ANALISIS REGRESI MULTILEVEL TERHADAP DATA KEPADATAN PENDUDUK PROVINSI LAMPUNG TAHUN 2016 DENGAN METODE *MAXIMUM LIKELIHOOD*

(Skripsi)

Oleh

FIETRA LISTIANA KARTIKA MURTI

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2018

ABSTRACT

MULTILEVEL REGRESSION ANALYSIS OF POPULATION DENSITY OF LAMPUNG PROVINCE IN 2016 WITH MAXIMUM LIKELIHOOD METHOD

By

Fietra Listiana Kartika Murti

Hierarchical data is data collected from two or more levels where lower level nest at higher level. Initially the analysis performed without regard to information at a higher level this resulted in dissatisfaction on the results of analysis and caused heteroskedastisity in error then the solve of the problem is used multilevel regression model. The purpose of this study is to determine the factors affecting population density in Lampung province in 2016 at district and sub-district level and to know the diversity that can be explained at the district and sub-district level. The results showed that the best multilevel model is a multilevel model that includes the variables at the district level and the factors affecting the population density in Lampung province that is the rate of economic growth, the sex ratio, population growth, and the original income of the sub-district. The diversity that can be explained at the district level is 58% while for sub-district level is 14%.

Key words: hierarchical data, population density, multilevel regression

ABSTRAK

ANALISIS REGRESI MULTILEVEL TERHADAP DATA KEPADATAN PENDUDUK PROVINSI LAMPUNG TAHUN 2016 DENGAN METODE *MAXIMUM LIKELIHOOD*

Oleh

Fietra Listiana Kartika Murti

Data hirarki adalah data dikumpulkan dari dua atau lebih level dimana level yang lebih rendah bersarang pada level yang lebih tinggi. Awalnya analisis yang dilakukan tanpa memperhatikan informasi pada level yang lebih tinggi hal ini mengakibatkan ketidakpuasan pada hasil analisisnya dan menimbulkan heteroskedastisitas pada galat maka untuk mengatasi masalah tersebut digunakan model regresi multilevel. Tujuan dari penelitian ini adalah untuk mengetahui faktor-faktor yang mempengaruhi kepadatan penduduk di provinsi Lampung tahun 2016 pada level kabupaten dan level kecamatan dan mengetahui keragaman yang dapat dijelaskan di level kabupaten dan di level kecamatan. Dari hasil analisis diperoleh bahwa model multilevel terbaik yaitu model multilevel yang mengikutsertakan variabel pada level kabupaten dan faktor-faktor yang mempengaruhi kepadatan penduduk di provinsi Lampung yaitu laju pertumbuhan ekonomi, rasio jenis kelamin, pertambahan penduduk, dan pendapatan asli kecamatan. Keragaman yang dapat dijelaskan pada level kabupaten sebesar 58% sedangkan untuk level kecamatan adalah sebesar 14% .

Kata kunci: data hirarki, kepadatan penduduk, regresi multilevel

ANALISIS REGRESI MULTILEVEL TERHADAP DATA KEPADATAN PENDUDUK PROVINSI LAMPUNG TAHUN 2016 DENGAN METODE *MAXIMUM LIKELIHOOD*

Oleh

FIETRA LISTIANA KARTIKA MURTI

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA SAINS

pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2018

Judul Skripsi : **ANALISIS REGRESI MULTILEVEL TERHADAP DATA KEPADATAN PENDUDUK PROVINSI LAMPUNG TAHUN 2016 DENGAN METODE *MAXIMUM LIKELIHOOD***

Nama Mahasiswa : *Fietra Listiana Kartika Murti*

Nomor Pokok Mahasiswa : 1417031052

Program Studi : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Ir. Netti Herawati, M.Sc., Ph.D.	**Drs. Eri Setiawan, M.Si.**
NIP. 19650125 199003 2 001	NIP. 19581101 198803 1 002

2. Ketua Jurusan Matematika

Prof. Dra. Wamiliana, M.A., Ph.D.
NIP. 19631108 198902 2 001

MENGESAHKAN

1. Tim Penguji

Ketua : **Ir. Netti Herawati, M.Sc., Ph.D.** ....

Sekretaris : **Drs. Eri Setiawan, M.Si.** ....

Penguji
Bukan Pembimbing : **Prof. Drs. Mustofa Usman, M.A., Ph.D.** ....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Prof. Warsito, S.Si., D.E.A., Ph.D.
NIP. 19710212 199512 1 001

Tanggal Lulus Ujian Skripsi : **5 Februari 2018**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini :

Nama	: **Fietra Listiana Kartika Murti**
Nomor Pokok Mahasiswa	: **1417031052**
Jurusan	: **Matematika**
Judul Skripsi	: **Analisis Regresi Multilevel Terhadap Data Kepadatan Penduduk Provinsi Lampung Tahun 2016 dengan Metode *Maximum Likelihood***

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, Februari 2018

Yang Menyatakan

Fietra Listiana Kartika Murti

RIWAYAT HIDUP

Penulis bernama lengkap Fietra Listiana Kartika Murti, putri bungsu Bapak Teguh Purnomo dan Ibu Tati Sumiati serta adik dari Fredy Kurniawan. Penulis lahir di Gedong Tataan pada tanggal 13 Februari 1997.

