat digunakan pada regresi logistik ordinal diantaranya adalah dengan metode kemungkinan maksimum.

Metode ini dapat dilakukan jika antara amatan yang satu dengan yang lain diasumsikan saling bebas. Fungsi kemungkinan-nya dapat dinyatakan sebagai (Hosmer & Lemeshow 2000) :

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left[ \pi_1(x_i)^{z_{0i}} \pi_2(x_i)^{z_{1i}} \dots \pi_j(x_i)^{z_{ji}} \right]$$

dengan

$$z_{ji} = \begin{cases} 1 & \text{untuk } y = j \\ 0 & \text{untuk } y \neq j \end{cases}$$

Sedangkan fungsi *log kemungkinan* -nya adalah :

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n z_{0i} \ln[\pi_1(x_i)] + \dots + z_{ji} \ln[\pi_j(x_i)]$$

Selanjutnya, untuk memperoleh penduga parameter dari regresi logistik ordinal adalah dengan memaksimumkan fungsi log kemungkinan tersebut terhadap parameter-nya.

Pada analisis regresi ordinal terdapat lima pilihan fungsi hubung (*link function*) seperti tercantum pada Tabel 1. Penggunaannya tergantung dari sebaran data yang dianalisis. Logit digunakan pada sebagian besar sebaran data, *complementary log-log* digunakan untuk data yang mempunyai kecenderungan bernilai tinggi, *negative log-log* digunakan untuk data yang mempunyai kecenderungan bernilai rendah, probit digunakan jika peubah laten menyebar secara normal, sedangkan cauchit digunakan jika peubah laten mempunyai nilai yang ekstrim. Analisis regresi ordinal yang telah dijelaskan sebelumnya adalah analisis regresi ordinal dengan fungsi hubung logit atau sering disebut regresi logistik ordinal (Norusis 2010).

Tabel 1 Fungsi hubung pada regresi ordinal

Fungsi hubung	Bentuk Fungsi
Logit	$\text{Log} \left( \frac{\pi}{1 - \pi} \right)$
<i>Complementary log-log</i>	$\text{Log}(-\text{Log}(1-\pi))$
<i>Negative log-log</i>	$-\text{Log}(-\text{Log}(\pi))$
Probit	$\Phi^{-1}(\pi)$
Chauchit	$\tan(\text{phi}(\pi-0.5))$

**Regresi Logistik Ordinal Multilevel**

Jika data mempunyai struktur hirarkhi, maka hubungan antara peubah respon yang berskala ordinal dengan peubah-peubah penjelasnya dapat dimodelkan dengan regresi logistik ordinal multilevel. Model multilevel ordinal dapat dinyatakan dalam bentuk linier melalui peubah laten untuk mempermudah pendugaan. Misalkan  $j$  adalah indeks kelompok pada level dua dan  $i$  adalah indeks individu pada level satu, maka peubah respon yang berskala ordinal ( $Y$ ) dengan  $S$  kategori ( $s = 1, 2, \dots, S$ ) dapat dinyatakan melalui peubah laten berskala kontinu  $\tilde{Y}$  yang mengikuti model :

$$\tilde{Y}_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \delta_j + \varepsilon_{ij}$$

dengan  $\mathbf{x}'_{ij}$  merupakan vektor peubah bebas (termasuk 1 untuk intersep),  $\boldsymbol{\beta}$  merupakan vektor parameter regresi,  $\delta_j$  merupakan galat pada level dua yang diasumsikan menyebar  $N(0, \sigma_\delta^2)$ , dan  $\varepsilon_{ij}$  merupakan galat pada level satu yang menyebar  $N(0, \sigma^2)$ ,  $\varepsilon_{ij}$  dan  $\delta_j$  diasumsikan saling bebas (Grilli & Pratesi 2002).

Peubah respon berskala ordinal ( $Y$ ) terhubung dengan peubah laten ( $\tilde{Y}$ ) melalui hubungan sebagai berikut :

$$Y_{ij} = s \Leftrightarrow \gamma_{s-1} < \tilde{Y}_{ij} \leq \gamma_s$$

dengan batasan :

$$-\infty = \gamma_0 \leq \gamma_1 \leq \dots \leq \gamma_{s-1} \leq \gamma_s = \infty$$

dengan  $\gamma_s$  menyatakan *thresholds*. Sehingga model peluang bersyarat untuk respon mahasiswa  $i$  pada kelas  $j$  dapat dinyatakan sebagai :

$$\begin{aligned} P(Y_{ij} = s | \delta_j) &= P(\gamma_{s-1} < \tilde{Y}_{ij} \leq \gamma_s | \delta_j) \\ &= P(\tilde{Y}_{ij} \leq \gamma_s | \delta_j) - P(\tilde{Y}_{ij} \leq \gamma_{s-1} | \delta_j) \end{aligned}$$

