

Johan Harlan

YT

YT

ANALISIS MULTILEVEL

YT



YT

PENERBIT GUNADARMA

ANALISIS MULTILEVEL

Johan Harlan



Penerbit Gunadarma

Analisis Multilevel

Penulis : Johan Harlan

Cetakan Pertama, Agustus 2016

Disain cover : Joko Slameto

Diterbitkan pertama kali oleh Gunadarma

Jl. Margonda Raya No. 100, Pondokcina, Depok 16424

Telp. +62-21-78881112, 7863819 Faks. +62-21-7872829

e-mail : sektor@gunadarma.ac.id

Hak Cipta dilindungi undang-undang. Dilarang mengutip atau memperbanyak dalam bentuk apapun sebagian atau seluruh isi buku tanpa ijin tertulis dari penerbit.

KATA PENGANTAR

Analisis multilevel merupakan salah satu hasil perpaduan antara perkembangan ilmu Statistika dengan kemajuan teknologi informatika, terutama dalam beberapa dekade terakhir. Perkembangan ilmu Statistika yang dimaksudkan adalah perkembangan dan perluasan “*linear models*” menjadi “*generalized linear mixed models*” beserta metodenya, sedangkan kemajuan teknologi informatikanya ialah peningkatan kemampuan komputasi dengan komputer mutakhir, baik dari segi kuantitas data yang mampu diolah maupun kecepatan pengolahan datanya.

Penerapan analisis multilevel pada saat ini telah mencakup berbagai bidang ilmu, seperti psikologi, kependidikan, sosiologi, kedokteran, bisnis, ekonomi, dan sebagainya. Metode analisis multilevel dapat digunakan untuk data kelompok (data kluster), data longitudinal dan hasil pengukuran berulang, model Bayesian, dan lain-lain. Analisis multilevel juga dapat digunakan dalam *structural equation modeling* (SEM), walaupun belum semua tekniknya dapat diterapkan pada multilevel SEM. Analisis multilevel merupakan teknik statistik yang dapat dikatakan belum sepenuhnya mencapai kesempurnaan dan masih terus berkembang di waktu mendatang.

Semua contoh yang dibahas dalam buku ini diolah dengan paket statistik STATA. Kemampuan awal yang dibutuhkan dari pembaca untuk memahami uraian dalam buku ini adalah pengetahuan dasar mengenai analisis regresi dan SEM.

Jakarta, Agustus 2016

Penulis

DAFTAR ISI

Kata Pengantar	v
Daftar Isi	vii
Bab 1 Pendahuluan	1
Level, Kluster, dan Grup	1
Contoh 1.1 Macam Level dan Grup	3
Efek Fixed, Efek Random, dan Efek Mixed	6
Estimasi Parameter	8
Bab 2 Tipe Model Linear	10
Model Linear	10
Generalized Linear Models	11
Generalized Linear Mixed Models	13
Model Multilevel	15
Ukuran Sampel pada Model Multilevel	16
Bab 3 Linear Models	18
Karakteristik Linear Models	18
Model dan Sintaks	18
Contoh 3.1 Kinerja Akademik Sekolah	19

Bab 4	Generalized Linear Models	25
	Karakteristik Generalized Linear Models	25
	Model dan Sintaks	25
	Contoh 4.1 Berat Badan Lahir Rendah	26
Bab 5	Linear Mixed Models	31
	Karakteristik Linear Mixed Models	31
	Model Umum	31
	A. Hanya Konstante Bervariasi Antar-Grup	32
	Contoh 5.1 Berat Badan Lahir Bayi (1)	32
	B. Konstante dan Koefisien Regresi <i>age</i> dan <i>childsex</i> Bervariasi Antar-Grup	39
	Contoh 5.2 Berat Badan Lahir Bayi (2)	39
Bab 6	Generalized Linear Mixed Models	44
	Karakteristik Generalized Linear Mixed Models	44
	Model Umum	45
	A. Hanya Konstante Bervariasi Antar-Grup	45
	Contoh 6.1 Penggunaan Kontrasepsi (1)	45
	B. Konstante dan Koefisien Regresi <i>age</i> Bervariasi Antar-Grup	52
	Contoh 6.2 Penggunaan Kontrasepsi (2)	53

Bab 7	Multilevel Linear Mixed Models	59
	Karakteristik Multilevel Linear Mixed Models	59
	Model Umum	59
	A. Hanya Konstante Bervariasi Antar-Grup	60
	Contoh 7.1 Popularitas Siswa (1)	60
	B. Konstante dan Koefisien Regresi <i>texp</i> Bervariasi Antar-Grup	66
	Contoh 7.2 Popularitas Siswa (2)	67
 Bab 8	 Multilevel Generalized Linear Mixed Models	 72
	Karakteristik Multilevel GLMM	72
	Model Umum	73
	A. Hanya Konstante Bervariasi Antar-Grup	73
	Contoh 8.1 Pengulangan Kelas Siswa (1)	73
	B. Konstante dan Koefisien Regresi <i>pped</i> Bervariasi Antar-Grup	80
	Contoh 8.2 Pengulangan Kelas Siswa (2)	80
 Bab 9	 Analisis Multilevel untuk Data Longitudinal	 85
	Analisis Data Longitudinal	85
	Contoh 9.1 IPK Mahasiswa (1)	86
	A. Hanya Konstante Bervariasi Antar-Grup	88
	B. Konstante dan Koefisiensi Regresi <i>occas</i> Bervariasi Antar-Grup	92
	Contoh 9.2 IPK Mahasiswa (2)	93

Bab 10	Multilevel SEM I: Analisis Jalur	98
	Tipe Analisis Jalur Multilevel	98
	A. Hanya Konstante Bervariasi Antar-Grup	99
	Contoh 10.1 Gaji Karyawan (Random Intercept)	100
	B. Koefisien Regresi Bervariasi Antar-Grup	108
	Contoh 10.2 Gaji Karyawan (Random Slope)	110
	C. Konstante dan Koefisien Regresi Bervariasi Antar-Grup	115
	Contoh 10.3 Gaji Karyawan (Random Intercept and Random Slope)	116
Bab 11	Multilevel SEM II: Analisis Faktor Konfirmatorik	123
	Model Multilevel CFA	123
	Contoh 11.1 Kemampuan Matematika Siswa	125
	Kepustakaan	136
	Lampiran 1: Ukuran Sampel	138
	Lampiran 2: Beberapa Prinsip Tentang Nilai-P	145

BAB 1

PENDAHULUAN

Level, Klaster, dan Grup

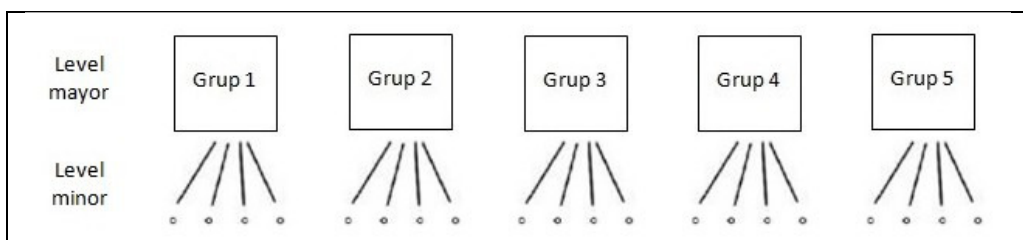
Penggunaan istilah ‘level’ dan ‘kluster’ berawal mula dari rancangan studi untuk pengumpulan data secara kategorik. Jika data dikumpulkan dari 2 atau lebih kategori yang memiliki hirarki bertingkat, kategori tersebut dinamakan ‘**level**’. Pembahasan selanjutnya di sini akan dibatasi hanya untuk 2 level. Pada pengumpulan data dengan 2 level, level yang lebih tinggi secara hirarkis dinamakan ‘**level makro**’, sedangkan level yang lebih rendah dinamakan ‘**level mikro**’. Misalnya pada pengumpulan data guru dan siswa di sekolah, data guru dapat dikumpulkankan pada level kelas (level makro) dan data siswa dikumpulkan pada level siswa (level mikro).

Jika pada suatu tingkatan/level terdapat beberapa kategori yang secara hirarki setingkat, kategori tersebut adalah ‘**klaster**’. Misalnya data siswa yang dikumpulkan dari beberapa kelas, tiap kelas merupakan 1 klaster. Pada level makro akan didapatkan lebih daripada 1 klaster.

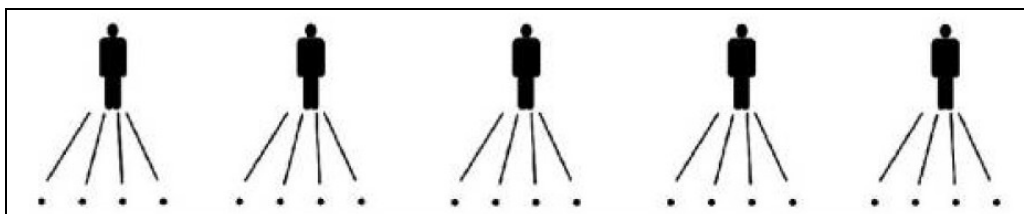
Grup memiliki pengertian yang sama dengan klaster. Istilah klaster lebih banyak digunakan dalam tahap pengumpulan data

yang biasanya menggunakan proses sampling acak kluster, sedangkan istilah grup lebih lazim dipakai pada tahap analisis dengan analisis multilevel (Gambar 1.1).

Jika data diperoleh dari lebih daripada 1 level, maka digunakan **Analisis Multilevel**. Data longitudinal juga dapat diolah dengan Analisis Multilevel, yaitu ada level subjek/objek penelitian yang menjalani pengukuran dan ada level pengukuran (Gambar 1.2).



Gambar 1.1 Skema sampling untuk data multilevel



Gambar 1.2 Skema pengumpulan data longitudinal

Contoh 1.1 Macam Level dan Grup

1. Satu grup

Diambil sampel beranggotakan n siswa di sebuah kelas 5 pada sebuah sekolah. Variabel yang diukur untuk tiap siswa adalah:

mat : Nilai matematika,

ipa : Nilai IPA

Level tunggal yang perlu diperhitungkan di sini adalah level siswa, yang berasal dari 1 kelas (1 grup).

Model regresi:

$$mat_i = \beta_0 + \beta_1 ipa_i + \varepsilon_i \quad (1.1)$$

$i = 1, 2, \dots, n$ menyatakan nomor urut siswa

Karena data hanya dikumpulkan dari 1 grup yaitu 1 kelas, level kelas tidak perlu diperhitungkan. Analisis data untuk satu grup ini **tidak** menggunakan analisis multilevel.

2. Dua level, beberapa grup pada level makro, tidak ada variabel diukur pada level makro

Di sebuah sekolah dimiliki 3 kelas 5 paralel, yaitu kelas 5A, 5B, dan 5C. Dari tiap kelas diambil sampel beranggotakan masing-masing n_1 , n_2 , dan n_3 siswa. Dari tiap siswa diukur nilai **mat** dan **ipa** -nya.

Di sini terdapat 2 level, yaitu level kelas dan level siswa, tetapi tidak ada variabel yang nilainya dikumpulkan pada

level kelas, yaitu level kelas hanya untuk gruping (pengelompokan) siswa. Terdapat 3 grup pada level kelas, yaitu kelas 5A, 5B, dan 5C.

Model regresi:

$$mat_{ij} = \beta_0 + \beta_1 ipa_{ij} + u_{0j} + u_{1j} ipa_{ij} + \varepsilon_{ij} \quad (1.2)$$

$i = 1, 2, \dots, n_j$ menyatakan nomor urut siswa pada kelas (grup) ke- j

$j = 1, 2, 3$ menyatakan nomor urut kelas (grup)

Analisis data untuk dua level, beberapa grup pada level makro tanpa pengukuran variabel pada level makro ini dilakukan **dengan** analisis multilevel.

Dalam tiap kelas (grup) mungkin hanya didapat intersep yang berbeda jika $u_{1j} = 0$, mungkin hanya koefisien regresi nilai IPA yang berbeda jika $u_{0j} = 0$, atau mungkin baik intersep maupun koefisien regresi nilai IPA-nya berbeda jika $u_{0j} \neq 0$ dan $u_{1j} \neq 0$.

3. Multi-level

Di sebuah sekolah dimiliki 10 kelas V paralel, yaitu kelas VA, VB, . . . , VJ. Tiap kelas masing-masing diajar oleh 1 orang

guru yang berbeda untuk tiap kelas, sehingga ada 10 orang guru. Untuk tiap guru diukur nilai:

tahun : Lama pengalaman mengajar dalam tahun

Dari tiap kelas juga diambil sampel beranggotakan masing-masing n_1, n_2, \dots, n_{10} siswa. Dari tiap siswa diukur nilai **mat** dan **ipa** -nya.

Karena pada tiap kelas hanya ada 1 orang guru yang berbeda untuk tiap kelas, istilah ‘level kelas’ identik dengan ‘level guru’. Di sini terdapat dua level, yaitu level siswa (level terbawah, level mikro) dan level kelas / guru (level teratas, level makro).

Pengukuran dilakukan baik pada level mikro (siswa) maupun level makro (guru), namun **variabel dependen selalu diukur pada level mikro**.

Model regresi:

$$\mathbf{mat}_{ij} = \beta_0 + \beta_1 \mathbf{ipa}_{ij} + u_{0j} + u_{1j} \mathbf{ipa}_{ij} + u_{2j} \mathbf{tahun}_{ij} + \varepsilon_{ij} \quad (1.3)$$

$i = 1, 2, \dots, n_j$ menyatakan nomor urut siswa pada kelas (grup) ke- j

$j = 1, 2, 3$ menyatakan nomor urut kelas / guru (nomor grup)

mat adalah variabel yang diukur pada level siswa, sedangkan **tahun** diukur pada level guru (1 guru pada tiap kelas). Data

tahun akan bernilai sama untuk tiap grup (kelas/guru), namun berbeda antar grup.

Analisis data di sini dilakukan **dengan** analisis multilevel. Perhatikan bahwa pada analisis multilevel digunakan lambang yang berbeda untuk koefisien regresi di kedua level ini, yaitu β_i untuk level siswa dan u_{ij} untuk level guru. Perhatikan juga bahwa kedua level masing-masing memiliki intersep sendiri yang berbeda, yaitu β_0 dan u_{0j} .

Efek Fixed, Efek Random, dan Efek Mixed

Efek sebuah prediktor dalam model regresi ditentukan oleh koefisien regresinya. Sebuah prediktor dikatakan memiliki **efek fixed**, jika koefisien regresinya bernilai sama bagi seluruh anggota sampel. **Model fixed** adalah model regresi yang seluruh prediktornya memiliki efek fixed.

Sebuah prediktor dikatakan memiliki **efek random**, jika nilai koefisien regresinya berbeda antar 2 atau lebih subkelompok (baca: grup) anggota sampel. **Model mixed** adalah model yang memiliki prediktor dengan efek fixed maupun prediktor dengan efek random dalam 1 model.

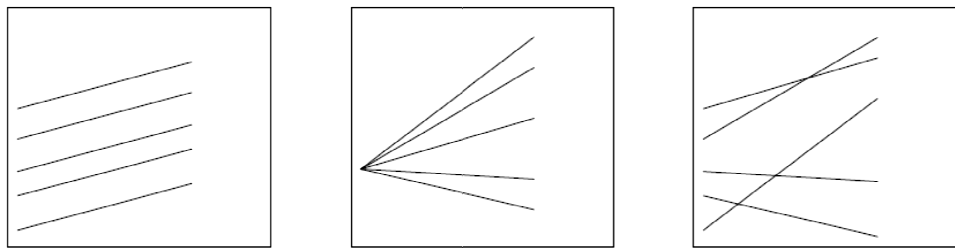
Dalam pembahasan mengenai efek fixed dan efek random pada model regresi ini, intersep juga dianggap sebagai koefisien

regresi bagi salah satu prediktor untuk respons, yaitu $X_{0i} = 1, i = 1, 2, \dots, n$; n menyatakan jumlah anggota sampel. Dengan demikian, model yang hanya berbeda nilai intersep-nya antar-grup ini juga tergolong dalam bentuk model dengan efek random.

Selanjutnya model dengan efek random (dan efek mixed) demikian dibedakan menjadi model dengan *random intercept* dan model dengan *random slope*. Model dengan ***random intercept*** adalah model yang hanya nilai intersep-nya berbeda antar grup, sedangkan model dengan ***random slope*** adalah model yang nilai koefisien regresinya (termasuk intersep-nya) berbeda pada tiap grup.

Gambaran model regresi linear multilevel dengan 1 prediktor demikian diperlihatkan pada gambar 1.3.

Dalam praktik, model dengan (hanya) *random slope* relatif jarang ditemukan, sehingga pembahasan selanjutnya terutama ditujukan pada model dengan (hanya) *random intercept* serta model dengan *random intercept* dan *random slope*.



Gambar 1.3 Gambaran model regresi linear multilevel dengan 1 prediktor.

Kiri: Model dengan *random intercept*. Tengah: Model dengan *random slope*. Kanan: Model dengan *random intercept* dan *random slope*.

Estimasi Parameter

Pada model regresi linear konvensional, estimasi parameter dilakukan dengan **Metode Kuadrat Terkecil** (*ordinary least square*; OLS), tetapi metode ini tak dapat digunakan pada analisis multilevel. Pada analisis multilevel, estimasi parameter yang paling lazim digunakan adalah **Metode Likelihood Maksimum** (*maximum likelihood*; ML).

Dua metode *maximum likelihood* yang dapat digunakan untuk analisis multilevel yaitu *Full Maximum Likelihood* (FML) dan *Restricted Maximum Likelihood* (RML). FML yang relatif lebih mudah dari segi komputasi, lazim digunakan untuk mengestimasi efek fixed, sedangkan untuk efek random lebih baik digunakan RML. Walaupun demikian, perbedaan hasil antara

kedua metode relatif kecil, dan untuk sampel besar perbedaan hasil antara keduanya dapat diabaikan.

Dalam program statistik STATA yang digunakan pada contoh-contoh selanjutnya, metode *default* adalah FML, walaupun opsi RML juga tersedia sebagai metode *non-default*.

BAB 2

TIPE MODEL LINEAR

Model Linear

Dalam pembahasan Statistika setengah abad lampau, yang dimaksud dengan ‘**model linear**’ adalah analisis regresi beserta bentuk variasinya yaitu analisis variansi dan analisis kovariansi. Pada model untuk ketiga bentuk analisis ini selalu didapatkan variabel respons kontinu, yang sekurang-kurangnya berskala interval.

Dalam perkembangan lebih lanjut, ditemukan berbagai teknik pemodelan Statistika untuk meregresikan bentuk-bentuk variabel respons lain terhadap himpunan prediktornya. Beberapa model regresi yang dikembangkan untuk berbagai bentuk variabel responsnya antara lain yaitu:

- Analisis regresi logistik untuk variabel respons **biner**,
- Analisis regresi logistik ordinal untuk variabel respons **ordinal**,
- Analisis regresi multinomial untuk variabel respons **nominal**,
- Analisis regresi Poisson dan regresi binomial negatif untuk variabel respons berupa **data cacah** (*count data*), dan lain-lain.

Generalized Linear Models

Keseluruhan model regresi dengan berbagai bentuk variabel respons ini dikelompokkan bersama sebagai ‘**Generalized Linear Models**’ (GLM). Ruas kiri persamaan model regresi tidak berisikan variabel responsnya sendiri, melainkan fungsi dari variabel respons tersebut, yang dinamakan ‘*link function*’, yaitu:

$$\eta_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} \quad (2.1)$$

dengan: $\eta_i = f(Y_i)$ adalah *link function*

atau dalam bentuk matriks:

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} \quad (2.1.a)$$

Model regresi linear menjadi salah satu anggota GLM dengan fungsi variabel respons yang sama dengan variabel responsnya sendiri [$f(Y_i) = Y_i$], sehingga *link function*-nya dinamakan **fungsi identitas** (*identity function*), hanya pada ruas kanan model analisis regresi didapatkan suku galat:

$$\eta_i = Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i \quad (2.2)$$

Selain model regresi linear, anggota GLM lainnya antara lain yaitu:

1. Model regresi logit:

Variabel respons Y berskala biner (*binary*); $Y_i = 0, 1$.

$$\eta_i = \ln \frac{\pi_i}{1-\pi_i} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi}$$

(2.3)

dengan: $\pi_i = E(y_i)$ (2.3.a)

2. **Model regresi logit ordinal:**

Variabel respons Y berskala ordinal.

3. **Model regresi logit multinomial:**

Variabel respons Y berskala nominal.

4. **Model regresi Poisson:**

Variabel respons Y adalah data cacah (*count data*) dengan eki-distensi (asumsi distribusi Poisson: variansi sama besar dengan rerata).

$$\eta_i = \ln Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi}$$

(2.4)

dengan $Var(Y_i) = E(Y_i)$.

5. **Model regresi binomial negatif:**

Variabel respons Y adalah data cacah (*count data*) dengan over-distensi (variansi lebih besar daripada rerata).

$$\eta_i = \ln Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi}$$

dengan $Var(Y_i) > E(Y_i)$.

Generalized Linear Mixed Models

Dalam tahap lebih lanjut, dikembangkan pula metode untuk mengestimasi parameter model regresi dengan data yang dikumpulkan dari beberapa level dan/atau grup. Pemodelan untuk tipe data tersebut terakhir ini melibatkan keberadaan efek random, sehingga kelompok *Generalized Linear Models* diperluas menjadi ‘**Generalized Linear Mixed Models**’ (GLMM).