Penulis menempuh pendidikan dasar di SD Negeri 8 Gadingrejo Kabupaten Pringsewu dari tahun 2002 – 2008. Kemudian melanjutkan pendidikan di SMP Negeri 1 Gadingrejo Kabupaten Pringsewu dan lulus pada tahun 2011. Kemudian menempuh pendidikan di SMA Negeri 1 Gadingrejo Kabupaten Pringsewu dan lulus pada tahun 2014.

Pada tahun 2014 penulis diterima sebagai mahasiswi di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN). Pada tahun 2017 penulis melakukan Kerja Praktik di Kantor Badan Pusat Statistik Kabupaten Pringsewu dan sebagai salah satu bentuk pengabdian kepada masyarakat penulis melaksanakan Kuliah Kerja Nyata di Desa Gayam, Kecamatan Penengahan, Kabupaten Lampung Selatan.

KATA INSPIRASI

Allah akan meninggikan orang-orang beriman di antara kamu dan orang-orang yang diberikan ilmu pengetahuan beberapa derajat.

(Q.S. Al Mujadilah ayat 11)

Barang siapa yang menempuh perjalanan untuk mencari ilmu, niscaya Allah subhanahu wata'ala menyediakan jalan untuknya menuju surga.

(H.R. Abu Daud, Tirmidzi, dan Ibnu Majjah)

Orang yang mengeluh adalah orang yang tidak bersyukur.

(Fietra Listiana Kartika Murti)

PERSEMBAHAN

Karyaku yang kecil dan sederhana ini ku persembahkan kepada:

Papa dan Mama

Terima kasih kepada Papa dan Mama yang senantiasa sabar membimbing, memberikan semangat, dukungan moril dan materil serta do'a dan kasih sayang yang tiada henti.

Kakakku Fredy dan Kakak Iparku Dwi

Terima kasih kepada kak Fredy dan kak Dwi yang selalu memberikan do'a, motivasi, saran, serta nasehat selama ini.

Sahabat-sahabatku Risky, Linda, Reka, Mona dan Indah

Terima kasih kepada para sahabatku yang tiada henti memberikan do'a, semangat, motivasi, dan saran padaku serta telah memberikan pengalaman yang berharga selama ini.

Almamater dan Negeriku

SANWACANA

Puji syukur kehadirat Allah SWT yang telah memberikan limpahan rahmat, taufik serta hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi ini, dengan judul Analisis Regresi Multilevel Terhadap Data Kepadatan Penduduk Provinsi Lampung Tahun 2016 dengan Metode *Maximum Likelihood*.

Penulis menyadari bahwa skripsi ini tidak akan terwujud tanpa adanya dukungan bimbingan, bantuan, saran, serta do'a dari berbagai pihak sehingga skripsi ini dapat terselesaikan. Oleh karena itu, dalam kesempatan ini penulis menyampaikan terima kasih kepada:

1. Ibu Ir. Netti Herawati, M.Sc., Ph.D., selaku dosen pembimbing satu yang telah memberikan bimbingan, pengarahan, dan saran kepada penulis dalam mengerjakan skripsi.
2. Bapak Drs. Eri Setiawan, M.Si., selaku pembimbing dua yang telah memberikan saran serta pembelajaran yang sangat bermanfaat dalam menyelesaikan skripsi.
3. Bapak Prof. Drs. Mustofa Usman, M.A., Ph.D., selaku pembahas dan penguji skripsi yang telah memberikan evaluasi, arahan, dan saran demi perbaikan skripsi.
4. Bapak Agus Sutrisno, S.Si., M.Si., selaku dosen pembimbing akademik.

5. Ibu Prof. Dra. Wamiliana, M.A., Ph.D., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Seluruh dosen Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Papa dan mama tercinta yang selalu berjuang dan berdo'a demi kesuksesan penulis.
8. Kak Fredy Kurniawan dan kak Dwi Nurhajizah terima kasih atas segala dukungan, motivasi, saran, dan nasehat.
9. Risky, Linda, Reka, Mona, dan Indah sahabat-sahabat tercinta terima kasih atas segala do'a dan dukungan kepada penulis.
10. Uti, Ulfa, Jelly, Rium, Aldo, Alvin, dan Kodir teman-teman satu bimbingan terima kasih atas dukungan serta saran selama penyelesaian skripsi.
11. Kiki, Priska, Malik, dan Anggi teman-teman KKN Desa Gayam terima kasih atas segala do'a dan dukungan kepada penulis.
12. Teman-teman mahasiswa Jurusan Matematika angkatan 2014.

Penulis menyadari bahwa dalam penulisan skripsi ini masih jauh dari sempurna, sehingga informasi tambahan, saran, dan kritik untuk pengembangan lebih lanjut sangatlah penulis harapkan. Semoga skripsi ini dapat bermanfaat bagi penulis dan bagi kita semua.

Bandar Lampung,   Februari 2018

Penulis

Fietra Listiana Kartika Murti

DAFTAR ISI

DAFTAR TABEL

Tabel Halaman

DAFTAR GAMBAR

Gambar Halaman

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Di era modern seperti saat ini, banyak penelitian yang dilakukan dengan menggunakan data yang diambil dari kelompok-kelompok yang membawahi unit-unit pengamatan. Sebagai contoh penelitian dalam bidang sosial dan kependudukan, kepadatan penduduk di suatu kecamatan dipengaruhi pula oleh keadaan sosial, ekonomi, maupun kesehatan di kabupaten dimana kecamatan itu berada. Data yang digunakan pada penelitian di atas disebut sebagai data berstruktur hirarki.