dengan

$$\begin{aligned} P(\tilde{Y}_{ij} \leq \gamma_s | \delta_j) &= P(\varepsilon_{ij} \leq \gamma_s - [\mathbf{x}'_{ij}\boldsymbol{\beta} + \delta_j] | \delta_j) \\ &= F\left(\frac{\gamma_s}{\sigma} - \left[\frac{1}{\sigma}\mathbf{x}'_{ij}\boldsymbol{\beta} + \frac{\delta_j}{\sigma}\right]\right) \\ &= F(\gamma_{\delta,s} - [\mathbf{x}'_{ij}\boldsymbol{\beta}_\sigma + \delta_j]) \end{aligned}$$

dimana  $F(\gamma_{\delta,s} - [\mathbf{x}'_{ij}\boldsymbol{\beta}_\sigma + \delta_j])$  merupakan fungsi sebaran dari galat level satu yang telah dibakukan ( $\frac{\varepsilon_{ij}}{\sigma}$ ), yang juga merupakan *invers link function* dari model ordinal (Grilli & Pratesi 2002).

Jika peluang kumulatif untuk peubah respon dari mahasiswa ke- $i$  pada kelas ke- $j$  dinyatakan dengan  $\pi_{ijs} = P(Y_{ij} \leq s)$ , maka model regresi logistik ordinal multilevel dengan satu efek acak (*single random effect*) dapat dinyatakan melalui fungsi logit kumulatif, yaitu :

$$\begin{aligned} L_s(x) &= \text{logit}\{P(Y_{ij} \leq s | x)\} = \log\left(\frac{\pi_{ijs}}{1 - \pi_{ijs}}\right) \\ &= \gamma_s - [\mathbf{x}'_{ij}\boldsymbol{\beta} + \delta_j] \end{aligned}$$

Secara umum, model kumulatif logit dengan banyak efek acak (*multiple random effect*) dapat dituliskan sebagai berikut :

$$L_s(x) = \log\left(\frac{\pi_{ijs}}{1 - \pi_{ijs}}\right) = \gamma_s - (\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}_{ij}\boldsymbol{\delta}_j)$$

untuk  $s = 1, \dots, S-1$ , sedangkan  $\mathbf{z}_{ij}$  merupakan vektor peubah bebas lainnya (termasuk 1 untuk intersep) dan  $\boldsymbol{\delta}_j$  merupakan vektor galat pada level dua yang diasumsikan menyebar  $N(\mathbf{0}, \boldsymbol{\Omega})$  dengan  $\mathbf{T}\mathbf{T}' = \boldsymbol{\Omega}$  (dekomposisi Cholesky). Model di atas dapat dituliskan dalam bentuk baku dengan mendefinisikan  $\boldsymbol{\delta}_j = \mathbf{T}\boldsymbol{\theta}_j$ , sehingga (Hedeker 2007) :

$$L_s(x) = \log\left(\frac{\pi_{ijs}}{1 - \pi_{ijs}}\right) = \gamma_s - (\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}_{ij}\mathbf{T}\boldsymbol{\theta}_j)$$
 Nilai

dugaan untuk  $\pi_{ijs} = P(Y_{ij} \leq s)$  dapat diperoleh dengan melakukan transformasi kebalikan terhadap fungsi logit kumulatifnya, yaitu :

$$\pi_{ijs} = P(Y_{ij} \leq s | x) = \frac{1}{1 + \exp(-L_s(x))}$$

Metode pendugaan parameter yang digunakan adalah metode kemungkinan maksimum. Fungsi kemungkinan bersyarat untuk vektor respon  $\mathbf{y}_i$  adalah :

$$L(\mathbf{y}_i | \boldsymbol{\theta}) = \prod_{j=1}^{n_i} \prod_{s=1}^S [P(Y_{ij} = s | \boldsymbol{\theta}_j)]^{d_{ijs}}$$

$$\text{dengan } d_{ijs} = \begin{cases} 1 & \text{jika } y_{ij} = s \\ 0 & \text{jika } y_{ij} \neq s \end{cases}$$

Selanjutnya dicari fungsi kemungkinan maksimum marjinal yaitu :

$$h(\mathbf{y}_i) = \int_{\boldsymbol{\theta}} l(\mathbf{y}_i | \boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

dengan  $g(\boldsymbol{\theta})$  merupakan fungsi kepekatan peluang dari sebaran normal baku (Hedeker 2007). Dugaan parameter diperoleh dengan memaksimalkan fungsi di atas melalui iterasi dengan bantuan PROC GLIMMIX pada *software SAS*.

### Pemilihan Model Terbaik

Strategi pemilihan model terbaik dilakukan dengan tahapan sebagai berikut (Hox 2002):

1. Memilih struktur efek tetap, yaitu :
  - a. Menganalisis model tanpa peubah penjelas.
  - b. Menganalisis model dengan menambahkan seluruh peubah penjelas di level kesatu.
  - c. Menganalisis model dengan menambahkan seluruh peubah penjelas di level kedua.
2. Memilih struktur kemiringan (*slope*) acak dengan cara menguji keragaman kemiringan setiap peubah penjelas pada level individu.
3. Menyusun model terbaik dengan cara menambahkan interaksi antara peubah penjelas level kedua dan peubah penjelas level kesatu yang memiliki keragaman kemiringan yang nyata.

