SKRIPSI

PEMODELAN REGRESI MULTILEVEL MENGGUNAKAN METODE *RESTRICTED MAXIMUM LIKELIHOOD*

Disusun dan diajukan oleh

SITTI CHAERUNNISA AULIA M

H121 16 504

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2021

PEMODELAN REGRESI MULTILEVEL MENGGUNAKAN METODE *RESTRICTED MAXIMUM LIKELIHOOD*

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

SITTI CHAERUNNISA AULIA M

H121 16 504

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2021

LEMBAR PENGESAHAN SKRIPSI

PEMODELAN REGRESI MULTILEVEL MENGGUNAKAN METODE *RESTRICTED MAXIMUM LIKELIHOOD*

Disusun dan diajukan oleh:

SITTI CHAERUNNISA AULIA M

H121 16 504

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin pada tanggal 31 Mei 2021 dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,	Pembimbing Pendamping
Dr. Erna Tri Herdiani, M.Si.	**Dr. Nirwan Ilyas, M.Si.**
NIP. 19750429 200003 2 001	**NIP. 19630306 198702 1002**

Ketua Program Studi

KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN UNIVERSITAS HASANUDDIN DEPARTEMEN STATISTIKA FAKULTAS MIPA

Dr. Nurtiti Sunusi, M.Si.

NIP. 19720117 199703 2002

PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini :

Nama	: Sitti Chaerunnisa Aulia M
NIM	: H 121 16 504
Program Studi	: Statistika
Jenjang	: S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

Pemodelan Regresi Multilevel Menggunakan Metode *Restricted Maximum Likelihood*

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain bahwa skripsi yang saya tulis ini benar benar merupakan hasil karya saya sendiri.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut

Makassar, 31 Mei 2021

Yang menyatakan

Sitti Chaerunnisa Aulia M

KATA PENGANTAR

Segala puji bagi Allah *Subhanahu Wata'ala*, sholawat dan salam semoga selalu dilimpahkan kepada baginda Rasulullah *Shallallahu'alaihi Wasallam* nabi yang paling mulia, pemimpin orang-orang takwa, dan kepada para keluarga dan sahabat beliau. Alhamdulillahirobbil'alamiin, berkat rahmat dan hidayah-NYA sehingga penulis dapat menyelesaikan skripsi dengan judul **"Pemodelan Regresi Multilevel Menggunakan Metode *Restricted Maximum Likelihood*"** yang disusun sebagai salah satu syarat akademik untuk meraih gelar sarjana pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin. Penulis berharap skripsi ini dapat memberikan tambahan pengetahuan bagi pembelajar statistika dan bagi pembaca secara umum.

Penulis menyadari dalam penulisan skripsi ini tidak lepas dari hambatan dan masalah namun dapat terselesaikan berkat dukungan dan bantuan berbagai pihak. Oleh karena itu, penulis haturkan rasa terima kasih yang setulus-tulusnya serta penghargaan yang setinggi-tingginya untuk orang tua penulis, Ayahanda dan Ibunda tercinta **Muliharram** dan **Marliati**, yang telah membesarkan dan mendidik dengan penuh kesabaran, memberikan cinta dan limpahan kasih sayang, dukungan dan doa yang tulus tanpa henti kepada penulis. Rasa terima kasih juga kepada saudara-saudara tersayang **Luthfiah** dan **Chosiah**, serta **Keluarga Besar** atas doa, dukungan, semangat, dan bantuannya kepada penulis.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada :

1. **Ibu Prof. Dr. Dwia Aries Tina Palubuhu, MA** selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin, M.Si.** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Nurtiti Sunusi, S.Si., M.Si**. selaku Ketua Departemen Statistika, segenap Dosen Pengajar dan Staf Departemen Statistika yang telah membekali ilmu dan kemudahan-kemudahan kepada Penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
4. **Ibu Dr. Erna Tri Herdiani, M.Si.** selaku Pembimbing Utama dan Penasihat Akademik atas saran, nasehat, motivasi, dan dukungan yang telah diberikan

kepada penulis selama menjadi mahasiswa dan meluangkan waktu untuk membimbing dan memberikan masukan dalam penulisan skripsi ini. **Bapak Dr. Nirwan Ilyas, M.Si**. selaku Pembimbing Pendamping yang dengan sabar meluangkan waktu dan pemikirannya untuk membimbing dan memberikan masukan dalam penulisan skripsi ini.

5. **Ibu Sitti Sahriman, S.Si., M.Si.** dan **Bapak Siswanto, S.Si., M.Si.** selaku Tim Penguji atas saran dan kritikan yang membangun dalam penyempurnaan penyusunan skripsi ini serta waktu yang telah diberikan kepada Penulis.
6. Sahabat Mipaku **Asti, Diki, Fahmi, Grace, Mamik, Risma, Rizki** Terima kasih telah menjadi teman perkuliahan sejak semester 1, tempat belajarku sekaligus tempat pergibahanku yang menjadi pewarna selama perkuliahan.
7. Sahabat 576 ku **Ade, Adel, Anggi, Caca, Cindy, Dea, Deby, Ines, Itin, Ulfa.** Terima kasih telah menjadi tempat hahahihihuhu selama menyusun tugas akhir dan 11 tahun terakhir.
8. +1 ku **Akhmad** Terima kasih selalu mendorong untuk tidak menunda-nunda skripsi dan selalu mengingatkan *deadline* penulis sekaligus mejadi tempat berkeluh kesahku.
9. Sobat-Sobat Labku **Aco, Andis, Idman, Ita, Mila, Nidar, Sultan, Surit, Ririn, Tari, Agung, Atiek** terimakasih mewarnai hari-hariku di lab dengan nanyian-nyanyian nada terjun bebas selama di lab dan membantu ketika ada sesuatu yang sulit penulis kerjakan terutama untuk **Aco** yang telah banyak membantu dalam penyususnan skripsi.
10. Teman seperjuangan di **Statistika 2016**, terima kasih atas ilmu, kebersamaan, dan dukungannya selama ini.
11. Teman-teman **KKN Bantaeng Kelurahan Bonto Jaya Gelombang 102**. Terima kasih telah menjadi teman sekaligus keluarga selama sebulan lebih, semoga silaturahmi tetap terjalin.
12. Semua pihak yang telah banyak berpartisipasi, baik secara langsung maupun tidak langsung yang tak sempat penulis sebutkan satu per satu. Terima kasih untuk segala bantuan dan dukungannya.