Jika data dikumpulkan dari dua atau lebih kategori yang memiliki hirarki bertingkat maka kategori-kategori tersebut dinamakan level. Terdapat dua level pada contoh di atas, level yang lebih tinggi secara hirarkis dinamakan level makro atau level 1 yaitu kabupaten, sedangkan level yang lebih rendah dinamakan level mikro atau level 2 yaitu kecamatan (Harlan, 2016).

Awalnya analisis yang dilakukan tanpa memperhatikan informasi pada level makro, hal ini mengakibatkan ketidakpuasan pada hasil analisisnya dan menimbulkan heteroskedastisitas pada galat (Tantular, 2009).

Untuk mengatasi masalah-masalah di atas model regresi yang digunakan adalah model regresi multilevel yang diperkenalkan oleh Goldstein. Untuk data Gaussian yaitu variabel dependen berjenis variabel kontinu dan berdistribusi normal dengan efek campuran maka model multilevel yang diterapkan yaitu model linear campuran (*linear mixed models*) dan pendugaan parameter menggunakan Metode Kemungkinan Maksimum (*Maximum Likelihood Methods*). Oleh karena itu, pada penelitian ini akan dibahas mengenai faktor-faktor yang mempengaruhi kepadatan penduduk di provinsi Lampung tahun 2016 pada level kabupaten dan level kecamatan menggunakan analisis regresi multilevel dengan metode *Maximum Likelihood*.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Menentukan model regresi multilevel terhadap faktor-faktor yang mempengaruhi kepadatan penduduk di provinsi Lampung tahun 2016.
2. Menguraikan keragaman yang dapat dijelaskan oleh level kabupaten dan level kecamatan terhadap kepadatan penduduk di provinsi Lampung tahun 2016.

1.3 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Dapat mengetahui model regresi multilevel terhadap faktor-faktor yang mempengaruhi kepadatan penduduk di provinsi Lampung tahun 2016.

2. Dapat mengetahui keragaman yang dapat dijelaskan oleh level kabupaten dan level kecamatan terhadap kepadatan penduduk di provinsi Lampung tahun 2016.

II. TINJAUAN PUSTAKA

2.1 Model Linear Campuran (*Linear Mixed Model*)

Model linear campuran (*linear mixed models*) merupakan suatu model yang menggabungkan efek tetap (*fixed effect*) dan efek acak (*random effect*) ke dalam suatu persamaan. Sebuah variabel independen dikatakan memiliki efek tetap jika koefisien regresinya bernilai sama bagi seluruh anggota sampel dan sebuah variabel independen dikatakan memiliki efek acak jika nilai koefisien regresinya berbeda antar dua atau lebih grup anggota sampel (Harlan, 2016). Secara umum model linear campuran menurut Bryck dan Raudenbush (1987) meliputi tiga hal yaitu sebagai berikut:

1. Efek acak, yaitu efek yang ditimbulkan oleh adanya pengaruh dari suatu variabel yang nilainya berasal dari sampel acak.
2. Efek hirarki, yaitu efek yang ditimbulkan oleh adanya pengaruh peubah yang diukur pada level yang berbeda.
3. Pengukuran berulang, dalam hal ini pengamatan-pengamatan berkaitan dengan pengamatan sebelumnya.

Menurut Rencher dan Schaalje (2007), persamaan model linear campuran dalam bentuk sederhana adalah

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z_1 u_1} + \boldsymbol{Z_2 u_2} + \cdots + \boldsymbol{Z_m u_m} + \boldsymbol{\varepsilon} \tag{2.1}$$

dengan

Y = vektor respon (n x 1)

X = matriks prediktor efek tetap (n x p)

$\boldsymbol{\beta}$ = vektor parameter efek tetap (p x 1)

$\boldsymbol{Z_i}$ = matriks prediktor efek acak (n x r_i)

$\boldsymbol{u_i}$ = vektor parameter efek acak (r_i x 1)

ε = vektor galat (r_i x 1)

$E(\boldsymbol{\varepsilon}) = \mathbf{0}$ dan $Cov(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I_n}$ untuk $i = 1, \dots, m$

$E(\boldsymbol{u_i}) = \mathbf{0}$ dan $Cov(\boldsymbol{u_i}) = \sigma^2 \boldsymbol{I}_{r_i}$ untuk $i = 1, \dots, m$

$Cov(\boldsymbol{u_i}, \boldsymbol{u_j}) = \mathbf{O}$ untuk $i \neq j$, dimana **O** adalah $\boldsymbol{r_i} \times \boldsymbol{r_j}$

$Cov(\boldsymbol{u_i}, \varepsilon) = \mathbf{O}$ untuk semua i, dimana **O** adalah $\boldsymbol{r_i} \times \boldsymbol{n}$.