Pembandingan dua model dilakukan dengan menggunakan nilai deviansi (*Deviance*). Statistik uji yang digunakan adalah (West *et al.* 2007) :

$$\begin{aligned} D &= -2\log\left(\frac{L_{\text{tersarang}}}{L_{\text{penuh}}}\right) \\ &= -2\log(L_{\text{tersarang}}) - (-2\log(L_{\text{penuh}})) \end{aligned}$$

dengan  $L_{\text{tersarang}}$  adalah nilai fungsi kemungkinan pada model tersarang,  $L_{\text{penuh}}$  adalah fungsi kemungkinan pada model penuh. Nilai  $D$  yang besar mengindikasikan model penuh lebih sesuai dibandingkan model tersarang.

### Interpretasi Koefisien

Interpretasi untuk model regresi logistik ordinal dapat dilakukan dengan menggunakan nilai rasio oddsnya. Misalkan untuk peubah X yang berskala nominal ( $x_1$  dan  $x_2$ ), rasio odds pada kategori  $Y \leq s$  merupakan perbandingan antara  $x_1$  dan  $x_2$  yang dirumuskan sebagai berikut (Agresti 1990) :

$$L_s(x_1) - L_s(x_2) = \log \left( \frac{P[Y \leq s | x_1 / Y > s | x_1]}{P[Y \leq s | x_2 / Y > s | x_2]} \right) = \beta_i(x_1 - x_2)$$

dengan  $i = 1, 2, \dots, p$  ( $p$  merupakan banyaknya peubah penjelas) dan  $s = 1, 2, \dots, S-1$ . Parameter  $\beta_i$  diartikan sebagai perubahan nilai fungsi logit yang disebabkan oleh perubahan satu unit peubah penjelas ke- $i$  yang disebut log odds, (misalnya antara  $x_1$  dan  $x_2$ ) yang dinotasikan sebagai :

$$\log[\psi(x_1, x_2)] = \beta_i(x_1 - x_2)$$

Sehingga didapatkan penduga untuk rasio odds ( $\hat{\psi}$ ) sebagai berikut (Agresti 1990):

$$\hat{\psi} = \exp[\beta_i(x_1 - x_2)]$$

Untuk peubah bebas kategorik, jika rasio odds bernilai  $> 1$ , maka odds saat  $x_1$  lebih besar daripada odds saat  $x_2$  atau dengan kata lain  $P[Y \leq s | x_1]$  akan selalu lebih besar dari  $P[Y > s | x_1]$ . Sehingga dapat dikatakan, saat  $x_1$  peluang untuk  $Y \leq s$  lebih besar daripada saat  $x_2$ .

Untuk peubah penjelas  $x$  berskala kontinu, odds saat  $x$  mengalami kenaikan 1 unit adalah sebesar  $\exp[\beta(x_1 - x_2)]$  kali odds saat  $x$  belum mengalami kenaikan. Jika nilai rasio odds tersebut bernilai  $> 1$ , maka peluang untuk  $Y \leq s$  saat  $x$  mengalami kenaikan adalah lebih besar dari saat  $x$  belum mengalami kenaikan. Untuk peubah kontinu berskala besar, diperlukan perubahan unit sebesar  $c$  untuk interpretasinya, dengan rasio odds sebesar  $\exp[c\beta]$ .

## METODOLOGI

### Data

Data yang digunakan dalam penelitian ini adalah data nilai akhir mahasiswa dalam mata kuliah Metode Statistika yang berupa huruf mutu pada tahun 2008/2009 dari tujuh departemen di FMIPA IPB. Peubah respon berupa huruf mutu, yaitu A, B, C, D, dan E. Skor untuk masing-masing kategori adalah  $A = 5, B = 4, C = 3, D = 2$  dan  $E = 1$ . Sedangkan peubah penjelas yang terdapat pada setiap level meliputi :

Level kesatu (mahasiswa) :

1. IPK TPB mahasiswa (skala 0 – 4)
2. Jenis kelamin (JK) mahasiswa  
0 : perempuan  
1 : laki-laki
3. Asal daerah (AD)  
0 : Jawa  
1 : Luar Jawa

Level kedua (kelas paralel):

1. Persentase nilai mutu Pengantar Matematika (PM) minimal B
2. Jumlah mahasiswa tiap kelas paralel (JMLH)

### Metode

Tahapan yang dilakukan pada penelitian ini adalah:

1. Melakukan analisis statistika deskriptif terhadap data.
2. Mencari nilai dugaan parameter regresi logistik ordinal satu level untuk semua kelas paralel dengan menggunakan program SAS PROC GENMOD.
3. Mencari nilai dugaan parameter regresi logistik ordinal multilevel dengan menggunakan program SAS PROC GLIMMIX.
4. Mengonfirmasi hasil regresi logistik ordinal satu level dan multilevel.
5. Menentukan model ordinal multilevel terbaik yang dapat memodelkan hubungan antara nilai akhir mahasiswa dalam mata kuliah Metode Statistika dengan peubah-peubah penjelasnya.
6. Menduga komponen ragam nilai akhir mahasiswa dalam mata kuliah Metode Statistika.
7. Menghitung nilai ketepatan klasifikasi dari model yang diperoleh.