Penulis menyadari bahwa masih banyak kekurangan dalam skripsi ini. Untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi pembaca.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Makassar, 31 Mei 2021

Sitti Chaerunnisa Aulia M

ABSTRAK

Dalam penelitian bidang pendidikan, pembentukan dan perkembangan perilaku individu dalam memperoleh pendidikan dipengaruhi oleh lingkungannya. Secara umum lingkungan dan individu memiliki stuktur berjenjang (hierarki). Untuk menganalisis struktur data hierarki dapat dilakukan dengan pemodelan regresi multilevel. Pada penelitian ini digunakan Ujian Nasional (UN) sebagai sistem evaluasi standar pendidikan di Indonesia. Peubah yang digunakan pada level individu adalah jumlah peserta UN, jumlah peserta didik dan jumlah guru tiap-tiap sekolah sedangkan pada level lingkungan adalah Angka Partitispasi Kasar (APK) dan Angka Partisipasi Murni (APM) kabupaten/kota di Sulawesi Selatan. Pembentukan model regresi multilevel pada penelitian ini untuk menentukan faktor-faktor yang mempengaruhi nilai UN SMP dengan memperhitungkan adanya varians antar sekolah dan kabupaten/kota di Provinsi Sulawesi Selatan. Metode estimasi parameter yang digunakan pada pemodelan ini adalah *Restricted Maximum Likelihood* (REML).

Kata Kunci : Ujian Nasional, hierarki, regresi multilevel, *Restricted Maximum Likelihood*

ABSTRACK

In educational research, self development of individual behavior in obtaining education is influenced by their environment. In general, environments and individuals have a tiered structure (hierarchy). To analyze the hierarchical structure of the data, multilevel regression modeling can be done. In this study, the National Examination (UN) is used as an evaluation system for education standards in Indonesia. The variables used at the individual level are the number of UN participants, the number of students and the number of teachers in each school, while at the environmental level are the Gross Participation Rate (GPR) and Net Enrollment Rate (NER) of districts/cities in South Sulawesi. The parameter estimation method used in this modeling is the Restricted Maximum Likelihood (REML). From the modeling results obtained a random coefficient model is the best model and the factors that affect the average school UN score are the number of UN participants, the number of students, the number of teachers, APK and APM with the variance of the average UN score explained by the variable level-1 and level-2 are 6.1% and 70.1%, respectively.

Keywords: Ujian Nasional, hierarchy, regresi multilevel, *Restricted Maximum Likelihood*

DAFTAR ISI

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi linear adalah metode statistik yang digunakan untuk melihat hubungan antara dua atau lebih peubah. Analisis regresi linear merupakan metode yang dapat digunakan sebagai pengambilan keputusan tentang hubungan antara peubah penjelas dengan peubah respon. Hubungan yang dimaksud dalam hal ini, berupa hubungan sebab-akibat antar kedua peubah tersebut (Drapper & Smith, 1998). Terdapat dua macam analisis regresi linear yaitu analisis regresi linear sederhana dan analisis regresi berganda. Perkembangan metode analisis regresi telah mengalami berbagai kemajuan yang pesat, baik dalam aspek metode estimasi maupun variasi data yang digunakan. Salah satu pengembangan dari analisis regresi linear sederhana yaitu model multilevel.

Model multilevel diperkenalkan oleh Goldstein (1995) yang menyebutkan bahwa model multilevel dapat mengatasi masalah yang muncul dari data dengan struktur hierarki atau berjenjang. Dalam struktur hierarki, individu-individu dalam kelompok yang sama memiliki karakteristik yang cenderung mirip, dengan kata lain antar amatan pada level yang lebih rendah tidak saling bebas, jika pelanggaran asumsi ini diabaikan maka akan mengakibatkan terjadinya pelanggaran asumsi kebebasan dalam pendekatan statistika konvensional (Hox, 2018). Pada pemodelan multilevel, peubah respon diukur pada level terendah, sedangkan peubah penjelas dapat didefinisikan pada setiap level. Model multilevel yang paling sederhana adalah model dua level dengan level kesatu adalah data individu dan level kedua adalah data kelompok (West et al. 2007).

Dalam penelitian sosial, kesehatan, ekonomi dan pendidikan sering terkonsentrasi pada masalah untuk menelusuri hubungan antar individu dan lingkungannya. Misalnya dalam penelitian bidang pendidikan, dalam memperoleh pendidikan suatu individu berkorelasi dengan lingkungan, misalnya, pada kota/kabupaten tempat mereka tinggal merupakan salah satu faktor yang mempengaruhi pembentukan dan perkembangan perilaku individu, baik fisik maupun sosial-psikologis (Tantular, 2009). Secara umum, individu dan kota/kabupaten merupakan struktur yang berjenjang atau hierarki. Populasi yang berjenjang mengindikasikan terdapatnya tingkatan atau level dalam suatu data. Salah satu bentuk struktur data berjenjang pada data pendidikan adalah rata-rata skor Ujian Nasional (Zulvia, 2017).

UN merupakan salah satu sistem evaluasi standar pendidikan di Indonesia. Pada peraturan pemerintah Nomor 13 Tahun 2015 (Puspendik, 2015) hasil Ujian Nasional (UN) dapat digunakan sebagai salah satu bahan pertimbangan untuk pembinaan dan pemberian bantuan kepada satuan pendidikan dalam upaya peningkatan mutu pendidikan. Dengan adanya hasil UN diharapkan akan memberikan gambaran peta mutu pendidikan pada tingkat nasional, provinsi, kabupaten/kota, dan sekolah. Sedangkan harapan secara khusus, hasil yang diperoleh dari masing-masing mata pelajaran dapat menjadi perbaikan proses pembelajaran pada tingkat sekolah. Rata-rata skor UN yang diperoleh pada setiap sekolah dapat dipengaruhi oleh faktor internal dan eksternal. Faktor internal seperti kualitas siswa, kompetensi guru, sarana dan prasarana sekolah. Sedangkan faktor eksternal dapat dilihat dari kondisi daerah seperti, Angka Partisipasi Kasar (APK), Angka Partisipasi Murni (APM), Angka Kelulusan (AL), Angka Melek Huruf (AMH), Rata-Rata Lama Sekolah (RLS) dan Angka Harapan Lama Lulus (AHLS) (Kemendikbud, 2013).