Maka $E(\boldsymbol{Y}) = \boldsymbol{X\beta}$ dan $Cov(\boldsymbol{Y}) = \sum = \sum_{i=1}^{m} \boldsymbol{\sigma_i^2}\, \boldsymbol{Z_i Z_i'} + \boldsymbol{\sigma^2 I_n}$

2.2. Model Regresi

Secara umum regresi adalah metode yang digunakan untuk meramalkan nilai harapan yang bersyarat (Kutner, dkk., 2004). Analisis regresi pertama kali dikembangkan oleh Sir Francis Galton pada abad ke-19. Bentuk hubungan antara variabel dependen dengan variabel independen dapat dinyatakan dalam bentuk persamaan atau model regresi. Persamaan regresi linear biasa ditulis sebagai berikut:

$$\boldsymbol{Y} = \boldsymbol{\beta X} + \boldsymbol{\varepsilon} \quad \text{dengan } \varepsilon \sim N(0, \sigma^2) \tag{2.2}$$

dengan

$\boldsymbol{Y}$ = variabel dependen

$\boldsymbol{X}$ = variabel independen

$\boldsymbol{\beta}$ = koefisien variabel independen

$\boldsymbol{\varepsilon}$ = galat

Analisis regresi dengan satu variabel independen dan satu variabel dependen disebut analisis regresi linear sederhana. Persamaan umumnya adalah:

$$Y = \alpha + \beta X + \varepsilon \tag{2.3}$$

dengan

Y = variabel dependen

X = variabel independen

α = intersep atau konstanta regresi

β = slope atau koefisien regresi

ε = galat

Sedangkan analisis regresi yang melibatkan lebih dari satu variabel independen dengan satu variabel dependen disebut analisis regresi linear berganda. Persamaan umumnya adalah:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots. + \beta_n X_n + \varepsilon \tag{2.4}$$

Dalam suatu persamaan regresi terdapat koefisien yang merupakan nilai duga perameter atau kondisi yang sebenarnya. Kuat tidaknya hubungan antara variabel independen dengan variabel dependen dapat diukur menggunakan koefisien korelasi, sedangkan besarnya pengaruh variabel independen terhadap variabel dependen dapat diukur menggunakan koefisien determinasi.

Asumsi-asumsi pada analisis regresi adalah sebagai berikut:

1. Galat menyebar normal $\varepsilon_i \sim N(0, \sigma^2)$;
2. Ragam galat homogen $Var(\varepsilon_i) = \sigma^2$; $i = 1,2, \dots, n$;
3. Nilai ε_i adalah bebas satu dengan yang lainnya $E(\varepsilon_i) = 0$ dan $E(\varepsilon_i{}^2) = \sigma^2$;
4. X dan Y terkait secara linear, untuk setiap nilai X yang dihubungkan dengan nilai Y maka akan membentuk garis lurus.

Dalam melakukan analisis regresi, asumsi-asumsi tersebut perlu dipenuhi terlebih dahulu (Myers, 1990).

2.3. Model Regresi Multilevel

Model regresi multilevel diperkenalkan oleh Goldstein pada tahun 1995 yang bertujuan untuk mengatasi masalah pada data yang berstruktur hirarki yaitu data yang dianalisis berasal dari beberapa level. Model regresi multilevel merupakan bagian dari model linear campuran yaitu menggabungkan efek tetap dan efek acak ke dalam suatu persamaan.

Pada regresi multilevel variabel dependen diukur pada level mikro atau level 2, sedangkan variabel independen dapat didefinisikan pada setiap level. Bentuk sederhana model regresi multilevel adalah regresi dua level. Jika ada lebih dari satu variabel independen maka diasumsikan ada banyaknya banyaknya variabel independen (Z) pada level 1 sebanyak q (q = 1, 2, 3, ..., s) dan variabel independen (X) pada level 2 sebanyak p (p = 1, 2, 3, ..., r) maka persamaan regresi multilevel tanpa variabel Z adalah sebagai berikut:

Level 1:

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{2.5}$$

sedangkan untuk persamaan regresi multilevel dengan variabel Z adalah sebagai berikut:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_{1j} + \gamma_{02}Z_{2j} + \cdots + \gamma_{0q}Z_{qj} + u_{0j}$$

atau dapat dituliskan sebagai berikut:

$$\beta_{0j} = \gamma_{00} + \sum_{q=1}^{s} \gamma_{0q}Z_{qj} + u_{0j} \tag{2.6}$$

dengan

β_{0j} = variabel dependen untuk kelompok ke-j pada level 1

γ_{00} = intersep level 1

γ_{0q} = efek tetap untuk variabel independen ke-q level 1

Z_{qj} = variabel independen ke-q untuk kelompok ke-j level 1

u_{0j} = efek acak, $u_{0j} \sim N(0, \sigma_u^2)$

Level 2:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \cdots + \beta_{pj}X_{pij} + \varepsilon_{ij}$$

atau dapat dituliskan sebagai berikut:

$$Y_{ij} = \beta_{0j} + \sum_{p=1}^{r} \beta_{pj}X_{pij} + \varepsilon_{ij} \tag{2.7}$$

dengan

Y_{ij} = variabel dependen dari individu ke-i dalam kelompok ke-j

β_{0j} = intersep

β_{pj} = koefisien regresi atau slope regresi

X_{pij} = variabel independen dari level 2 ke-*i* dalam level 1 ke-*j*

ε_{ij} = galat, $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$

i = indeks pada level 2 ($i = 1, 2, \ldots, n_j$)

j = indeks pada level 1 ($j = 1, 2, \ldots, J$)

