

## Pemodelan Regresi 2-Level Dengan Metode *Iterative Generalized Least Square* (IGLS)

(Studi Kasus: Tingkat Pendidikan Anak di Kabupaten Semarang)

Dyan Anggun Krismala<sup>1</sup>, Dwi Ispriyanti<sup>2\*</sup>, Moch. Abdul Mukid<sup>3</sup>

<sup>1</sup> Mahasiswa Jurusan Statistika FSM UNDIP

<sup>2,3</sup> Staff Pengajar Jurusan Statistika FSM UNDIP

### Abstract

In a research, data was used often hierarchical structure. Hierarchical data is data obtained through multistage sampling from a population with independent variables can be defined within each level and dependent variable can be defined at the lowest level. One analysis that can be used for data with a hierarchical structure is a multilevel regression analysis. Multilevel regression analysis is the most simple regression analysis 2-levels. 2-level regression analysis will be used to construct a regression model the education level of children in Semarang where children (level-1) nested on the districts (level-2) with the factors that influence. Estimation of parameter in 2-level regression model can use some methods, one of them is Iterative Generalized Least Square (IGLS). From the results of the discussion indicates that the factors which affect the level of education of children in Semarang is the mother's education, father's education, and percentage of farm families. The diversity level of the education of children in Semarang caused more variation among children than the variation between districts.

**Keywords** : Hierarchical structure, multistage sampling, 2-level regression

## 1. PENDAHULUAN

### Latar Belakang

Dalam Suatu Penelitian terkadang diperoleh struktur data hirarki atau berjenjang. Data yang berstruktur hirarki merupakan data yang timbul karena individu-individu terkumpul dalam kelompok-kelompoknya, dimana individu dalam kelompok yang sama memiliki karakteristik yang cenderung sama. Data hirarki dapat diperoleh melalui *multistage sampling* dari populasi berjenjang, dimana variabel-variabelnya dapat didefinisikan dari setiap level dengan level yang lebih rendah tersarang pada level yang lebih tinggi (Hox, 1995). Pada data hirarki, individu-individu dalam kelompok yang sama cenderung mempunyai karakteristik yang sama dibandingkan dengan individu-individu dalam kelompok yang berbeda, sehingga data yang diperoleh tidak *independent*. Hal ini menjadi alasan mengapa analisis regresi linier berganda kurang tepat digunakan pada data yang mempunyai struktur hirarki. Salah satu analisis yang dapat digunakan untuk menganalisis data hirarki adalah analisis regresi multilevel.

Dalam beberapa sumber terdapat metode yang berbeda untuk menaksir parameter model regresi multilevel, diantaranya metode *Iterative Generalized Least Square* (IGLS), *Two Stage Least Square* (2SLS), *Generalized Estimating Equation* (GEE), dan *Restricted Maximum Likelihood* (REML). Bentuk data hirarki dapat dijumpai dalam berbagai bidang penelitian, antara lain dalam bidang pendidikan. Dalam penelitian tugas akhir ini akan dibahas mengenai pemodelan regresi multilevel pada data hirarki untuk data hirarki 2-level.

## Tujuan Penulisan

Tujuan dalam penulisan tugas akhir ini adalah

1. Mengestimasi parameter model regresi 2-level dengan metode IGLS.
2. Menentukan model terbaik
3. Mengetahui faktor-faktor yang berpengaruh terhadap tingkat pendidikan anak di Kabupaten Semarang dengan model regresi 2-level.

## 2. TINJAUAN PUSTAKA

### 2.1. Definisi Pendidikan dan Faktor yang Mempengaruhinya

Pendidikan adalah aktivitas dan usaha manusia untuk meningkatkan kepribadiannya dengan jalan membina potensi-potensi pribadinya, yaitu rohani (pikir, karsa, rasa, cipta dan budi pekerti). Pendidikan juga berarti lembaga yang bertanggung jawab menetapkan cita-cita (tujuan) pendidikan, isi, sistem, dan organisasi pendidikan. Lembaga-lembaga ini meliputi keluarga, sekolah dan masyarakat (Ihsan Fuad, 2005).