Penelitian mengenai data pendidikan telah dilakukan oleh (Tantular, 2009) yang menganalisis tanpa memperhatikan informasi pada level lingkungan sehingga menimbulkan heteroskedastisitas pada galat. Pada penelitian yang dilakukan oleh Safitri *et al*., (2015) yaitu model linear campuran dan data panel dengan mengabaikan tingkatan atau level dari sekolah terhadap kabupaten/kota memperoleh hasil penelitian yang memperlihatkan kuadrat tengah galat masih relatif besar. Hal ini mengindikasikan adanya struktur data berjenjang yang berarti data berasal dari beberapa level-level yang lebih rendah tersarang pada level yang lebih tinggi dan akan menimbulkan masalah dalam konseptual dan uji statistik (Hox, 2018). Hal inilah yang menjadi salah satu alasan diperlukannya analisis multilevel pada data dengan struktur hierarki (Jones & Steenbergen, 1997).

Dalam mengestimasi parameter-parameter model multilevel terdapat beberapa metode estimasi parameter seperti yang dilakukan Ringdal (1992) membuat pendekatan sederhana menggunakan metode kuadrat terkecil dua tahap

(*Two Stage* OLS). Pada *Two Stage* OLS membutuhkan asumsi bahwa setiap kelompok pada level-2 mempunyai sampel yang sama sehingga penduga ini bukan penduga yang baik apabila ukuran sampel tiap kelompok pada level-2 tidak sama. Metode lain dapat dilakukan dengan menggunakan metode metode *Restricted Maximum Likelihood* (REML). REML menghasilkan penduga koefisien regresi dan komponen varians yang efisien dan konsisten apabila sampel yang digunakan dalam jumlah besar sehingga pelanggaran asumsi dapat diabaikan (Cahyati, 2020). Namun dalam mengestimasi koefisien regresi dapat menggunakan metode *Maximum Likelihood* (MLE) walaupun perbedaan hasil antara kedua metode relatif kecil, dan untuk sampel besar perbedaan hasil antara keduanya dapat diabaikan (Hox, 1995). Penduga REML adalah penduga parameter dengan mengoptimalkan fungsi kemungkinan (*likelihood*) melalui proses iterasi sehingga mendapatkan nilai penduga yang konvergen (West *et al.* 2007). Oleh karena itu berdasarkan latar belakang diatas pada penelitian ini akan dilakukan Pemodelan Regresi Multilevel menggunakan REML pada rata-rata nilai UN SMP tiap-tiap sekolah dengan memperhitungkan adanya varians antar kabupaten/kota di Provinsi Sulawesi Selatan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang diperoleh, maka didapatkan rumusan masalah sebagai berikut:

1. Bagaimana mengestimasi parameter model regresi multilevel menggunakan metode REML?
2. Bagaimana model regresi linear multilevel pada nilai UN SMP dengan memperhitungkan adanya varians antar sekolah dan kabupaten/kota di Provinsi Sulawesi Selatan?

1.3 Batasan Masalah

Agar penelitian ini tidak menyimpang maka perlu diberi batasan masalah yaitu pada penelitian ini akan dilakukan pemodelan regresi 2-level yang diasumsikan normal dan tidak mengandung multikolinearitas dengan koefisien regresi dan komponen varians akan diduga menggunakan metode REML.

1.4 Tujuan Penelitian

Adapun tujuan penelitian ini sebagai berikut:

1. Mendapatksn hasil estimasi parameter model regresi multilevel menggunakan metode REML.
2. Mendapatkan model regresi linear multilevel pada nilai UN SMP dengan memperhitungkan adanya varians antar sekolah dan kabupaten/kota di Provinsi Sulawesi Selatan.

1.5 Manfaat Penelitian

Manfaat penelitian ini sebagai berikut:

1. Bagi penulis

   Menambah pengetahuan tentang model regresi multilevel dengan estimasi parameter menggunakan metode REML.
2. Bagi Pembaca

   Sebagai bentuk informasi dan referensi model regresi multilevel dengan dengan estimasi parameter menggunakan metode REML.

BAB II

TINJAUAN PUSTAKA

2.1 Model Regresi Linear

Analisis regresi adalah teknik yang digunakan dalam statistik untuk menyelidiki dan memodelkan hubungan antara peubah respon dengan beberapa peubah penjelas (Drapper & Smith, 1998). Jika persamaan model regresi hanya memiliki satu peubah penjelas maka model tersebut disebut sebagai regresi linear sederhana namun jika memiliki dua atau lebih peubah penjelas model tersebut disebut sebagai regresi linear berganda. Teknik regresi linier berganda digunakan untuk mengetahui ada tidaknya pengaruh signifikan dua atau lebih peubah penjelas $(X_1, X_2, \dots, X_k)$ terhadap peubah respon (Y). Secara umum, model regresi linear sebagai berikut:

$$Y_i = \beta_0 + \sum_{j=1}^{k} X_j\beta_j + \varepsilon_i, \qquad \begin{array}{l} i = 1,2, \dots, n, \\ j = 1,2, \dots, k. \end{array} \tag{2.1}$$

dengan

Y_i : Peubah respon pada pengamatan ke – i

X_j : Peubah bebas ke-j

β_0 dan β_j : Parameter regresi

ε_i : galat pengamatan ke – i

Atau dapat juga dituliskan dalam bentuk matriks seperti berkut:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.2}$$

2.2 Model Regresi Multilevel

Model multilevel merupakan suatu pemodelan untuk menduga hubungan antar peubah yang diamati pada level-level yang berbeda dalam stuktur data berjenjang. Model regresi multilevel mengasumsikan bahwa data yang digunakan memiliki struktur data yang berjenjang, terdiri dari atas amatan (individu) yang tersarang di dalam kelompok. Model regresi multilevel merupakan bagian dari model linear campuran karena terdapat dua parameter yaitu parameter efek tetap dan efek acak yang digabungkan kedalam satu persamaan (Goldstein, 1995).