Beberapa anggota keluarga *Generalized Linear Mixed Models* ini adalah:

1. *Linear models* (LM)

Data Gaussian; efek *fixed*; satu grup

Model (dalam bentuk persamaan matriks): $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
(2.5)

2. *Generalized linear models* (GLM)

Data non-Gaussian; efek *fixed*; satu grup

Model: $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$
(2.6)

3. *Linear mixed models* (LMM)

Data Gaussian; efek *mixed*; dua level-beberapa grup tanpa pengukuran variabel pada level makro

Model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$
(2.7)

4. *Generalized linear mixed models* (GLMM)

Data non-Gaussian; efek *mixed*; dua level-beberapa grup tanpa pengukuran variabel pada level makro

$$\text{Model: } \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \quad (2.8)$$

5. *Multilevel linear mixed models* (Multilevel LMM)

Data Gaussian; efek *mixed*; multi-level

$$\text{Model: } \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \quad (2.9)$$

6. *Multilevel generalized linear mixed models* (Multilevel GLMM)

Data non-Gaussian; efek *mixed*; multi-level

$$\text{Model: } \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \quad (2.10)$$

Catatan

- Untuk menyederhanakan pembahasan, “parameter” selanjutnya digunakan dalam arti (sebenarnya) “statistik” yang diperoleh dari analisis data sampel.
- Istilah “data Gaussian dan non-Gaussian” mengacu pada variabel respons Y . Untuk model dengan variabel respons Y merupakan data non-Gaussian, ruas kanan persamaan tidak memiliki suku galat.
- $\boldsymbol{\eta}$ adalah *link function*. Untuk data Gaussian, $\boldsymbol{\eta} = \mathbf{y}$ (fungsi identitas); sedangkan untuk respons biner, $\boldsymbol{\eta} = \ln \frac{\pi_i}{1-\pi_i}$; $\pi_i = E(y_i)$.
- Pembahasan untuk GLM, GLMM, dan *Multilevel* GLMM selanjutnya dibatasi untuk respons biner (model logit), walaupun sebenarnya GLM, GLMM, dan *Multilevel* GLMM juga mencakup respons kategorik ordinal (model logit

ordinal), kategorik nominal (model logit multinomial), serta data cacah (*count data*; model Poisson dan binomial negatif).

Model Multilevel

Di antara keenam model pada *Generalized Linear Mixed Models* yang disebutkan di atas, yang tergolong dalam model multilevel adalah model 3) s.d. 6), yaitu LMM, GLMM, *Multilevel* LMM, dan *Multilevel* GLMM. Analisis data untuk model 3) s.d. 6) ini harus dilakukan dengan **analisis multilevel**, sedangkan model 1) dan 2) dapat dianalisis secara konvensional tanpa menggunakan analisis multilevel.

Yang membedakan model multilevel ini dengan model non-multilevel (LM dan GLM) yaitu pada model multilevel selalu didapatkan efek *mixed*, yang terdiri atas efek *fixed* dan efek random. Dengan demikian, dalam kepustakaan Statistika model multilevel ini dikenal juga sebagai **model mixed**, dan ada pula sumber kepustakaan yang menamakannya sebagai **model hierarkis** (*hierarchical models*).

Model multilevel dibedakan menjadi model multilevel dengan *Nested Groupings* (pengelompokan tersarang) dan model multilevel dengan *Non-Nested Groupings* (pengelompokan tak-tersarang). Contoh model multilevel dengan pengelompokan tersarang misalnya yaitu model 3 level yang terdiri atas level

siswa, level kelas, dan level sekolah. Siswa tersarang dalam kelas, sedangkan kelas tersarang dalam sekolah. Seorang siswa menjadi anggota 1 kelas, tak mungkin juga menjadi anggota kelas lain. Sebuah kelas termasuk dalam 1 sekolah, tak mungkin juga termasuk dalam sekolah lain. Pada data longitudinal, pengamatan (untuk tiap titik waktu) tersarang dalam subjek penelitian.

Contoh model multilevel dengan pengelompokan tak-tersarang misalnya yaitu pekerja yang dikelompokkan menurut jenis pekerjaan dan wilayah kediaman. Dalam 1 wilayah dapat ditemukan pekerja dengan jenis pekerjaan yang berbeda, sedangkan pekerja dengan jenis pekerjaan yang sama dapat berkediaman di wilayah yang berbeda. Di sini jenis pekerjaan tidak tersarang dalam wilayah kediaman, begitu pula sebaliknya.

Dalam pembahasan selanjutnya pada buku ini hanya akan dibahas model multilevel dengan pengelompokan tersarang.

Ukuran Sampel pada Model Multilevel

Ukuran sampel pada analisis multilevel memerlukan perhitungan yang rumit dan harus ditentukan untuk tiap level. Aturan umum yang berlaku yaitu untuk ukuran total sampel yang sama, ukuran sampel yang lebih besar untuk level yang lebih tinggi akan menghasilkan kekuatan uji (*power*) dan presisi yang lebih tinggi. Misalnya, untuk pengamatan yang diperoleh dari

1000 siswa yang masing-masing tersarang dalam sekolahnya, rancangan dengan 50 sekolah dan 20 siswa di tiap sekolah lebih baik daripada rancangan dengan 20 sekolah dan 50 siswa di tiap sekolah.

Van Breukelen dan Moerbeek (2013) memasukkan fungsi biaya pengumpulan sampel sebagai kendala untuk menghitung ukuran sampel optimal dengan kesimpulan yang sama, yaitu ukuran sampel yang lebih besar untuk level yang lebih tinggi (jumlah grup) akan menghasilkan kekuatan uji dan presisi yang lebih tinggi.

Hox et al (2013) menyimpulkan bahwa 50 grup pada rancangan 2 level sudah menghasilkan akurasi yang cukup dalam praktik, bahkan jika yang diminati hanya koefisien regresi, 20 grup sudah mencukupi.

Kreft mengajukan *rule of thumb*, yang dinamakan aturan 30/30. Untuk mencapai tujuan dengan aman, sebaiknya diupayakan sampel yang paling sedikit terdiri atas 30 kelompok dengan paling sedikit 30 individu per kelompok (Hox, 2010).

Pembahasan mengenai ukuran sampel untuk analisis multilevel dapat dilihat secara lebih rinci pada Lampiran 1.

BAB 3

LINEAR MODELS

Karakteristik Linear Models

Linear Models adalah model yang didapatkan pada analisis regresi linear biasa yang telah lama dikenal, baik regresi linear sederhana (*simple linear regression*) dengan satu variabel independen maupun regresi linear ganda (*multiple linear regression*) dengan lebih daripada satu variabel independen.

Karakteristik *Linear Model* (LM) yaitu:

- Data Gaussian, yaitu variabel dependen berskala kontinu dan berdistribusi normal.
- Efek *fixed*, yaitu parameter bernilai sama untuk seluruh anggota populasi (estimasi parameter bernilai sama untuk seluruh anggota sampel).
- Satu grup. Populasi dan sampel hanya berasal dari satu grup.

Model dan Sintaks

Model umum (dalam bentuk matriks) pada *Linear Model* adalah:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3.1)$$

y : Vektor variabel dependen

β : Vektor parameter (koefisien regresi; termasuk intersep)

X : Matriks variabel independen

ϵ : Vektor galat

Pada analisis data dengan STATA, sintaks yang digunakan adalah:

. regress depvar indepvars

depvar : Variabel dependen

indepvars : Variabel independen

Contoh 3.1: Kinerja Akademik Sekolah

Data: model-01_elemapi.dta

. use "D:\Analisis Multilevel\Data\model-01_elemapi"

. summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
api00	400	647.6225	142.249	369	940
meals	315	71.99365	24.38557	6	100
acs_k3	398	18.54774	5.004933	-21	25
full	400	66.0568	40.29793	.42	100

Variabel:

- **api00** : Kinerja akademik sekolah tahun 2000
- **meals** : Persentase siswa yang mendapat makanan gratis (indikator kemiskinan)

- ***acs_k3*** : Rerata ukuran kelas TK s.d. kelas 3
- ***full*** : Persentase guru yang memiliki akreditasi penuh untuk mengajar

Keterangan:

File ini memuat data 400 sekolah di sebuah area di AS, dengan kinerja akademik tiap sekolah pada tahun 2000 (***api00***) sebagai variabel dependen. Prediktor adalah persentase siswa yang mendapat makanan gratis di sekolah (***meals***), rerata ukuran kelas TK s.d. kelas 3 (***acs_k3***), dan persentase guru yang memiliki akreditasi penuh untuk mengajar (***full***).

Perhatikan bahwa unit sampling di sini adalah sekolah dan seluruh variabel diukur pada level sekolah, tidak ada variabel yang diukur pada level siswa.

Hanya prediktor ***full*** yang memiliki data lengkap untuk 400 sekolah. Untuk prediktor ***meals*** dan ***acs_k3*** didapatkan nilai-nilai kosong (*missing data*), masing-masing yaitu 85 nilai kosong untuk ***meals*** dan 2 nilai kosong untuk ***acs_k3***.

. list in 1/10

	api00	meals	acs_k3	full
1.	693	67	16	76.00
2.	570	92	15	79.00
3.	546	97	17	68.00
4.	571	90	20	87.00
5.	478	89	18	87.00
6.	858	.	20	100.00
7.	918	.	19	100.00
8.	831	.	20	96.00
9.	860	.	20	100.00
10.	737	29	21	96.00

Model: $api00_i = \beta_0 + \beta_1 acs_k3_i + \beta_2 meals_i + \beta_3 full_i + \varepsilon_i$

Perintah Stata:

. regress api00 acs_k3 meals full

Source	SS	df	MS	Number of obs =	313
Model	2634884.26	3	878294.754	F(3, 309) =	213.41
Residual	1271713.21	309	4115.57673	Prob > F =	0.0000
Total	3906597.47	312	12521.1457	R-squared =	0.6745
				Adj R-squared =	0.6713
				Root MSE =	64.153

api00	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
acs_k3	-2.681508	1.393991	-1.92	0.055	-5.424424	.0614073
meals	-3.702419	.1540256	-24.04	0.000	-4.005491	-3.399348
full	.1086104	.090719	1.20	0.232	-.0698947	.2871154
_cons	906.7392	28.26505	32.08	0.000	851.1228	962.3555

Model estimasi:

$$api00_i = 906.739 - 2.682acs_k3_i - 3.702meals_i + 0.109full_i + \varepsilon_i$$

Untuk mendapatkan nilai-nilai koefisien regresi terstandarisasi, perintah STATA adalah:

. regress api00 acs_k3 meals full, beta

api00	Coef.	Std. Err.	t	P> t	Beta
acs_k3	-2.681508	1.393991	-1.92	0.055	-.0635654
meals	-3.702419	.1540256	-24.04	0.000	-.8075094
full	.1086104	.090719	1.20	0.232	.0408765
_cons	906.7392	28.26505	32.08	0.000	.

Jika yang ingin ditampilkan hanya nilai-nilai koefisien regresi beserta koefisien terstandarisasinya, perintah Stata adalah:

. listcoef, help

regress (N=313): Unstandardized and Standardized Estimates

Observed SD: 111.89793

SD of Error: 64.152761

	api00	b	t	P> t	bStdX	bStdY	bStdXY	SDofX
acs_k3		-2.68151	-1.924	0.055	-7.1128	-0.0240	-0.0636	2.6526
meals		-3.70242	-24.038	0.000	-90.3586	-0.0331	-0.8075	24.4053
full		0.10861	1.197	0.232	4.5740	0.0010	0.0409	42.1138

b = raw coefficient

t = t-score for test of b=0

P>|t| = p-value for t-test

bStdX = x-standardized coefficient

bStdY = y-standardized coefficient

bStdXY = fully standardized coefficient

SDofX = standard deviation of X

Untuk memprediksi nilai-nilai **api00** ; berdasarkan model tersebut, perintah Stata adalah:

. predict yhat

(option xb assumed; fitted values)

(87 missing values generated)

Berikut diperlihatkan 10 nilai-nilai pertama prediksi $\mathbf{api00}_i$, yaitu:

. list $\mathbf{api00}$ \mathbf{yhat} in 1/10

	+-----+	
	api00	yhat

1.	693	624.0273
2.	570	534.4742
3.	546	509.4043
4.	571	529.3403
5.	478	538.4058

6.	858	.
7.	918	.
8.	831	.
9.	860	.
10.	737	753.4839
	+-----+	

Prediksi untuk subjek No. 6 s.d. 9 tidak ada, karena nilai 1 atau lebih prediktornya kosong (*missing values*).

BAB 4

GENERALIZED LINEAR MODELS

Karakteristik Generalized Linear Models

Karakteristik *Generalized Linear Models* (GLM) adalah:

- Data non-Gaussian, yaitu variabel dependen tidak berdistribusi normal, bahkan tak berskala kontinu. Variabel dependen dapat berskala biner, kategorik nominal, kategorik ordinal, atau data cacah. Di sini hanya akan dibahas GLM dengan variabel dependen berskala biner.
- Efek *fixed*, yaitu parameter bernilai sama untuk seluruh anggota populasi (estimasi parameter bernilai sama untuk seluruh anggota sampel).
- Satu grup. Populasi dan sampel hanya berasal dari satu grup.

Model dan Sintaks

Model umum untuk *Generalized Linear Model* adalah:

$$\eta = X\beta$$

Dengan membatasi pembahasan hanya pada model logit, sintaks **Stata** adalah:

. logit depvar indepvars

depvar : Variabel dependen

indepvars : Variabel independen

Contoh 4.1: Berat Badan Lahir Rendah

Data: model-02_lbw.dta

. use "D:\Analisis Multilevel\Data\model-02_lbw"

(Hosmer & Lemeshow data)

. summarize

Variable		Obs	Mean	Std. Dev.	Min	Max
id		189	121.0794	63.30363	4	226
low		189	.3121693	.4646093	0	1
age		189	23.2381	5.298678	14	45
lwt		189	129.8201	30.57515	80	250
race		189	1.846561	.9183422	1	3
smoke		189	.3915344	.4893898	0	1
ptl		189	.1957672	.4933419	0	3
ht		189	.0634921	.2444936	0	1
ui		189	.1481481	.3561903	0	1

Variabel:

- **low** : Berat badan lahir bayi kurang daripada 2500 gram;
1: ya, 0: tidak
- **age** : Usia ibu
- **lwt** : Berat badan ibu pada haid terakhir

- **race** : Ras; 1: *white*, 2: *black*, 3: *other*
- **smoke** : Ibu merokok pada waktu hamil; 1: ya, 0: tidak
- **ptl** : Riwayat kelahiran prematur; skor 0 s.d. 3
- **ht** : Riwayat hipertensi; 1: ya, 0: tidak
- **ui** : Iritabilitas uterus; 1: ada, 0: tidak ada

Keterangan:

File ini memuat data 189 bayi baru lahir dengan berat badan lahir bayi (**low**) sebagai variabel dependen biner, yaitu berat badan lahir rendah (kurang daripada 2500 g; **low** = 1) atau normal (2500 g atau lebih; **low** = 0). Prediktornya adalah usia ibu (**age**), berat badan ibu pada haid terakhir sebelum kehamilan (**lwt**), ras ibu (**race**), kebiasaan merokok ibu (**smoke**), riwayat kelahiran prematur ibu (**ptl**), riwayat hipertensi ibu (**ht**), dan iritabilitas uterus pada ibu (**ui**).

. list in 1/10

	id	low	age	lwt	race	smoke	ptl	ht	ui
1.	85	0	19	182	black	0	0	0	1
2.	86	0	33	155	other	0	0	0	0
3.	87	0	20	105	white	1	0	0	0
4.	88	0	21	108	white	1	0	0	1
5.	89	0	18	107	white	1	0	0	1
6.	91	0	21	124	other	0	0	0	0
7.	92	0	22	118	white	0	0	0	0
8.	93	0	17	103	other	0	0	0	0
9.	94	0	29	123	white	1	0	0	0

```
10. | 95      0      26     113    white      1      0      0      0 |
    +-----+-----+-----+-----+-----+-----+-----+-----+-----+
```

Model:

$$\text{logit } low_i = \beta_0 + \beta_1 age_i + \beta_2 lwt_i + \beta_3 race2_i + \beta_4 race3_i + \beta_5 smoke_i + \beta_6 ptl_i + \beta_7 ht_i + \beta_8 ui_i$$

Perintah Stata:

```
. logit low age lwt i.race smoke ptl ht ui
```

```
Iteration 0: log likelihood =   -117.336
Iteration 1: log likelihood = -101.28644
Iteration 2: log likelihood = -100.72617
Iteration 3: log likelihood =   -100.724
Iteration 4: log likelihood =   -100.724
Logistic regression              Number of obs   =       189
                                LR chi2(8)        =       33.22
                                Prob > chi2        =     0.0001
Log likelihood = -100.724      Pseudo R2        =     0.1416
```

low	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-.0271003	.0364504	-0.74	0.457	-.0985418	.0443412
lwt	-.0151508	.0069259	-2.19	0.029	-.0287253	-.0015763
race						
black	1.262647	.5264101	2.40	0.016	.2309024	2.294392
other	.8620792	.4391532	1.96	0.050	.0013548	1.722804
smoke	.9233448	.4008266	2.30	0.021	.137739	1.708951
ptl	.5418366	.346249	1.56	0.118	-.136799	1.220472
ht	1.832518	.6916292	2.65	0.008	.4769494	3.188086
ui	.7585135	.4593768	1.65	0.099	-.1418484	1.658875
_cons	.4612239	1.20459	0.38	0.702	-1.899729	2.822176

Model estimasi:

$$\text{logit } low_i = 0.461 - 0.027age_i - 0.015lwt_i + 1.263race2_i + \\ 0.862race3_i + 0.923smoke_i + 0.542ptl_i + 1.833ht_i + \\ 0.759ui_i$$

Untuk mendapatkan nilai-nilai rasio odds, digunakan perintah STATA berikut:

. logistic low age lwt i.race smoke ptl ht ui

low	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
age	.9732636	.0354759	-0.74	0.457	.9061578 1.045339
lwt	.9849634	.0068217	-2.19	0.029	.9716834 .9984249
race					
black	3.534767	1.860737	2.40	0.016	1.259736 9.918406
other	2.368079	1.039949	1.96	0.050	1.001356 5.600207
smoke	2.517698	1.00916	2.30	0.021	1.147676 5.523162
ptl	1.719161	.5952579	1.56	0.118	.8721455 3.388787
ht	6.249602	4.322408	2.65	0.008	1.611152 24.24199
ui	2.1351	.9808153	1.65	0.099	.8677528 5.2534
_cons	1.586014	1.910496	0.38	0.702	.1496092 16.8134

Untuk memperoleh hanya nilai-nilai koefisien regresi beserta rasio odds-nya, perintah STATA adalah:

. list coef, help

logit (N=189): Factor Change in Odds

Odds of: 1 vs 0

low	b	z	P> z	e^b	e^bStdX	SDofX
age	-0.02710	-0.743	0.457	0.9733	0.8662	5.2987
lwt	-0.01515	-2.188	0.029	0.9850	0.6292	30.5752
2.race	1.26265	2.399	0.016	3.5348	1.5466	0.3454
3.race	0.86208	1.963	0.050	2.3681	1.5121	0.4796
smoke	0.92334	2.304	0.021	2.5177	1.5713	0.4894
ptl	0.54184	1.565	0.118	1.7192	1.3064	0.4933
ht	1.83252	2.650	0.008	6.2496	1.5652	0.2445
ui	0.75851	1.651	0.099	2.1351	1.3102	0.3562

b = raw coefficient

z = z-score for test of b=0

P>|z| = p-value for z-test

e^b = exp(b) = factor change in odds for unit increase
in X

e^bStdX = exp(b*SD of X) = change in odds for SD increase
in X

SDofX = standard deviation of X

BAB 5

LINEAR MIXED MODELS

Karakteristik Linear Mixed Models

Karakteristik *Linear Mixed Models* (LMM) adalah:

- Data Gaussian, yaitu variabel dependen berskala kontinu dan berdistribusi normal.
- Efek *mixed*, yaitu sebagian parameter bernilai sama untuk seluruh anggota populasi dan sebagian parameter lain bernilai berbeda antar-grup.
- Dua level dan beberapa grup, namun semua variabel diukur pada level mikro, tidak ada yang diukur pada level makro. Data nomor urut grup ada dalam basis data, tetapi bukan merupakan nilai variabel melainkan hanya menyatakan pengenalan untuk masing-masing grup.

Model Umum

Model umum untuk *Linear Mixed Models* adalah:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

A. Hanya konstante bervariasi antar-grup:

Sintaks Stata:

```
. mixed depvar fe_equation || grp_var:
```

depvar : Variabel dependen

fe_equation : Variabel independen dengan efek *fixed*

grp_var : Variabel grup

Contoh 5.1: Berat Badan Lahir Bayi (1)

Data: model-03_nmihs.dta

```
. use "D:\Analisis Multilevel\Data\model-03_nmihs"
```

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
idnum	9,946	1.25e+07	1468507	1.01e+07	1.52e+07
stratan	9,946	3.960487	1.685638	1	6
age	9,946	25.6106	5.784402	15	46
vagbleed	9,946	.056304	.2305195	0	1
miscar	9,946	.1512166	.3582779	0	1
childsex	9,946	1.493766	.4999863	1	2
birthwgt	9,946	2845.094	983.476	227	5810

Variabel:

- **birthwgt** : Berat badan lahir bayi dalam gram
- **stratan** : Indikator grup, bernilai 1 s.d. 6
- **age** : Usia ibu dalam tahun
- **vagbleed** : Riwayat perdarahan vagina; 1=ya, 0=tidak
- **miscar** : Riwayat abortus; 1=ya, 0=tidak
- **childsex** : Jenis kelamin bayi

. tabulate **stratan**

Group indicator 1-6	Freq.	Percent	Cum.
1	841	8.46	8.46
2	803	8.07	16.53
3	3,578	35.97	52.50
4	710	7.14	59.64
5	714	7.18	66.82
6	3,300	33.18	100.00
Total	9,946	100.00	

Keterangan:

File ini memuat data berat badan lahir 9,946 bayi dalam gram (**birthwgt**) sebagai variabel dependen. Prediktornya adalah usia ibu (**age**), riwayat perdarahan vagina pada ibu selama kehamilan (**vagbleed**), riwayat abortus ibu (**miscar**), dan jenis kelamin bayi (**childsex**).

Pengumpulan data dilakukan pada 6 grup (***stratan***) yang tidak dijelaskan dasar pengelompokannya (mungkin saja misalnya berupa 6 lokasi atau 6 RS). Seluruh pengukuran variabel dilakukan pada level bayi dan ibunya. Tidak ada variabel yang diukur pada level di atasnya.

Tampak bahwa hanya ada 6 grup (***stratan***), sedangkan anggota tiap grup berkisar antara 710 s.d. 3,578. Ukuran sampel yang kecil pada level 2 (ada 6 grup) ini akan menghasilkan kekuatan uji dan presisi yang relatif rendah.

