

# PENERAPAN *HIERARCHICAL LINEAR MODELING* UNTUK MENGANALISIS DATA MULTILEVEL

Dewi Wulandari<sup>1</sup>, Ali Shodiqin<sup>2</sup>, dan Aurora Nur Aini<sup>3</sup>

<sup>1</sup> Universitas PGRI Semarang, dewiwulandari1611@yahoo.com,

<sup>2</sup> Universitas PGRI Semarang, alishodiqin81@gmail.com

<sup>3</sup> Universitas PGRI Semarang, aurora.nuraini@gmail.com

**Abstract.** Multilevel data are data that are nested within the other data which are in the higher level. As an example is students are nested in the classes. The student is the level-1 variable and the class is the level-2 variable. Multilevel data are not restricted only level-2 but also more than it. As an example we have taken, school is the level-3 variable, region is the level-4 variable etc. Students in one class will be different from another class, classes in one school will be different from another school, etc. Because of this variation then we need Hierarchical Linear Modeling (HLM) to analyze it. This method is a complex form of OLS (Ordinary Least Square) regression. In estimating the parameters we use GLS (Generalized Least Square). In this research, we use mathematics score of Nasima junior high school student Semarang. From the analysis result which are got by using software HLM student version we can conclude that there's no significant variation within groups or classes, then it's enough using OLS regression to analyze the factors affecting mathematics score. Hypothesis of the reason of this is the amount of unit in level-2 variables are not enough, they are only 4 units. To prove this hypothesis, we need another research.

**Keywords:** *HLM, regression, multilevel data, OLS, GLS*

**Abstrak.** Data multilevel merupakan data yang tersarang dalam data lain yang memiliki tingkatan lebih tinggi. Sebagai contoh adalah data siswa tersarang dalam data kelas. Data siswa sebagai data level 1 dan data kelas sebagai data level 2. Data multilevel tidak hanya terbatas sampai level 2 saja namun juga bisa lebih dari 2. Dalam contoh yang digunakan, level 3 dapat ditempati oleh data sekolah, level 4 dapat ditempati oleh data kabupaten, dst. Siswa dalam kelas yang satu akan berbeda dengan siswa pada kelas yang lain, kelas pada sekolah yang satu akan berbeda dengan kelas pada sekolah yang lain, dst. Karena adanya variasi tersebut maka diperlukan model pendekatan data multilevel untuk menganalisisnya. Salah satunya adalah *Hierarchical Linear Modeling* (HLM). Metode ini merupakan bentuk kompleks dari metode regresi OLS (*Ordinary Least Square*). Dalam melakukan estimasi untuk intersep dan koefisien regresi digunakan GLS (*Generalized Least Square*). Dalam penelitian ini digunakan data hasil belajar siswa yaitu nilai matematika UN tahun 2016 SMP Nasima Semarang. Dari hasil analisis penelitian ini dengan bantuan software HLM versi student menunjukkan bahwa tidak ada variasi yang signifikan antara kelas yang satu dengan kelas yang lain, sehingga cukup digunakan regresi OLS biasa untuk menganalisis faktor-faktor yang mempengaruhi nilai matematika siswa. Hal ini diduga karena jumlah kelas atau unit dalam data level 2 terlalu sedikit yaitu hanya 4 unit.

**Kata Kunci:** *HLM, regresi, data multilevel, OLS, GLS.*

## 1 Pendahuluan

Data multilevel merupakan data yang tersarang di dalam data lain [5]. Beberapa contohnya adalah data nilai siswa di suatu sekolah tersarang di dalam

data kelas kelas sekolah tersebut, data pasien di suatu rumah sakit tersarang di dalam data rumah sakit tersebut, data anggota keluarga tersarang didalam data keluarga tersebut, dll. Contoh-contoh tersebut merupakan contoh data dua-level. Misal untuk data nilai siswa di suatu sekolah. Data nilai siswa menduduki level pertama dan data kedua menduduki level kedua. Data tersebut dapat ditinjau lebih dalam lagi sehingga akan ditemukan data sebagai sarang yang tingkatannya lebih tinggi. Untuk data nilai siswa di suatu sekolah, maka akan dapat ditemukan data sekolah sebagai level ketiga.