Sebuah peubah bebas dikatakan memiliki efek tetap jika koefisien regresinya bernilai sama bagi seluruh anggota sampel dan sebuah peubah bebas dikatakan memiliki efek acak jika nilai koefisien regresinya berbeda antar dua atau lebih grup anggota sampel (Harlan, 2016). Adapun persamaan model linear campuran dalam bentuk sederhana adalah (Rencher dan Schaalje 2007) :

$$y = \boldsymbol{X}\beta + \boldsymbol{Z}u + \varepsilon \tag{2.3}$$

dengan:

y : vektor peubah respon

$\boldsymbol{X}$: matriks peubah penjelas untuk parameter tetap

β : vektor parameter efek tetep

$\boldsymbol{Z}$: matriks peubah penjelas untuk parameter acak

u : vektor efek acak

ε : vektor galat acak

dengan asumsi antar galat $\boldsymbol{\varepsilon}$ dan $\boldsymbol{u}$ tidak saling bebas dan berdistribusi normal dengan varians θ yang dapat ditulis sebagai berikut:

$$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \theta \begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R} \end{bmatrix}\right) \tag{2.4}$$

dengan $\boldsymbol{G} = \boldsymbol{G}(\gamma)$ dan $\boldsymbol{R} = \sigma^2\boldsymbol{\Sigma}$ dengan $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\phi)$. γ dan ϕ adalah vektor parameter varians yang terikat dengan u dan ε sedangkan σ^2 dan θ adalah parameter varians yang akan disebut sebagai parameter skala varians sisa dan parameter skala keseluruhan. Parameter varians tersebut yaitu adalah θ dan $\boldsymbol{\kappa} = (\gamma, \sigma^2, \phi)$. $\boldsymbol{G}, \boldsymbol{R}$ dan $\boldsymbol{\Sigma}$ merupakan matriks varians yang diasumsikan simetris dan definit positif.

$$\boldsymbol{y} \sim N(\boldsymbol{X}\beta, \theta\boldsymbol{H}) \tag{2.5}$$

dengan

$$\boldsymbol{H} = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}' + \boldsymbol{R} \tag{2.6}$$

2.3 Struktur Data Model 2-Level

Pada pemodelan multilevel dengan 2-level, terdapat sebanyak n individu yang berasal dari m kelompok. $Y_{1j}, Y_{2j}, \dots, Y_{n_i j}$ adalah peubah respon masing-masing n_j individu pada kelompok ke-j, $j = 1,2, \dots, m$ dan jika $X_{1j}, X_{2j}, \dots, X_{kj}$ adalah peubah penjelas pada level 1 untuk kelompok ke-j, serta $Z_1, Z_2, \dots, Z_l$ adalah peubah penjelas pada level 2, maka struktur pemodelan multilevel dengan 2-level dapat disajikan dalam Tabel 2.1 berikut.

Tabel 2.1 Struktur Data Model 2 Level

Klp.	Observasi	Peubah Respon	Peubah Penjelas Level 1				Peubah Penjelas Level 2			
		Y	X_1	X_2	...	X_k	Z_1	Z_2	...	Z_l
	1	y_{11}	x_{111}	x_{211}		x_{k11}				
	2	y_{21}	x_{121}	x_{221}		x_{k21}				
	⋮	⋮	⋮	⋮		⋮				
1	i	y_{i1}	x_{1i1}	x_{2i1}	...	x_{ki1}	z_{11}	z_{21}	...	z_{l1}
	⋮	⋮	⋮	⋮		⋮				
	n_j	$y_{n_j 1}$	$x_{1n_j 1}$	$x_{2n_j 1}$		$x_{kn_j 1}$				
	1	y_{12}	x_{112}	x_{212}		x_{k12}				
	2	y_{22}	x_{122}	x_{222}		x_{k22}				
	⋮	⋮	⋮	⋮		⋮				
2	i	y_{i2}	x_{1i2}	x_{2i2}	...	x_{ki2}	z_{12}	z_{22}	...	z_{l2}
	⋮	⋮	⋮	⋮		⋮				
	n_j	$y_{n_j 2}$	$x_{1n_j 2}$	$x_{2n_j 2}$		$x_{kn_j 2}$				
⋮	⋮	⋮	⋮	⋮	...	⋮	⋮	⋮	...	⋮
	i	y_{1m}	x_{11m}	x_{21m}		x_{k1m}				
	2	y_{2m}	x_{12m}	x_{22m}		x_{k2m}				
	⋮	⋮	⋮	⋮		⋮				
M	i	y_{im}	x_{1im}	x_{2im}	...	x_{kim}	z_{1m}	z_{2m}	...	z_{lm}
	⋮	⋮	⋮	⋮		⋮				
	n_j	$y_{n_j m}$	$x_{1n_j m}$	$x_{2n_j m}$		$x_{kn_j m}$				

2.4 Model Regresi 2-Level

Model multilevel mempunyai struktur data berjenjang dengan satu peubah respon yang hanya diukur pada level terendah saja dan peubah penjelas yang diukur pada semua level (Goldstein, 2011). Model multilevel yang paling sederhana adalah model 2-level dengan level 1 adalah data individu dan level 2 adalah data kelompok dengan level 1 sebagai level rendah yang tersarang pada level yang tinggi yaitu level 2. (Hox et al., 2018). Berdasarkan struktur data pada Tabel 2.1 pemodelan regresi multilevel dengan 2-level adalah sebagai berikut: (Goldstein, 2011).