Pada regresi multilevel masing-masing level memiliki nilai koefisien intersep β_{0j} dan koefisien slope β_{1j}, β_{2j}, sampai β_{pj}yang berbeda-beda sedangkan galat pada semua level diasumsikan sama dan dilambangkan dengan σ^2. Jika persamaan 2.5 dan 2.6 disubstitusikan ke persaman 2.7 maka:

$$Y_{ij} = \gamma_{00} + \sum_{p=1}^{r} \beta_{pj} X_{pij} + u_{0j} + \varepsilon_{ij} \qquad (2.8)$$

persamaan 2.8 merupakan persamaan regresi multilevel tanpa variabel Z sedangkan untuk persamaan regresi multilevel dengan variabel Z adalah sebagai berikut:

$$Y_{ij} = \gamma_{00} + \sum_{q=1}^{s} \gamma_{0q} Z_{qj} + \sum_{p=1}^{r} \beta_{pj} X_{pij} + u_{0j} + \varepsilon_{ij} \qquad (2.9)$$

Menurut Goldstein (1999) asumsi model regresi multilevel adalah sebagai berikut:

1. $E(e_{ij}) = E(u_{0j}) = E(u_{pj}) = 0$
2. $V(e_{ij}) = \sigma_e^2$, $V(u_{0j}) = \sigma_{u0}^2$, dan $V(u_{pj}) = \sigma_{up}^2$
3. $Cov\,(u_{0j}, u_{pj}) = \sigma_{u0p}^2$

2.4 Uji Asumsi

2.4.1 Uji Normalitas

Uji normalitas adalah uji untuk melihat apakah data berdistribusi normal atau tidak. Uji statistik yang sering digunakan untuk menghitung uji normalitas adalah uji *Kolmogorov-Smirnov*. Uji *Kolmogorov-Smirnov* bekerja dengan cara membandingkan dua distribusi atau sebaran data, yaitu distribusi yang dihipotesiskan dan distribusi yang teramati. Apabila distribusi yang teramati mirip dengan distribusi yang dihipotesiskan, maka dapat disimpulkan bahwa data yang diamati memiliki distribusi atau sebaran normal (Kurniawan, 2008). Selain uji *Kolmogorov-Smirnov* uji normalitas dapat dilakukan dengan melihat *Normal Probability Plot.*

2.4.2 Uji Multikolinearitas

Uji multikolinearitas adalah uji untuk melihat ada atau tidaknya korelasi yang tinggi antara variabel-variabel independen dalam suatu model regresi linear berganda. Uji statistik yang sering digunakan untuk menguji gangguan multikolinearitas adalah *Variance Inflation Factor* (VIF). VIF dapat menginterpretasikan akibat dari korelasi antar peubah prediktor ke-*i* pada ragam penduga koefisien regresi. Perhitungan VIF sebagai berikut:

$$VIF_i = \frac{1}{1-R_I^2} \text{ ; dimana } I = 1, 2, \dots, p-1$$

(2.10)

Pada uji statistik *Variance Inflation Factor* (VIF), apabila nilai VIF lebih besar dari sepuluh mengindikasi adanya multikolinearitas yang serius (Zulmi, 2011).

2.5 Metode Pendugaan Parameter

Pendugaan parameter pada regresi multilevel ini pada dasarnya mempunyai kegunaan yang sama pada regresi linear biasa, yaitu untuk mengetahui nilai-nilai parameter yang akan digunakan dalam proses regresi. Pada analisi multilevel, pendugaan parameter yang paling lazim digunakan adalah Metode Kemungkinan Maksimum atau *Maximum Likelihood Estimation*. Metode *Maximum Likelihood* mengestimasi parameter dengan cara memaksimumkan fungsi *likelihood* (Hox, 2010). Berikut ini adalah pendugaan parameter dengan metode *Maximum Likelihood.*

Diberikan persamaan model linear campuran sebagai berikut :

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \varepsilon \tag{2.11}$$

dimana $\boldsymbol{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \boldsymbol{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\sigma^2 I_n} = \sum_{\boldsymbol{i}}$ dan $\boldsymbol{E}(\boldsymbol{u_i}) = \mathbf{0}, \boldsymbol{cov}(\boldsymbol{u_i}) = \boldsymbol{\sigma^2 I_{r_i}} = \boldsymbol{D}$

atau $\begin{bmatrix} \mathbf{u}_{\boldsymbol{i}} \\ \varepsilon \end{bmatrix} \sim \boldsymbol{N}_{\boldsymbol{mq+n}}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{G} & \mathbf{0}_{\boldsymbol{mq \times n}} \\ \mathbf{0}_{\boldsymbol{n \times mq}} & \boldsymbol{R} \end{bmatrix}\right).$

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{D} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{D} \end{bmatrix} \text{ dan } \boldsymbol{R} = \begin{bmatrix} \sum_{\mathbf{1}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sum_{\mathbf{2}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sum_{\boldsymbol{n}} \end{bmatrix}$$

dengan

Y = vektor respon (n x 1)

X = matriks desain efek tetap (n x p)

$\boldsymbol{\beta}$ = vektor parameter efek tetap (p x 1)

Z = matriks desain efek acak (n x q)

u = vektor parameter efek acak (q x 1)