## HASIL DAN PEMBAHASAN

### Analisis Deskriptif

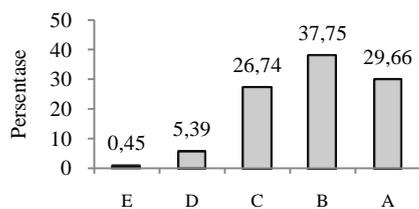
Persentase perolehan nilai akhir Metode Statistika berupa huruf mutu A, B, C, D, dan E disajikan pada Tabel 2. Tampak bahwa kelas STK, KIM, FIS, ILKOM, dan BKM cenderung mendapatkan nilai akhir yang baik. Hal ini terlihat dari tingginya perolehan nilai A dan B. Namun untuk kelas ILKOM, persentase mahasiswa dengan huruf mutu D dan E cukup besar, mencapai 11.24%. Sedangkan kelas GFM dan MTK cenderung memperoleh nilai akhir yang rendah, terlihat dari banyaknya mahasiswa yang mendapat nilai C dan D.

Tabel 2 Persentase perolehan huruf mutu Metode Statistika

Kelas	A	B	C	D	E
STK	45,45	34,85	18,18	1,52	0,00
GFM	4,08	30,61	57,14	8,16	0,00
KIM	61,33	36,00	2,67	0,00	0,00
MTK	12,33	20,55	53,42	13,70	0,00
ILKOM	28,09	42,70	17,98	8,99	2,25
FIS	18,75	46,88	34,38	0,00	0,00
BKM	22,95	57,38	18,03	1,64	0,00

Persentase perolehan nilai akhir Metode Statistika secara keseluruhan disajikan pada Gambar 1. Sebagian besar (37,75%) mahasiswa mendapatkan nilai B, sedangkan mahasiswa yang

mendapatkan nilai A, C, dan D berturut-turut sebesar 29,66%, 26,74%, dan 5,39%. Mahasiswa yang mendapatkan nilai E hanya dua orang atau sebesar 0,45%.



Gambar 1 Persentase perolehan huruf mutu Metode Statistika

Deskripsi nilai peubah penjelas untuk setiap kelas paralel dapat dilihat pada Lampiran 1. Rata-rata IPK TPB mahasiswa adalah sebesar 2,99, sedangkan rata-rata jumlah mahasiswa per kelas sebanyak 73 mahasiswa. Adapun rata-rata persentase nilai mutu Pengantar Matematika minimal B per kelas sebesar 65,39%. Dari Lampiran 1 terlihat bahwa hampir seluruh kelas didominasi oleh mahasiswa dengan jenis kelamin perempuan, kecuali MTK dan ILKOM. Sebagian besar mahasiswa (78,43%) berasal dari Jawa, sedangkan sisanya (21,57%) berasal dari luar Jawa.

Hubungan antara nilai akhir Metode Statistika dengan masing-masing peubah penjelas dapat dilihat pada Lampiran 2. Dari Lampiran 2 dapat dilihat bahwa mahasiswa perempuan cenderung mendapatkan nilai akhir yang lebih baik daripada mahasiswa laki-laki. Mahasiswa yang berasal dari Jawa mendapatkan nilai akhir sedikit lebih baik dibandingkan dengan mahasiswa yang berasal dari luar Jawa, walaupun ada dua mahasiswa dari Jawa yang mendapatkan nilai E.

Untuk analisis deskriptif, IPK TPB dibagi menjadi tiga kelompok, yaitu kelompok  $IPK < 2,75$ ,  $2,75 - 3,5$ , dan  $IPK \geq 3,5$ . Sedangkan untuk peubah persentase nilai Pengantar Matematika minimal B dan jumlah mahasiswa dikelompokkan menjadi dua, yaitu kelompok di atas nilai rata-rata dan di bawah nilai rata-rata.

Berdasarkan kelompok IPK dapat dilihat bahwa semakin besar nilai IPK TPB maka nilai akhir yang diperoleh semakin baik. Hal ini terlihat dari banyaknya mahasiswa dengan IPK TPB di atas 3,5 yang mendapatkan nilai A. Semakin besar persentase nilai Pengantar Matematika minimal B di suatu kelas paralel (di atas 65%) tidak memperbesar proporsi mahasiswa mendapatkan nilai yang lebih baik. Hal ini terlihat pada kelas KIM dengan persentase nilai Pengantar Matematika minimal B sebesar 58,67%, mahasiswanya cenderung memperoleh nilai A dan B. Sedangkan pada kelas MTK dengan persentase nilai Pengantar Matematika minimal B sebesar 76,71% (Lampiran 1), mahasiswanya mempunyai

kecenderungan mendapatkan nilai C dan D sebanyak 67,12% (Tabel 2). Kelas paralel yang memiliki jumlah mahasiswa di atas 73 mempunyai kecenderungan mendapatkan nilai yang lebih baik daripada kelas paralel yang jumlah mahasiswanya di bawah 73. Namun hal ini disebabkan karena sebagian besar kelas paralel merupakan kelas besar, sehingga tidak terlalu terlihat pengaruh jumlah mahasiswa terhadap nilai akhir Metode Statistika.