Secara umum seorang anak dalam mendapatkan pendidikan dipengaruhi oleh banyak faktor yang saling berkorelasi, antara lain faktor diri sendiri, faktor keluarga dan faktor lingkungan. Dari ketiga faktor tersebut faktor yang ada dalam diri sendiri merupakan faktor yang paling berpengaruh terhadap pendidikan anak, namun selain itu faktor keluarga juga mempunyai peran yang penting dalam mempengaruhi seorang anak untuk mendapatkan pendidikan, dalam hal ini adalah peran orang tua, yang menjadi pendorong terciptanya pendidikan untuk anak-anaknya. Karena itu pendidikan di keluarga yang mencerahkan dan mampu membentuk karakter anak yang kreatif adalah modal penting bagi kesuksesan anak di masa-masa yang akan datang (Badan Pusat Statistik, 2011). Faktor lingkungan juga dapat mempengaruhi tingkat pendidikan seseorang, antara lain tersedianya fasilitas pendidikan yang memadai. Selain itu peran masyarakat di sekitar lingkungan tempat tinggal juga dapat menjadi faktor lain yang mempengaruhi tingkat pendidikan seseorang.

### 2.2. Data Hirarki

Data yang berstruktur hirarki merupakan data yang timbul karena individu-individu terkumpul dalam kelompok-kelompoknya. Data hirarki yang diperoleh dalam suatu penelitian merupakan data yang diperoleh melalui *multistage sampling* dari populasi berjenjang yang variabel-variabelnya dapat didefinisikan dari setiap level, dimana level yang lebih rendah tersarang pada level yang lebih tinggi. Data hirarki disebut juga sebagai data multilevel (Hox, 2002).

### 2.3. Model Regresi 2-Level

Model regresi dua level adalah model regresi multilevel yang paling sederhana karena hanya terdiri dari 2 level saja, dimana untuk level satu diartikan sebagai level terendah (level individu) dan untuk level kedua diartikan sebagai level tertinggi (kelompok). Secara matematis, model regresi 2-level dengan satu variabel bebas level 1 dapat ditulis:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij} \quad (2.1)$$

$i$  = individu dalam kelompok ke- $j$ ,  $i = 1, 2, \dots, n_j$

$j$  = kelompok,  $j = 1, 2, \dots, J$

$Y_{ij}$  = nilai respon pada individu ke- $i$  dan kelompok ke- $j$ ,

$\beta_{0j}$  = intersep pada kelompok ke- $j$

$\beta_{1j}$  = kemiringan (slope) garis pada kelompok ke- $j$

Koefisien regresi  $\beta_{0j}$  (*intersep*) dan  $\beta_{1j}$  (*slope*) untuk setiap kelompok dalam model 2-level memiliki nilai yang berbeda. Asumsi yang mendasari model regresi 2-level yaitu  $\varepsilon_{ij}$  berdistribusi normal dengan rata-rata nol dan ragam  $\sigma_j^2$ . Hal ini menunjukkan bahwa ragam tiap kelompok berbeda (Hox, 2002).

Pada persamaan (2.1) koefisien  $\beta_0$  dan  $\beta_1$  dapat diperoleh dengan menganggap  $\beta_0$  dan  $\beta_1$  sebagai respon dari persamaan-persamaan berikut.:

$$\left. \begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}Z_j + \delta_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}Z_j + \delta_{1j} \end{aligned} \right\} \quad (2.2)$$

Dengan  $Z_j$  adalah variabel bebas pada level kedua, sedangkan  $\delta_{0j}$  dan  $\delta_{1j}$  adalah galat pada level 2.

Selanjutnya persamaan (2.2) disubstitusikan ke persamaan (2.1) sehingga akan diperoleh :

$$Y_{ij} = \gamma_{00} + \gamma_{01}Z_j + \gamma_{10}X_{1j} + \gamma_{11}Z_jX_{1j} + (\delta_{0j} + \delta_{1j}X_{1j} + \varepsilon_{ij}) \quad (2.3)$$

Dalam persamaan (2.12) pada ruas kanan yang tidak berada dalam kurung merupakan komponen tetap (*fixed part*), sedangkan bagian yang berada dalam kurung disebut komponen acak (*random part*). Persamaan tersebut dapat disederhanakan menjadi model sebagai berikut :

$$Y_{ij} = \gamma_{00} + \gamma_{01}Z_j + \gamma_{10}X_{1j} + \gamma_{11}Z_jX_{1j} + u_{ij} \quad (2.4)$$

Dengan  $u_{ij} = (\delta_{0j} + \delta_{1j}X_{1j} + \varepsilon_{ij})$  atau disebut sebagai galat total.