ε = vektor galat (n x 1)

D = matriks kovarian dari efek acak **u**

R = matriks kovarian dari vektor galat ε

Misalkan

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

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon^* \tag{2.12}$$

dengan

$\varepsilon^* := \boldsymbol{Zu} + \varepsilon$

$= (\boldsymbol{Z} \quad \boldsymbol{I}_{\boldsymbol{n}\times\boldsymbol{n}})\begin{pmatrix}\boldsymbol{u}\\ \varepsilon\end{pmatrix}$

$= \boldsymbol{A}\begin{pmatrix}\boldsymbol{u}\\ \varepsilon\end{pmatrix}$

$\varepsilon^* \sim \boldsymbol{N}_{\boldsymbol{n}}(\boldsymbol{0}, \boldsymbol{V})$

dimana

$\boldsymbol{V} = \boldsymbol{A}\begin{pmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}\end{pmatrix}\boldsymbol{A}^{\boldsymbol{t}}$

$= (\boldsymbol{Z} \quad \boldsymbol{I}_{\boldsymbol{n}\times\boldsymbol{n}})\begin{pmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}\end{pmatrix}\begin{pmatrix}\boldsymbol{Z}^{\boldsymbol{t}}\\ \boldsymbol{I}_{\boldsymbol{n}\times\boldsymbol{n}}\end{pmatrix}$

$= (\boldsymbol{ZG} \quad \boldsymbol{R})\begin{pmatrix}\boldsymbol{Z}^{\boldsymbol{t}}\\ \boldsymbol{I}_{\boldsymbol{n}\times\boldsymbol{n}}\end{pmatrix}$

$= \boldsymbol{ZGZ}^{\boldsymbol{t}} + \boldsymbol{R}$

$$\boldsymbol{V} = \boldsymbol{ZGZ}^{\boldsymbol{t}} + \boldsymbol{R}$$

Vektor parameter kovarian θ diduga oleh $\hat{\theta}_{ML}$ dengan memaksimumkan fungsi likelihood.

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

dengan $\begin{bmatrix}\mathbf{u}_i\\ \varepsilon\end{bmatrix} \sim \boldsymbol{N}_{\boldsymbol{mq+n}}\left(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \begin{bmatrix}\boldsymbol{G} & \boldsymbol{0}_{\boldsymbol{mq\times n}}\\ \boldsymbol{0}_{\boldsymbol{n\times mq}} & \boldsymbol{R}\end{bmatrix}\right)$.

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon^*$$

dengan $\varepsilon^{*} \sim \boldsymbol{N_n}(\mathbf{0}, \boldsymbol{V}(\boldsymbol{\theta}))$ dengan $\boldsymbol{V}(\boldsymbol{\theta}) = \boldsymbol{ZG}(\boldsymbol{\theta})\boldsymbol{Z^t} + \boldsymbol{R}(\boldsymbol{\theta})$.

Diberikan fungsi likelihood sebagai berikut:

$$\ln L(\boldsymbol{\beta}, \boldsymbol{\sigma^2}) = -\frac{1}{2}\{\ln|V(\theta)| + (\boldsymbol{Y} - \boldsymbol{X\beta})^{\boldsymbol{t}}\boldsymbol{V}(\boldsymbol{\theta})^{-\mathbf{1}}(\boldsymbol{Y} - \boldsymbol{X\beta})\} \tag{2.13}$$

dengan memaksimumkan log likelihood untuk θ tetap terhadap β maka diperoleh

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) \coloneqq (\boldsymbol{X^t V}(\boldsymbol{\theta})^{-\mathbf{1}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X^t V}(\boldsymbol{\theta})^{-\mathbf{1}}\boldsymbol{Y} \tag{2.14}$$

$$\boldsymbol{Y} \sim \boldsymbol{N_n}(\boldsymbol{X\beta},\ \boldsymbol{V}) \text{ dan } \boldsymbol{u} \sim \boldsymbol{N_{mq}}(\mathbf{0},\ \boldsymbol{G}).$$

$$\begin{aligned} Cov(\boldsymbol{Y}, \boldsymbol{u}) &= Cov\ (\boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}, \boldsymbol{u}) \\ &= Cov\ (\boldsymbol{X\beta}, \boldsymbol{u}) + Cov\ (\boldsymbol{Zu}, \boldsymbol{u}) + Cov\ (\boldsymbol{\varepsilon}, \boldsymbol{u}) \\ &= \mathbf{0} + \boldsymbol{ZG} + \mathbf{0} \\ &= \boldsymbol{ZG} \end{aligned}$$

$$Cov(\boldsymbol{Y}, \boldsymbol{u}) = \boldsymbol{ZG} \tag{2.15}$$

$$\begin{aligned} \boldsymbol{E}(\boldsymbol{u}|\boldsymbol{Y}) &= \boldsymbol{E}(\boldsymbol{u}) + \boldsymbol{cov}(\boldsymbol{u}, \boldsymbol{Y})[\boldsymbol{cov}(\boldsymbol{Y})]^{-\mathbf{1}}[\boldsymbol{Y} - \boldsymbol{E}(\boldsymbol{Y})] \\ &= \mathbf{0} + \boldsymbol{GZ^t V^{-1}}(\boldsymbol{Y} - \boldsymbol{X\beta}) \\ &= \boldsymbol{GZ^t V^{-1}}(\boldsymbol{Y} - \boldsymbol{X\beta}) \end{aligned}$$

$$\boldsymbol{E}(\boldsymbol{u}|\boldsymbol{Y}) = \boldsymbol{GZ^t V^{-1}}(\boldsymbol{Y} - \boldsymbol{X\beta}) \tag{2.16}$$

$\boldsymbol{GZ^t V^{-1}}(\boldsymbol{Y} - \boldsymbol{X\beta})$ adalah *Best Linear Unbiased Predictor* (BLUP) dari $\boldsymbol{u}$.