### Regresi Logistik Ordinal Satu Level

Hasil perbandingan model untuk analisis regresi logistik ordinal satu level memperlihatkan bahwa model dengan semua peubah penjelas (baik pada level mahasiswa maupun level kelas paralel) lebih baik dibandingkan model tanpa peubah penjelas dan model dengan peubah penjelas pada level mahasiswa. Hasil uji perbandingannya dapat dilihat pada Tabel 3. Adapun hasil dugaan parameter berdasarkan model dengan semua peubah penjelas disajikan pada Tabel 4.

Tabel 3 Hasil perbandingan model pada regresi logistik ordinal satu level

Model	-2 log L	Deviansi	Nilai p
tanpa peubah penjelas	1123,82		
dengan peubah penjelas satu level	973,83	150,00	0,00
dengan semua peubah penjelas	935,11	38,72	0,00

Tabel 4 Nilai dugaan parameter regresi logistik ordinal satu level

	Dugaan	Galat Baku	Nilai p
Intercept1	-1,441	1,094	0,1874
Intercept2	1,423	0,861	0,0983
Intercept3	4,088	0,871	<,0001
Intercept4	6,260	0,901	<,0001
JK(0)	-0,612	0,187	0,0011
AD(0)	-0,395	0,226	0,0801
IPK	-2,736	0,238	<,0001
PM	0,085	0,014	<,0001
JMLH	-0,023	0,005	<,0001

Berdasarkan Tabel 4 dapat dilihat bahwa semua peubah penjelas memberikan hasil yang nyata terhadap nilai akhir Metode Statistika pada taraf 5%, kecuali asal daerah yang nyata pada taraf 10%. Hasil ini mengonfirmasi pernyataan sebelumnya bahwa pengabaian struktur hirarkhi cenderung membawa kepada penolakan hipotesis nol.

Dari model yang diperoleh dapat dihitung nilai peluang untuk masing-masing kategori nilai sehingga dapat diketahui ketepatan klasifikasinya. Ketepatan klasifikasi untuk model regresi logistik ordinal satu level dapat dilihat pada Tabel 5. Secara

keseluruhan model yang terbentuk memiliki persentase ketepatan klasifikasi sebesar 51,46%.

Jika dilakukan analisis regresi ordinal dengan mencoba beberapa fungsi hubung, maka nilai ketepatan klasifikasi dengan fungsi hubung logit hanya sedikit lebih baik dibandingkan penggunaan fungsi hubung lainnya. Nilai ketepatan klasifikasi regresi ordinal dengan fungsi hubung *complementary log-log*, *negative log-log*, probit dan cauchit berturut-turut sebesar 51,01%, 49,66%, 51,24%, dan 50,11%. Dengan demikian dalam kasus ini pemilihan fungsi hubung tidak menjadi masalah yang kritis.

Tabel 5 Persentase ketepatan klasifikasi model regresi logistik ordinal satu level

Aktual	Prediksi					Persentase Benar
	E	D	C	B	A	
E	0	0	1	1	0	0,00
D	0	3	16	5	0	12,50
C	0	0	62	51	6	52,10
B	0	0	31	96	41	57,14
A	0	0	6	58	68	51,52
Persentase Benar Keseluruhan						51,46

#### Regresi Logistik Ordinal Multilevel

Pembentukan model terbaik dalam regresi logistik ordinal multilevel memerlukan beberapa tahapan. Tahapan pertama adalah memilih struktur efek tetap, kemudian memilih struktur kemiringan acak, dan tahapan yang terakhir adalah menyusun model terbaik dengan menambahkan interaksi antara peubah penjelas level kedua dan peubah penjelas level kesatu yang memiliki keragaman kemiringan yang nyata.

#### Pemilihan Struktur Efek Tetap

Metode pendugaan parameter yang digunakan dalam pemilihan struktur efek tetap adalah metode kemungkinan maksimum dengan pendekatan *Gauss - Hermite Quadrature*. Adapun tahapan dalam memilih struktur efek tetap adalah dengan membentuk model-model intersep acak sebagai berikut :

1. Model tanpa peubah penjelas (M1.1).
2. Model dengan menambahkan seluruh peubah penjelas di level kesatu (M1.2).
3. Model dengan menambahkan seluruh peubah penjelas, baik di level kesatu maupun di level kedua (M1.3).