Dalam persamaan (2.4) terlihat bahwa terdapat tiga komponen galat yaitu  $\delta_{0j}$ ,  $\delta_{1j}$  dan  $\varepsilon_{ij}$ . Asumsi yang mendasari model tersebut adalah :

1.  $E(\delta_{0j}) = E(\delta_{1j}) = E(\varepsilon_{ij}) = 0$
2.  $V(\delta_{0j}) = \sigma_{\delta_0}^2, V(\delta_{1j}) = \sigma_{\delta_1}^2, V(\varepsilon_{ij}) = \sigma_{\varepsilon}^2$
3.  $Cov(\delta_{0j}, \varepsilon_{ij}) = Cov(\delta_{1j}, \varepsilon_{ij}) = Cov(\varepsilon_{ij}, \varepsilon_{kl}) = 0$
4.  $Cov(\delta_{0j}, \delta_{1j}) = \sigma_{\delta_0\delta_1}$

Pada umumnya terdapat lebih dari satu variabel bebas pada level terendah (level-1), demikian pula pada level tertinggi (level-2). Jika diasumsikan terdapat P variabel bebas ( $X$ ) pada level-1, dan Q variabel bebas ( $Z$ ) pada level-2 maka persamaan 2.4 menjadi persamaan yang lebih umum sebagai berikut:

$$Y_{ij} = \gamma_{00} + \sum_{q=1}^Q \gamma_{0q} Z_{qj} + \sum_{p=1}^P \gamma_{p0} X_{pij} + \sum_{q=1}^Q \sum_{p=1}^P \gamma_{pq} Z_{qj} X_{pij} + \delta_{0j} + \sum_{p=1}^P \delta_{pj} X_{pij} + \varepsilon_{ij}$$

#### 2.4. Pendugaan Parameter

Pendugaan parameter yang digunakan adalah dengan metode *Iterative Generalized Least Square* (IGLS). Model yang digunakan adalah model dengan bentuk matriks sebagai berikut:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}$$

Langkah pertama dalam menaksir parameter dengan menggunakan metode IGLS adalah menaksir parameter tetap  $\boldsymbol{\beta}$ , dengan menggunakan metode *Generalized Least Square* (GLS), sehingga diperoleh persamaan sebagai berikut:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}\mathbf{V}^{-1}\mathbf{y}$$

Persamaan untuk penaksir  $\boldsymbol{\beta}$  tersebut masih mengandung unsur parameter yang nilainya tidak diketahui yaitu pada matriks  $\mathbf{V}$  yang merupakan matriks *block diagonal* dimana nilainya dapat diketahui dengan menggunakan metode *Generalized Least Square*. Selanjutnya hasil taksiran yang diperoleh digunakan untuk menaksir parameter acak ( $\sigma_{\delta_0}^2$  dan  $\sigma_{\varepsilon_0}^2$ ) dalam model menggunakan metode GLS. Prosedur penaksiran

parameter tetap dan parameter acak dilakukan berulang-ulang secara bergantian sampai mendapatkan hasil taksiran yang konvergen.

## 2.5. Korelasi Intraklas (*Intraclass Corelation*)

Jika mempunyai data dengan struktur hirarki sederhana yaitu 2-level, maka analisis regresi 2-level dapat digunakan untuk memberikan nilai dugaan bagi korelasi intraklas (*Intraclass Corelation*) (Hox, 2002). Korelasi intraklas menunjukkan proporsi keragaman yang dapat dijelaskan oleh struktur kelompok dalam populasi, yang dapat juga diinterpretasikan sebagai korelasi harapan antara dua unit yang dipilih secara acak yang berada dalam kelompok yang sama (Goldstein 1999, Hox 2002). Dengan menggunakan model intersep atau pun model intersep acak korelasi intraklas ( $\rho$ ) dapat dituliskan sebagai berikut:

$$\rho = \frac{\sigma_{\delta_0}^2}{\sigma_{\delta_0}^2 + \sigma_{\varepsilon}^2} \quad 0 \leq \rho \leq 1 \quad (2.5)$$

dengan  $\sigma_{\delta_0}^2$  adalah ragam dari galat pada level-2 dan  $\sigma_{\varepsilon}^2$  adalah ragam dari galat pada level-1.

## 2.6. Pemilihan dan Perbandingan Model

Menurut Hox (1995), dalam pembentukan model regresi 2-level terdapat beberapa langkah, yaitu:

1. Memilih Intersep acak
  - a. Menyusun model intersep acak tanpa variabel bebas
  - b. Menyusun model dengan menambahkan seluruh variabel bebas level-1
2. Memilih Kemiringan (slope) acak dengan cara menguji keragaman kemiringan pada setiap variabel bebas di level-1.
3. Memilih efek tetap dengan menambahkan variabel bebas level-2 ke dalam model
4. Menyusun model dengan menambahkan interaksi antar variabel bebas level-1 dan level-2 yang memiliki keragaman kemiringan yang signifikan.