Menurut West dkk (2007) efek tetap β dan efek acak u diduga oleh

$$\widehat{\boldsymbol{\beta}} \coloneqq (\boldsymbol{X^t}\widehat{\boldsymbol{V}}^{-\mathbf{1}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X^t}\widehat{\boldsymbol{V}}^{-\mathbf{1}}\boldsymbol{Y} \tag{2.17}$$

$$\widehat{\boldsymbol{u}} \coloneqq \boldsymbol{GZ^t}\widehat{\boldsymbol{V}}^{-\mathbf{1}}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) \tag{2.18}$$

dimana $\widehat{\boldsymbol{V}} = \boldsymbol{V}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{ML}}\right)$.

Metode *Maximum Likelihood* memiliki dua kelebihan dibandingkan dengan REML yaitu relatif lebih mudah dari segi komputasi dan metode ini lazim digunakan untuk mengestimasi efek tetap, sedangkan untuk efek acak lebih baik menggunakan REML. Walaupun demikian, perbedaan hasil antara kedua metode relatif kecil, dan untuk sampel besar perbedaan hasil antara keduanya dapat diabaikan (Hox, 1995).

2.6 Pengujian Hipotesis

Hipotesis untuk parameter level 1:

$$\mathrm{H}_0 : \gamma_{qj} = 0 \text{ vs } H_1 : \gamma_{qj} \neq 0$$

Hipotesis untuk parameter level 2:

$$\mathrm{H}_0 : \beta_{pj} = 0 \text{ vs } \mathrm{H}_1 : \beta_{pj} \neq 0$$

Menurut Jones dan Steenbergen (1997) pengujian hipotesis tersebut dilakukan dengan menggunakan uji statistik Wald dengan persamaan sebagai berikut:

$$t = \frac{\text{penduga}}{\text{galat baku penduga}} \tag{2.19}$$

$$\text{Level 1} : t = \frac{\hat{\gamma}_{qj}}{SE\ (\gamma_{qj})}$$

$$\text{Level 2} : t = \frac{\hat{\beta}_{pj}}{SE\ (\beta_{pj})}$$

dengan t mengikuti sebaran t *student*, db untuk penduga parameter level 1 = j-s-1 dan db untuk penduga parameter level 2 = i-r-1.

2.7 Pemilihan Model Terbaik

Dalam regresi multilevel, pemilihan model terbaik dilakukan dengan melihat nilai deviasi. Deviasi merupakan suatu ukuran yang dapat digunakan untuk menentukan kesesuaian suatu model. Menurut Tantular (2009), untuk memilih model regresi multilevel yang terbaik dapat menggunakan uji rasio *likelihood* atau disebut juga sebaran deviasi yaitu ukuran untuk menentukan cocok tidaknya suatu model. Perhitungan untuk pengujian ini adalah selisih nilai deviasi antara dua model atau *diff* yaitu sebagai berikut:

$$Diff = -2log_e\left(\frac{\lambda_0}{\lambda_1}\right) \qquad (2.20)$$

dengan

λ_0 = nilai deviasi untuk *null model*

λ_1 = nilai deviasi untuk *full model*

Menurut Hox (2010) , semakin kecil nilai deviasi pada model tersebut maka model tersebut dikatakan semakin cocok dan apabila nilai *diff* > $\chi^2_{(\alpha,db)}$ maka tolak H_0, dimana db merupakan selisih jumlah parameter dari kedua model. Sehingga dapat disimpulkan bahwa efek acak signifikan, artinya terdapat keragaman atau variasi variabel dependen yang signifikan antarkelompok.

2.8 Koefisien Korelasi Intraklas

Model regresi multilevel diasumsikan saling tidak bebas antar observasi. Menurut West dkk (2007) ketidakbebasan antar observasi diukur dengan nilai koefisien korelasi intraklas. Koefisien korelasi intraklas atau *Intraclass Correlation Coefficients* (ICC) dapat diformulasikan sebagai berikut :

$$\rho = \frac{\sigma^2_{u0j}}{\sigma^2_{u0j}+\sigma^2_{\varepsilon ij}} \qquad 0 \leq \rho \leq 1 \qquad (2.21)$$

dengan

σ^2_{u0j} = ragam galat pada level 1

$\sigma^2_{\varepsilon ij}$ = ragam galat pada level 2

2.9 Keragaman Model

Keragaman variabel dependen yang dapat dijelaskan oleh variabel independen dalam model disebut koefisien determinasi. Koefisien determinasi juga dapat diperoleh dalam model multilevel, meskipun dalam model multilevel akan didapatkan koefisien determinasi lebih dari satu (Tantular, 2009).