Analisis terhadap data siswa tidak bisa begitu saja mengabaikan level-level tersebut karena adanya pengaruh unsur-unsur dari data berlevel lebih tinggi. Unsur-unsur tersebut terkadang tidak mudah untuk diukur seperti misalnya cara/metode guru mengajar. Seandainya dapat diukur sekalipun tidak akan akan cocok diselesaikan dengan analisis regresi konvensional karena akan muncul banyak variabel Dummy sementara dalam regresi konvensional, variabel independen yang terlalu banyak akan mengurangi kevaliditasan model. Kelas dan sekolah juga akan menimbulkan munculnya dependensi antara data satu dengan data yang lain sementara kelas dan sekolah yang berbeda akan menimbulkan keheterogenan di dalam data. Padahal independensi dan kehomogenan data adalah beberapa asumsi yang harus dipenuhi untuk menggunakan analisis regresi konvensional. Oleh sebab itu untuk menganalisis data multilevel diperlukan *Hierarchical Linear Modeling* (HLM) [9].

HLM secara simultan menyelidiki hubungan di dalam dan diantara level-level hierarki dari data yang dikelompok-kelompokkan, sehingga akan lebih efisien untuk menghitung varians antar variabel pada level-level yang berbeda dibandingkan dengan analisis lainnya. Dalam penelitian ini dibahas mengenai metode estimasi dalam HLM. Selanjutnya, dilakukan studi kasus untuk menerapkan HLM yaitu analisis data nilai matematika siswa SMP di kota Semarang di mana akan diambil sampel di suatu sekolah yang terletak di kota Semarang. Data tersebut diambil karena seperti yang telah dijelaskan sebelumnya bahwa nilai dari seorang siswa dipengaruhi oleh kelas di mana siswa tersebut menerima materi pembelajaran.

## 2 Kajian Teori

### 2.1 Analisis Regresi Linier Sederhana

Analisis regresi merupakan salah satu alat yang digunakan untuk mengetahui kuat lemahnya pengaruh variabel prediktor (variabel independen, variabel bebas, dll) terhadap variabel *outcome* (variabel dependen, variabel terikat, dll) dan untuk memprediksi variabel *outcome* dengan menggunakan variabel prediktor [14], [8]. Secara umum, persamaan prediksi dapat ditulis sebagai berikut.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

dengan  $\hat{\beta}_0$  adalah intersep dan  $\hat{\beta}_1$  adalah slope.

## 2.2 Estimasi Parameter Regresi Linier Sederhana

Ada beberapa cara mengestimasi parameter regresi linier sederhana [8], yaitu:

1. *Maximum Likelihood Estimator* (MLE)

Berikut ini hasil estimasi parameter regresi dengan menggunakan MLE.

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \text{ dan } \hat{\beta}_1 = \frac{\sum_{i=1}^n Y_i(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}.$$

2. *Ordinary Least Square* (OLS)

Berikut ini adalah hasil estimasi parameter regresi dengan OLS.

$$(X^t X)^{-1} X^t Y.$$

3. *Generalized Least Square* (GLS)

Berikut ini adalah hasil estimasi parameter regresi dengan GLS.

$$(X^t \Omega^{-1} X)^{-1} X^t \Omega^{-1} Y.$$

## 2.3 Analisis Regresi Logistik

Regresi logistik ialah regresi dengan variabel respon terdiri dari dua kejadian, sukses atau gagal, disebut respon biner, sehingga hasil kejadian tersebut dapat didekati oleh distribusi Binomial [4]. Selanjutnya, yang dimodelkan ialah probabilitas terjadi sukses, dengan prediktor yang diduga berkontribusi terhadap kejadian sukses. Model regresi logistik dinyatakan dengan persamaan :

$$P(x_i) = \frac{1}{1 + e^{-x_i^T \beta}}, i = 1, 2, \dots, s.$$

dengan  $P(x_i)$  adalah probabilitas terjadi sukses pada kelompok ke- $i$  dan  $x_i^t = \beta_0 + \beta_1 x_i$ , bila digunakan satu prediktor.

## 3 Penerapan *Hierarchical Linier Modeling* untuk Menganalisis Data Multilevel

*Hierarchical Linear Modeling* (HLM) merupakan bentuk kompleks dari regresi OLS [10]. Model ini digunakan untuk menganalisis keragaman di dalam variabel *outcome* ketika variabel-variabel prediktor berada dalam level bertingkat yang bervariasi. Sebagai contoh, murid-murid yang berada dalam kelas-kelas yang berbeda di mana variansinya bergantung pada guru, kondisi kelas, dll. Sebelum berkembangnya HLM, data bertingkat masih ditaksir dengan menggunakan teknik regresi linear sederhana, namun teknik ini tidak cocok karena mengabaikan variansi dalam kelas-kelas [15]. Berikut ini merupakan model HLM untuk data bertingkat 2 level.