Untuk memilih model regresi 2-level yang terbentuk, digunakan nilai *deviance* yaitu ukuran untuk menentukan cocok tidaknya suatu model. Perhitungan untuk pengujian ini adalah selisih nilai *deviance* antara dua model (*diff*), dapat dituliskan sebagai berikut:

$$\begin{aligned} diff &= -2 \log \left( \frac{l_0}{l_1} \right) \\ &= -2 \log (l_0) - (-2 \log (l_1)) \end{aligned}$$

Dengan  $-2 \log (l_0)$  adalah nilai *deviance* untuk model yang lebih sederhana dan  $-2 \log (l_1)$  adalah nilai *deviance* untuk model yang melibatkan parameter yang diuji. Nilai statistik *diff* menyebar mengikuti sebaran chi kuadrat dengan derajat bebas selisih dari banyaknya parameter antara kedua model (West, 2007).

## 2.7. Eksplorasi Data

Eksplorasi data dilakukan untuk mendeteksi keberadaan interaksi antar variabel bebas dalam level yang berbeda. Eksplorasi interaksi bermanfaat dalam pemilihan model. Untuk mengetahui adanya interaksi antar level yang berbeda digunakan uji beda antar garis regresi yaitu dengan melakukan uji kesejajaran garis. Kesejajaran atau ketidaksejajaran garis ditentukan oleh sama tidaknya besar nilai gradien garis-garis yang diuji atau besarnya sudut kemiringan garis (Santoso, Ratno D, 1992).

### 3. METODOLOGI PENELITIAN

#### 3.1. Data

Data yang digunakan dalam penelitian tugas akhir ini adalah data sekunder yang bersumber dari Survei Sosial Ekonomi Nasional atau SUSENAS tahun 2011 dan data hasil Potensi Desa (PODES) Jawa Tengah tahun 2011. Pada penelitian data yang digunakan sebanyak 500 yang diambil secara acak dari 19 kecamatan di Kabupaten Semarang. Adapun variabel-variabel penelitian yang digunakan adalah sebagai berikut:

1. Variabel tak bebas/ respon

$Y$  : Pendidikan Anak (Tahun)

2. Variabel bebas level-1 (Anak)

$X_1$  : Jenis Kelamin Anak ( 0 = Perempuan, 1 = Laki-laki)

$X_2$  : Pendidikan Ibu (Tahun)

$X_3$  : Pendidikan Ayah (Tahun)

$X_4$  : Tempat tinggal ( 0 = perdesaan, 1 = perkotaan)

3. Variabel bebas level-2 (Kecamatan)

$Z_1$  : Banyak SD di kecamatan

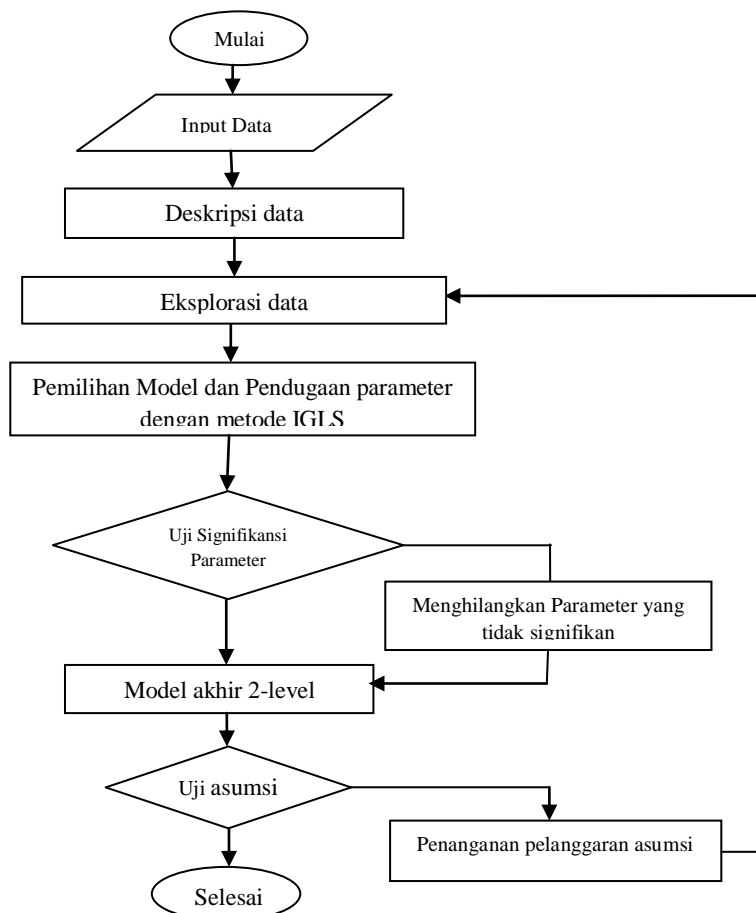
$Z_2$  : Banyak SMP di kecamatan

$Z_3$  : Banyak SMA di kecamatan

$Z_4$  : Prosentase keluarga petani di kecamatan

#### 3.2. Tahapan Analisis

Tahapan analisis yang dilakukan pada penelitian ini dapat dilihat pada diagram alir analisis berikut ini:



## 4. HASIL DAN PEMBAHASAN

### 4.1. Deskripsi dan Eksplorasi Data

Data Pendidikan anak di Kabupaten Semarang terdiri atas 2819 anak, dengan jumlah sampel diambil secara acak sebanyak 500 anak yang tersarang di dalam 19 kecamatan. Secara umum, deskripsi mengenai tingkat pendidikan seluruh kecamatan dapat dilihat pada Tabel 1, sebagai berikut:

**Tabel 1.** Statistika deskriptif tingkat pendidikan di seluruh kecamatan

| Statistik       | Nilai |
|-----------------|-------|
| Rataan          | 8,62  |
| Minimum         | 0,00  |
| Maksimum        | 17,00 |
| Q1              | 6,00  |
| Median          | 17,00 |
| Q3              | 12,00 |
| Standar Deviasi | 4,11  |

Rata-rata anak yang diteliti berjenis kelamin perempuan dengan prosentase sebesar 55.23% dan 55.78% anak yang diteliti sebagian besar tinggal di daerah perdesaan. Rata-rata prosentase keluarga petani per kecamatan sebesar 51%, dengan kecamatan yang paling banyak terdapat keluarga yang mata pencahariannya sebagai petani adalah kecamatan Bancok dengan 83%.

Eksplorasi dilakukan untuk mendeteksi keberadaan interaksi antar variabel bebas dalam level yang berbeda. Dari hasil pengujian secara grafis dan statistik dapat diketahui bahwa interaksi antar level terjadi pada variabel jumlah SD dengan pendidikan ibu, jumlah SD dengan pendidikan ayah, jumlah SMP dengan jenis kelamin, jumlah SMP dengan pendidikan ibu, jumlah SMP dengan pendidikan ayah, jumlah SMA dengan pendidikan ibu, jumlah SMA dengan pendidikan ayah, petani dengan jenis kelamin, petani dengan tempat tinggal, petani dengan pendidikan ibu dan petani dengan pendidikan ayah. Untuk interaksi antar variabel bebas lainnya tidak terlihat adanya interaksi dalam eksplorasi, karena menunjukkan garis sejajar.

### 4.2. Pemodelan Regresi 2-Level

Untuk mendapatkan model regresi 2-level yang terbaik, maka dilakukan tahapan-tahapan sebagai berikut:

#### Tahap 1 Pemilihan Struktur intersep acak

Pada tahap satu dilakukan pemilihan intersep acak dengan membandingkan dua model yaitu model intersep acak tanpa variabel bebas (M.1.1) dan model intersep acak dengan variabel bebas level-1 (M.1.2). Pemilihan struktur intersep acak digunakan untuk mengetahui apakah terdapat keragaman intersep antar anak dalam kecamatan dan antar kecamatan. Hasil perbandingan terhadap kedua model dapat disajikan pada Tabel 2 sebagai berikut:

**Tabel 2.** Hasil Perbandingan Model M.1.1 Dan M.1.2

| Model | Deviance | Diff    | Parameter | $\chi^2_{0,05(4)}$ |
|-------|----------|---------|-----------|--------------------|
| M.1.1 | 2812,753 |         | 3         |                    |
| M.1.2 | 2539,780 | 272,973 | 7         | 9,49               |

Berdasarkan Tabel 2 terlihat bahwa nilai *diff* lebih besar dari  $\chi^2_{0,05(4)}$ , sehingga model yang terbaik dalam tahap pemilihan struktur intersep acak adalah model M.1.2. Hal ini berarti variabel-variabel bebas level-1 dalam model memberikan pengaruh yang nyata terhadap pendidikan anak.