. list in 1/10

	idnum	stratan	age	vagbleed	miscar	childsex	birthwgt
1.	10600699	1	20	no bleed	nomiscar	1	1304
2.	11901902	1	17	no bleed	nomiscar	2	1474
3.	14507256	1	15	no bleed	nomiscar	1	950
4.	14202452	1	21	no bleed	nomiscar	2	1010
5.	14001909	1	22	no bleed	nomiscar	2	822
6.	10301232	1	22	no bleed	nomiscar	2	1040
7.	12103232	1	24	no bleed	nomiscar	2	1134
8.	13405707	1	23	no bleed	miscar	2	1389
9.	10900972	1	23	bleed	nomiscar	2	680
10.	14500218	1	22	no bleed	nomiscar	2	964

Model:

$$\mathbf{birthwgt}_{ij} = \beta_0 + \beta_1 \mathbf{age}_{ij} + \beta_2 \mathbf{vagbleed}_{ij} + \beta_3 \mathbf{miscar}_{ij} + \beta_4 \mathbf{childsex}_{ij} + u_j + \varepsilon_{ij}$$

Perintah Stata:

```
. mixed birthwgt age vagbleed miscar childsex || stratan:
```

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0: log likelihood = -74358.768

Iteration 1: log likelihood = -74358.768 (backed up)

Computing standard errors:

Mixed-effects ML regression	Number of obs	=	9,946
Group variable: stratan	Number of groups	=	6

Obs per group:

min	=	710
avg	=	1,657.7
max	=	3,578

Log likelihood = -74358.768	Wald chi2(4)	=	147.85
	Prob > chi2	=	0.0000

birthwgt	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	5.942812	.7655304	7.76	0.000	4.4424	7.443224
vagbleed	-68.74908	18.85292	-3.65	0.000	-105.7001	-31.79803
miscar	-15.88161	12.17143	-1.30	0.192	-39.73719	7.973961
childsex	-73.62037	8.556032	-8.60	0.000	-90.38988	-56.85086
_cons	2180.776	392.0378	5.56	0.000	1412.396	2949.156

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
stratan: Identity				
var(_cons)	918679.4	530454.9	296258.9	2848764
var(Residual)	181558.7	2575.365	176580.6	186677.1

LR test vs. linear model: $\text{chibar2}(01) = 16276.01$

Prob >= $\text{chibar2} = 0.0000$

Model estimasi:

$$\mathbf{birthwgt}_{ij} = 2180.776 + 5.943\mathbf{age}_{ij} - 68.749\mathbf{vagbleed}_{ij} - 15.882\mathbf{miscar}_{ij} - 73.620\mathbf{childsex}_{ij} + \mathbf{u}_j + \varepsilon_{ij}$$

atau:

$$\mathbf{birthwgt}_{ij} = [2180.776 + \mathbf{u}_j] + 5.943 \mathbf{age}_{ij} - 68.749 \mathbf{vagbleed}_{ij} - 15.882 \mathbf{miscar}_{ij} - 73.620 \mathbf{childsex}_{ij} + \varepsilon_{ij}$$

Perhatikan: Adanya suku \mathbf{u}_j menyebabkan nilai konstante bervariasi antar-grup.

```
. predict u0, reffects
```

```
. list idnum stratan birthwgt u0 in 1/10
```

```
+-----+
|      idnum      stratan      birthwgt      u0 |
+-----+
1. | 10600699          1         1304    -1157.001 |
2. | 11901902          1         1474    -1157.001 |
3. | 14507256          1          950    -1157.001 |
4. | 14202452          1         1010    -1157.001 |
5. | 14001909          1          822    -1157.001 |
+-----+
6. | 10301232          1         1040    -1157.001 |
7. | 12103232          1         1134    -1157.001 |
8. | 13405707          1         1389    -1157.001 |
9. | 10900972          1          680    -1157.001 |
10. | 14500218          1          964    -1157.001 |
+-----+
```

Nilai-nilai prediksi tidak ditampilkan pada jendela hasil,
tetapi langsung muncul pada basis-data.

```
. tabstat u0, by(stratan)
```

```
Summary for variables: u0
by categories of: stratan (Group indicator 1-6)
```

```
stratan |      mean
+-----+
1 | -1157.001
2 | -16.89059
3 | 1092.853
4 | -1176.904
5 | -4.047667
6 | 1261.99
+-----+
```


Misalnya:

- Untuk **stratan** = 1, model estimasi adalah:

$$\begin{aligned}
 \mathbf{birthwgt}_{ij} &= [2180.776 + \mathbf{u}_j] + 5.943 \mathbf{age}_{ij} - 68.749 \mathbf{vagbleed}_{ij} \\
 &\quad - 15.882 \mathbf{miscar}_{ij} - 73.620 \mathbf{childsex}_{ij} + \varepsilon_{ij} \\
 &= [2180.776 - 1157.001] + 5.943 \mathbf{age}_{ij} \\
 &\quad - 68.749 \mathbf{vagbleed}_{ij} - 15.882 \mathbf{miscar}_{ij} \\
 &\quad - 73.620 \mathbf{childsex}_{ij} + \varepsilon_{ij} \\
 &= 1023.775 + 5.943 \mathbf{age}_{ij} - 68.749 \mathbf{vagbleed}_{ij} \\
 &\quad - 15.882 \mathbf{miscar}_{ij} - 73.620 \mathbf{childsex}_{ij} + \varepsilon_{ij}
 \end{aligned}$$

- Untuk **stratan** = 2, model estimasi adalah:

$$\begin{aligned}
 \mathbf{birthwgt}_{ij} &= [2180.776 - 16.891] + 5.943 \mathbf{age}_{ij} \\
 &\quad - 68.749 \mathbf{vagbleed}_{ij} - 15.882 \mathbf{miscar}_{ij} \\
 &\quad - 73.620 \mathbf{childsex}_{ij} + \varepsilon_{ij} \\
 &= 2163.885 + 5.943 \mathbf{age}_{ij} - 68.749 \mathbf{vagbleed}_{ij} \\
 &\quad - 15.882 \mathbf{miscar}_{ij} - 73.620 \mathbf{childsex}_{ij} + \varepsilon_{ij}
 \end{aligned}$$

dan seterusnya.

B. Konstante dan koefisien regresi *age* dan *childsex* bervariasi antar-grup:

Sintaks Stata:

```
. mixed depvar fe_equation || grp_var: re_equation
```

depvar : Variabel dependen

fe_equation : Variabel independen dengan efek *fixed*

re_equation : Variabel independen dengan efek random

grp_var : Variabel grup

Contoh 5.2: Berat Badan Lahir Bayi (2)

Data: model-03_nmihs.dta

```
. use "D:\Analisis Multilevel\Data\model-03_nmihs, clear"
```

Model:

$$\begin{aligned} birthwgt_{ij} = & \beta_0 + \beta_1 age_{ij} + \beta_2 vagbleed_{ij} + \beta_3 miscar_{ij} \\ & + \beta_4 childsex_{ij} + u_{0j} + u_{1j}age_{ij} + u_{2j}childsex_{ij} + \varepsilon_{ij} \end{aligned}$$

Perintah Stata:

```
. mixed birthwgt age vagbleed miscar childsex || stratan:  
age childsex
```

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0: log likelihood = -74335.406

Iteration 1: log likelihood = -74335.406

Computing standard errors:

Mixed-effects ML regression	Number of obs	= 9,946
Group variable: stratan	Number of groups	= 6

Obs per group:	
min =	710
avg =	1,657.7
max =	3,578

Log likelihood = -74335.406	Wald chi2(4) =	22.70
	Prob > chi2 =	0.0001

birthwgt	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	4.084926	1.58981	2.57	0.010	.9689549	7.200897
vagbleed	-68.38018	18.80431	-3.64	0.000	-105.2359	-31.52441
miscar	-15.5745	12.13566	-1.28	0.199	-39.35996	8.210956
childsex	-30.81131	28.15984	-1.09	0.274	-86.00359	24.38098
_cons	2162.722	396.8262	5.45	0.000	1384.957	2940.487

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
stratan: Independent				
var(age)	10.37326	8.145393	2.226006	48.33972
var(childsex)	4081.625	2651.96	1142.302	14584.3
var(_cons)	939908.2	544915.6	301715.7	2928013
var(Residual)	180335.3	2559.364	175388.1	185422

LR test vs. linear model: $\chi^2(3) = 16322.73$
 Prob > $\chi^2 = 0.0000$

Note: LR test is conservative and provided only for reference.

Model estimasi:

$$\begin{aligned} \text{birthwgt}_{ij} = & 2162.722 + 4.085\text{age}_{ij} - 68.380\text{vagbleed}_{ij} \\ & - 15.575\text{miscar}_{ij} - 30.811\text{childsex}_{ij} + u_{0j} \\ & + u_{1j}\text{age}_{ij} + u_{2j}\text{childsex}_{ij} + \varepsilon_{ij} \end{aligned}$$

atau:

$$\begin{aligned} \text{birthwgt}_{ij} = & [2162.722 + u_{0j}] + [4.085 + u_{1j}] \text{age}_{ij} \\ & - 68.380\text{vagbleed}_{ij} - 15.575\text{miscar}_{ij} \\ & - [30.811 + u_{2j}] \text{childsex}_{ij} + \varepsilon_{ij} \end{aligned}$$

Untuk memperoleh nilai-nilai u_{0j} , u_{1j} , dan u_{2j} bagi tiap grup (**stratan**), digunakan perintah Stata berikut:

. predict u0 u1 u2, reffects

. list idnum stratan birthwgt u0 u1 u2 in 1/10

	idnum	stratan	birthwgt	u0	u1	u2
1.	10600699	1	1304	-1.211372	14.1959	-1148.379
2.	11901902	1	1474	-1.211372	14.1959	-1148.379
3.	14507256	1	950	-1.211372	14.1959	-1148.379
4.	14202452	1	1010	-1.211372	14.1959	-1148.379
5.	14001909	1	822	-1.211372	14.1959	-1148.379

6.		10301232	1	1040	-1.211372	14.1959	-1148.379	
7.		12103232	1	1134	-1.211372	14.1959	-1148.379	
8.		13405707	1	1389	-1.211372	14.1959	-1148.379	
9.		10900972	1	680	-1.211372	14.1959	-1148.379	
10.		14500218	1	964	-1.211372	14.1959	-1148.379	

+-----+

. tabstat u0 u1 u2, by(stratan)

Summary statistics: mean
by categories of: stratan (Group indicator 1-6)

stratan		u0	u1	u2
+-----+				
1		-1.211372	14.1959	-1148.379
2		-1.893447	75.80069	-91.11327
3		3.992543	-84.04402	1120.231
4		-1.917933	37.03446	-1178.019
5		-2.455102	31.92701	13.9417
6		3.48531	-74.91404	1283.339
+-----+				
Total		2.024227	-42.83418	641.2442
+-----+				

Misalnya:

- **stratan = 1**

$$\begin{aligned}
 \text{birthwgt}_{ij} = & [2162.722 + u_{0j}] + [4.085 + u_{1j}] \text{ age}_{ij} \\
 & - 68.380 \text{vagbleed}_{ij} - 15.575 \text{miscar}_{ij} \\
 & - [30.811 + u_{2j}] \text{childsex}_{ij} + \varepsilon_{ij}
 \end{aligned}$$

$$\begin{aligned}
 \text{birthwgt}_{ij} = & [2162.722 - 1.211] + [4.085 + 14.196] \text{ age}_{ij} \\
 & - 68.380 \text{vagbleed}_{ij} - 15.575 \text{miscar}_{ij} \\
 & - [30.811 - 1148.379] \text{childsex}_{ij} + \varepsilon_{ij}
 \end{aligned}$$

$$= 2161.511 + 18.281 \text{ *age*_{ij}} - 68.380 \text{ *vagbleed*_{ij}} \\ - 15.575 \text{ *miscar*_{ij}} + 1117.568 \text{ *childsex*_{ij}} + \varepsilon_{ij}$$

- stratan = 2

$$\text{*birthwgt*_{ij}} = [2162.722 - 1.893] + [4.085 + 75.801] \text{ *age*_{ij}} \\ - 68.380 \text{ *vagbleed*_{ij}} - 15.575 \text{ *miscar*_{ij}} \\ - [30.811 - 91.113] \text{ *childsex*_{ij}} + \varepsilon_{ij} \\ = 2160.829 + 79.886 \text{ *age*_{ij}} - 68.380 \text{ *vagbleed*_{ij}} \\ - 15.575 \text{ *miscar*_{ij}} + 60.302 \text{ *childsex*_{ij}} + \varepsilon_{ij}$$

dan seterusnya.

BAB 6

GENERALIZED LINEAR MIXED MODELS

Karakteristik Generalized Linear Mixed Models

Karakteristik *Generalized Linear Mixed Models* (GLMM) adalah:

- Data non-Gaussian, yaitu variabel dependen tidak berdistribusi normal, bahkan tidak berskala kontinu. Variabel dependen dapat berskala biner, kategorik nominal, kategorik ordinal, atau data cacah. Di sini hanya akan dibahas GLMM dengan variabel dependen berskala biner.
- Efek *mixed*, yaitu sebagian parameter bernilai sama untuk seluruh anggota populasi dan sebagian parameter lain bernilai berbeda antar-grup.
- Dua level dan beberapa grup, namun pengumpulan data hanya dilakukan pada level mikro, tidak ada variabel yang dikumpulkan datanya pada level makro.

Model Umum

Model umum untuk *Generalized Linear Mixed Model* adalah:

$$\eta = X\beta + Zu \quad (6.1)$$

A. Hanya konstante bervariasi antar-grup:

Sintaks Stata (model logit):

```
. melogit depvar fe_equation || grp_var:
```

depvar : Variabel dependen

fe_equation : Variabel independen dengan efek *fixed*

grp_var : Variabel grup

Contoh 6.1: Penggunaan Kontrasepsi (1)

Data: model-04_bangladesh.dta

```
. use "D:\Analisis Multilevel\Data\model-04_bangladesh"
```

(Bangladesh Fertility Survey, 1989)

. summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
district	1,934	29.35367	17.95983	1	61
c_use	1,934	.3924509	.4884225	0	1
urban	1,934	.2905895	.4541518	0	1
age	1,934	.0020662	9.013392	-13.5599	19.44
child1	1,934	.1830403	.3867996	0	1
child2	1,934	.1587384	.3655264	0	1
child3	1,934	.3841779	.4865261	0	1

Variabel:

- **district** : Distrik domisili
- **c_use** : Menggunakan kontrasepsi; 1 = ya, 0 = tidak
- **urban** : Penduduk urban (perkotaan) atau rural (pedesaan);
1 = urban, 0 = rural
- **age** : Usia, dihitung terhadap nilai rerata
- **child1** : Memiliki 1 anak
- **child2** : Memiliki 2 anak
- **child3** : Memiliki 3 anak atau lebih

Keterangan:

File ini muat data tentang 1,934 ibu usia subur di sejumlah distrik di negara Bangladesh. Variabel dependen adalah penggunaan kontrasepsi (**c_use**), **c_use** = 1 jika ibu menggunakan kontrasepsi dan **c_use** = 0 jika tidak. Prediktornya

adalah daerah domisili ibu (**urban**), usia ibu yang dihitung terhadap nilai rerata (**age**), dan jumlah anak (**child***); 1, atau 2; atau 3 atau lebih.

Grup adalah distrik domisili ibu (**district**). Seluruh data diperoleh dari level responden, tidak ada yang diukur pada level distrik. Pengukuran data dilakukan pada 61 grup (**district**), jumlah grup ini cukup memadai untuk memperoleh kekuatan uji dan presisi yang tinggi.

. list in 1/10

	district	c_use	urban	age	child1	child2	child3
1.	1	no	urban	18.44	0	0	1
2.	1	no	urban	-5.56	0	0	0
3.	1	no	urban	1.44	0	1	0
4.	1	no	urban	8.44	0	0	1
5.	1	no	urban	-13.56	0	0	0
6.	1	no	urban	-11.56	0	0	0
7.	1	no	urban	18.44	0	0	1
8.	1	no	urban	-3.56	0	0	1
9.	1	no	urban	-5.56	1	0	0
10.	1	no	urban	1.44	0	0	1

Model:

$$\text{logit } c_use_{ij} = \beta_0 + \beta_1 \text{urban}_{ij} + \beta_2 \text{age}_{ij} + \beta_3 \text{child1}_{ij} \\ + \beta_4 \text{child2}_{ij} + \beta_5 \text{child3}_{ij} + u_j$$

Perintah Stata:

. melogit *c_use urban age child || *district*:**

Fitting fixed-effects model:

Iteration 0: log likelihood = -1229.5485
Iteration 1: log likelihood = -1228.5268
Iteration 2: log likelihood = -1228.5263
Iteration 3: log likelihood = -1228.5263
Refining starting values:

Grid node 0: log likelihood = -1219.2681

Fitting full model:

Iteration 0: log likelihood = -1219.2681 (not concave)
Iteration 1: log likelihood = -1207.5978
Iteration 2: log likelihood = -1206.8428
Iteration 3: log likelihood = -1206.8322
Iteration 4: log likelihood = -1206.8322

Mixed-effects logistic regression	Number of obs	= 1,934
Group variable: district	Number of groups	= 60

Obs per group:

min	=	2
avg	=	32.2
max	=	118

Integration method: mvaghermite	Integration pts.	= 7
	Wald chi2(5)	= 109.60

Log likelihood = -1206.8322 Prob > chi2 = 0.0000

c_use	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
urban	.7322765	.1194857	6.13	0.000	.4980888	.9664641
age	-.0264981	.0078916	-3.36	0.001	-.0419654	-.0110309
child1	1.116001	.1580921	7.06	0.000	.8061465	1.425856
child2	1.365895	.1746691	7.82	0.000	1.02355	1.70824
child3	1.344031	.1796549	7.48	0.000	.9919139	1.696148
_cons	-1.68929	.1477591	-11.43	0.000	-1.978892	-1.399687
<hr/>						
district						
var(_cons)	.215618	.0733222			.1107208	.4198954

LR test vs. logistic model: chibar2(01) = 43.39

Prob >= chibar2 = 0.0000

Model estimasi:

$$\begin{aligned} \text{logit } c_use_{ij} = & -1.689 + 0.732urban_{ij} - 0.026age_{ij} \\ & + 1.116child1_{ij} + 1.366child2_{ij} + 1.344child3_{ij} \\ & + u_j \end{aligned}$$

atau:

$$\begin{aligned} \text{logit } c_use_{ij} = & [-1.689 + u_j] + 0.732urban_{ij} - 0.026age_{ij} \\ & + 1.116child1_{ij} + 1.366child2_{ij} + 1.344child3_{ij} \end{aligned}$$

Untuk mendapatkan nilai-nilai rasio odds, digunakan perintah sebagai berikut:

. melogit *c_use urban age child || *district:* , or**

c_use	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
urban	2.07981	.2485075	6.13	0.000	1.645573	2.628633
age	.9738498	.0076852	-3.36	0.001	.958903	.9890297
child1	3.052624	.4825958	7.06	0.000	2.239262	4.16142
child2	3.919229	.6845681	7.82	0.000	2.783057	5.519239
child3	3.834469	.6888813	7.48	0.000	2.69639	5.452903
_cons	.1846507	.0272838	-11.43	0.000	.1382223	.2466742
district						
var(_cons)	.215618	.0733222			.1107208	.4198954

LR test vs. logistic model: $\chi^2(01) = 43.39$

Prob >= $\chi^2 = 0.0000$

Untuk mendapatkan nilai konstante bagi tiap distrik:

. predict u0, reffects

(calculating posterior means of random effects)
(using 7 quadrature points)

. list district c_use u0 in 1/10

	district	c_use	u0
1.	1	no	-.7281059
2.	1	no	-.7281059
3.	1	no	-.7281059
4.	1	no	-.7281059
5.	1	no	-.7281059
6.	1	no	-.7281059
7.	1	no	-.7281059
8.	1	no	-.7281059
9.	1	no	-.7281059
10.	1	no	-.7281059

Selanjutnya hanya akan diperlihatkan nilai **u0** untuk 5 distrik pertama (ada 61 distrik):

. tabstat u0 if district<=5, by(district)

Summary for variables: u0

by categories of: district (District)

district	mean
1	-.7281059
2	-.0409355
3	.2077022
4	.1873026
5	.051294
Total	-.3748661

Untuk **district** = 1:

$$\text{logit } c_use_{ij} = [-1.689 + u_j] + 0.732urban_{ij} - 0.026age_{ij} \\ + 1.116child1_{ij} + 1.366child2_{ij} + 1.344child3_{ij}$$

$$\text{logit } c_use_{ij} = [-1.689 - 0.728] + 0.732urban_{ij} - 0.026age_{ij} \\ + 1.116child1_{ij} + 1.366child2_{ij} + 1.344child3_{ij}$$

$$\text{logit } c_use_{ij} = -2.417 + 0.732urban_{ij} - 0.026age_{ij} \\ + 1.116child1_{ij} + 1.366child2_{ij} + 1.344child3_{ij}$$

Untuk **district** = 2:

$$\text{logit } c_use_{ij} = [-1.689 - 0.041] + 0.732urban_{ij} - 0.026age_{ij} \\ + 1.116child1_{ij} + 1.366child2_{ij} + 1.344child3_{ij}$$

$$\text{logit } c_use_{ij} = -1.730 + 0.732urban_{ij} - 0.026age_{ij} \\ + 1.116child1_{ij} + 1.366child2_{ij} + 1.344child3_{ij}$$

dan seterusnya.