Pemilihan model terbaik dilakukan dengan membandingkan model-model yang terbentuk, yaitu menggunakan nilai *Deviansi*. Hasil uji perbandingannya disajikan pada Tabel 6.

Tabel 6 memperlihatkan bahwa hasil perbandingan antara M1.1 dengan M1.2 menghasilkan nilai p sebesar 0,000 sehingga model

yang dipilih adalah M1.2. Sementara itu, hasil perbandingan antara M1.2 dengan M1.3 menghasilkan nilai p sebesar 0,4045. Hal ini berarti peubah penjelas level kedua yang ditambahkan pada M1.3 tidak memberikan hasil yang nyata. Oleh karena itu, model terbaik pada tahapan pemilihan struktur efek tetap adalah M1.2 adalah model dengan peubah penjelas pada level kesatu yang meliputi jenis kelamin, asal daerah, dan IPK TPB.

Tabel 6 Hasil perbandingan model dalam pemilihan struktur efek tetap

Model	-2 log L	Deviansi	Nilai p
M1.1	1025,80		
M1.2	867,97	157,83	0,0000
M1.3	866,16	1,81	0,4045

#### Pemilihan Struktur Kemiringan Acak

Setelah diperoleh model intersep acak terbaik, tahapan selanjutnya adalah memilih model dengan menambahkan efek kemiringan acak pada M1.2. Metode pendugaan parameter yang digunakan masih sama seperti dalam pemilihan struktur efek tetap, yaitu metode *Kemungkinan maksimum* dengan pendekatan *Gauss - Hermite Quadrature*. Model-model yang dibentuk pada tahapan ini adalah :

1. Model dengan intersepanya saja yang acak (M2.1).
2. Model dengan intersep acak dan kemiringan asal daerah acak (M2.2).
3. Model dengan intersep acak dan kemiringan jenis kelamin acak (M2.3).
4. Model dengan intersep acak dan kemiringan IPK TPB acak (M2.4).

Tabel 7 Hasil perbandingan model dalam pemilihan struktur kemiringan acak

Model	Deviansi	Nilai p
M2.1 dengan M2.2	867,97-867,54 = 0,43	0,8065
M2.1 dengan M2.3	867,97-866,60 = 1,37	0,5041
M2.1 dengan M2.4	867,97-863,81 = 4,16	0,1249

Tabel 7 menunjukkan bahwa hasil perbandingan ketiganya tidak ada yang nyata. Sehingga dapat disimpulkan model yang terbaik sampai tahap ini adalah M2.1, yaitu model dengan peubah penjelas pada level kesatu dan intersepanya saja yang acak.

#### Pemilihan Model Terbaik

Tahapan selanjutnya adalah menambahkan interaksi antara peubah penjelas level kedua dan peubah penjelas level kesatu yang memiliki keragaman kemiringan nyata. Dari langkah kedua diperoleh hasil bahwa tidak ada satu pun

kemiringan acak yang nyata. Oleh karena itu, tidak ada interaksi yang dimasukkan ke dalam model.

Dari tahapan pemilihan model yang telah dilakukan, diperoleh model akhir terbaik yaitu model dengan peubah penjelas pada level kesatu dan intersepnya saja yang acak. Hasil pendugaan parameternya disajikan pada Tabel 8. Dari Tabel 8 dapat dilihat bahwa peubah penjelas jenis kelamin dan IPK TPB memberikan hasil yang nyata pada taraf 5%. Karena masih terdapat peubah penjelas yang tidak nyata, yaitu asal daerah maka dilakukan pereduksian terhadap model akhir dengan cara mengeliminasi peubah asal daerah dari model.

Tabel 8 Nilai dugaan parameter pada model akhir

	Dugaan	Galat Baku	Nilai p
Intercept 1	1,654	1,028	0,1586
Intercept 2	4,597	0,803	0,0012
Intercept 3	7,612	0,854	0,0001
Intercept 4	10,142	0,915	<,0001
AD (0)	-0,283	0,236	0,2303
JK (0)	-0,489	0,195	0,0124
IPK	-2,752	0,247	<,0001
$\sigma_{u_{0j}}^2$	1,164		

Hasil perbandingan antara model hasil reduksi dengan model akhir dapat dilihat pada Tabel 9. Uji perbandingan tersebut menghasilkan nilai p sebesar 0,2301, sehingga model terbaik yang dipilih adalah model hasil reduksi, dengan nilai dugaan parameter tercantum pada Tabel 10.

Tabel 9 Hasil perbandingan model reduksi dengan model akhir

Model	-2 log L	Deviansi	Nilai p
Reduksi	867,97		
Akhir	869,41	1,44	0,2301

Tabel 10 Nilai dugaan parameter pada model regresi logistik ordinal multilevel terbaik

	Dugaan	Galat Baku	Nilai p
Intercept 1	1,464	1,016	0,1996
Intercept 2	4,395	0,785	0,0014
Intercept 3	7,411	0,836	0,0001
Intercept 4	9,934	0,897	<,0001
JK( 0)	-0,483	0,195	0,0134
IPK	-2,756	0,247	<,0001
$\sigma_{u_{0j}}^2$	1,184		

Berdasarkan Tabel 10, peubah jenis kelamin dan IPK TPB memberikan pengaruh yang nyata

terhadap nilai akhir Metode Statistika. Model terbaik di atas dapat dituliskan sebagai:

$$\hat{L}_s(x) = \gamma_s - 0,4834JK - 2,7557IPK + u_{0j}$$

dengan  $s = 1, 2, 3, 4$  dan  $j = 1, 2, 3, 4, 5, 6, 7$ .