1. Model level-1

Defenisi model level 1 adalah model yang disusun tanpa memperhatikan pengaruh dari level kelompok. Pemodelan multilevel untuk tiap kelompok dapat ditulis sebagai berikut:

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \cdots + \beta_{kj}X_{kij} + \varepsilon_{ij}$$
$$y_{ij} = \beta_{0j} + \sum_{q=1}^{k} \beta_{qj}X_{qij} + \varepsilon_{ij} \quad (2.7)$$

dengan $q = 1,2, \dots, k,$ individu-$i = 1,2, \dots, n_j$ dan kelompok-$j = 1,2, \dots, m$ atau dinyatakan dalam matriks berikut

$$\boldsymbol{y}_j = \boldsymbol{X}_j\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j \qquad \boldsymbol{\varepsilon}_j \sim N(0, \sigma^2 I) \quad (2.8)$$

dengan

$\boldsymbol{y}_j = [y_{1j} \quad y_{2j} \quad \cdots \quad y_{n_jj}]',$

$$\boldsymbol{X}_j = \begin{bmatrix} 1 & x_{11j} & x_{21j} & \dots & x_{k1j} \\ 1 & x_{12j} & x_{22j} & \dots & x_{k2j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{kn_jj} & x_{kn_jj} & \dots & x_{kn_jj} \end{bmatrix},$$

$\boldsymbol{\beta}_j = [\boldsymbol{\beta}_{0j} \quad \boldsymbol{\beta}_{1j} \quad \cdots \quad \boldsymbol{\beta}_{kj}]',$

$\boldsymbol{\varepsilon}_j = [\boldsymbol{\varepsilon}_{1j} \quad \boldsymbol{\varepsilon}_{2j} \quad \cdots \quad \boldsymbol{\varepsilon}_{n_jj}]'$

keterangan:

$\boldsymbol{y}_j$: vektor peubah respon

$\boldsymbol{X}_j$: matriks peubah penjelas untuk parameter tetap

$\boldsymbol{\beta}_j$: vektor parameter efek tetap

$\boldsymbol{\varepsilon}_j$: vektor galat

2. Model level-2

Koefisien regresi pada level-1, β_{pj} dengan $p = 0,1,2,\dots,k$ dalam model level-1 memiliki nilai yang berbeda antar kelompok. Variasi nilai β_{pj} akan dijelaskan dengan membentuk model level 2. Pembentukan model level 2 dilakukan untuk setiap koefisien regresi sebagai respon ke-p dengan menggunakan peubah penjelas pada level-2. Bentuk pemodelan pada level-2 dapat ditulis sebagai berikut:

$$\beta_{pj} = \gamma_{p0} + \gamma_{1p}Z_{1j} + \gamma_{2p}Z_{2j} + \cdots + \gamma_{lp}Z_{lj} + u_{pj}$$

$$\beta_{pj} = \gamma_{p0} + \sum_{p=1}^{k} \gamma_{pl}Z_{lj} + u_{pj} \tag{2.9}$$

dengan $p = 1,2,\dots,k$, dan $l = 1,2,\dots,n_j$ atau dinyatakan dalam matriks berikut

$$\boldsymbol{\beta}_p = \boldsymbol{Z}\boldsymbol{\gamma}_p + \boldsymbol{u}_p \qquad \boldsymbol{u}_p \sim N(0, \sigma_\tau^2 I) \tag{2.10}$$

dengan

$\boldsymbol{\beta}_p = [\boldsymbol{\beta}_{p1} \quad \boldsymbol{\beta}_{p2} \quad \cdots \quad \boldsymbol{\beta}_{pm}]'$,

$$\boldsymbol{X}_j = \begin{bmatrix} 1 & Z_{11} & Z_{21} & \dots & Z_{l1} \\ 1 & Z_{12} & Z_{22} & \dots & Z_{l2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & Z_{1m} & Z_{2m} & \dots & Z_{lm} \end{bmatrix},$$

$\boldsymbol{\gamma}_p = [\boldsymbol{\gamma}_{0p} \quad \boldsymbol{\gamma}_{1p} \quad \cdots \quad \boldsymbol{\gamma}_{lp}]'$,

$\boldsymbol{u}_p = [\boldsymbol{u}_{p1} \quad \boldsymbol{u}_{p2} \quad \cdots \quad \boldsymbol{u}_{pm}]'$

Jika persamaan 2.8 disubstitusikan pada persamaan 2.10 maka diperoleh model sebagai berikut:

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \cdots + \beta_{kj}X_{kij} + \varepsilon_{ij}$$

$$\begin{aligned} y_{ij} = {} & \gamma_{00} + \gamma_{10}Z_{1j} + \gamma_{20}Z_{2j} + \cdots + \gamma_{l0}Z_{lj} + u_{0j} + \big(\gamma_{01} + \gamma_{11}Z_{1j} + \\ & \gamma_{21}Z_{2j} + \cdots + \gamma_{l1}Z_{lj} + u_{1j}\big)X_{1ij} + \cdots + \big(\gamma_{02} + \gamma_{12}Z_{1j} + \\ & \gamma_{22}Z_{2j} + \cdots + \gamma_{l2}Z_{lj} + u_{2j}\big)X_{2ij} + \cdots + \big(\gamma_{0k} + \gamma_{1k}Z_{1j} + \\ & \gamma_{2k}Z_{2j} + \cdots + \gamma_{lk}Z_{lj} + u_{kj}\big)X_{kij} + \varepsilon_{ij} \end{aligned}$$

$$\begin{aligned} y_{ij} = {} & \gamma_{00} + \textstyle\sum_{q=1}^{l} \gamma_{0q}Z_{qj} + \sum_{p=1}^{k} \gamma_{0p}Z_{pij} + \sum_{p=1}^{k}\sum_{q=1}^{l} \gamma_{qp}Z_{qj}X_{pij} + \\ & u_{0j} + \textstyle\sum_{p=1}^{k} u_{pj}X_{pij} + \varepsilon_{ij} \end{aligned} \tag{2.11}$$

atau dinyatakan dalam bentuk matriks sebagai berikut:

$$\boldsymbol{y}_j = \boldsymbol{X}_j\boldsymbol{Z}_j\boldsymbol{\gamma} + \boldsymbol{X}_j\boldsymbol{u}_j + \boldsymbol{\varepsilon}_j \tag{2.12}$$

dengan $\boldsymbol{X}_j\boldsymbol{Z}_j\boldsymbol{\gamma}$ adalah efek tetap dan $\boldsymbol{X}_j\boldsymbol{u}_j + \boldsymbol{\varepsilon}_j$ adalah efek acak.