B. Konstante dan koefisien regresi *age* bervariasi antar-grup:

Sintaks Stata (model logit):

```
. melogit depvar fe_equation || grp_var: re_equation
```

depvar : Variabel dependen

fe_equation : Variabel independen dengan efek *fixed*

re_equation : Variabel independen dengan efek random

grp_var : Variabel grup

Contoh 6.2: Penggunaan Kontrasepsi (2)

Data: model-04_bangladesh.dta

```
. use "D:\Analisis Multilevel\Data\model-04_bangladesh,  
clear"
```

(Bangladesh Fertility Survey, 1989)

Model:

$$\begin{aligned}\text{logit } c_use_{ij} = & \beta_0 + \beta_1 \text{urban}_{ij} + \beta_2 \text{age}_{ij} + \beta_3 \text{child1}_{ij} \\ & + \beta_4 \text{child2}_{ij} + \beta_5 \text{child3}_{ij} + u_{0j} + u_{1j} \text{age}_{ij}\end{aligned}$$

Perintah Stata:

```
. melogit c_use urban age child* || district: age
```

Fitting fixed-effects model:

Iteration 0: log likelihood = -1229.5485

Iteration 1: log likelihood = -1228.5268

Iteration 2: log likelihood = -1228.5263

Iteration 3: log likelihood = -1228.5263

Refining starting values:

Grid node 0: log likelihood = -1353.6948

Fitting full model:

Iteration 0: log likelihood = -1353.6948 (not concave)

Iteration 1: log likelihood = -1342.0158 (not concave)

Iteration 2: log likelihood = -1258.2688 (not concave)

Iteration 3: log likelihood = -1234.8111 (not concave)

Iteration 4: log likelihood = -1207.4924

Iteration 5: log likelihood = -1206.5178
 Iteration 6: log likelihood = -1206.5084
 Iteration 7: log likelihood = -1206.5083

Mixed-effects logistic regression Number of obs = 1,934
 Group variable: district Number of groups = 60

Obs per group:
 min = 2
 avg = 32.2
 max = 118

Integration method: mvaghermite Integration pts. = 7

Log likelihood = -1206.5083 Wald chi2(5) = 108.97
 Prob > chi2 = 0.0000

c_use	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
urban	.7376083	.1202172	6.14	0.000	.5019868	.9732297
age	-.0266392	.0082769	-3.22	0.001	-.0428617	-.0104168
child1	1.125424	.1591084	7.07	0.000	.8135772	1.437271
child2	1.371254	.1753508	7.82	0.000	1.027573	1.714936
child3	1.350024	.1803648	7.48	0.000	.9965158	1.703533
_cons	-1.697742	.1489046	-11.40	0.000	-1.989589	-1.405894
-----+-----						
district						
var(age)	.0002463	.0003492			.0000153	.0039661
var(_cons)	.2183637	.0742389			.1121477	.4251778

LR test vs. logistic model: chi2(2) = 44.04
 Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

Model estimasi:

$$\begin{aligned}\text{logit } c_use_{ij} = & -1.698 + 0.738urban_{ij} - 0.027age_{ij} \\ & + 1.125child1_{ij} + 1.371child2_{ij} + 1.350child3_{ij} \\ & + u_{0j} + u_{1j}age_{ij}\end{aligned}$$

atau:

$$\begin{aligned}\text{logit } c_use_{ij} = & [-1.698 + u_{0j}] + 0.738urban_{ij} \\ & + [-0.027 + u_{1j}]age_{ij} + 1.125child1_{ij} + \\ & 1.371child2_{ij} + 1.350child3_{ij}\end{aligned}$$

Untuk mendapatkan nilai-nilai rasio odds, digunakan perintah STATA:

. melogit *c_use urban age child || *district: age, or***

c_use	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
urban	2.090929	.2513656	6.14	0.000	1.652	2.646478
age	.9737125	.0080593	-3.22	0.001	.9580439	.9896373
child1	3.081523	.4902963	7.07	0.000	2.255964	4.209192
child2	3.94029	.6909332	7.82	0.000	2.794276	5.556318
child3	3.857519	.6957607	7.48	0.000	2.708827	5.49332
_cons	.1830965	.0272639	-11.40	0.000	.1367516	.2451478
district						
var(age)	.0002463	.0003492			.0000153	.0039661
var(_cons)	.2183637	.0742389			.1121477	.4251778

LR test vs. logistic model: $\chi^2(2) = 44.04$

Prob > $\chi^2 = 0.0000$

Untuk mendapatkan nilai-nilai koefisien regresi pada tiap distrik:

. predict u0 u1, reffects

(calculating posterior means of random effects)
(using 7 quadrature points)

. list district c_use u0 u1 in 1/10

	district	c_use	u0	u1
1.	1	no	-.0039822	-.7307618
2.	1	no	-.0039822	-.7307618
3.	1	no	-.0039822	-.7307618
4.	1	no	-.0039822	-.7307618
5.	1	no	-.0039822	-.7307618
6.	1	no	-.0039822	-.7307618
7.	1	no	-.0039822	-.7307618
8.	1	no	-.0039822	-.7307618
9.	1	no	-.0039822	-.7307618
10.	1	no	-.0039822	-.7307618

Di sini hanya akan disajikan nilai-nilai **u0** dan **u1** untuk 5 distrik pertama dari keseluruhan 61 distrik:

. tabstat u0 u1 if district<=5, by(district)

Summary statistics: mean
by categories of: district (District)

district		u0	u1
1		-.0039822	-.7307618
2		.0026013	-.0394008
3		-.0016016	.2091366
4		-.0001785	.188184
5		-.0057079	.0501268
Total		-.0031012	-.3762904

Untuk **district = 1**:

$$\begin{aligned} \text{logit } c_use_{ij} = & [-1.698 + u_{0j}] + 0.738urban_{ij} \\ & + [-0.027 + u_{1j}]age_{ij} + 1.125child1_{ij} \\ & + 1.371child2_{ij} + 1.350child3_{ij} \end{aligned}$$

$$\begin{aligned} \text{logit } c_use_{ij} = & [-1.698 - 0.004] + 0.738urban_{ij} \\ & + [-0.027 - 0.731]age_{ij} + 1.125child1_{ij} \\ & + 1.371child2_{ij} + 1.350child3_{ij} \end{aligned}$$

$$\begin{aligned} \text{logit } c_use_{ij} = & -1.702 + 0.738urban_{ij} - 0.758age_{ij} \\ & + 1.125child1_{ij} + 1.371child2_{ij} + 1.350child3_{ij} \end{aligned}$$

Untuk **district** = 2:

$$\begin{aligned}\text{logit } c_use_{ij} = & [-1.698 - 0.003] + 0.738urban_{ij} \\ & + [-0.027 - 0.039]age_{ij} + 1.125child1_{ij} \\ & + 1.371child2_{ij} + 1.350child3_{ij}\end{aligned}$$

$$\begin{aligned}\text{logit } c_use_{ij} = & -1.701 + 0.738urban_{ij} - 0.066age_{ij} + \\ & 1.125child1_{ij} + 1.371child2_{ij} + 1.350child3_{ij}\end{aligned}$$

dan seterusnya.

BAB 7

MULTILEVEL LINEAR MIXED MODELS

Karakteristik Multilevel Linear Mixed Models

Karakteristik *Multilevel Linear Mixed Models* (Multilevel LMM) adalah:

- Data Gaussian, yaitu variabel dependen berskala kontinu dan berdistribusi normal.
- Efek *mixed*, yaitu sebagian parameter bernilai sama untuk seluruh anggota populasi dan sebagian parameter lain bernilai berbeda antar-grup.
- Multi-level, yaitu pengumpulan data pada lebih daripada satu level.

Model Umum

Model umum untuk *Multilevel Linear Mixed Model* adalah:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \quad (7.1)$$

A. Hanya konstante bervariasi antar-grup:

Sintaks Stata:

```
. mixed depvar fe_equation || macr_level:
```

depvar : Variabel dependen

fe_equation : Variabel independen dengan efek *fixed*

macr_level : Level makro

Contoh 7.1: Popularitas Siswa (1)

Data: model-05_pop1.dta

```
. use "D:\Analisis Multilevel\Data\model-05_pop1"
```

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
pupil	2,000	10.649	5.968217	1	26
school	2,000	50.3695	29.07782	1	100
extrav	2,000	5.215	1.262368	1	10
texp	2,000	14.263	6.551816	2	25
popular	2,000	5.308	1.225923	2	9
sex	2,000	.487	.499956	0	1

Variabel:

- **pupil** : Nomor urut siswa
- **school** : Nomor sekolah, berfungsi sebagai grup
- **extrav** : Sifat extrovert siswa; dinyatakan dalam kategori 1 s.d. 10
- **texp** : Pengalaman mengajar guru dalam tahun, untuk tiap sekolah hanya diambil 1 guru (yang mengajar siswa)
- **popular** : Skala popularitas siswa, dinyatakan dalam kategori 2 s.d. 9 (8 kategori)
- **sex** : Jenis kelamin siswa; 1 = wanita, 0 = pria

Keterangan:

File ini memuat data 2000 orang siswa (**pupil**) yang berasal dari 100 sekolah (**school**), yang berfungsi sebagai grup. Tiap sekolah dapat diwakili oleh 1 s.d. 26 orang siswa. Semua siswa dari 1 sekolah diajar oleh 1 orang guru, sehingga tiap sekolah hanya diwakili oleh 1 orang guru. Variabel dependen adalah popularitas siswa menurut penilaian guru (**popular**), yang dinyatakan dalam skala dengan 8 kategori. Karena jumlah kategori cukup banyak, variabel dependen dapat dianggap sebagai variabel numerik yang diasumsikan berdistribusi normal (data Gaussian).

Prediktor adalah sifat ekstrovert siswa (**extrav**), jenis kelamin siswa (**sex**), dan pengalaman (lama) mengajar guru (**texp**). Di sini terdapat 2 level, level siswa dan level guru/sekolah. Data siswa diperoleh dari level siswa dan data guru diperoleh dari level guru/sekolah.

Di sini terdapat 100 grup (**school**), sehingga ukuran sampel bagi jumlah grup mencukupi untuk mendapatkan kekuatan uji dan presisi yang memadai.

. list in 1/10

	pupil	school	extrav	texp	popular	sex
1.	1	1	5	24	8	1
2.	2	1	7	24	7	0
3.	3	1	4	24	7	1
4.	4	1	3	24	9	1
5.	5	1	5	24	8	1
6.	6	1	4	24	7	0
7.	7	1	5	24	7	0
8.	8	1	4	24	7	0
9.	9	1	5	24	7	0
10.	10	1	5	24	8	0

Model:

$$popular_{ij} = \beta_0 + \beta_1 extrav_{ij} + \beta_2 sex_{ij} + \beta_3 texp_{ij} + u_j + \varepsilon_{ij}$$

Perintah Stata:

. mixed *popular extrav sex texp* || *school*:

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0: log likelihood = -2206.0551

Iteration 1: log likelihood = -2206.0551

Computing standard errors:

Mixed-effects ML regression	Number of obs	=	2000
Group variable: school	Number of groups	=	100

Obs per group: min	=	16
avg	=	20.0
max	=	26

	Wald chi2(3)	=	844.83
Log likelihood = -2206.0551	Prob > chi2	=	0.0000

popular	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
extrav	.057945	.014245	4.07	0.000	.0300252	.0858648
sex	.8347056	.0309263	26.99	0.000	.7740911	.8953201
texp	.097754	.0106657	9.17	0.000	.0768496	.1186584
_cons	3.201788	.1894814	16.90	0.000	2.830411	3.573165

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
school: Identity				
var(_cons)	.463758	.0689343	.3465504	.6206066
var(Residual)	.4563093	.0148056	.4281943	.4862704

LR test vs. linear regression: $\text{chibar2}(01) = 1077.71$

Prob >= $\text{chibar2} = 0.0000$

Model estimasi:

$$\text{popular}_{ij} = 3.202 + 0.058\text{extrav}_{ij} + 0.835\text{sex}_{ij} + 0.098\text{texp}_{ij} + u_j + \varepsilon_{ij}$$

atau:

$$\text{popular}_{ij} = [3.202 + u_j] + 0.058\text{extrav}_{ij} + 0.835\text{sex}_{ij} + 0.098\text{texp}_{ij} + \varepsilon_{ij}$$

Untuk memperoleh nilai-nilai **u0** dan intersep pada tiap sekolah, perintah STATA adalah:

. predict u0, reffects

. list pupil school u0 in 1/10

	pupil	school	u0
1.	1	1	1.242597
2.	2	1	1.242597
3.	3	1	1.242597
4.	4	1	1.242597
5.	5	1	1.242597
6.	6	1	1.242597
7.	7	1	1.242597
8.	8	1	1.242597
9.	9	1	1.242597
10.	10	1	1.242597

Karena ada 100 sekolah, hanya diperlihatkan nilai **u0** untuk 5 sekolah pertama:

. tabstat u0 if school<=5, by(school)

Summary for variables: u0
by categories of: school (school)

school	mean
1	1.242597
2	-1.113929
3	1.423839
4	.5631154
5	.183042
Total	.4411571

Untuk **school** = 1:

$$\begin{aligned} \text{popular}_{ij} &= [3.202 + u_j] + 0.058\text{extrav}_{ij} + 0.835\text{sex}_{ij} \\ &\quad + 0.098\text{texp}_{ij} + \varepsilon_{ij} \end{aligned}$$

$$\begin{aligned} \text{popular}_{ij} &= [3.202 + 1.243] + 0.058\text{extrav}_{ij} + 0.835\text{sex}_{ij} \\ &\quad + 0.098\text{texp}_{ij} + \varepsilon_{ij} \end{aligned}$$

$$\begin{aligned} \text{popular}_{ij} &= 4.445 + 0.058\text{extrav}_{ij} + 0.835\text{sex}_{ij} + 0.098\text{texp}_{ij} \\ &\quad + \varepsilon_{ij} \end{aligned}$$

Untuk **school** = 2:

$$\begin{aligned} \text{popular}_{ij} &= [3.202 - 1.114] + 0.058\text{extrav}_{ij} + 0.835\text{sex}_{ij} \\ &\quad + 0.098\text{texp}_{ij} + \varepsilon_{ij} \end{aligned}$$

$$\begin{aligned} \text{popular}_{ij} &= 2.088 + 0.058\text{extrav}_{ij} + 0.835\text{sex}_{ij} + 0.098\text{texp}_{ij} \\ &\quad + \varepsilon_{ij} \end{aligned}$$

B. Konstante dan koefisien regresi *texp* bervariasi antar-grup:

Sintaks Stata:

```
. mixed depvar fe_equation || macr_level: re_equation
```

depvar : Variabel dependen

fe_equation : Variabel independen dengan efek *fixed*

re_equation : Variabel independen dengan efek random

macr_level : Level makro

Contoh 7.2: Popularitas Siswa (2)

Data: model-05_pop1.dta

. use “D:\Analisis Multilevel\Data\model-05_pop1, clear”

Model:

$$\begin{aligned} popular_{ij} = & \beta_0 + \beta_1 extrav_{ij} + \beta_2 sex_{ij} + \beta_3 texp_{ij} + u_{0j} \\ & + u_{1j} texp_{ij} + \varepsilon_{ij} \end{aligned}$$

Perintah Stata:

. mixed popular extrav sex texp || school: texp

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0: log likelihood = -2205.0802

Iteration 1: log likelihood = -2205.0362

Iteration 2: log likelihood = -2205.0362

Computing standard errors:

Mixed-effects ML regression	Number of obs	=	2000
Group variable: school	Number of groups	=	100
	Obs per group: min	=	16
	avg	=	20.0
	max	=	26
	Wald chi2(3)	=	843.20
Log likelihood = -2205.0362	Prob > chi2	=	0.0000

popular	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
extrav	.0583171	.0142322	4.10	0.000	.0304225 .0862117
sex	.8345927	.0309248	26.99	0.000	.7739812 .8952041
texp	.0964077	.0105727	9.12	0.000	.0756855 .1171298
_cons	3.217236	.1781405	18.06	0.000	2.868087 3.566385

Random-effects Parameters | Estimate Std. Err. [95% Conf. Interval]

school: Independent	
var(texp)	.0005188 .0003875 .00012 .0022424
var(_cons)	.3346841 .0955196 .1912935 .585558
var(Residual)	.4563108 .0148057 .4281956 .486272

LR test vs. linear regression: chi2(2) = 1079.75

Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

Model estimasi:

$$\mathbf{popular}_{ij} = 3.217 + 0.058\mathbf{extrav}_{ij} + 0.835\mathbf{sex}_{ij} + 0.096\mathbf{texp}_{ij} + \mathbf{u}_{0j} + \mathbf{u}_{1j}\mathbf{texp}_{ij} + \varepsilon_{ij}$$

atau:

$$\mathbf{popular}_{ij} = [3.217 + \mathbf{u}_{0j}] + 0.058\mathbf{extrav}_{ij} + 0.835\mathbf{sex}_{ij} + [0.096 + \mathbf{u}_{1j}]\mathbf{texp}_{ij} + \varepsilon_{ij}$$

Untuk mendapatkan nilai-nilai **u0** dan **u1** di tiap sekolah, perintah STATA adalah:

. predict u0 u1, reffects

. list pupil school u0 u1 in 1/10

```

+-----+
| pupil   school      u0      u1 |
+-----+
1. |      1      1   .0250209   .6724996 |
2. |      2      1   .0250209   .6724996 |
3. |      3      1   .0250209   .6724996 |
4. |      4      1   .0250209   .6724996 |
5. |      5      1   .0250209   .6724996 |
+-----+
6. |      6      1   .0250209   .6724996 |
7. |      7      1   .0250209   .6724996 |
8. |      8      1   .0250209   .6724996 |
9. |      9      1   .0250209   .6724996 |
10. |     10      1   .0250209   .6724996 |
+-----+

```


Selanjutnya akan ditampilkan nilai-nilai **u0** dan **u1** untuk 5 sekolah pertama dan perhitungan intersep-nya:

. tabstat u0 u1 if school<=5, by(school)

Summary statistics: mean
by categories of: school (school)

school	u0	u1
1	.0250209	.6724996
2	-.0184615	-.850629
3	.0226256	1.122686
4	.0110209	.3554588
5	.0012696	.163794
Total	.0080254	.2770685

Untuk **school = 1**

$$\text{popular}_{ij} = [3.217 + u_{0j}] + 0.058\text{extrav}_{ij} + 0.835\text{sex}_{ij} \\ + [0.096 + u_{1j}]\text{texp}_{ij} + \varepsilon_{ij}$$

$$\text{popular}_{ij} = [3.217 + 0.025] + 0.058\text{extrav}_{ij} + 0.835\text{sex}_{ij} \\ + [0.096 + 0.672]\text{texp}_{ij} + \varepsilon_{ij}$$

$$\text{popular}_{ij} = 3.242 + 0.058\text{extrav}_{ij} + 0.835\text{sex}_{ij} + 0.768\text{texp}_{ij} \\ + \varepsilon_{ij}$$

Untuk **school** = 2

$$\begin{aligned} \mathbf{popular}_{ij} = & [3.217 - 0.018] + 0.058\mathbf{extrav}_{ij} + 0.835\mathbf{sex}_{ij} \\ & + [0.096 - 0.851]\mathbf{texp}_{ij} + \varepsilon_{ij} \end{aligned}$$

$$\begin{aligned} \mathbf{popular}_{ij} = & 3.199 + 0.058\mathbf{extrav}_{ij} + 0.835\mathbf{sex}_{ij} - 0.755\mathbf{texp}_{ij} \\ & + \varepsilon_{ij} \end{aligned}$$

dan seterusnya.

BAB 8

MULTILEVEL GENERALIZED LINEAR MIXED MODELS

Karakteristik Multilevel GLMM

Karakteristik *Multilevel Generalized Linear Mixed Models* (*Multilevel GLMM*) adalah:

- Data non-Gaussian, yaitu variabel dependen tidak berdistribusi normal, bahkan tidak berskala kontinu. Variabel dependen dapat berskala biner, kategorik nominal, kategorik ordinal, atau data cacah. Di sini hanya akan dibahas *Multilevel GLMM* dengan variabel dependen berskala biner.
- Efek *mixed*, yaitu sebagian parameter bernilai sama untuk seluruh anggota populasi dan sebagian parameter lain bernilai berbeda antar-grup.
- Multi-level, yaitu pengumpulan data pada lebih daripada satu level.

Model Umum

Model umum untuk *Multilevel Generalized Linear Mixed Model* adalah:

$$\eta = X\beta + Zu \quad (8.1)$$

A. Hanya konstante bervariasi antar-grup:

Sintaks Stata:

```
. melogit depvar fe_equation || macr_level:
```

depvar : Variabel dependen

fe_equation : Variabel independen dengan efek *fixed*

macr_level : Level makro

Contoh 8.1: Pengulangan Kelas Siswa (1)

Data: model-06_gthai1.dta

```
. use "D:\Analisis Multilevel\Data\model-06_gthai1"
```

. summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
schoolid	612	60812.75	31131.47	10101	110204
gender	612	.5081699	.5003422	0	1
pped	612	.5179739	.5000856	0	1
repeat	612	.5163399	.5001417	0	1
trial	612	6.279412	4.418978	1	30
msesc	612	1.456291	3.294062	-.64	9

Variabel:

- **schoolid** : Nomor sekolah, berfungsi sebagai grup
- **gender** : Jenis kelamin siswa; 1: laki-laki, 2: perempuan
- **pped** : Perolehan pendidikan pra-SD (PAUD/TK);
1: ya, 0: tidak
- **repeat** : Pernah mengulang (tidak naik kelas) di SD;
1: ya, 0: tidak
- **msesc** : Rerata tingkat sosial ekonomi sekolah, diukur pada level sekolah

Keterangan:

File ini memuat data tentang 612 orang siswa yang berasal dari sejumlah sekolah (**schoolid**), tiap sekolah diwakili oleh 2 s.d. 4 orang siswa. Variabel dependennya adalah pernah tidaknya siswa mengulang kelas selama di SD (**repeat**), yang merupakan variabel biner; **repeat** = 1 jika ya dan **repeat** = 0 jika tidak.

Prediktor adalah jenis kelamin siswa (**gender**) dan perolehan pendidikan pra-SD (**pped**) yang diukur pada level siswa, serta rerata tingkat sosial ekonomi (**msesc**) yang diukur pada level sekolah.

Di sini terdapat 612 grup (**schoolid**; No. 10101 s.d. 11204), sehingga ukuran sampel bagi jumlah grup sudah mencukupi.