## Tahap 2. Pemilihan Struktur Kemiringan Acak

Dalam tahap pemilihan Kemiringan acak, variabel bebas yang digunakan adalah variabel bebas yang diperoleh pada langkah struktur intersep acak (M.2.1), yaitu jenis kelamin, pendidikan ibu, pendidikan ayah, dan tempat tinggal. Dengan pengujiannya dilakukan satu per satu setiap variabel, sehingga diperoleh 4 model yang dibandingkan yaitu:

1. Model intersep acak dan kemiringan acak pada variabel jenis kelamin (M.2.2)
2. Model intersep acak dan kemiringan acak pada variabel pendidikan ibu (M.2.3)
3. Model intersep acak dan kemiringan acak pada variabel pendidikan ayah (M.2.4)
4. Model intersep acak dan kemiringan acak pada variabel tempat tinggal (M.2.5)

**Tabel 3.** Hasil Perbandingan Model Kemiringan Acak

| Model              | <i>diff</i> | db | $\chi^2_{0,05(2)}$ |
|--------------------|-------------|----|--------------------|
| M.2.1 dengan M.2.2 | 1,256       | 2  | 5,99               |
| M.2.1 dengan M.2.3 | 0,988       | 2  | 5,99               |
| M.2.1 dengan M.2.4 | 7,099       | 2  | 5,99               |
| M.2.1 dengan M.2.5 | 8,836       | 2  | 5,99               |

Tabel 3 menunjukkan hasil perbandingan keempat model, dapat dilihat bahwa variabel bebas yang memiliki kemiringan acak yang signifikan terhadap model adalah kemiringan pendidikan ayah dan tempat tinggal, dikarenakan nilai *diff* dari Model M.2.4 dan M.2.5 lebih besar dari nilai  $\chi^2_{0,05(2)}$ , maka dapat dikatakan bahwa model terbaik untuk tahap ini adalah model M.2.4 dan M.2.5. Sehingga dengan menambahkan kemiringan acak pendidikan ayah dan tempat tinggal pada model M.1.2

## Tahap 3 Pemilihan Struktur Efek Tetap

Pemilihan struktur efek tetap bertujuan untuk mendapatkan variabel-variabel bebas yang memiliki pengaruh besar terhadap pendidikan anak dengan cara memasukkan variabel-variabel bebas setiap levelnya pada model. Dalam tahap ini terdapat dua model yang dibandingkan yaitu:

1. Model M.1.2 ditambah kemiringan acak yang signifikan (M.3.1)
2. Model M.1.2 ditambah kemiringan acak yang signifikan dan variabel bebas level-2 (M.3.2)

**Tabel 4.** Hasil perbandingan model M.3.1 dan M.3.2

| Model | <i>Deviance</i> | <i>diff</i> | Parameter | $\chi^2_{0,05(4)}$ |
|-------|-----------------|-------------|-----------|--------------------|
| M.3.1 | 2524,360        |             | 12        |                    |
| M.3.2 | 2512,411        | 11.849      | 16        | 9,49               |

Berdasarkan tabel 4 terlihat bahwa model yang dipilih adalah model M.3.2 dikarenakan nilai *diff* lebih besar dari nilai  $\chi^2_{0,05(4)}$ . Jadi model yang baik pada tahap ini adalah model M.3.2 dengan variabel bebas pada level-1 meliputi jenis kelamin, pendidikan ibu, pendidikan ayah, dan tempat tinggal serta variabel bebas pada level-2 yaitu jumlah SD, jumlah SMP, jumlah SMA dan prosentase keluarga petani.

#### Tahap 4 Penyusunan Model

Tahap akhir dalam pemilihan model, yaitu tahap penyusunan model dengan efek tetap dan efek acak yang signifikan ditambah dengan interaksi antar level (**M.4.1**). Berdasarkan eksplorasi data dan pemilihan kemiringan acak yang signifikan, interaksi yang ditambahkan dalam model adalah jumlah SD dengan pendidikan ayah, jumlah SMP dengan pendidikan ayah, jumlah SMA dengan pendidikan ayah, prosentase keluarga petani dengan pendidikan ayah dan prosentase keluarga petani dengan tempat tinggal, sehingga diperoleh model :

- **Model Level-1 (Anak)**

$$Y_{ij} = \beta_{0j} + \beta_1 X_1 + \beta_2 X_2 + \beta_{3j} X_3 + \beta_{4j} X_4 + \varepsilon_{ij}$$

- **Model Level-2 ( Kecamatan )**

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_1 + \gamma_{02} Z_2 + \gamma_{03} Z_3 + \gamma_{04} Z_4 + \delta_{0j}$$

$$\beta_1 = \gamma_{10}$$

$$\beta_2 = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31} Z_1 + \gamma_{32} Z_2 + \gamma_{33} Z_3 + \gamma_{34} Z_4 + \delta_{3j}$$

$$\beta_{4j} = \gamma_{40} + \gamma_{44} Z_4 + \delta_{4j}$$

Kedua model tersebut dapat digabungkan dengan mensubstitusikan model level-2 pada model level 1, sehingga diperoleh:

$$Y_{ij} = \gamma_{00} + \gamma_{01} Z_1 + \gamma_{02} Z_2 + \gamma_{03} Z_3 + \gamma_{04} Z_4 + \gamma_{10} X_1 + \gamma_{20} X_2 + \gamma_{30} X_3 + \gamma_{31} Z_1 * X_3 + \gamma_{32} Z_2 * X_3 + \gamma_{33} Z_3 * X_3 + \gamma_{34} Z_4 * X_3 + \gamma_{40} X_4 + \gamma_{44} Z_4 * X_4 + \delta_{0j} + \delta_{3j} X_3 + \delta_{4j} X_4 + \varepsilon_{ij}$$