Hox (2010) matriks $\boldsymbol{Z}$ dalam persamaan 2.12 merupakan peubah perantara untuk menghubungkan $\boldsymbol{\gamma}$ dan $\boldsymbol{X}$. Oleh karena itu, variasi hubungan antara $\boldsymbol{\gamma}$ dan $\boldsymbol{X}$ tergantung pada $\boldsymbol{R}$. Interpretasi dari koefisien regresi pada model level 1 dan koefisien regresi model level 2 terhadap y tergantung pada tanda positif dan negatif dari kedua koefisien regresi tersebut. Jika koefisien regresi model level 1 dan koefisien regresi model level 2 bertanda sama berarti kedua koefisien regresi tersebut memiliki pengaruh berbanding lurus terhadap y. Pada $\boldsymbol{X}_j\boldsymbol{Z}_j$ merupakan faktor interaksi dalam model sebagai varians slope peubah $\boldsymbol{X}$.

2.5 *Intraclass Correlation*

Model regresi multilevel dapat menghasilkan nilai dugaan bagi *intraclass correlation* (Hox 2002). Nilai dugaan tersebut dapat diperoleh menggunakan model tanpa peubah penjelas atau dikenal *intercept-only model* atau *null model*. *Null model* adalah model regresi yang hanya terdiri dari intersep saja tanpa memasukkan pengaruh peubah bebas. Null model pada level-1 dapat dituliskan sebagai berikut:

$$Y_{ij} = \beta_{oj} + \varepsilon_{ij} \tag{2.13}$$

Sedangkan pada level-2 persamaannya sebagai berikut:

$$\beta_{oj} = \gamma_{00} + u_{0j} \tag{2.14}$$

Pada persamaan di atas $u_{0j} \sim N(0, \sigma^2_{\mathrm{u0}})$ adalah varians galat pada level kelompok dan $\varepsilon_{ij} \sim N(0, \sigma^2_{\varepsilon})$ adalah varians galat dari level individu. Sehingga dari *null model* inilah diperoleh *intraclass correlation* (ρ) dapat dirumuskan sebagai berikut:

$$\rho = \frac{\sigma^2_{u_{oj}}}{\sigma^2_{u_{oj}} + \sigma^2_{\varepsilon_{ij}}} \tag{2.15}$$

Dengan $\sigma^2_{u_{oj}}$ merupakan varians level-2 dan $\sigma^2_{\varepsilon_{ij}}$ merupakan varians level-1. Semakin besar nilai ρ menunjukkan semakin tinggi korelasi antar individu sehingga tidak bisa dilakukan analisis regresi biasa dan diperlukan analisis regresi multilevel.

2.6 Estimasi Parameter

Estimasi dalam model linear campuran melibatkan dua proses terkait yaitu akan diestimasi efek tetap dan efek acak ($\boldsymbol{\beta}$ dan $\boldsymbol{u}$**)** dan parameter varians ($\boldsymbol{\theta}$ dan $\boldsymbol{\kappa}$**)**. Penaksir konvensional dari efek tetap disebut sebagai estimator, sedangkan penaksir efek acak disebut prediktor. Estimator efek tetap ($\widehat{\boldsymbol{\beta}}$) terbaik disebut *Best Linear Unbias Estimator* (BLUE) dan prediktor efek acak ($\widetilde{\boldsymbol{u}}$**)** terbaik disebut *Best Linear Unbias Predictor* (BLUP). Estimasi ini dilakukan dengan menggunakan metode yang pertama kali ditetapkan oleh Henderson (1950) dan Henderson (1963) dengan metode yang kekar untuk estimasi parameter varians dalam model linier campuran adalah estimasi REML (Patterson dan Thompson (1971).

2.6.1 Estimasi Parameter Efek tetap dan Efek acak

Handerson (1950) mengusulkan metode untuk mendapatkan solusi estimasi $\boldsymbol{\beta}$ dan $\boldsymbol{u}$ dengan memaksimalkan likelihood gabungan dari $(\boldsymbol{y}, \boldsymbol{u})$. Kepadatan peluang gabungan dari $\boldsymbol{y}$ dan $\boldsymbol{u}$ dapat ditulis dengan mengalikan kepadatan kondisional dari $\boldsymbol{y}$ terhadap $\boldsymbol{u}$ dan fungsi kepadan peluang dari **u**

$$f(\boldsymbol{y}, \boldsymbol{u}; \boldsymbol{\beta}, \theta, \boldsymbol{\kappa}) = f(\boldsymbol{y}|\boldsymbol{u}; \boldsymbol{\beta}, \theta, \varphi) f(\boldsymbol{u}; \theta, \gamma) \tag{2.16}$$

dengan

$$\boldsymbol{y}|\boldsymbol{u} \sim N(\boldsymbol{X\beta} + \boldsymbol{Zu}, \theta\boldsymbol{R})$$
$$\boldsymbol{u} \sim N(\boldsymbol{0}, \theta\boldsymbol{G})$$

Distribusi kondisional dari $\boldsymbol{y}$ terhadap $\boldsymbol{u}$

$$\boldsymbol{y}|\boldsymbol{u} \sim N(\boldsymbol{X}\beta + \boldsymbol{Zu}, \theta\boldsymbol{R})$$

Maka distribusi log dari $\boldsymbol{y}$ terhadap $\boldsymbol{u}$ sebagai berikut:

$$\begin{aligned}\log f(\boldsymbol{y}, \boldsymbol{u}) &= \log f(\boldsymbol{y}|\boldsymbol{u}) + \log f(\boldsymbol{u}) \\ &= -\frac{1}{2}\left\{n \log\theta + \log\boldsymbol{R} + (\boldsymbol{y} - \boldsymbol{X\beta} + \boldsymbol{Zu})\boldsymbol{R}^{-1}\frac{(\boldsymbol{y}-\boldsymbol{X\beta}+\boldsymbol{Zu})}{\theta}\right\} - \frac{1}{2}\left\{q\log\theta + \log\boldsymbol{G} + \boldsymbol{u}'\boldsymbol{G}^{-1}\frac{\boldsymbol{u}}{\theta}\right\} \\ &= \frac{1}{2}\left\{(n+q)\log\theta + \log\boldsymbol{R} + \log\boldsymbol{G} + (\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{R}^{-1}\frac{(\boldsymbol{y}-\boldsymbol{X\beta})}{\theta}\right\} - \frac{1}{2\theta}\boldsymbol{u}'(\boldsymbol{ZR}^{-1}\boldsymbol{Z}' + \boldsymbol{G}^{-1})\boldsymbol{u} - 2(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{R}^{-1}\boldsymbol{Zu}\end{aligned} \tag{2.17}$$

(Lee & Nelder, 2001).