. list in 1/10

	schoolid	gender	pped	repeat	trial	msesc
1.	10101	0	1	0	15	9
2.	10101	1	1	0	4	9
3.	10102	0	0	0	1	9
4.	10102	0	1	0	10	9
5.	10102	1	1	0	13	9
6.	10103	0	0	0	2	.88
7.	10103	0	1	0	4	.88
8.	10103	1	1	1	11	.88
9.	10104	0	0	0	7	.2
10.	10104	0	1	0	8	.2

Model:

$$\text{logit } repeat_{ij} = \beta_0 + \beta_1 gender_{ij} + \beta_2 pped_{ij} + \beta_3 msesc_{ij} + u_j$$

Perintah Stata:

. melogit *repeat gender pped msesc* || *schoolid*:

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -416.90461
Iteration 1:   log likelihood = -416.62368
Iteration 2:   log likelihood = -416.62365
```

Refining starting values:

```
Grid node 0:   log likelihood =  -408.3934
```

Fitting full model:

```
Iteration 0:   log likelihood =  -408.3934
Iteration 1:   log likelihood =  -408.1699
Iteration 2:   log likelihood = -408.16967
Iteration 3:   log likelihood = -408.16967
```

Mixed-effects logistic regression	Number of obs	=	612
Group variable: schoolid	Number of groups	=	195

Obs per group:

min	=	1
avg	=	3.1
max	=	4

Integration method: mvaghermite	Integration pts.	=	7
---------------------------------	------------------	---	---

Log likelihood = -408.16967	Wald chi2(3)	=	14.96
	Prob > chi2	=	0.0019

repeat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
gender	.5850916	.1833719	3.19	0.001	.2256893	.9444939
pped	-.4103378	.1850878	-2.22	0.027	-.7731032	-.0475723
msesc	.0034711	.0353665	0.10	0.922	-.0658461	.0727882
_cons	.0129252	.1832727	0.07	0.944	-.3462827	.3721331
schoolid						
var(_cons)	.9821336	.3636554			.4753303	2.029297

LR test vs. logistic model: $\chi^2(01) = 16.91$

Prob >= $\chi^2 = 0.0000$

Model estimasi:

$$\text{logit } repeat_{ij} = 0.013 + 0.585gender_{ij} - 0.410pped_{ij} \\ + 0.003msesc_{ij} + u_j$$

atau:

$$\text{logit } repeat_{ij} = [0.013 + u_j] + 0.585gender_{ij} - 0.410pped_{ij} \\ + 0.003msesc_{ij}$$

Untuk mengestimasi nilai-nilai di tiap sekolah:

. predict u0, reffects

(calculating posterior means of random effects)

(using 7 quadrature points)

. list schoolid repeat u0 in 1/10

	schoolid	repeat	u0
1.	10101	0	-.6773751
2.	10101	0	-.6773751
3.	10102	0	-.9032223
4.	10102	0	-.9032223
5.	10102	0	-.9032223
6.	10103	0	-.2757971
7.	10103	0	-.2757971
8.	10103	1	-.2757971
9.	10104	0	-1.13526
10.	10104	0	-1.13526

Nilai- nilai **u0** untuk 5 sekolah pertama adalah:

. tabstat u0 if schoolid<=10105, by(schoolid)

Summary for variables: u0
by categories of: schoolid

schoolid	mean
10101	-.6773751
10102	-.9032223
10103	-.2757971
10104	-1.13526
10105	.4890398
Total	-.4672931

Untuk **schoolid** = 10101:

$$\text{logit } \textit{repeat}_{ij} = [0.013 + u_j] + 0.585\textit{gender}_{ij} - 0.410\textit{pped}_{ij} \\ + 0.003\textit{msesc}_{ij}$$

$$\text{logit } \textit{repeat}_{ij} = [0.013 - 0.677] + 0.585\textit{gender}_{ij} - 0.410\textit{pped}_{ij} \\ + 0.003\textit{msesc}_{ij}$$

$$\text{logit } \textit{repeat}_{ij} = -0.664 + 0.585\textit{gender}_{ij} - 0.410\textit{pped}_{ij} \\ + 0.003\textit{msesc}_{ij}$$

Untuk **schoolid** = 10102:

$$\text{logit } \textit{repeat}_{ij} = [0.013 - 0.903] + 0.585\textit{gender}_{ij} - 0.410\textit{pped}_{ij} \\ + 0.003\textit{msesc}_{ij}$$

$$\text{logit } \textit{repeat}_{ij} = -0.890 + 0.585\textit{gender}_{ij} - 0.410\textit{pped}_{ij} \\ + 0.003\textit{msesc}_{ij}$$

dan seterusnya.

B. Konstante dan koefisien regresi *pped* bervariasi antar-grup:

Sintaks Stata:

```
. melogit depvar fe_equation || macr_level: re_equation
```

depvar : Variabel dependen

fe_equation : Variabel independen dengan efek *fixed*

re_equation : Variabel independen dengan efek random

macr_level : Level makro

Contoh 8.2: Pengulangan Kelas Siswa (2)

Data: model-06_gthai1.dta

```
. use "D:\Analisis Multilevel\Data\model-06_gthai1, clear"
```

Model:

$$\begin{aligned} \text{logit } repeat_{ij} = & \beta_0 + \beta_1 gender_{ij} + \beta_2 pped_{ij} + \beta_3 msesc_{ij} \\ & + u_{0j} + u_{1j} pped_{ij} \end{aligned}$$

Perintah Stata:

```
. melogit repeat gender pped msesc || schoolid: pped
```

Fitting fixed-effects model:

```
Iteration 0:    log likelihood = -416.90461  
Iteration 1:    log likelihood = -416.62368  
Iteration 2:    log likelihood = -416.62365
```

Refining starting values:

```
Grid node 0:    log likelihood = -404.18709
```

Fitting full model:

```
Iteration 0:    log likelihood = -404.18709  
Iteration 1:    log likelihood = -401.69178  
Iteration 2:    log likelihood = -401.30728  
Iteration 3:    log likelihood = -401.30226  
Iteration 4:    log likelihood = -401.30225
```

Mixed-effects logistic regression	Number of obs	=	612
Group variable: schoolid	Number of groups	=	195

Obs per group:	
min	= 1
avg	= 3.1
max	= 4

Integration method: mvaghermite	Integration pts.	=	7
---------------------------------	------------------	---	---

Log likelihood = -401.30225	Wald chi2(3)	=	14.24
	Prob > chi2	=	0.0026

repeat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
gender	.6758112	.2034251	3.32	0.001	.2771052	1.074517
pped	-.4786013	.2573502	-1.86	0.063	-.9829985	.0257959
msesc	-.0036121	.0399565	-0.09	0.928	-.0819253	.0747011
_cons	-.037039	.1884096	-0.20	0.844	-.4063151	.3322371
schoolid						
var(pped)	3.770743	1.755714			1.513893	9.392018
var(_cons)	.8871386	.464355			.3180146	2.474776

LR test vs. logistic model: $\chi^2(2) = 30.64$ Prob > $\chi^2 = 0.0000$

Note: LR test is conservative and provided only for reference.

Model estimasi:

$$\text{logit } repeat_{ij} = -0.037 + 0.676gender_{ij} - 0.479pped_{ij} \\ - 0.004msesc_{ij} + u_{0j} + u_{1j} pped_{ij}$$

atau:

$$\text{logit } repeat_{ij} = [-0.037 + u_{0j}] + 0.676gender_{ij} \\ + [-0.479 + u_{1j}] pped_{ij} - 0.004msesc_{ij}$$

Untuk mengestimasi nilai-nilai u_{0j} dan u_{1j} pada tiap sekolah:

. predict u0 u1, reffects

(calculating posterior means of random effects)
(using 7 quadrature points)

. list schoolid u0 u1 in 1/10

	schoolid	u0	u1
1.	10101	-1.456552	-.3427255
2.	10101	-1.456552	-.3427255
3.	10102	-1.328793	-.6260999
4.	10102	-1.328793	-.6260999
5.	10102	-1.328793	-.6260999
6.	10103	.2839322	-.3096711
7.	10103	.2839322	-.3096711
8.	10103	.2839322	-.3096711
9.	10104	-1.209414	-.9346197
10.	10104	-1.209414	-.9346197

Berikut ditampilkan nilai-nilai u_{0j} dan u_{1j} pada 5 sekolah pertama:

. tabstat u0 u1 if schoolid<=10105, by(schoolid)

Summary statistics: mean
by categories of: schoolid

schoolid	u0	u1
10101	-1.456552	-.3427255
10102	-1.328793	-.6260999
10103	.2839322	-.3096711
10104	-1.209414	-.9346197
10105	1.693405	.2067509
Total	-.2569826	-.4002649

Untuk **schoolid** = 10101:

$$\begin{aligned}\text{logit } \textit{repeat}_{ij} &= [-0.037 + u_{0j}] + 0.676\textit{gender}_{ij} \\ &\quad + [-0.479 + u_{1j}]\textit{pped}_{ij} - 0.004\textit{msesc}_{ij} \\ \text{logit } \textit{repeat}_{ij} &= [-0.037 - 1.457] + 0.676\textit{gender}_{ij} \\ &\quad + [-0.479 - 0.343]\textit{pped}_{ij} - 0.004\textit{msesc}_{ij} \\ \text{logit } \textit{repeat}_{ij} &= -1.494 + 0.676\textit{gender}_{ij} - 0.822\textit{pped}_{ij} \\ &\quad - 0.004\textit{msesc}_{ij}\end{aligned}$$

Untuk **schoolid** = 10102:

$$\begin{aligned}\text{logit } \textit{repeat}_{ij} &= [-0.037 - 1.329] + 0.676\textit{gender}_{ij} \\ &\quad + [-0.479 - 0.626]\textit{pped}_{ij} - 0.004\textit{msesc}_{ij} \\ \text{logit } \textit{repeat}_{ij} &= -1.366 + 0.676\textit{gender}_{ij} - 1.105\textit{pped}_{ij} \\ &\quad - 0.004\textit{msesc}_{ij}\end{aligned}$$

dan seterusnya.

BAB 9

ANALISIS MULTILEVEL UNTUK DATA LONGITUDINAL

Analisis Data Longitudinal

Pada data longitudinal, jika jarak antar-sesi sama dan dimiliki data lengkap untuk seluruh anggota sampel pada tiap sesi, analisis datanya dapat dilakukan dengan Analisis Variansi (ANOVA) untuk pengukuran berulang. Analisis data dapat juga dilakukan menggunakan Analisis Regresi dengan metode *Generalized Estimating Equation* (GEE).

Jika jarak antar-sesi tidak seluruhnya sama dan/atau tidak dimiliki data lengkap untuk seluruh anggota sampel pada tiap sesi, dapat digunakan Analisis Multilevel untuk data longitudinal. Pada Analisis Multilevel untuk data longitudinal, tiap subjek/objek yang menjalani pengukuran berulang dianggap dan diperlakukan sebagai 1 grup. Contoh yang diberikan di sini hanya model dengan variabel dependen berupa data Gaussian.

Contoh 9.1: IPK Mahasiswa (1)

Data: model-07_gpa2long.dta

```
. use "D:\Analisis Multilevel\Data\model-07_gpa2long"  
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
student	1,200	100.5	57.75838	1	200
occas	1,200	2.5	1.708537	0	5
gpa	1,200	2.865	.3930484	1.7	4
job	1,200	2.1075	.4275076	1	3
sex	1,200	.525	.4995828	0	1
highgpa	1,200	2.9875	.5948854	2	4

Variabel:

- **student** : No identitas mahasiswa
- **occas** : Sesi pengumpulan data; nilai **occas** 0 s.d. 5 (6 sesi pengumpulan data)
- **gpa** : *Grade point average* (= indeks prestasi kumulatif)
- **job** : Status pekerjaan; jumlah jam kerja/minggu pada sesi tertentu, nilai **job** 1 s.d. 3
- **sex** : Jenis kelamin mahasiswa
- **highgpa** : Nilai IPK siswa di SMA

Keterangan

File ini memuat data Indeks Prestasi Kumulatif (**gpa**) 200 orang mahasiswa, yang masing-masing menjalani 6 sesi pengumpulan data (**occas**). Dalam tiap sesi dikumpulkan status pekerjaan mahasiswa dalam jumlah jam kerja/minggu (**job**), jenis kelamin mahasiswa (**sex**), dan nilai IPK siswa sewaktu di SMA (**highgpa**). Tiap mahasiswa (**student**) dianggap sebagai 1 grup.

Jumlah grup di sini adalah jumlah mahasiswa (**student**), yaitu sebanyak 200 orang, yang memenuhi syarat untuk jumlah grup minimum.

. list in 1/10

	student	occas	gpa	job	sex	highgpa
1.	1	0	2.3	2	1	2.8
2.	1	1	2.1	2	1	2.8
3.	1	2	3	2	1	2.8
4.	1	3	3	2	1	2.8
5.	1	4	3	2	1	2.8
6.	1	5	3.3	2	1	2.8
7.	2	0	2.2	2	0	2.5
8.	2	1	2.5	3	0	2.5
9.	2	2	2.6	2	0	2.5
10.	2	3	2.6	2	0	2.5

A. Hanya konstante bervariasi antar-grup:

Sintaks Stata:

```
. mixed depvar fe_equation || macr_level:
```

depvar : Variabel dependen

fe_equation : Variabel independen dengan efek *fixed*

macr_level : Level makro

Model:

$$gpa_{ij} = \beta_0 + \beta_1 occas_{ij} + \beta_2 job_{ij} + \beta_3 highgpa_{ij} + \beta_4 sex_{ij} + u_j + \varepsilon_{ij}$$

Perintah Stata:

```
. mixed gpa occas job highgpa sex || student:
```

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0: log likelihood = -141.37984

Iteration 1: log likelihood = -141.37984

Computing standard errors:

Mixed-effects ML regression	Number of obs	=	1,200
Group variable: student	Number of groups	=	200

```

Obs per group:
min = 6
avg = 6.0
max = 6
Wald chi2(4) = 839.56
Log likelihood = -141.37984 Prob > chi2 = 0.0000

```

	gpa	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	occas	.1024519	.0039898	25.68	0.000	.0946321	.1102716
	job	-.1722102	.0180633	-9.53	0.000	-.2076136	-.1368067
	highgpa	.0846949	.0277593	3.05	0.002	.0302876	.1391022
	sex	.1472521	.033053	4.46	0.000	.0824693	.2120349
	_cons	2.64147	.0975222	27.09	0.000	2.45033	2.83261

Random-effects Parameters		Estimate	Std. Err.	[95% Conf. Interval]	
student: Identity					
	var(_cons)	.0449748	.0055226	.0353547	.0572126
	var(Residual)	.0551389	.0024737	.0504976	.0602068

```

LR test vs. linear model: chibar2(01) = 329.05
Prob >= chibar2 = 0.0000

```

Model Estimasi:

$$\mathbf{gpa}_{ij} = 2.641 + 0.102\mathbf{occas}_{ij} - 0.172\mathbf{job}_{ij} + 0.085\mathbf{highgpa}_{ij} \\ + 0.147\mathbf{sex}_{ij} + \mathbf{u}_j + \varepsilon_{ij}$$

atau:

$$\mathbf{gpa}_{ij} = [2.641 + \mathbf{u}_j] + 0.102\mathbf{occas}_{ij} - 0.172\mathbf{job}_{ij} \\ + 0.085\mathbf{highgpa}_{ij} + 0.147\mathbf{sex}_{ij} + \varepsilon_{ij}$$

Untuk mengestimasi nilai-nilai **u0** bagi tiap **student**:

. predict u0, reffects

(calculating posterior means of random effects)
(using 7 quadrature points)

. list student gpa u0 in 1/10

	student	gpa	u0
1.	1	2.3	-.128074
2.	1	2.1	-.128074
3.	1	3	-.128074
4.	1	3	-.128074
5.	1	3	-.128074
6.	1	3.3	-.128074
7.	2	2.2	-.0992651
8.	2	2.5	-.0992651
9.	2	2.6	-.0992651
10.	2	2.6	-.0992651

Nilai- nilai **u0** untuk 5 **student** pertama adalah:

. **tabstat u0 if student<=5, by(student)**

Summary for variables: u0
by categories of: student (student id)

student	mean
1	-.128074
2	-.0992651
3	.0690837
4	-.1630098
5	.0699695
Total	-.0502591

Untuk **student = 1**:

$$gpa_{ij} = [2.641 + u_j] + 0.102occas_{ij} - 0.172job_{ij} + 0.085highgpa_{ij} + 0.147sex_{ij} + \varepsilon_{ij}$$

$$gpa_{ij} = [2.641 - 0.128] + 0.102occas_{ij} - 0.172job_{ij} + 0.085highgpa_{ij} + 0.147sex_{ij} + \varepsilon_{ij}$$

$$gpa_{ij} = 2.513 + 0.102occas_{ij} - 0.172job_{ij} + 0.085highgpa_{ij} + 0.147sex_{ij} + \varepsilon_{ij}$$

Untuk **student** = 2:

$$gpa_{ij} = [2.641 - 0.099] + 0.102occas_{ij} - 0.172job_{ij} + 0.085highgpa_{ij} + 0.147sex_{ij} + \varepsilon_{ij}$$

$$gpa_{ij} = 2.542 + 0.102occas_{ij} - 0.172job_{ij} + 0.085highgpa_{ij} + 0.147sex_{ij} + \varepsilon_{ij}$$

dan seterusnya.

B. Konstante dan koefisien regresi *occas* bervariasi antar-grup:

Sintaks Stata:

. mixed *depvar fe_equation* || *macr_level*: *re_equation*

depvar : Variabel dependen

fe_equation : Variabel independen dengan efek *fixed*

re_equation : Variabel independen dengan efek random

macr_level : Level makro

Contoh 9.2: IPK Mahasiswa (2)

Data: model-07_gpa2long.dta

. use "D:\Analisis Multilevel\Data\model-07_gpa2long,
clear"

Model:

$$gpa_{ij} = \beta_0 + \beta_1 occas_{ij} + \beta_2 job_{ij} + \beta_3 highgpa_{ij} + \beta_4 sex_{ij} \\ + u_{0j} + u_{1j} occas_{ij} + \varepsilon_{ij}$$

Perintah Stata:

. mixed *gpa occas job highgpa sex* || *student: occas*

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0: log likelihood = -86.571483

Iteration 1: log likelihood = -86.57148

Computing standard errors:

Mixed-effects ML regression	Number of obs	=	1,200
Group variable: student	Number of groups	=	200

Obs per group:

min	=	6
avg	=	6.0
max	=	6

Log likelihood = -86.57148	Wald chi2(4)	=	472.23
	Prob > chi2	=	0.0000

	gpa	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	occas	.1034018	.0053827	19.21	0.000	.0928519	.1139517
	job	-.1298556	.0173044	-7.50	0.000	-.1637715	.0959397
	highgpa	.0891193	.0264402	3.37	0.001	.0372975	.1409412
	sex	.1067106	.0314925	3.39	0.001	.0449865	.1684348
	_cons	2.557899	.0923978	27.68	0.000	2.376803	2.738995

Random-effects Parameters		Estimate	Std. Err.	[95% Conf. Interval]	
student: Independent					
	var(occas)	.0033444	.0005267	.0024562	.0045539
	var(_cons)	.0333268	.0049942	.0248448	.0447046
	var(Residual)	.0423533	.0021125	.0384088	.0467028

LR test vs. linear model: $\chi^2(2) = 438.67$

Prob > $\chi^2 = 0.0000$

Note: LR test is conservative and provided only for reference.

Model Estimasi:

$$gpa_{ij} = 2.558 + 0.103occas_{ij} - 0.130job_{ij} + 0.089highgpa_{ij} + 0.107sex_{ij} + u_{0j} + u_{1j}occas_{ij} + \varepsilon_{ij}$$

atau:

$$gpa_{ij} = [2.558 + u_{0j}] + [0.103 + u_{1j}]occas_{ij} - 0.130job_{ij} + 0.089highgpa_{ij} + 0.107sex_{ij} + \varepsilon_{ij}$$

Untuk mengestimasi nilai-nilai u_{0j} dan u_{1j} bagi tiap **student**:

. predict u0 u1, reffects

(calculating posterior means of random effects)
(using 7 quadrature points)

. list student gpa u0 u1 u1 in 1/10

	student	gpa	u0	u1	u1
1.	1	2.3	.0463166	-.2025036	-.2025036
2.	1	2.1	.0463166	-.2025036	-.2025036
3.	1	3	.0463166	-.2025036	-.2025036
4.	1	3	.0463166	-.2025036	-.2025036
5.	1	3	.0463166	-.2025036	-.2025036
6.	1	3.3	.0463166	-.2025036	-.2025036
7.	2	2.2	-.0033897	-.1095118	-.1095118
8.	2	2.5	-.0033897	-.1095118	-.1095118
9.	2	2.6	-.0033897	-.1095118	-.1095118
10.	2	2.6	-.0033897	-.1095118	-.1095118

Nilai- nilai **u0** dan **u1** untuk 5 **student** pertama adalah:

. **tabstat u0 u1 if student<=5, by(student)**

Summary for variables: mean
by categories of: student (student id)

student	u0	u1
1	.0463166	-.2025036
2	-.0033897	-.1095118
3	.0414806	-.0013151
4	-.0482231	-.085117
5	-.0172603	.090929
Total	.0037848	-.0615037

Untuk **student = 1**:

$$gpa_{ij} = [2.558 + u_{0j}] + [0.103 + u_{1j}]occas_{ij} - 0.130job_{ij} + 0.089highgpa_{ij} + 0.107sex_{ij} + \varepsilon_{ij}$$

$$gpa_{ij} = [2.558 + 0.046] + [0.103 - 0.203]occas_{ij} - 0.130job_{ij} + 0.089highgpa_{ij} + 0.107sex_{ij} + \varepsilon_{ij}$$

$$gpa_{ij} = 2.604 - 0.100occas_{ij} - 0.172job_{ij} + 0.085highgpa_{ij} + 0.147sex_{ij} + \varepsilon_{ij}$$

Untuk **student = 2**:

$$\mathbf{gpa}_{ij} = [2.558 - 0.003] + [0.103 - 0.110]\mathbf{occas}_{ij} - 0.130\mathbf{job}_{ij} \\ + 0.089\mathbf{highgpa}_{ij} + 0.107\mathbf{sex}_{ij} + \varepsilon_{ij}$$

$$\mathbf{gpa}_{ij} = 2.555 - 0.007\mathbf{occas}_{ij} - 0.172\mathbf{job}_{ij} + 0.085\mathbf{highgpa}_{ij} \\ + 0.147\mathbf{sex}_{ij} + \varepsilon_{ij}$$

dan seterusnya.

BAB 10

MULTILEVEL SEM I:

ANALISIS JALUR

Beberapa bentuk SEM (*Structural Equation Modeling*; Pemodelan Persamaan Struktural) antara lain yaitu Analisis Jalur (*Path Analysis*), Analisis Faktor Konfirmatorik (*Confirmatory Factor Analysis*), Model Regresi Struktural (Model Hibrid), dan *Generalized SEM*. Dalam 2 bab berikut hanya akan dibahas Analisis Multilevel untuk Analisis Jalur dan Analisis Faktor Konfirmatorik.