Koefisien determinasi pertama pada level 1 bertujuan untuk menilai rasio ragam galat terhadap ragam total dirumuskan sebagai berikut:

$$R_1^2 = \frac{\sigma^2_{u0} - \sigma^2_{u0q}}{\sigma^2_{u0}} \tag{2.22}$$

dengan

σ^2_{u0q} = ragam galat level 1 dengan q variabel independen

σ^2_{u0} = penduga ragam galat level 1 tanpa variabel independen

Koefisien determinasi peda level 2 dirumuskan sebagai berikut:

$$R_2^2 = \frac{\sigma^2_{e0} - \sigma^2_{ep}}{\sigma^2_{e0}} \tag{2.23}$$

dengan

σ^2_{ep} = ragam galat level 2 dengan p variabel independen

σ^2_{e0} = ragam galat level 2 tanpa variabel independen

Koefisien pada persamaan 2.22 dan persamaan 2.23 juga merupakan persentase keragaman yang dapat dijelaskan pada setiap level terhadap variabel dependen (Hox, 2010).

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun ajaran 2017/2018 bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder hasil survei Badan Pusat Statistik Provinsi Lampung tahun 2016. Variabel dependen yaitu berupa kepadatan penduduk per kecamatan di Provinsi Lampung. Sedangkan variabel independen yang terdapat pada setiap levelnya meliputi:

Level 1 (kabupaten):

1. Indeks pembangunan manusia (Z_1)
2. Laju pertumbuhan ekonomi (Z_2)

Level 2 (kecamatan):

1. Rasio jenis kelamin (X_1)
2. Pertambahan penduduk per kecamatan (X_2)
3. Pendapatan asli kecamatan (X_3)

3.3 Metode Penelitian

Penelitian ini dilaksanakan dengan studi pustaka yaitu dengan pengkajian secara teoritis dan praktik komputasi. Adapun langkah-langkah yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Menginput data hasil survei BPS tahun 2016 pada *Microsoft Excel* dan mengelompokkannya berdasarkan level kabupaten dan level kecamatan.
2. Melakukan analisis deskriptif pada masing-masing variabel pada level kabupaten dan level kecamatan.
3. Melakukan uji asumsi pada data.
   a. Uji normalitas menggunakan *Normal Probability Plot* dan Uji Kolmogorov-Smirnov.
   b. Uji multikolinearitas dengan melihat nilai *Variance Inflation Factor* (VIF).
4. Memodelkan menggunakan analisis regresi multilevel untuk mengetahui faktor-faktor yang mempengaruhi kepadatan penduduk di provinsi Lampung tahun 2016 pada level kabupaten dan level kecamatan dengan metode *Maximum Likelihood*.
5. Membandingkan model-model regresi multilevel menggunakan nilai deviasi dan *diff* untuk mendapatkan model terbaik.
6. Menghitung nilai koefisien korelasi intraklass.
7. Menguraikan keragaman yang dapat dijelaskan oleh level kabupaten dan level kecamatan terhadap kepadatan penduduk di provinsi Lampung tahun 2016.

DAFTAR PUSTAKA

Badan Pusat Statistik. 2017. Provinsi Lampung dalam Angka 2017. CV. Jaya Wijaya, Bandar Lampung.

Bryck, A.S. and Raudenbush, S.W. 1987. Applying the Hierarchical Linear Models to Measurement of Change Problems. *Psychological Bulletin*. **101**: 147-158.

Goldstein, H. 1999. *Multilevel Statistical Model*. Ed. ke-2. E-Book of Arnold, London.

Harlan, J. 2016. *Analisis Multilevel*. Gunadarma. Depok.

Hox, J.J. 1995. *Applied Multilevel Analysis*. TT-Publikaties, Amsterdam.

Hox, J.J. 2010. *Multilevel Analysis: Techniques and Applications*. Ed ke-2. Routledge, New York.

Jones, B.S. and Steenbergen, M.R. 1997. Modeling Multilevel Data Structures. *American Journal of Political Science*. **46**(1): 218-237.

Kurniawan, D. 2008. *Regresi Linear*. R Development Core Team, Australia.

Kutner, M.H., Nachtsheim, C.J., and Neter, J. 2004. *Applied Linear Regression Models*. Ed ke-4. McGraw-Hill Companies, New York.

Myers, R.H. 1990. *Classical and Modern Regression with Application*. Ed ke-2. PWS Kent, Boston.

Rencher, A.C. and Schaalje, G.B. 2007. *Linear Model in Statistics*. Ed ke-2. John Wiley & Sons, Inc., New Jersey.

Tantular, B. 2009. Penerapan Model Regresi Linier Multilevel Pada Data Pendidikan dan Data Nilai Ujian. Tesis. Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Pertanian Bogor, Bogor.

West, B.T., Welch, K.B. and Galechi, A.T. 2007. *Linear Mixed Models: A Practical Guide Using Stastical Software*. Chapman & Hall, Boca Raton.

Zulmi, R. 2011. Pengaruh Luas Lahan, Tenaga Kerja, Penggunaan Benih, dan Pupuk Terhadap Produksi Padi di Jawa Tengah Tahun 1994-2008. Skripsi. Jurusan Ilmu Ekonomi dan Studi Pembangunan Fakultas Ekonomi Universitas Diponegoro, Semarang.