2.6.2 Estimasi Parameter Varians

Verbyla (1990) menganggap matiks non-singular $\boldsymbol{L} = [\boldsymbol{L_1}\ \boldsymbol{L_2}]$ dengan $\boldsymbol{L_1}$berukuran $n \times t$ dan $\boldsymbol{L_2}$ berukuran $n \times (n - t)$ yang dipilih untuk memenuhi $\boldsymbol{L_1'X} = \boldsymbol{I_t}$ dan $\boldsymbol{L_2'X} = 0$. Data tersebut ditransformasi dari $\boldsymbol{y}$ menjadi $\boldsymbol{L'y}$

Dengan

$$\boldsymbol{L'y} = \begin{bmatrix} \boldsymbol{L}_1'\boldsymbol{y} \\ \boldsymbol{L}_2'\boldsymbol{y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} \tag{2.18}$$

Transformasi data diatas berdistribusi normal dengan varians θ yang dapat ditulis sebagai berikut:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{0} \end{bmatrix}, \theta \begin{bmatrix} \boldsymbol{L}_1'\boldsymbol{HL}_1 & \boldsymbol{L}_1'\boldsymbol{HL}_2 \\ \boldsymbol{L}_2'\boldsymbol{HL}_1 & \boldsymbol{L}_2'\boldsymbol{HL}_2 \end{bmatrix}\right)$$

Fungsi kepadatan peluang dari $\boldsymbol{L'y}$ dapat ditulis sebagai perkalian dari kondisional kepadatan dari $\boldsymbol{y}_1|\boldsymbol{y}_2$ dan kepadatan marginal $\boldsymbol{y}_2$. Distribusi marginal $\boldsymbol{y}_2$ dituliskan sebagai berikut:

$$\boldsymbol{y}_2 \sim N(0, \theta\boldsymbol{L}_2'\boldsymbol{HL}_2)$$

dari distribusi diatas dapat dibentuk fungsi kepadatan peluangnya sebagai berikut:

$$f(\boldsymbol{y}_2) = \frac{1}{(2\pi\theta)^{\frac{n-t}{2}} |\boldsymbol{L}_2'\boldsymbol{HL}_2|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2\theta}\boldsymbol{y}_2'(\boldsymbol{L}_2'\boldsymbol{HL}_2)^{-1}\boldsymbol{y}_2\right\} \tag{2.19}$$

Selanjutnya dapat dituliskan fungsi log-likelihood sebagai berikut:

$$\begin{aligned}
(\theta, \kappa, \boldsymbol{y}_2) &= -\frac{1}{2}\left((n-t)\log\theta + \log|\boldsymbol{L}_2'\boldsymbol{HL}_2| + \frac{\boldsymbol{y_2'}(\boldsymbol{L}_2'\boldsymbol{HL}_2)^{-1}\boldsymbol{y}_2}{\theta}\right) \\
&= -\frac{1}{2}\left((n-t)\log\theta + \log|\boldsymbol{L}_2'\boldsymbol{HL}_2| + \frac{\boldsymbol{y}'\boldsymbol{L}_2(\boldsymbol{L}_2'\boldsymbol{HL}_2)^{-1}\boldsymbol{L_2'y}}{\theta}\right) \\
&= -\frac{1}{2}\left((n-t)\log\theta + \log|\boldsymbol{L}_2'\boldsymbol{HL}_2| + \frac{\boldsymbol{y'Py}}{\theta}\right)
\end{aligned} \tag{2.20}$$

Dengan

$$\boldsymbol{P} = \boldsymbol{L}_2(\boldsymbol{L}_2'\boldsymbol{HL}_2)^{-1}\boldsymbol{L_2'} = \boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{X'H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X'H}^{-1} \tag{2.21}$$

Sehingga distribusi kondisional $\boldsymbol{y_1}|\boldsymbol{y_2}$ dapat dituliskan sebagai berikut:

$$\boldsymbol{y_1}|\boldsymbol{y_2} \sim N\big(\beta + \boldsymbol{L}_1'\boldsymbol{HL}_2(\boldsymbol{L}_2'\boldsymbol{HL}_2)^{-1}\boldsymbol{y_2}, \boldsymbol{\theta}(\boldsymbol{L}_1'\boldsymbol{HL}_1 - \boldsymbol{L}_1'\boldsymbol{HL}_2(\boldsymbol{L}_2'\boldsymbol{HL}_2)^{-1}\boldsymbol{L}_2'\boldsymbol{HL}_1)\big)$$

dengan matriks kondisional varians sebagai berikut:

$$\begin{aligned}
&\boldsymbol{L}_1'\boldsymbol{HL}_1 - \boldsymbol{L}_1'\boldsymbol{HL}_2(\boldsymbol{L}_2'\boldsymbol{HL}_2)^{-1}\boldsymbol{L}_2'\boldsymbol{HL}_1 \\
&= \boldsymbol{L}_1'(\boldsymbol{H} - \boldsymbol{HL}_2(\boldsymbol{L}_2'\boldsymbol{HL}_2)^{-1}\boldsymbol{L}_2'\boldsymbol{H})\boldsymbol{L}_1
\end{aligned} \tag{2.22}$$

$= \boldsymbol{L}_1'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{H}^{-1}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{L}_{\mathbf{1}}$

$= (\boldsymbol{X}'\boldsymbol{H}^{-1}\boldsymbol{X})^{-\mathbf{1}}$, dengan $\boldsymbol{L}_1'\boldsymbol{X} = \boldsymbol{I}_t$.