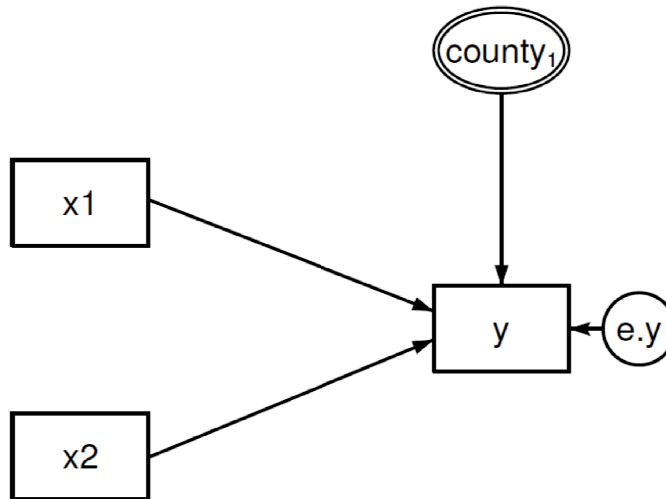
Tipe Analisis Jalur Multilevel

Beberapa tipe Analisis Multilevel untuk Analisis Jalur yaitu:

- Hanya konstante bervariasi antar grup (*random intercept*)
- Koefisien regresi bervariasi antar grup (*random slope*)
- Konstante dan koefisien regresi bervariasi antar grup (*random intercept dan random slope*)

A. Hanya konstante bervariasi antar-grup (*random intercept*)

Contoh Model:



x1 dan **x2** adalah prediktor untuk **y**. **county** dalam lingkaran ganda menyatakan variabel laten pada level **county** yang konstan dalam (*within*) **county** dan bervariasi antar (*between*) **county**. Perhatikan bahwa **county** dalam lingkaran ganda ini **tidak** menyatakan nomor **county** seperti yang ada dalam basis data. Variabel laten ini akan diberi nama lain, biasanya **M1[*county*]** atau **M1** saja.

Perintah Stata:

. sem (x1 x2 *M1[**county**]* -> y)

Variabel laten untuk **county** dalam lingkaran ganda dinamakan *M1[**county**]*.

Model Matematik:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_{2j} + \beta_3 M_{1,c} + \varepsilon$$

Program STATA secara otomatis akan menetapkan koefisien regresi β_3 bernilai sama dengan 1, sehingga model menjadi:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_{2j} + M_{1,c} + \varepsilon$$

dan $(\beta_0 + M_{1,c})$ menjadi intersep yang bervariasi antar grup.

Contoh 10.1 Gaji Karyawan

(*random intercept*):

Data: model-08_gsem-nlsy.dta

. use "D:\Analisis Multilevel\Data\model-08_gsem-nlsy"

(NLSY 1968)

. summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
idcode	2,763	249.0894	147.1098	1	499
year	2,763	1977.936	6.447717	1968	1988
grade	2,763	12.82519	2.282903	0	18
union	1,904	.2268908	.4189314	0	1
ln_wage	2,763	1.77696	.4535444	.0044871	4.49981

Variabel:

- **idcode** : Nomor identitas NLS (National Longitudinal Survey)
- **year** : Tahun wawancara
- **grade** : Grade terakhir yang diselesaikan
- **union** : 1 jika menjadi anggota union
- **ln_wage** : ln (wage/GNP deflator)

Keterangan:

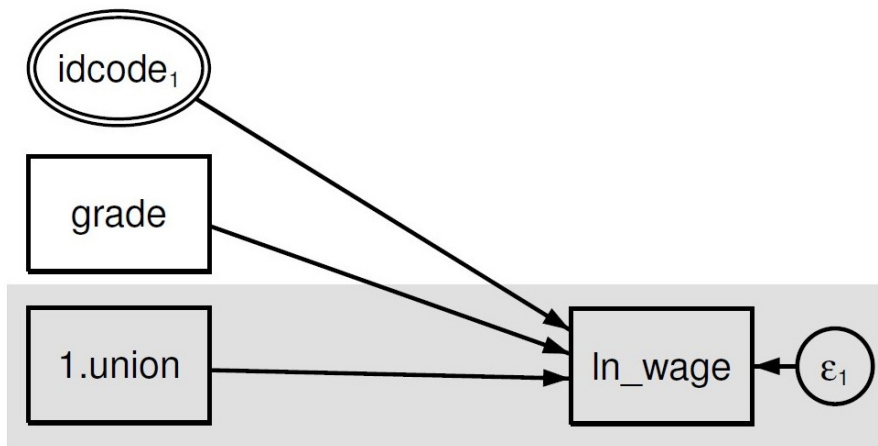
File ini memuat data runtun waktu 499 orang responden (**idcode**), dengan ln gaji/deflator GNP (**ln_wage**) sebagai variabel dependen. Prediktor adalah tahun wawancara (**year**), **grade** pendidikan terakhir yang diselesaikan responden pada tahun wawancara (**grade**), keanggotaan responden dalam organisasi buruh pada tahun wawancara (**union**), **union** = 1 jika ya dan **union** = 0 jika tidak.

Seperti pada analisis multilevel untuk data longitudinal, identitas responden (***idcode***) di sini menjadi grup. Untuk variabel ***union*** terdapat sejumlah *missing data*, sehingga pengamatan lengkap untuk seluruh responden tersisa menjadi 1,904.

. list in 1/20, sepby(*idcode*)

	idcode	year	grade	union	ln_wage
1.	1	1970	12	.	1.451214
2.	1	1971	12	.	1.02862
3.	1	1972	12	1	1.589977
4.	1	1973	12	.	1.780273
5.	1	1975	12	.	1.777012
6.	1	1977	12	0	1.778681
7.	1	1978	12	.	2.493976
8.	1	1980	12	1	2.551715
9.	1	1983	12	1	2.420261
10.	1	1985	12	1	2.614172
11.	1	1987	12	1	2.536374
12.	1	1988	12	1	2.462927
13.	2	1971	12	0	1.360348
14.	2	1972	12	.	1.206198
15.	2	1973	12	.	1.549883
16.	2	1975	12	.	1.832581
17.	2	1977	12	1	1.726721
18.	2	1978	12	1	1.68991
19.	2	1980	12	1	1.726964
20.	2	1982	12	1	1.808289

Model:



Pengumpulan data dilakukan pada 2 level: **ln_wage** dan **union** yang bervariasi pada level mikro (level pengamatan), sedangkan **grade** bervariasi pada level makro (level subjek).

idcode dalam lingkaran-ganda menyatakan variabel laten pada level **idcode** yang konstan **dalam** (*within*) kode identifikasi dan bervariasi **antar** (*between*) kode identifikasi. Dalam sintaks STATA, **idcode** dalam lingkaran-ganda ini dinyatakan dengan **M1[idcode]**.

Model Matematik:

$$\ln_wage = \beta_0 + \beta_1 \text{1.union} + \beta_2 \text{grade} + M1[idcode] + \varepsilon$$

Perintah Stata:

```
. gsem (ln_wage <- 1.union grade M1[idcode])
```

Fitting fixed-effects model:

Iteration 0: log likelihood = -925.06629

Iteration 1: log likelihood = -925.06629

Refining starting values:

Grid node 0: log likelihood = -763.3769

Fitting full model:

Iteration 0: log likelihood = -763.3769

Iteration 1: log likelihood = -622.04625
(backed up)

Iteration 2: log likelihood = -613.54948

Iteration 3: log likelihood = -607.56242

Iteration 4: log likelihood = -607.49246

Iteration 5: log likelihood = -607.49233

Iteration 6: log likelihood = -607.49233

Generalized structural equation model Number of obs = 1,904

Response : ln_wage

Family : Gaussian

Link : identity

Log likelihood = -607.49233

```
( 1) [ln_wage]M1[idcode] = 1
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ln_wage <-						
1.union	.1637408	.0227254	7.21	0.000	.1191998	.2082818
grade	.0767919	.0067923	11.31	0.000	.0634791	.0901046
M1[idcode]	1	(constrained)				
_cons	.7774129	.0906282	8.58	0.000	.5997848	.955041
var(M1[idcode])	.080247	.0073188			.0671113	.0959537
var(e.ln_wage)	.078449	.0028627			.0730342	.0842653

Model Estimasi:

$$\ln_wage = 0.777 + 0.164 \mathbf{1.union} + 0.077 \mathbf{grade} + \mathbf{M1[idcode]} + \varepsilon$$

atau:

$$\ln_wage = [0.777 + \mathbf{M1[idcode]}] + 0.164 \mathbf{1.union} + 0.077 \mathbf{grade} + \varepsilon$$

Untuk memprediksi nilai-nilai **M1[idcode]**, perintah STATA adalah:

```
. predict stub1, latent
```

```
(option ebmeans assumed)
(using 7 quadrature points)
(80 missing values generated)
```

stub* dalam sintaks Stata menyatakan variabel baru yang sebelumnya tidak ada dalam basis-data. Perintah **predict stub*** digunakan untuk memprediksi nilai-nilai variabel teramati, sedangkan perintah **predict stub*, latent** adalah untuk memprediksi nilai-nilai variabel laten.

. list idcode ln_wage stub1 in 1/10

	idcode	ln_wage	stub1
1.	1	1.451214	.385988
2.	1	1.02862	.385988
3.	1	1.589977	.385988
4.	1	1.780273	.385988
5.	1	1.777012	.385988
6.	1	1.778681	.385988
7.	1	2.493976	.385988
8.	1	2.551715	.385988
9.	1	2.420261	.385988
10.	1	2.614172	.385988

Untuk memperlihatkan nilai-nilai **stub1** bagi 5 **idcode** pertama:

. tabstat stub1 if idcode<=5, by(idcode)

Summary for variables: stub1
by categories of: idcode (NLS ID)

idcode	mean
1	.385988
2	-.0934264
3	-.0923483
4	-.2168231
5	.1719501
Total	.0267526

Untuk **idcode** = 1:

$$\ln_wage = [0.777 + M1[idcode]] + 0.164 \mathbf{1.union} + 0.077 \mathbf{grade} + \varepsilon$$

$$\ln_wage = [0.777 + 0.386] + 0.164 \mathbf{1.union} + 0.077 \mathbf{grade} + \varepsilon$$

$$\ln_wage = 1.163 + 0.164 \mathbf{1.union} + 0.077 \mathbf{grade} + \varepsilon$$

Untuk **idcode** = 2:

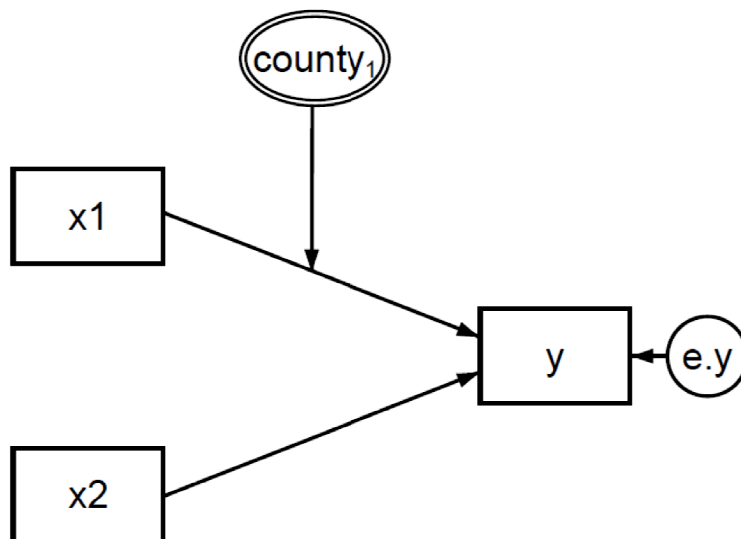
$$\ln_wage = [0.777 - 0.093] + 0.164 \mathbf{1.union} + 0.077 \mathbf{grade} + \varepsilon$$

$$\ln_wage = 0.684 + 0.164 \mathbf{1.union} + 0.077 \mathbf{grade} + \varepsilon$$

dan seterusnya.

B. Koefisien regresi bervariasi antar-grup (*random slope*)

Contoh Model:



Contoh model ini hampir sama seperti pada contoh model A, dengan perbedaan bahwa panah dari ***M1[*county*]*** tidak tertuju kepada ***y***, melainkan ke arah panah dari ***x1*** ke ***y***. Ini menunjukkan bahwa yang dipengaruhi oleh ***M1[*county*]*** adalah koefisien regresi (*slope*) ***y*** terhadap ***x1***. Dalam analisis statistik, panah dari ***M1[*county*]*** ke arah panah dari ***x1*** ke ***y*** diinterpretasikan sebagai interaksi antara ***M1[*county*]*** dengan ***x1***.

Perintah Stata:

```
(y <- x1 c.x1#M1[county] x2)
```

Model Matematik:

$$y = \beta_0 + \beta_1 x1 + \beta_2 x2_j + \beta_3 M_{1,C} x1 + \varepsilon$$

Program STATA secara otomatis akan menetapkan koefisien regresi β_3 bernilai sama dengan 1, sehingga model menjadi:

$$y = \beta_0 + \beta_1 x1 + \beta_2 x2_j + M_{1,C} x1 + \varepsilon$$

dan $(\beta_1 + M_{1,C})$ menjadi koefisien regresi untuk ***x1*** yang bervariasi antar grup.

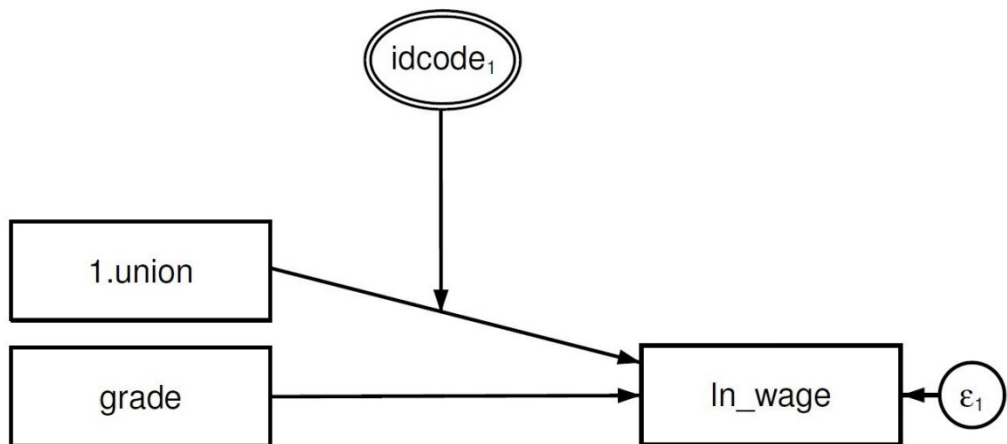
Contoh 10.2 Gaji Karyawan (*random slope*):

Data: model-08_gsem-nlsy.dta

```
. use "D:\Analisis Multilevel\Data\model-08_gsem-nlsy,  
clear"
```

```
(NLSY 1968)
```

Model:



Model Matematik:

$$\ln_wage = \beta_0 + \beta_1 1.union + \beta_2 grade + M1[idcode] \# 1.union + \varepsilon$$

Perintah Stata:

```
. gsem (ln_wage <- 1.union grade 1.union#M1[idcode])
```

Fitting fixed-effects model:

Iteration 0: log likelihood = -925.06629

Iteration 1: log likelihood = -925.06629

Refining starting values:

Grid node 0: log likelihood = -1006.2178

Fitting full model:

Iteration 0: log likelihood = -1006.2178 (not concave)

Iteration 1: log likelihood = -949.05484 (not concave)

Iteration 2: log likelihood = -902.42792

Iteration 3: log likelihood = -898.82814

Iteration 4: log likelihood = -898.46757

Iteration 5: log likelihood = -898.46628

Iteration 6: log likelihood = -898.46628

Generalized structural equation model Number of obs = 1,904

Response : ln_wage

Family : Gaussian

Link : identity

Log likelihood = -898.46628

```
( 1) [ln_wage]1.union#M2[idcode] = 1
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ln_wage <-						
1.union	.1122345	.0292696	3.83	0.000	.0548672	.1696017
grade	.0762541	.0042882	17.78	0.000	.0678493	.0846588
union#M2[idcode]						
1	1	(constrained)				
_cons	.8224346	.0567733	14.49	0.000	.711161	.9337083
var(M2[idcode])	.0457248	.0108563			.0287113	.07282
var(e.ln_wage)	.1437018	.0047919			.1346102	.1534074

Model Estimasi:

$$\ln_wage = 0.822 + 0.112\mathbf{1.union} + 0.076\mathbf{grade} \\ + \mathbf{M1[idcode]\#1.union} + \varepsilon$$

atau:

$$\ln_wage = 0.822 + [0.112 + \mathbf{M1[idcode]}]\mathbf{1.union} \\ + 0.077\mathbf{grade} + \varepsilon$$

Untuk memprediksi nilai-nilai **M1[idcode]**, perintah STATA adalah:

. predict stub1, latent

```
(option ebmeans assumed)
(using 7 quadrature points)
(80 missing values generated)
```

. list idcode ln_wage stub1 in 1/10

	idcode	ln_wage	stub1
1.	1	1.451214	.3365639
2.	1	1.02862	.3365639
3.	1	1.589977	.3365639
4.	1	1.780273	.3365639
5.	1	1.777012	.3365639
6.	1	1.778681	.3365639
7.	1	2.493976	.3365639
8.	1	2.551715	.3365639
9.	1	2.420261	.3365639
10.	1	2.614172	.3365639

Untuk memperlihatkan nilai-nilai **stub1** bagi 5 **idcode** pertama:

. tabstat stub1 if idcode<=5, by(idcode)

Summary for variables: stub1
by categories of: idcode (NLS ID)

idcode	mean
1	.3365639
2	-.0439834
3	-1.21e-18
4	.0052894
5	-1.21e-18
Total	.0585106

Untuk **idcode** = 1:

$$\begin{aligned} \mathbf{ln_wage} = & 0.822 + [0.112 + \mathbf{M1[idcode]}] \mathbf{1.union} \\ & + 0.077 \mathbf{grade} + \varepsilon \end{aligned}$$

$$\begin{aligned} \mathbf{ln_wage} = & 0.822 + [0.112 + 0.337] \mathbf{1.union} + 0.077 \mathbf{grade} \\ & + \varepsilon \end{aligned}$$

$$\mathbf{ln_wage} = 0.822 + 0.449 \mathbf{1.union} + 0.077 \mathbf{grade} + \varepsilon$$

Untuk **idcode** = 2:

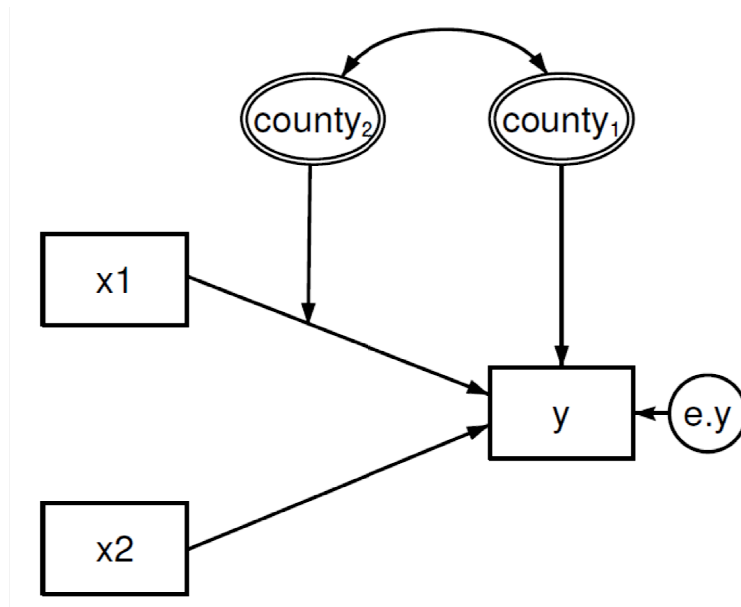
$$\begin{aligned} \mathbf{ln_wage} = & 0.822 + [0.112 - 0.044] \mathbf{1.union} + 0.077 \mathbf{grade} \\ & + \varepsilon \end{aligned}$$

$$\mathbf{ln_wage} = 0.822 + 0.068 \mathbf{1.union} + 0.077 \mathbf{grade} + \varepsilon$$

dan seterusnya.

C. Konstante dan koefisien regresi bervariasi antar-grup (*random intercept and random slope*)

Contoh Model:



Model ini dapat dianggap sebagai gabungan antara model A dan model B di atas. **$M1[*county*]$** ($county_1$ dalam lingkaran ganda) adalah variabel laten yang mempengaruhi intersep dan **$M2[*county*]$** ($county_2$ dalam lingkaran ganda) adalah variabel laten yang mempengaruhi *slope*.

Perintah Stata:

```
(y <- x1 x2 c.x1#M2[county] M1[county])
```

Model Matematik:

$$y = \beta_0 + \beta_1 x1 + \beta_2 x2_j + \beta_3 M_{2,C} x1 + \beta_4 M_{1,C} + \varepsilon$$

Program STATA secara otomatis akan menetapkan koefisien regresi β_3 dan β_4 bernilai sama dengan 1, sehingga model menjadi:

$$y = \beta_0 + \beta_1 x1 + \beta_2 x2_j + M_{2,C} x1 + M_{1,C} + \varepsilon$$

sehingga $(\beta_0 + M_{1,C})$ menjadi intersep dan $(\beta_1 + M_{2,C})$ menjadi koefisien regresi untuk $x1$ yang bervariasi antar grup.