Nilai negatif pada koefisien regresi peubah jenis kelamin menunjukkan bahwa mahasiswa perempuan cenderung memperoleh huruf mutu yang lebih tinggi dibandingkan mahasiswa laki-laki. Sedangkan nilai negatif pada koefisien regresi peubah IPK TPB berarti semakin tinggi IPK TPB maka mahasiswa cenderung mendapatkan huruf mutu yang lebih tinggi. Model di atas menunjukkan adanya perbedaan antar kelas paralel dengan keragaman intersep pada fungsi logit sebesar 1,184. Adapun nilai dugaan untuk intersep acak dari masing-masing kelas dapat dilihat selengkapnya pada Tabel 11.

Tabel 11 Nilai dugaan intersep acak

	Kelas	Dugaan
Intersep	STK	-0,3771
Intersep	GFM	1,0212
Intersep	KIM	-1,9353
Intersep	MTK	1,6610
Intersep	ILKOM	0,0366
Intersep	FIS	-0,0573
Intersep	BKM	-0,3355

Dari Tabel 11 dapat dilihat bahwa nilai dugaan intersep acak untuk kelas STK, KIM, FIS, dan BKM memiliki nilai koefisien negatif. Artinya, mahasiswa di kelas ini mempunyai peluang yang lebih tinggi untuk mendapatkan nilai dengan kategori baik. Hal sebaliknya terjadi untuk kelas-kelas dengan koefisien intersep positif. Hasil ini sesuai dengan hasil yang didapatkan dari analisis deskriptif pada Tabel 2.

Hasil analisis regresi logistik ordinal multilevel memperlihatkan bahwa peubah penjelas yang berpengaruh secara nyata terhadap nilai akhir Metode Statistika adalah peubah jenis kelamin dan IPK TPB. Hasil ini berbeda dengan hasil analisis regresi logistik ordinal satu level, dimana hampir semua peubah penjelas yang dimasukkan ke dalam model memberikan pengaruh yang nyata.

Dari model yang diperoleh dapat dihitung nilai peluang untuk masing-masing kategori nilai sehingga dapat diketahui ketepatan klasifikasinya. Hasil klasifikasinya dapat dilihat pada Tabel 12.

Dari Tabel 12 dapat diketahui bahwa secara keseluruhan model yang terbentuk memiliki persentase ketepatan klasifikasi sebesar 56,85%. Nilai ini lebih besar dibandingkan ketepatan klasifikasi pada model regresi logistik ordinal satu level (51,46%), sehingga model logistik ordinal multilevel memberikan hasil yang lebih baik dibandingkan model regresi logistik ordinal satu

level. Masih besarnya salah klasifikasi diduga karena ada peubah-peubah penjelas lain yang juga berpengaruh terhadap nilai akhir Metode Statistika yang tidak dimasukkan ke dalam model.

Tabel 12 Persentase ketepatan klasifikasi model regresi logistik ordinal multilevel

Aktual	Prediksi					Persentase Benar
	E	D	C	B	A	
E	0	0	1	1	0	0,00
D	0	5	16	3	0	20,83
C	0	2	68	45	4	57,14
B	0	0	28	101	39	60,12
A	0	0	3	50	79	59,85
Persentase Benar Keseluruhan						56,85

Hasil pengklasifikasian di atas menunjukkan bahwa kesalahan klasifikasi yang terjadi cenderung mengarah pada kategori nilai yang lebih tinggi. Hal ini dikarenakan peubah penjelas yang paling berpengaruh adalah IPK TPB, dimana semakin tinggi IPK TPB maka mahasiswa akan cenderung mendapatkan kategori nilai yang lebih baik. Dengan rata-rata IPK TPB yang cukup tinggi, yaitu sebesar 2,99, ada kecenderungan pengklasifikasian ke kategori yang lebih tinggi. Sebagai contoh, dua orang mahasiswa yang mendapatkan nilai akhir E mempunyai IPK TPB sebesar 2,22 dan 3,11. Hasil pengklasifikasian dari kedua mahasiswa tersebut adalah C dan B.

### Interpretasi Koefisien

Interpretasi koefisien untuk model regresi logistik ordinal dapat dilakukan dengan menggunakan nilai rasio oddsnya. Nilai dan selang kepercayaan rasio odds disajikan pada Tabel 13.