Untuk *mengestimasi* nilai $\boldsymbol{\kappa}$ diperlukan distribusi marginal $\boldsymbol{y}_{\mathbf{2}}$ pada persamaan 2.20. Marginal log-likelihood dari $\boldsymbol{y}_{\mathbf{2}}$ yang disebut sebagai fungsi log-likelihood REML adalah sebagai berikut:

$$l_R(\boldsymbol{\theta}; \boldsymbol{\kappa}; \boldsymbol{y}_{\mathbf{2}}) = -\frac{1}{2}\left\{(n-t)\log\theta + \log|\boldsymbol{L}_{\mathbf{2}}'\boldsymbol{H}^{-1}\boldsymbol{L}_{\mathbf{2}}| + \frac{\boldsymbol{y}'\boldsymbol{P}\boldsymbol{y}}{\theta}\right\} \quad (2.23)$$

Log-likelihood dari $\boldsymbol{L}'\boldsymbol{y}$ dapat ditulis sebagai jumlah log-likelihood kondisional dari $\boldsymbol{y}_{\mathbf{1}}$ terhadap $\boldsymbol{y}_{\mathbf{2}}$ sehingga marginal log-likelihood dapat ditulis:

$$l(\boldsymbol{\theta}; \boldsymbol{\kappa}; \boldsymbol{L}'\boldsymbol{y}) = \boldsymbol{l}\big(\boldsymbol{\theta}; \boldsymbol{\kappa}; \boldsymbol{y}_{\mathbf{1}|}\boldsymbol{y}_{\mathbf{2}}\big) + \boldsymbol{l}(\boldsymbol{\kappa}; \boldsymbol{y}_{\mathbf{2}}) \quad (2.24)$$

Persamaan diatas dapat ditulis menjadi log-determinan sebagai berikut:

$$\log|\boldsymbol{L}'\boldsymbol{H}\boldsymbol{L}| = \log\left|(\boldsymbol{X}'\boldsymbol{H}^{-\mathbf{1}}\boldsymbol{X})^{-1}\right| + \log|\boldsymbol{L}_{\mathbf{2}}'\boldsymbol{H}\boldsymbol{L}_{\mathbf{2}}| \quad (2.25)$$

Maka

$$\log|\boldsymbol{L}_{\mathbf{2}}'\boldsymbol{H}\boldsymbol{L}_{\mathbf{2}}| = \log|\boldsymbol{L}'\boldsymbol{L}| + \log|\boldsymbol{H}| + \log\left|\boldsymbol{X}'\boldsymbol{H}^{-\mathbf{1}}\boldsymbol{X}\right| \quad (2.26)$$

Sehingga fungsi log-likelihood REML pada persamaan 2.23 dapat ditulis sebagai berikut tanpa memperhatikan konstanta

$$l_R(\boldsymbol{\theta}; \boldsymbol{\kappa}; \boldsymbol{y}_{\mathbf{2}}) = -\frac{1}{2}\left\{(n-t)\log\theta + \log|\boldsymbol{H}| + \log|\boldsymbol{X}\boldsymbol{H}^{-1}\boldsymbol{X}| + \frac{\boldsymbol{y}'\boldsymbol{P}_{\boldsymbol{y}}}{\theta}\right\} \quad (2.27)$$

(Searle et al., 2006).

2.7 Pemilihan Model Terbaik

Dalam pemilihan model terbaik untuk menentukan cocok tidaknya suatu model juga dapat dilihat dari *deviance*, untuk menguji signifikansi perbedaan nilai *deviance*, statistik uji yang digunakan dapat menggunakan uji rasio likelihood atau *likelihood ratio test (LRT)* untuk menentukan cocok tidaknya suatu model (Tantular, 2009). Dari fungsi *likelihood*, dapat dihitung statistik yang disebut dengan *deviance* yang menunjukkan seberapa baik model tersebut cocok dengan data yang digunakan. *Deviance* dapat diformulasikan sebagai berikut:

$$D = -2\log(L_1) \quad (2.28)$$

dengan

L_1 : nilai fungsi *likelihood*

Sedangkan untuk uji *likelihood ratio test (LRT)* diformulasikan sebagai berikut:

$$LRT = -2\log\left(\frac{L_0}{L_1}\right) \quad (2.29)$$

dengan

L_0 : nilai fungsi *likelihood* pada model sebelumnya

L_1 : nilai fungsi *likelihood* pada model yang diuji

Perbandingan nilai *likelihood* tersebut menyebar *chi-square* dengan derajat bebas sebesar selisih dari parameter kedua model tersebut (Hox dan Wijngaards 2013). Semakin kecil nilai *deviance* model tersebut dikatakan semakin cocok. Akan tetapi tidak ada ketentuan yang pasti berapa besar ukuran untuk nilai *deviance*. Sehingga untuk mengetahui suatu model cocok atau tidak harus dibandingkan dengan model lain (Hox, 2017).

2.8 Koefisien Determinasi

Nilai koefisien determinasi dalam analisis regresi multilevel dapat diperoleh pada setiap level (Hox, 2017). Kisaran nilai koefisien determinasi mulai dari 0% sampai 100% semakin besar nilai koefisien determinasi berarti model semakin mampu menerangkan perilaku peubah respon. Pada model multilevel nilai koefisien determinasi yang didapatkan akan lebih dari satu.

Koefisien determinasi pada level satu dapat dirumuskan sebagai berikut:

$$R_1^2 = \left(\frac{\hat{\sigma}_{e_0}^2 - \hat{\sigma}_{e_p}^2}{\hat{\sigma}_{e_0}^2}\right) \quad (2.30)$$

dengan:

R_1^2 : nilai koefisien determinasi untuk level-1.

$\hat{\sigma}_{ep}^2$: penduga varians dari residual pada level ke-satu dengan p peubah penjelas.

$\hat{\sigma}_{e_0}^2$: penduga varians dari residual pada level kesatu tanpa peubah penjelas.

Koefisien determinasi kedua didefinisikan pada level ke-dua sebagai berikut:

$$R_2^2 = \left(\frac{\hat{\sigma}_{u_0}^2 - \hat{\sigma}_{u_{0p}}^2}{\hat{\sigma}_{u_0}^2}\right) \quad (2.31)$$

dengan:

R_2^2 : nilai koefisien determinasi untuk level-2.

$\hat{\sigma}_{u_{0p}}^2$: penduga varians dari residual pada level kedua dengan p peubah penjelas.

$\hat{\sigma}_{u_0}^2$: penduga varians dari residual pada level kedua tanpa peubah penjelas.