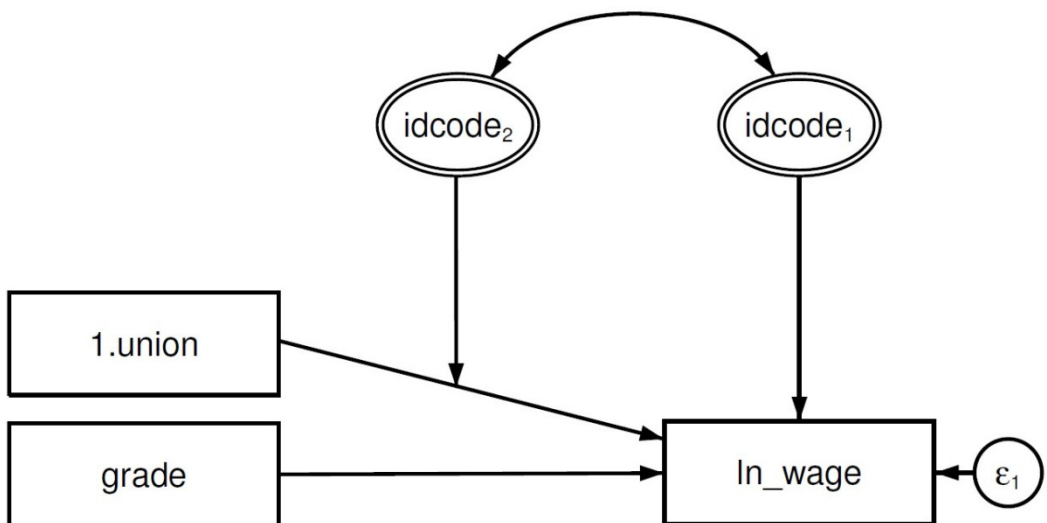
Contoh 10.3 Gaji Karyawan (*random intercept and random slope*):

Data: model-08_gsem-nlsy.dta

```
. use "D:\Analisis Multilevel\Data\model-08_gsem-nlsy,  
clear"
```

```
(NLSY 1968)
```

Model:



Model Matematik:

$$\ln_wage = \beta_0 + \beta_1 1.union + \beta_2 grade + M2[idcode]1.union + M1[idcode] + \varepsilon$$

Perintah Stata:

```
. gsem (ln_wage <- 1.union grade M1[idcode]  
1.union#M2[idcode])
```

Fitting fixed-effects model:

```
Iteration 0: log likelihood = -925.06629  
Iteration 1: log likelihood = -925.06629
```


Refining starting values:

Grid node 0: log likelihood = -869.92254

Fitting full model:

Iteration 0: log likelihood = -869.92254 (not concave)
Iteration 1: log likelihood = -727.21757 (not concave)
Iteration 2: log likelihood = -711.81244 (not concave)
Iteration 3: log likelihood = -684.4227 (not concave)
Iteration 4: log likelihood = -665.95677 (not concave)
Iteration 5: log likelihood = -609.85439
Iteration 6: log likelihood = -591.37219
Iteration 7: log likelihood = -586.604
Iteration 8: log likelihood = -581.53956
Iteration 9: log likelihood = -581.3076
Iteration 10: log likelihood = -581.30551
Iteration 11: log likelihood = -581.30551

Generalized structural equation model Number of obs = 1,904

Response : ln_wage
Family : Gaussian
Link : identity
Log likelihood = -581.30551

```
( 1)  [ln_wage]M1[idcode] = 1
( 2)  [ln_wage]1.union#M2[idcode] = 1
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
ln_wage <-					
1.union	.1459555	.028876	5.05	0.000	.0893595 .2025515
grade	.0766554	.0065295	11.74	0.000	.0638578 .089453
M1[idcode]	1	(constrained)			
union#					
M2[idcode]					
1	1	(constrained)			
_cons	.7760526	.0874931	8.87	0.000	.6045694 .9475359
var(M1[idcode])	.0927952	.0088244			.0770158 .1118074
var(M2[idcode])	.0825137	.0186016			.0530437 .1283566
cov(M2[idcode],					
M1[idcode])	-.0550905	.0115985	-4.75	0.000	-.0778231 -.0323578
var(e.ln_wage)	.0720854	.0027134			.0669586 .0776047

Model Estimasi:

$$\ln_wage = 0.776 + 0.1461.union + 0.077 grade \\ + M2[idcode]1.union + M1[idcode] + \varepsilon$$

atau:

$$\ln_wage = [0.776 + M1[idcode]] \\ + [0.146 + M2[idcode]]1.union + 0.077 grade \\ + \varepsilon$$

Untuk memprediksi nilai-nilai ***M1[idcode]*** dan ***M2[idcode]***, perintah STATA adalah:

. predict stub1 stub2, latent

(option ebmeans assumed)
 (using 7 quadrature points)
 (80 missing values generated)

. list idcode ln_wage stub1 stub2 in 1/10

	idcode	ln_wage	stub1	stub2
1.	1	1.451214	.1672946	.2654232
2.	1	1.02862	.1672946	.2654232
3.	1	1.589977	.1672946	.2654232
4.	1	1.780273	.1672946	.2654232
5.	1	1.777012	.1672946	.2654232
6.	1	1.778681	.1672946	.2654232
7.	1	2.493976	.1672946	.2654232
8.	1	2.551715	.1672946	.2654232
9.	1	2.420261	.1672946	.2654232
10.	1	2.614172	.1672946	.2654232

Untuk memperlihatkan nilai-nilai **stub*** bagi 5 **idcode** pertama:

. tabstat stub* if idcode<=5, by(idcode)

Summary statistics: mean
by categories of: idcode (NLS ID)

idcode	stub1	stub2
-----+-----		
1	.1672946	.2654232
2	-.1829917	.126367
3	-.091287	.0541951
4	-.4246151	.3970935
5	.1790876	-.1063204
-----+-----		
Total	-.069811	.1428347

Untuk **idcode** = 1:

$$\begin{aligned} \ln_wage = & [0.776 + M1[idcode]] \\ & + [0.146 + M2[idcode]]1.union + 0.077 grade \\ & + \varepsilon \end{aligned}$$

$$\begin{aligned} \ln_wage = & [0.776 + 0.167] + [0.146 + 0.265]1.union \\ & + 0.077 grade + \varepsilon \end{aligned}$$

$$\ln_wage = `0.943 + 0.4111.union + 0.077 grade + \varepsilon$$

Untuk **idcode** = 2:

$$\begin{aligned} \mathbf{ln_wage} = & [0.776 - 0.183] + [0.146 + 0.126]\mathbf{1.union} \\ & + 0.077 \mathbf{grade} + \varepsilon \end{aligned}$$

$$\mathbf{ln_wage} = 0.593 + 0.272\mathbf{1.union} + 0.077 \mathbf{grade} + \varepsilon$$

dan seterusnya.

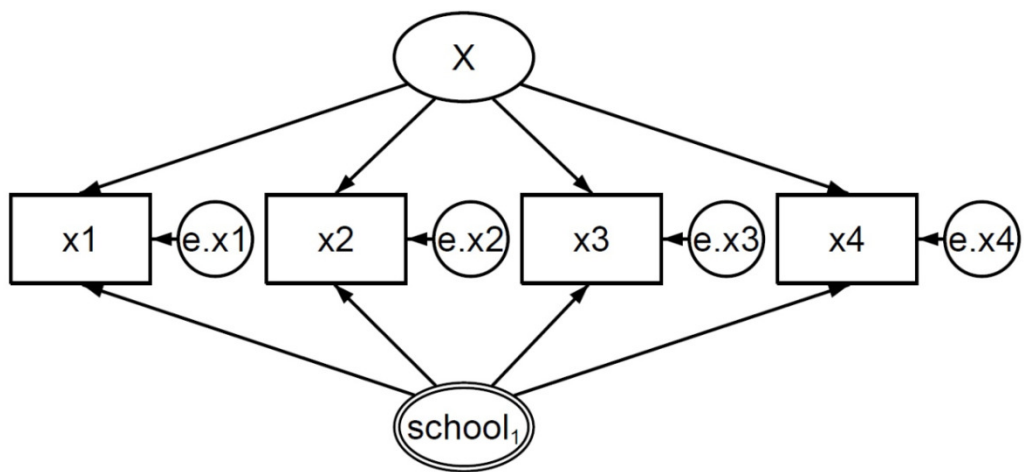
BAB 11

MULTILEVEL SEM II: ANALISIS FAKTOR KONFIRMATORIK

Model Multilevel CFA

Di bawah ini diperlihatkan sebuah contoh model multilevel CFA. X adalah variabel laten dengan 4 indikator, x_1 s.d. x_4 . Pengukuran dilakukan di beberapa sekolah, sehingga sekolah berfungsi sebagai grup. **school** dalam lingkaran ganda menyatakan variabel laten pada level **school** yang konstan **dalam** (*within*) satu sekolah dan bervariasi **antar** (*between*) antar sekolah. **school** dalam lingkaran ganda **tidak** menyatakan nomor **school**, melainkan variabel laten yang biasa dinamakan **M1[school]**.

Model CFA menggunakan analisis faktor dan bukan analisis regresi, sehingga pada Model Multilevel CFA ini tidak dikenal bentuk-bentuk *random intercept* dan *random slope*.



Perintah Stata:

. sem (X M1[school] -> x1 x2 x3 x4)

Variabel laten untuk **school** dalam lingkaran ganda dinamakan **M1[school]**.

Model Persamaan:

$$x1 = \alpha_1 + \beta_1 X + \gamma_1 M_{1,C} + \varepsilon_{x1}$$

$$x2 = \alpha_2 + \beta_2 X + \gamma_2 M_{1,C} + \varepsilon_{x2}$$

$$x3 = \alpha_3 + \beta_3 X + \gamma_3 M_{1,C} + \varepsilon_{x3}$$

$$x4 = \alpha_4 + \beta_4 X + \gamma_4 M_{1,C} + \varepsilon_{x4}$$

atau:

$$x_c = \alpha_c + \beta_c X + \gamma_c M_{1,c} + \varepsilon_{xc}$$

$c = 1, 2, \dots, C$ menyatakan nomor urut sekolah (grup).

Perhatikan bahwa pada Analisis Multilevel untuk model CFA ini γ_c tidak selalu bernilai sama dengan satu. Model *default* adalah $\gamma_1 = 1$ dan $\beta_2 = 1$, yaitu $M_{1,c}$ terjangkar pada (*anchored to*) x_1 dan X terjangkar pada x_2 , kecuali jika dinyatakan lain dalam perintah STATA.

Contoh 11.1: Kemampuan Matematika Siswa

Data: model-09_gsem-cfa.dta

```
. use "D:\Analisis Multilevel\Data\model-09_gsem-cfa"
```

```
(Fictional math abilities data)
```


. summarize

Variable		Obs		Mean	Std. Dev.	Min	Max
school		500		10.5	5.772056	1	20
id		500	50681.71	29081.41	71	100000	
q1		500		.506	.5004647	0	1
q2		500		.394	.4891242	0	1
q3		500		.534	.4993423	0	1
q4		500		.424	.4946852	0	1
q5		500		.49	.5004006	0	1
q6		500		.434	.4961212	0	1
q7		500		.52	.5001002	0	1
q8		500		.494	.5004647	0	1

. notes

_dta:

1. Fictional data on math ability of 500 students from 20 schools.
2. Variables q1-q8 are incorrect/correct (0/1) on individual math questions.

Keterangan:

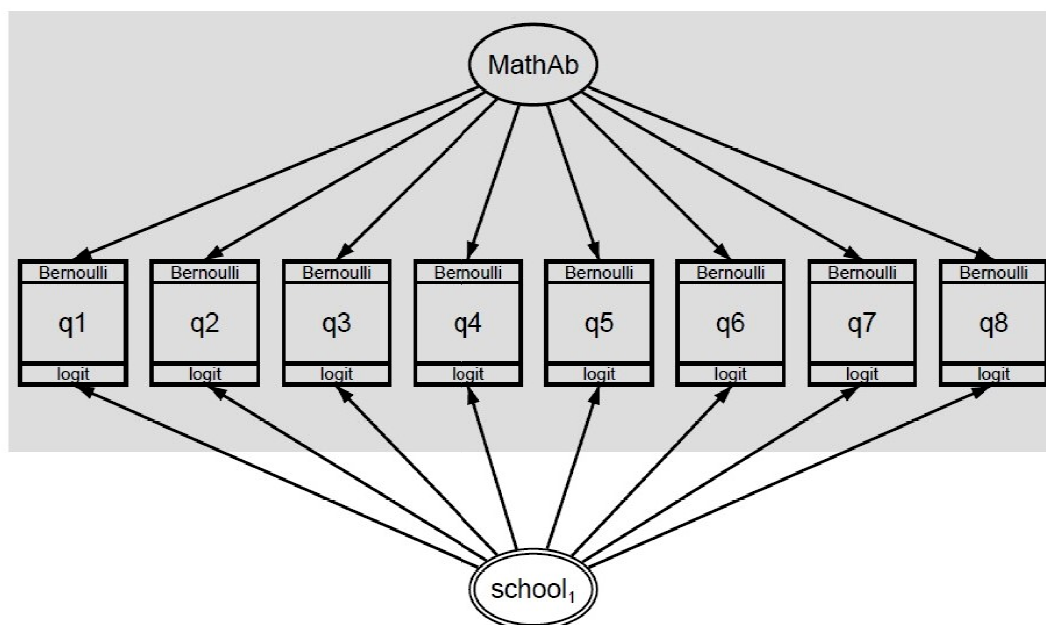
File ini memuat data fiktif kemampuan matematika 500 orang siswa dari 20 sekolah sebagai variabel laten dengan

indikatornya adalah 8 variabel **q1** s.d. **q8**, yang masing-masing merupakan variabel biner, sehingga regresi di sini akan dilakukan dengan model logit.

. list school id q1 q2 q3 in 1/10

	school	id	q1	q2	q3
1.	1	77764	Incorrect	Correct	Incorrect
2.	2	1843	Incorrect	Correct	Incorrect
3.	3	80226	Correct	Incorrect	Incorrect
4.	4	42412	Correct	Incorrect	Incorrect
5.	5	84980	Incorrect	Incorrect	Incorrect
6.	6	67589	Correct	Incorrect	Correct
7.	7	32921	Correct	Incorrect	Correct
8.	8	60192	Correct	Incorrect	Correct
9.	9	99227	Incorrect	Incorrect	Correct
10.	10	10017	Incorrect	Incorrect	Incorrect

Model:



Model Matematik:

$$\text{logit } q_c = \alpha_c + \beta_c \text{MathAb} + \gamma_c M1[\text{school}]$$

Perhatikan bahwa model regresi logit tidak memiliki suku galat di ruas kanan persamaan.

Perintah Stata:

. gsem (MathAb M1[school] -> q1-q8), logit

Fitting fixed-effects model:

Iteration 0: log likelihood = -2750.3114
Iteration 1: log likelihood = -2749.3709
Iteration 2: log likelihood = -2749.3708

Refining starting values:

Grid node 0: log likelihood = -2649.0033

Fitting full model:

Iteration 0: log likelihood = -2649.0033 (not concave)
Iteration 1: log likelihood = -2645.0613 (not concave)
Iteration 2: log likelihood = -2641.9755 (not concave)
Iteration 3: log likelihood = -2634.3857
Iteration 4: log likelihood = -2631.1111
Iteration 5: log likelihood = -2630.7898
Iteration 6: log likelihood = -2630.2477
Iteration 7: log likelihood = -2630.2402
Iteration 8: log likelihood = -2630.2074
Iteration 9: log likelihood = -2630.2063
Iteration 10: log likelihood = -2630.2063

Generalized structural equation model Number of obs = 500
Log likelihood = -2630.2063

```
( 1) [q1]M1[school] = 1
( 2) [q2]MathAb = 1
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
<hr/>						
q1 <-						
M1[school]	1 (constrained)					
MathAb	2.807515	.9468682	2.97	0.003	.9516878	4.663343
_cons	.0388021	.1608489	0.24	0.809	-.276456	.3540602
<hr/>						
q2 <-						
M1[school]	.6673925	.3058328	2.18	0.029	.0679712	1.266814
MathAb	1 (constrained)					
_cons	-.4631159	.1201227	-3.86	0.000	-.698552	-.2276798
<hr/>						
q3 <-						
M1[school]	.3555867	.3043548	1.17	0.243	-.2409377	.9521111
MathAb	1.455529	.5187786	2.81	0.005	.4387416	2.472316
_cons	.1537831	.1070288	1.44	0.151	-.0559894	.3635556
<hr/>						
q4 <-						
M1[school]	.7073241	.3419273	2.07	0.039	.037159	1.377489
MathAb	.8420897	.3528195	2.39	0.017	.1505762	1.533603
_cons	-.3252735	.1202088	-2.71	0.007	-.5608784	-.0896686
<hr/>						
q5 <-						
M1[school]	.7295553	.3330652	2.19	0.028	.0767595	1.382351
MathAb	2.399529	.8110973	2.96	0.003	.8098079	3.989251
_cons	-.0488674	.1378015	-0.35	0.723	-.3189533	.2212185
<hr/>						

q6 <-							
M1[school]		.484903	.2844447	1.70	0.088	-.0725983	1.042404
MathAb		1.840627	.5934017	3.10	0.002	.6775813	3.003673
_cons		-.3139302	.1186624	-2.65	0.008	-.5465042	-.0813563
<hr/>							
q7 <-							
M1[school]		.3677241	.2735779	1.34	0.179	-.1684787	.903927
MathAb		2.444023	.8016872	3.05	0.002	.8727449	4.015301
_cons		.1062164	.1220796	0.87	0.384	-.1330552	.3454881
<hr/>							
q8 <-							
M1[school]		.5851299	.3449508	1.70	0.090	-.0909612	1.261221
MathAb		1.606287	.5367614	2.99	0.003	.5542541	2.65832
_cons		-.0261962	.1189835	-0.22	0.826	-.2593995	.2070071
<hr/>							
var (M1[school])		.2121216	.1510032			.052558	.8561121
var (MathAb)		.2461246	.1372513			.0825055	.7342217
<hr/>							

Model Estimasi:

$$\text{logit } q_1 = 0.039 + 2.808\text{MathAb} + M1[\text{school}]$$

$$\text{logit } q_2 = -0.463 + \text{MathAb} + 0.667M1[\text{school}]$$

$$\text{logit } q_3 = 0.154 + 1.456\text{MathAb} + 0.356M1[\text{school}]$$

$$\text{logit } q_4 = -0.325 + 0.842\text{MathAb} + 0.707M1[\text{school}]$$

$$\text{logit } q_5 = -0.049 + 2.4\text{MathAb} + 0.73M1[\text{school}]$$

$$\text{logit } q_6 = -0.314 + 1.841\text{MathAb} + 0.485M1[\text{school}]$$

$$\text{logit } q_7 = 0.106 + 2.444\textit{MathAb} + 0.368\textit{M1}[\textit{school}]$$

$$\text{logit } q_8 = -0.026 + 1.606\textit{MathAb} + 0.585\textit{M1}[\textit{school}]$$

. predict stub*, latent

(option ebmeans assumed)

(using 7 quadrature points)

. list school stub1 stub2 in 1/10

	+	-----	+
		school	stub1 stub2

1.		1 1.030031	-.5106067
2.		2 .1873977	-.1208757
3.		3 -.060478	.4987507
4.		4 .2634546	-.3698531
5.		5 .2942677	-.473594

6.		6 .0873652	-.1947995
7.		7 -.5579194	.4085083
8.		8 .1776904	.3511564
9.		9 .3780018	.0950235
10.		10 .3415898	-.6816544

	+	-----	+

Berikut ini diperlihatkan hasil prediksi nilai-nilai **M1[school]** untuk 5 grup (**school**) pertama:

. tabstat stub1 if school<=5, by(school)

Summary for variables: stub1
by categories of: school (School id)

school	mean
-----+-----	
1	1.030031
2	.1873977
3	-.060478
4	.2634546
5	.2942677
-----+-----	
Total	.3429347

Model Estimasi untuk sekolah pertama:

$$\begin{aligned}
 \text{logit } q_1 &= 0.039 + 2.808\text{MathAb} + \text{M1[school]} \\
 &= 0.039 + 2.808\text{MathAb} + 1.030 \\
 &= 1.069 + 2.808\text{MathAb}
 \end{aligned}$$

$$\begin{aligned}
 \text{logit } q_2 &= -0.463 + \text{MathAb} + 0.667(1.030) \\
 &= -0.463 + \text{MathAb} + 0.687 \\
 &= 0.224 + \text{MathAb}
 \end{aligned}$$

$$\begin{aligned}\text{logit } q_3 &= 0.154 + 1.456\text{MathAb} + 0.356(1.030) \\ &= 0.521 + 1.456\text{MathAb}\end{aligned}$$

$$\begin{aligned}\text{logit } q_4 &= -0.325 + 0.842\text{MathAb} + 0.707(1.030) \\ &= 0.403 + 0.842\text{MathAb}\end{aligned}$$

$$\begin{aligned}\text{logit } q_5 &= -0.049 + 2.4\text{MathAb} + 0.73(1.030) \\ &= 0.703 + 2.4\text{MathAb}\end{aligned}$$

$$\begin{aligned}\text{logit } q_6 &= -0.314 + 1.841\text{MathAb} + 0.485(1.030) \\ &= 0.186 + 1.841\text{MathAb}\end{aligned}$$

$$\begin{aligned}\text{logit } q_7 &= 0.106 + 2.444\text{MathAb} + 0.368(1.030) \\ &= 0.485 + 2.444\text{MathAb}\end{aligned}$$

$$\begin{aligned}\text{logit } q_8 &= -0.026 + 1.606\text{MathAb} + 0.585(1.030) \\ &= 0.577 + 1.606\text{MathAb}\end{aligned}$$

Model Estimasi untuk sekolah kedua:

$$\begin{aligned}\text{logit } q_1 &= 0.039 + 2.808\text{MathAb} + 0.187 \\ &= 0.226 + 2.808\text{MathAb}\end{aligned}$$

$$\begin{aligned}\text{logit } q_2 &= -0.463 + \text{MathAb} + 0.667(0.187) \\ &= -0.338 + \text{MathAb}\end{aligned}$$

$$\begin{aligned}\text{logit } q_3 &= 0.154 + 1.456\textit{MathAb} + 0.356(0.187) \\ &= 0.221 + 1.456\textit{MathAb}\end{aligned}$$

$$\begin{aligned}\text{logit } q_4 &= -0.325 + 0.842\textit{MathAb} + 0.707(0.187) \\ &= -0.193 + 0.842\textit{MathAb}\end{aligned}$$

$$\begin{aligned}\text{logit } q_5 &= -0.049 + 2.4\textit{MathAb} + 0.73(0.187) \\ &= 0.088 + 2.4\textit{MathAb}\end{aligned}$$

$$\begin{aligned}\text{logit } q_6 &= -0.314 + 1.841\textit{MathAb} + 0.485(0.187) \\ &= -0.223 + 1.841\textit{MathAb}\end{aligned}$$

$$\begin{aligned}\text{logit } q_7 &= 0.106 + 2.444\textit{MathAb} + 0.368(0.187) \\ &= 0.175 + 2.444\textit{MathAb}\end{aligned}$$

$$\begin{aligned}\text{logit } q_8 &= -0.026 + 1.606\textit{MathAb} + 0.585(0.187) \\ &= 0.084 + 1.606\textit{MathAb}\end{aligned}$$

dan seterusnya.