Pendugaan parameter untuk model **M.4.1** dapat dilihat pada Tabel 5 . Dari Tabel 5 terlihat bahwa parameter variabel bebas yang berpengaruh terhadap pendidikan anak adalah pendidikan ibu, pendidikan ayah dan prosentase keluarga petani.

**Tabel 5.** Nilai Pendugaan Parameter Model Regresi 2-level (M.4.1)

| Parameter           | Dugaan | Standar Error | P_value |
|---------------------|--------|---------------|---------|
| <i>Fixed effect</i> |        |               |         |
| $\gamma_{00}$       | 5,827  | 2,113         | 0,006   |
| $\gamma_{10}$       | 0,307  | 0,268         | 0,254   |
| $\gamma_{20}$       | 0,357  | 0,054         | 0,000   |
| $\gamma_{30}$       | 0,472  | 0,055         | 0,001   |
| $\gamma_{40}$       | -0,483 | 1,055         | 0,647   |
| $\gamma_{01}$       | -0,017 | 0,047         | 0,719   |
| $\gamma_{02}$       | -0,039 | 0,144         | 0,783   |
| $\gamma_{03}$       | 0,223  | 0,269         | 0,407   |
| $\gamma_{04}$       | -0,095 | 0,039         | 0,016   |
| $\gamma_{31}$       | 0,001  | 0,007         | 0,843   |
| $\gamma_{32}$       | -0,001 | 0,022         | 0,966   |
| $\gamma_{33}$       | 0,036  | 0,044         | 0,411   |
| $\gamma_{34}$       | 0,002  | 0,006         | 0,707   |
| $\gamma_{44}$       | 0,029  | 0,039         | 0,465   |

Karena terdapat variabel bebas yang tidak signifikan, maka model yang terbentuk yaitu model **M.3.2** dimodifikasi dengan menghilangkan variabel yang tidak signifikan, yaitu jenis kelamin, tempat tinggal, jumlah SD, jumlah SMP dan jumlah SMA. Selain itu juga



tidak menyertakan interaksi antar variabel bebas pada level yang berbeda, sehingga diperoleh model akhir sebagai berikut:

**Tabel 6.** Hasil Pendugaan Parameter pada Model Akhir

| Parameter           | Dugaan | P-value |
|---------------------|--------|---------|
| <i>Fixed effect</i> |        |         |
| $\gamma_{00}$       | 5,862  | 0,000   |
| $\gamma_{20}$       | 0,346  | 0,000   |
| $\gamma_{30}$       | 0,486  | 0,000   |
| $\gamma_{04}$       | -0,058 | 0,000   |

Berdasarkan hasil estimasi parameter yang disajikan pada Tabel 6 dapat dibentuk model untuk menggambarkan studi kasus tingkat pendidikan anak di Kabupaten Semarang, dengan model yang terbentuk sebagai berikut:

**Model Level-1 :**

$$\hat{Y}_{ij} = \hat{\beta}_{0j} + \hat{\beta}_2 X_2 + \hat{\beta}_{3j} X_3$$

**Model Level-2 :**

$$\hat{\beta}_{0j} = 5,862 - 0,058Z_4$$

$$\hat{\beta}_2 = 0,346$$

$$\hat{\beta}_{3j} = 0,486$$

Dalam bentuk persamaan yang disubstitusikan adalah sebagai berikut:

$$\hat{Y}_{ij} = 5,862 - 0,058Z_4 + 0,346X_2 + 0,486X_3$$

Dari model tersebut dapat dijelaskan bahwa apabila prosentase keluarga petani meningkat satu persen, maka akan memberikan dampak penurunan tingkat pendidikan anak sebesar 0,058, sebaliknya apabila terjadi peningkatan pendidikan ibu dan ayah sebesar masing-masing satu tahun akan memberikan pengaruh positif terhadap peningkatan pendidikan anak di Kabupaten Semarang masing-masing sebesar 0,346 dan 0,486.