Tabel 13 Rasio odds model regresi logistik ordinal multilevel

Peubah	Dugaan Rasio Odds	SK 95% Rasio Odds	
		Lower	Upper
JK	0,617	0,421	0,904
IPK	0,064	0,039	0,103

Berdasarkan selang kepercayaan 95% dari nilai rasio odds, dapat dijelaskan bahwa mahasiswa laki-laki mempunyai peluang untuk mendapatkan nilai akhir yang lebih baik antara 0,421 - 0,904 kali dibandingkan mahasiswa perempuan. Untuk setiap penurunan IPK TPB sebesar 0.01 satuan, peluang mahasiswa untuk mendapatkan nilai akhir yang lebih baik menjadi 0,039 - 0,103 kali dibandingkan sebelum mengalami penurunan.

## KESIMPULAN

Berdasarkan regresi logistik ordinal multilevel, peubah penjelas yang berpengaruh nyata terhadap nilai akhir Metode Statistika adalah IPK TPB dan jenis kelamin, dengan keragaman intersep antar kelas paralel pada fungsi logit sebesar 1.184. Model regresi logistik ordinal multilevel memberikan ketepatan klasifikasi yang lebih tinggi dibandingkan model regresi logistik ordinal satu level. Untuk meningkatkan ketepatan klasifikasi, perlu dicari peubah-peubah lain yang dapat menjelaskan keragaman perolehan nilai Metode Statistika.

## DAFTAR PUSTAKA

- Agresti A. 1990. *Categorical Data Analysis*. New Jersey : John Wiley and Sons.
- Grilli L, Pratesi M. 2002. Weighted Estimation in Multilevel Ordinal Models to Allow for Informativeness of the Sampling Design. <http://www.ds.unifi.it/ricerca/publicazioni.pdf> [14 Mei 2010].
- Hedeker D. 2007. Multilevel models for ordinal and nominal variables. Di dalam: Leeuw J de, Meijer E, editor. *Handbook of Multilevel Analysis*. New York : Springer. hlm 239-276.
- Hosmer DW, Lemeshow S. 2000. *Applied Logistic Regression*. Ed ke-2. New York : John Wiley and Sons.
- Hox J. 2002. *Multilevel Analysis Techniques and Applications*. New Jersey : Lawrence Erlbaum Associates, Inc.
- Husniyati I. 2010. Penerapan regresi logistik biner multilevel terhadap nilai akhir metode statistika tahun 2008/2009 [skripsi]. Bogor : Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.
- Norusis MJ. 2010. SPSS Statistics Guides: Ordinal Regression. [http://www.norusis.com/pdf/ASPC\\_v13.pdf](http://www.norusis.com/pdf/ASPC_v13.pdf) [20 Agustus 2010].
- West BT, Welch KB, Galecki AT. 2007. *Linear Mixed Models: A Practical Guide Using Statistical Software*. New York : Chapman & Hall.
- Widiyani W. 2009. Pengkajian model regresi dua level terhadap capaian nilai akhir metode statistika tahun 2008/2009 [skripsi]. Bogor : Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.

Lampiran 1 Deskripsi peubah penjelas untuk setiap kelas paralel

Kelas	Rata-rata IPK TPB	Persentase PM $\geq$ B	Jumlah Mahasiswa	Jenis Kelamin (%)		Asal Daerah (%)	
				P	L	Jawa	Luar Jawa
STK	3,20	71,21	70	62,12	37,88	74,24	25,76
GFM	2,76	51,02	49	51,02	48,98	57,14	42,86
KIM	2,99	58,67	78	61,64	38,36	86,30	13,70
MTK	2,98	76,71	79	39,33	60,67	83,15	16,85
ILKOM	3,00	71,91	115	50,00	50,00	59,38	40,63
FIS	2,89	56,25	38	55,74	44,26	80,33	19,67
BKM	2,95	60,66	79	62,67	37,33	89,33	10,67
Rataan	2,99	65,39	73	54,61	45,39	78,43	21,57

Lampiran 2 Tabulasi silang antara nilai akhir Metode Statistika dengan peubah penjelas dalam persen (%).

Peubah penjelas	Huruf Mutu				
	E	D	C	B	A
<b>Jenis Kelamin</b>					
Perempuan	0,4	4,1	23,5	37,9	34,2
Laki-laki	0,5	6,9	30,7	37,6	24,3
<b>Asal Daerah</b>					
Jawa	0,6	4,6	26,1	36,4	32,4
Luar Jawa	0,0	8,3	29,2	42,7	19,8
<b>IPK TPB</b>					
< 2,75	0,8	14,5	45,2	29,8	9,7
2,75-3,5	0,4	2,4	24,0	43,7	29,5
$\geq$ 3,5	0,0	0,0	0,3	29,9	67,2
<b>Persen PM minimal B</b>					
< 65%	0,0	2,3	24	42,4	31,3
$\geq$ 65%	0,9	8,3	29,4	33,3	28,1
<b>Jumlah mahasiswa</b>					
< 73	0,0	3,4	34,7	36,1	25,9
$\geq$ 73	0,7	6,4	22,8	38,6	31,5