KEPUSTAKAAN

- Brown H, Prescott R. **Applied Mixed Models in Medicine**, 3rd Ed. Chichester: John Wiley & Sons, 2015.
- Gelman A, Hill J. **Data Analysis Using Regression and Multilevel/Hierarchical Models**. Cambridge: Cambridge University Press, 2007.
- Gill J, Womack AJ. **The Multilevel Model Framework**. In: MA Scott et al (eds). *The SAGE Handbook of Multilevel Modeling*. Los Angeles: SAGE Publications, 2013, pp 3-20.
- Goldstein H. **Multilevel Statistical Models**, 4th Ed. Chichester: John Wiley & Sons, 2011.
- Hox JJ. **Multilevel Analysis: Techniques and Applications**, 2nd Ed. New York: Routledge, 2010.
- Hox J, van de Schoot R. **Robust Methods for Multilevel Analysis**. In: MA Scott et al (eds). *The SAGE Handbook of Multilevel Modeling*. Los Angeles: SAGE Publications, 2013, pp 387-402.
- Moerbeek M, Teerenstra S. **Power Analysis of Trials with Multilevel Data**. Boca Raton: CRC Press, 2016.
- Rabe-Hesketh S, Skrondal A, Zheng X. **Multilevel Structural Equation Modeling**. In: RH Hoyle (ed). *Handbook of Structural Equation Modeling*. New York: The Guilford Press, 2012, pp 512-531.
- Rindskopf D. **Multilevel Models in the Social and Behavioral Sciences**. In: MA Scott et al (eds). *The SAGE Handbook of Multilevel Modeling*. Los Angeles: SAGE Publications, 2013, pp 521-539.

- Schuetz CG. **Multilevel Business Processes: Modeling and Data Analysis**. Wiesbaden: Springer, 2015.
- Snijders TAB, Bosker RJ. **Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling**. London: SAGE Publication, 2003.
- Stroup WW. **Generalized Linear Mixed Models: Modern Concepts, Methods and Applications**. Boca Raton: CRC Press, 2013.
- Van Breukelen G, Moerbeek M. **Design Considerations in Multilevel Studies**. In: MA Scott et al (eds). *The SAGE Handbook of Multilevel Modeling*. Los Angeles: SAGE Publications, 2013, pp 183-199.
- Verbeke G, Molenberghs G. **Linear Mixed Models for Longitudinal Data**. New York: Springer-Verlag, 2000.
- Wasserstein RL, Lazar NA. "The ASA's Statement on p -Values: Context, Process, and Purpose". **The American Statistician**, Vol 70, Issue 2, 2016, pp 129-133.

LAMPIRAN 1

UKURAN SAMPEL

Misalkan dimiliki model regresi linear sederhana:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (1)$$

dengan X berskala biner; $X_i = 1, 2$; untuk membandingkan 2 kelompok perlakuan [$X_i = 1$ vs $X_i = 2$].

Misalkan uji signifikansi antar kedua kelompok perlakuan akan dilakukan dengan tingkat signifikansi α dan diharapkan memiliki kekuatan uji (*power*) sebesar $(1 - \beta)$. Misalkan pula variansi kedua kelompok sama [$\sigma_1^2 = \sigma_2^2 = \sigma^2$] dan ukuran kedua kelompok juga sama [$n_1 = n_2 = n$]. Maka ukuran sampel minimum 1 kelompok n untuk mendeteksi selisih efek sebesar $(\mu_1 - \mu_2)$ dengan prosedur sampling acak sederhana adalah:

$$n = \frac{2\sigma^2 (Z_\alpha + Z_\beta)^2}{(\mu_1 - \mu_2)^2} \quad (2)$$

Ukuran sampel seluruhnya adalah $2n$ yang dibulatkan ke atas ke bilangan genap terdekat. Jika sampel diperoleh dengan prosedur sampling acak klaster dua-tahap, maka model regresi linearnya menjadi:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \mu_{0j} + \mu_{1j} X_{ij} + \varepsilon_{ij} \quad (3)$$

j menyatakan nomor klaster (grup);

$$\text{atau: } Y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j}) X_{ij} + \varepsilon_{ij} \quad (3.a)$$

Untuk menyederhanakan pembahasan, digunakan model yang bervariasi hanya pada intersep-nya, yaitu:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \mu_{0j} + \varepsilon_{ij} \quad (4)$$

atau:
$$Y_{ij} = (\beta_0 + \mu_{0j}) + \beta_1 X_{ij} + \varepsilon_{ij} \quad (4.a)$$

Di sini ada ukuran grup (jumlah subjek per grup) n_1 dan ada jumlah grup n_2 , sedemikian hingga ukuran sampel seluruhnya n adalah sama dengan $n_1 \times n_2$.

Pengaruh penggunaan prosedur sampling acak kluster terhadap ukuran sampel dinamakan “**efek desain**” (*design effect*), yaitu:

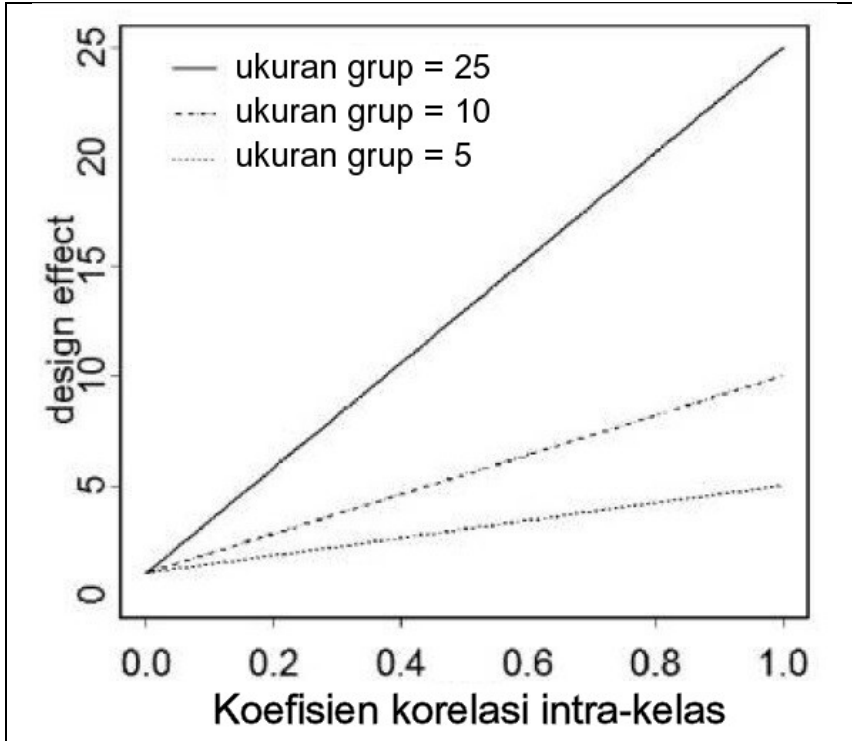
$$de = 1 + (n_1 - 1) \rho \quad (5)$$

ρ adalah **koefisien korelasi intra-kelas**, yaitu:

$$\rho = \frac{\sigma_{uo}^2}{\sigma_{u0}^2 + \sigma_e^2} \quad (6)$$

$$\sigma_{uo}^2 = \text{Var}(\mu_{0j}) \quad \text{dan} \quad \sigma_e^2 = \text{Var}(\varepsilon_{ij})$$

Jika $\rho = 1$, maka seluruh subjek dalam sebuah grup memberi respons identik; jika $\rho = 0$ maka subjek dalam sebuah grup tak lebih berkorelasi daripada subjek antar-grup.



Gambar 1 Efek desain sebagai fungsi koefisien korelasi intra-kelas dan ukuran grup

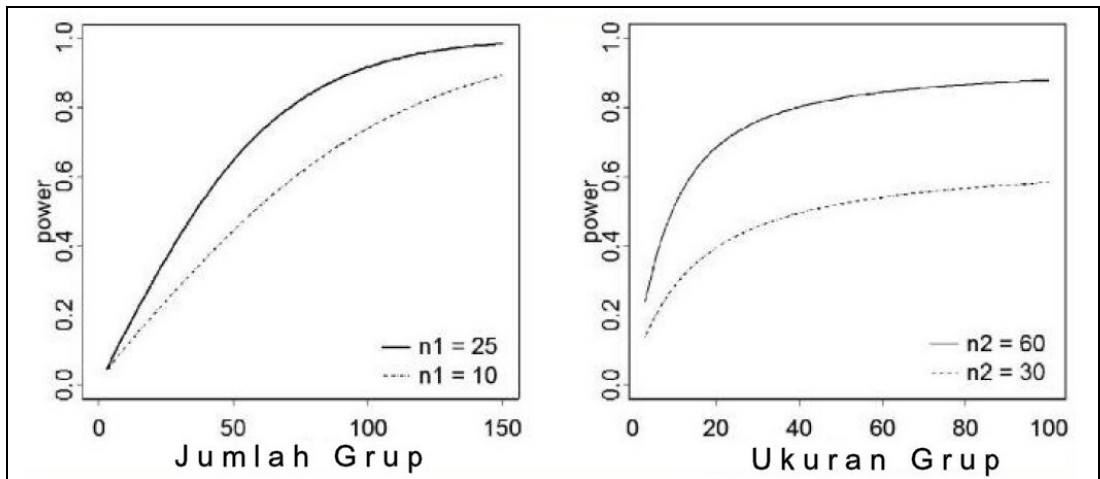
Jika dengan prosedur sampling acak sederhana ukuran sampling minimum yang dibutuhkan adalah n , maka dengan prosedur sampling acak kluster dua-tahap di atas ukuran sampel minimum yang dibutuhkan harus diperbesar dengan perkalian dengan efek desain, yaitu $[1 + (n_1 - 1) \rho] n$.

Jumlah grup minimum n_2 yang dibutuhkan pada sampling acak kluster dua-tahap adalah:

$$\begin{aligned}
 n_2 &= 4 \frac{\sigma_e^2 + n_1 \sigma_u^2}{n_1} \left(\frac{Z_\alpha + Z_\beta}{\mu_1 - \mu_2} \right)^2 \\
 &= 4 \frac{1 + (n_1 - 1) \rho}{n_1} \left(\frac{Z_\alpha + Z_\beta}{Z_{\mu_1 - \mu_2}} \right)^2 \quad (7)
 \end{aligned}$$

Jika jumlah grup terbatas, maka yang perlu dihitung adalah ukuran grup:

$$\begin{aligned}
 n_1 &= \frac{4\sigma_e^2}{\left(\frac{\mu_1 - \mu_2}{Z_\alpha + Z_\beta}\right)^2 n_2 - 4\sigma_u^2} \\
 &= \frac{4(1-\rho)}{\left(\frac{Z_{\mu_1 - \mu_2}}{Z_\alpha + Z_\beta}\right)^2 n_2 - 4\rho} \quad (8)
 \end{aligned}$$



Gambar 2 Jumlah grup, ukuran grup, dan kekuatan uji pada regresi multilevel

Contoh 1

Misalkan hendak diteliti efektivitas program intervensi yang bertujuan mengurangi masalah pada anak dari ibu dengan riwayat pencarian perlindungan dari kekerasan domestik. Diambil sampel 1 anak dari tiap ibu dengan sampling acak sederhana. Selisih rerata efek minimum yang diharapkan untuk dideteksi antara kelompok intervensi dan non-intervensi adalah 5.5.

Variansi kedua kelompok dapat dianggap sama dengan reratanya adalah 213.03. Kekuatan uji yang diinginkan adalah 0.90 dengan kesalahan tipe I maksimum 0.05.

$$\alpha = 0.05 \quad \text{dan} \quad 1 - \beta = 0.90$$

$$Z_{\alpha} = 1.64 \quad \text{dan} \quad Z_{\beta} = 1.28$$

$$(\mu_1 - \mu_2) = 5.5 \quad \sigma_1^2 = \sigma_2^2 = \sigma^2 = 213.03$$

Ukuran sampel minimum untuk 1 kelompok dengan sampling acak sederhana adalah:

$$\begin{aligned} n &= \frac{2\sigma^2(Z_{\alpha} + Z_{\beta})^2}{(\mu_1 - \mu_2)^2} \\ &= \frac{2(213.03)(1.64 + 1.28)^2}{5.5^2} = 120.50 \approx 121 \end{aligned}$$

Contoh 2

Lihat kembali data pada Contoh 1. Misalkan pengumpulan ibu untuk sampel dilakukan dengan sampling acak kluster dua-tahap. Dari tiap grup diambil 10 orang ibu. Jika koefisien korelasi intra-kelas $\rho = 0.05$, maka efek desain adalah:

$$\begin{aligned} de &= 1 + (n_1 - 1)\rho \\ &= 1 + (10 - 1)0.05 = 1.45 \end{aligned}$$

Pada contoh 1 telah dihitung ukuran sampel minimum 1 kelompok dengan sampling acak sederhana, yaitu 121 ibu. Dengan sampling acak kluster dua-tahap, ukuran minimum 1 kelompok adalah:

$$n = (1.45)(121) = 175.45 \approx 176$$

Contoh 3

Misalkan hendak dipelajari efektivitas intervensi luar-sekolah terhadap sikap siswa mengenai kebiasaan merokok. Dengan sampling acak kluster dua-tahap, variansi respons pada level siswa dan level sekolah masing-masing adalah 62 dan 8. Selisih rerata respons antara kelompok intervensi dan non-intervensi yang diharapkan untuk dideteksi adalah 2. Akan dilakukan uji signifikansi multilevel dengan kekuatan uji yang diinginkan adalah 0.8 dan tingkat signifikansi 0.05.

$$\sigma_e^2 = 62 \qquad \sigma_u^2 = 8$$

Koefisien korelasi intra-kelas adalah:

$$\begin{aligned}\rho &= \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \\ &= \frac{8}{8+62} = 0.11\end{aligned}$$

Selisih rerata respons yang diharapkan untuk dideteksi dan selisih terstandarisasinya masing-masing adalah:

$$\begin{aligned}(\mu_1 - \mu_2) &= 2 \\ Z_{(\mu_1 - \mu_2)} &= \frac{\mu_1 - \mu_2}{\sqrt{\sigma_u^2 + \sigma_e^2}} \\ &= \frac{2}{\sqrt{8+62}} = 0.24\end{aligned}$$

Jika 80 sekolah akan direkrut untuk penelitian, maka jumlah siswa yang akan diambil per sekolah adalah:

$$n_1 = \frac{4(1-\rho)}{\left(\frac{Z_{\mu_1 - \mu_2}}{Z_\alpha + Z_\beta} \right)^2} n_2 - 4\rho$$

$$= \frac{4(1-0.11)}{\left(\frac{0.24}{1.64+0.84}\right)^2 80-4(0.11)} = 11.63 \approx 12$$

Contoh 4

Lihat kembali data pada contoh 3. Seandainya jumlah siswa per sekolah yang ditentukan sebanyak 20 orang, sedangkan jumlah sekolah yang akan direkrut ditentukan melalui perhitungan, maka perhitungannya adalah sebagai berikut:

$$n_1 = 20$$

$$\begin{aligned} n_2 &= 4 \frac{1+(n_1-1)\rho}{n_1} \left(\frac{Z_\alpha + Z_\beta}{Z_{\mu_1 - \mu_2}} \right)^2 \\ &= 4 \frac{1+(20-1)(0.11)}{20} \left(\frac{1.64+0.84}{0.24} \right)^2 = 66.26 \approx 68 \end{aligned}$$

Perhatikan bahwa untuk perbandingan dua kelompok, jumlah grup harus dibulatkan ke atas ke bilangan genap terdekat.

LAMPIRAN 2

BEBERAPA PRINSIP TENTANG NILAI-P

(*American Statistical Association*, 2016)

Atas dasar banyaknya interpretasi yang salah mengenai nilai- p serta pemahaman kemaknaan statistik yang tidak benar dalam literatur ilmiah, *American Statistical Association* (ASA; 2016) telah mengeluarkan pernyataan resmi mengenai kemaknaan statistik dan nilai- p . Beberapa prinsip yang dikemukakan mengenai nilai- p yaitu:

1. **Nilai- p dapat mengindikasikan seberapa jauh data tak kompatibel dengan model statistik yang dispesifikasikan.**

Nilai- p merupakan salah satu pendekatan untuk meringkas inkompatibilitas antara himpunan data tertentu dengan model yang diajukan untuk data tersebut. Konteks yang paling umum adalah sebuah model, yang disusun di bawah sejumlah asumsi, bersama dengan yang disebut sebagai “hipotesis nol”. Seringkali hipotesis nol mempostulatkan ketiadaan suatu efek, seperti tidak adanya perbedaan antara dua kelompok, atau tidak adanya hubungan antara faktor dengan respons. Semakin kecil nilai- p , semakin besar inkompatibilitas data dengan hipotesis nol, jika asumsi-asumsi yang mendasari perhitungan nilai- p benar. Inkompatibilitas ini dapat diinterpretasikan sebagai pernyataan keragu-raguan atau bukti penentangan terhadap hipotesis nol ataupun asumsi-asumsi yang mendasarinya.

2. **Nilai- p tidak mengukur probabilitas bahwa hipotesis studi benar, atau probabilitas bahwa data semata dihasilkan oleh peluang acak.**

Peneliti acapkali berkeinginan mentransformasikan nilai- p menjadi pernyataan mengenai kebenaran hipotesis nol, atau probabilitas bahwa data yang diamati dihasilkan oleh peluang acak. Nilai- p bukan merupakan keduanya ini. Nilai- p merupakan pernyataan tentang data sehubungan dengan hipotesis penjelasan yang diajukan, dan bukan merupakan pernyataan tentang penjelasan itu sendiri.

3. **Kesimpulan ilmiah dan keputusan bisnis atau kebijakan tidak boleh semata didasarkan atas fakta apakah nilai- p melampaui suatu ambang tertentu.**

Praktik mereduksi analisis data atau inferensi statistik menurut aturan mekanistik dengan “batas-nyata” (seperti “ $p < 0.05$ ”) untuk membenarkan klaim atau kesimpulan ilmiah dapat menyebabkan kepercayaan yang salah atau pengambilan keputusan yang buruk. Suatu kesimpulan akan tidak langsung menjadi “benar” di satu sisi pembagian dan “salah” di sisi lainnya. Peneliti harus menyajikan berbagai faktor kontekstual sebelum sampai pada inferensi ilmiah, termasuk desain studi, kualitas pengukuran, bukti-bukti eksternal tentang fenomena yang dipelajari, serta validitas asumsi yang mendasari analisis data. Pertimbangan pragmatik sering membutuhkan keputusan biner, “ya-tidak”, namun hal ini tak berarti bahwa nilai- p semata dapat memastikan apakah suatu keputusan benar atau salah. Penggunaan “kemaknaan statistik” secara meluas (umumnya diinterpretasikan sebagai “ $p \leq 0.05$ ”) sebagai lisensi untuk mengklaim temuan ilmiah (atau kebenaran tersirat) menyebabkan distorsi yang parah terhadap proses ilmiah.

4. **Inferensi yang benar memerlukan pelaporan lengkap dan transparansi.**

Nilai- p dan analisis yang berkaitan tak boleh dilaporkan secara selektif. Melakukan analisis ganda terhadap data dan hanya melaporkan yang memiliki nilai- p tertentu (secara tipikal yang melampaui ambang kemaknaan) menyebabkan nilai- p yang dilaporkan secara esensial tak dapat diinterpretasikan. *Cherry-picking* (menampilkan hanya butir-butir yang mendukung pendapat peneliti) yang menjanjikan temuan, yang juga dikenal dengan berbagai istilah seperti *data dredging* (mencari butir-butir bermakna tanpa terlebih dahulu mengajukan pendapat peneliti sendiri), *significance chasing* (perburuan kemaknaan), *significance questing* (pencarian kemaknaan), *selective inference* (inferensi selektif), dan “*p-hacking*” (peretasan nilai- p agar sesuai pendapat peneliti), membawa pada eksese hasil-hasil yang bermakna secara statistik namun penuh kepalsuan dalam literatur yang dipublikasikan dan harus sungguh-sungguh dihindari. Agar masalah ini tidak terjadi, kita tidak perlu secara formal melaksanakan analisis ganda: Apabila seorang peneliti memilih apa yang akan dipresentasikan berdasarkan hasil statistik, validitas interpretasi hasilnya akan sangat menurun jika pembaca tak diinformasikan mengenai pilihan dan dasarnya. Peneliti harus menjelaskan seluruh hipotesis yang dieksplorasi dalam studi, seluruh keputusan pengumpulan data, seluruh analisis statistik yang dilakukan, dan seluruh nilai- p yang dihitung. Kesimpulan ilmiah yang valid berdasarkan nilai- p dan statistik yang berkaitan tak dapat ditarik tanpa paling sedikit mengetahui seberapa banyak dan analisis apa saja yang dikerjakan, serta bagaimana analisis ini (termasuk nilai- p -nya) dipilih untuk pelaporan.

5. Nilai- p , atau kemaknaan statistik, tidak mengukur besar efek atau derajat kepentingan suatu hasil.

Kemaknaan statistik tidak ekuivalen dengan kemaknaan ilmiah, kemanusiaan, ataupun ekonomi. Nilai- p yang lebih kecil tidak harus menyiratkan keberadaan efek yang lebih besar atau lebih penting, sedangkan nilai- p yang lebih besar tidak menyiratkan kurangnya atau bahkan tidak adanya efek. Tiap efek, sebagaimana kecil pun, dapat menghasilkan nilai- p yang kecil jika ukuran sampel atau presisi pengukuran cukup tinggi, dan efek yang besar dapat menghasilkan nilai- p yang tak bermakna jika ukuran sampel kecil atau pengukuran tidak tepat. Begitu pula, estimasi efek yang identik akan menunjukkan nilai- p yang berbeda-beda jika presisi estimasinya berbeda-beda.

6. Secara sendiri, nilai- p bukan merupakan ukuran pembuktian yang memadai mengenai suatu model atau hipotesis.

Peneliti seharusnya menyadari bahwa nilai- p tanpa konteks atau bukti lain hanya memberikan informasi terbatas. Misalnya, nilai- p yang mendekati 0.05 semata hanya memberikan bukti lemah untuk menentang hipotesis nol. Demikian pula, nilai- p yang relatif besar tidak menyiratkan bukti untuk mendukung hipotesis nol; ada banyak hipotesis lain yang mungkin sama atau lebih konsisten dengan data yang diamati. Berdasarkan alasan-alasan ini, analisis data tidak boleh diakhiri dengan perhitungan nilai- p jika masih ada pendekatan-pendekatan lain yang relevan dan layak.