Selain diketahui nilai parameternya juga dapat diketahui nilai dugaan komponen acaknya berdasarkan model akhir yang terbentuk seperti disajikan pada tabel 7, sebagai berikut:

**Tabel 7.** Hasil Pendugaan Parameter Acak Model Akhir

| Parameter                 | Dugaan | P-value |
|---------------------------|--------|---------|
| <i>Random effect</i>      |        |         |
| $\sigma_{\delta_0}^2$     | 1,019  | 0,139   |
| $\sigma_{int:Pend\_ayah}$ | -0,107 | 0,202   |
| $\sigma_{\delta_3}^2$     | 0,011  | 0,355   |
| $\sigma_{\epsilon}^2$     | 8,973  | 0,000   |

Berdasarkan tabel 7 terlihat bahwa nilai varian (keragaman) pendidikan anak dalam kecamatan lebih tinggi dibandingkan nilai varian pendidikan anak antar kecamatan. Nilai dugaan kovarian antara intersep dan kemiringan pendidikan ayah sebesar -0,107 menandakan adanya hubungan negatif antara intersep dan kemiringan pendidikan ayah. Artinya untuk kecamatan-kecamatan dengan intersep rendah, pengaruh pendidikan ayah lebih besar dari pada kecamatan-kecamatan yang memiliki intersep tinggi, demikian pula sebaliknya. Model diatas menunjukkan adanya perbedaan antar kecamatan dengan keragaman intersep sebesar 1,109.

### 4.3. Korelasi Intraklas

Nilai koefisien korelasi intraklas diperoleh dari pendugaan varian pada setiap level untuk model tanpa melibatkan variabel bebas (*Intersept Only Models*). Hasil yang diperoleh disajikan pada Tabel 8.

**Tabel 8.** Hasil Pendugaan Parameter Acak

| Parameter                | Dugaan | P-value                  |
|--------------------------|--------|--------------------------|
| $\sigma_{\delta_0}^2$    | 1,224  | $\sigma_{\delta_0}^2$    |
| $\sigma_{\varepsilon}^2$ | 15,587 | $\sigma_{\varepsilon}^2$ |

Diperoleh nilai dugaan korelasi intraklas untuk regresi 2-level sebesar 0,0728. Sehingga korelasi antar anak dalam kecamatan yang mengukur proporsi dari total variansi antar kecamatan, yaitu sebesar 0,0728 atau dapat diartikan korelasi antara dua anak dalam satu kecamatan sebesar 7,28%.

## 5. KESIMPULAN

1. Model Regresi 2-level digunakan untuk memodelkan data dengan struktur hirarki. Model Regresi 2-level mengatasi masalah-masalah yang muncul dari data berstruktur hirarki yaitu masalah heteroskedastisitas.
2. Faktor-faktor yang mempengaruhi pendidikan anak di Kabupaten Semarang adalah pendidikan ibu dan pendidikan ayah. Dan faktor lain yang berpengaruh yang diukur pada tingkat kecamatan adalah prosentase keluarga petani.
3. Model regresi 2-level yang terbentuk adalah sebagai berikut:

**Model Level-1 :**

$$\hat{Y}_{ij} = \hat{\beta}_{0j} + \hat{\beta}_2 X_2 + \hat{\beta}_{3j} X_3$$

**Model Level-2 :**

$$\hat{\beta}_{0j} = 5,862 - 0,058Z_4$$

$$\hat{\beta}_2 = 0,346$$

$$\hat{\beta}_{3j} = 0,486$$

4. Nilai Korelasi Intraklas atau korelasi antara dua anak dalam satu kecamatan sebesar 7,28%.

## DAFTAR PUSTAKA

- Badan Pusat Statistik. 2011. *Kabupaten Semarang Dalam Angka Tahun 2011*. BPS. Semarang.
- Goldstein, H. 1999. *Multilevel Statistical Models*. Institut of Education. London.
- Hox, J. J. 1995. *Applied Multilevel Analysis*. TT-Publikaties. Amsterdam.
- Hox, J. J. 2002. *Multilevel Analysis: Techniques and Applications*. Lawrence Erlbaum Associates Publishers. New Jersey.
- Ihsan, Fuad. 2005. *Dasar-Dasar Kependidikan*. Rineka Cipta. Jakarta
- Jones, Bradford. S., Steenbergen, Marco. (1997). *Modeling Multilevel Data Structure*. Sociological Methods and Research, Vol 22. 3, pp. 283-299
- Santoso, Ratno D., Kusnadi M. H. 1992. *Analisis Regresi*. Andi Offset. Yogyakarta.
- West, Brady T., Welch K.B., Galecki, A.T. 2007. *Linear Mixed Model: A Partical Guide Using Statistical Software*. Chapman & Hall. New York.