## 4 Studi Kasus

Data yang digunakan adalah data nilai matematika UN SMP Nasima Semarang dimana dalam data ini siswa tersarang dalam kelas. Terdapat 4 kelas dalam data ini dengan rata-rata siswa per kelas adalah 30 siswa. Data ini dipilih karena merupakan data multilevel dimana siswa menempati level 1 dan kelas menempati level 2. Pertama, ditentukan apakah HLM diperlukan atau tidak. Kemudian bagaimana pengaruh variabel-variabel independen terhadap variabel dependen. Variabel independen level 1 adalah pendidikan terakhir orang tua siswa, variabel independen level 2 adalah gaya mengajar guru, sedangkan variabel dependennya adalah nilai UN matematika siswa.

Ada 4 model HLM yang digunakan untuk menganalisis data ini. Model-model tersebut adalah null model, model intersep acak, *means as outcomes model* serta model intersep acak dan slope. Null model merupakan model paling awal yang harus disusun karena model ini bertujuan mengetahui apakah HLM diperlukan untuk menganalisis data yang akan dianalisis.

Tabel 1 berikut ini menampilkan hasil estimasi komponen varians. Terlihat bahwa p-value yang didapatkan adalah lebih dari 0.001. Hal tersebut menunjukkan bahwa tidak terdapat variasi atau perbedaan yang signifikan antara kelas satu dengan kelas yang lainnya. Hal ini diduga karena terlalu sedikit kelas atau unit dalam level 2. Dengan demikian, penulis membuat simulasi data dengan unit yang lebih banyak. Variabel independennya adalah nilai matematika siswa. Variabel independen level 1 dalam data simulasi ini adalah jam belajar dan untuk level 2 adalah lama pengalaman guru mengajar.

Tabel 2 berikut ini menunjukkan bahwa HLM memang diperlukan karena ditemukan perbedaan yang cukup signifikan antar kelas. Hal tersebut diindikasikan dengan p-value pada estimasi komponen varians.

Tabel 1: Estimasi Komponen Varians

<i>Random Effect</i>	<i>Standard Deviation</i>	<i>Variance Component</i>	d.f	$\chi^2$	p-value
INTRCPT1, $u_0$	2.22058	4.93096	3	4.21049	0.238
level-1, r	18.80133	353.48983			

Tabel 2: Estimasi Komponen Varians untuk Data Simulasi

<i>Random Effect</i>	<i>Standard Deviation</i>	<i>Variance Component</i>	d.f	$\chi^2$	p-value
INTRCPT1, $u_0$	5.37918	28.93554	9	55.19123	0.001
level-1, r	12.8679	165.50274			

Selanjutnya dengan model intersep acak dan *means as outcomes model* ditentukan apakah variabel jam belajar dan pengalaman guru memberikan pengaruh yang cukup signifikan atau tidak. Tabel 3 berikut ini menunjukkan bahwa variabel jam belajar cukup memberikan pengaruh yang signifikan ditunjukkan dengan p-value 0.003. Jika diambil taraf signifikansi 0.005, maka jelas bahwa nilai tersebut kurang dari 0.005. Pada Tabel 4 berikut ini untuk pengalaman guru tidak memberikan pengaruh yang cukup signifikan karena p-value tidak kurang dari 0.001 maupun 0.005.

Tabel 3: Estimasi Parameter Model Intersep Acak

<i>Fixed Effect</i>	<i>Coefficient</i>	<i>Standard Error</i>	<i>t-ratio</i>	<i>Approx. d. f.</i>	<i>p-value</i>
For INTRCPT1, $\beta_0$					
INTRCPT2, $\gamma_{00}$	75.740542	1.860005	40.721	9	< 0.001
For JBSlope, $\beta_1$					
INTRCPT2, $\gamma_{10}$	2.378251	0.606234	3.923	9	0.003

Tabel 4: Estimasi Parameter *Means as Outcomes Model*

<i>Fixed Effect</i>	<i>Coefficient</i>	<i>Standard Error</i>	<i>t-ratio</i>	<i>Approx. d. f.</i>	<i>p-value</i>
For INTRCPT1, $\beta_0$					
INTRCPT2, $\gamma_{00}$	75.729185	1.173361	64.540	8	< 0.001
TEACHER, $\gamma_{01}$	3.729295	0.972283	3.836	8	0.005

## 5 Kesimpulan

Beberapa simpulan yang dapat ditarik dari pembahasan-pembahasan sebelumnya adalah terdapat 4 model HLM yaitu null model, model intersep acak, *means as outcomes model* serta model intersep acak dan slope. Terdapat 3 langkah yaitu estimasi *fixed effects* (efek tetap), estimasi random *effects* dan estimasi komponen varians. Namun dalam penelitian ini untuk estimasi komponen varians belum didapatkan di mana seharusnya dapat digunakan *full maximum likelihood*. Untuk mendapatkan estimasi *fixed effects* digunakan GLS, sedangkan untuk mendapatkan estimasi random *effects* digunakan OLS. Model HLM tidak diperlukan untuk menganalisis data multilevel nilai matematika UN SMP Nasima Semarang. Hal tersebut ditunjukkan oleh nilai p-value komponen varians dalam null model yang melebihi 0.001. Hal ini berarti tidak ada variasi atau perbedaan yang cukup signifikan antara kelas satu dengan kelas yang lain sehingga cukup digunakan regresi OLS biasa. Hal ini diduga karena banyaknya kelas cukup sedikit. Oleh sebab itu, penulis melanjutkan percobaan dengan membangkitkan data multilevel (menggunakan Ms.Excel) dengan unit kelas yang lebih banyak yaitu 10 kelas. Diperoleh hasil bahwa metode HLM cocok digunakan menganalisis data simulasi tersebut. Dengan begitu, penulis menyimpulkan bahwa tidak semua data multilevel memiliki perbedaan atau variasi yang cukup signifikan antar kelompok atau kelasnya.

## Daftar Pustaka

- [1] Bliese, P. 2006. *Multilevel Models in R (2.2)*. R Development Core Team.
- [2] Dempster, A. P., Rubin, D. B., dan Tsutakawa, R. K. 1981. Estimation in Covariance Components Models. *Journal of the American Statistical Association*. 76(374):341-353.
- [3] Dempster, A. P., Laird, N. M., dan Rubin, D. B. 1977. Maximum Likelihood From Incomplete Data Via The EM Algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*. 39(1):1-38.
- [4] Dewi, A.L. 2008. *Estimasi Parameter Logistik Multilevel*. Skripsi. Universitas Indonesia, Depok, Indonesia.
- [5] Goldstein, H. 1995. *Multilevel Statistical Models 2nd Ed*. London. Arnold London.

- [6] Kramer, M. 2005. *R<sup>2</sup> Statistics for Mixed Models*. Published Paper in Biometrical Consulting Service, ARS (Beltsville, MD), USDA.
- [7] Krismala, D.A. 2014. Pemodelan Regresi 2-Level dengan Metode *Iterative Generalized Least Square* (IGLS). *Jurnal Gaussian*. 3(1):51-60.
- [8] Myers, R.H. 1990. *Classical and Modern Regression with Applications*. PWS-Kent Publishing Company, Boston.
- [9] Raudenbush, S. W., dan Bryk, A. S. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods, Second Edition*. Newbury Park, CA: Sage.
- [10] Ringdal, K. 1992. Methods for Multilevel Analysis. *Acta Sociologica*. 35:235-243.
- [11] Sudjana, N. 2005. *Penilaian Hasil Proses Belajar Mengajar*. PT. Remaja Rosdikarya. Bandung.
- [12] Sudjana, N. 2009. *Penilaian Hasil Proses Belajar Mengajar*. PT. Remaja Rosdikarya. Bandung.
- [13] Tantular, B. 2009. *Penerapan Model Regresi Linier Multilevel pada Data Pendidikan dan Data Nilai Ujian*. Tesis. Institut Pertanian Bogor, Bogor, Indonesia.
- [14] West, B. T., Welch, K. B., dan Galecki, A. T. 2007. *Linear Mixed Models : A Practical Guide Using Statistical Software*. Chapman and Hall, Boca Raton.
- [15] Woltman H., Feldstain A., MacKay J.C., dan Rocchi M. 2012. An Introduction to Hierarchical Linear Modeling. *Tutorials in Quantitative Methods for Psychology*.8(1):52-69.