ANALISIS REGRESI MULTILEVEL DENGAN METODE *MAXIMUM LIKELIHOOD*
(Studi Kasus: Kepadatan Penduduk di Provinsi Sumatera Selatan Tahun 2020)

(Skripsi)

Oleh

SITI BUNGA ROHIYATUN NUFUS

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022

ABSTRACT

MULTILEVEL REGRESSION ANALYSIS WITH MAXIMUM LIKELIHOOD METHOD
(Case Study: Population Density of Sumatera Selatan Province in 2020)

By

Siti Bunga Rohiyatun Nufus

Multilevel regression is a statistical technique for performing regression analysis on hierarchically structured data. The parameter estimation method that is generally applied for multilevel regression analysis is the maximum likelihood method. In this study, we will analyze multilevel regression using the maximum likelihood method to determine the factors that affect population density in Sumatera Selatan Province in 2020 at the sub-district and district levels, as well as determine the value of diversity described at each level. Based on the results of the analysis, it is known that the best model is a multilevel regression model with variable z, namely a model that includes variables at the district level. Factors that affect population density in South Sumatra Province are the human development index variable, gross regional domestic product per capita, the number of sub-district residents, and the sub-district sex ratio. The diversity of population density can be explained by a level 1 (sub-district) of 73.03%, while the value of diversity that can be explained by a level 2 (district) is 24.23%.

Keywords: Multilevel Regression, Population Density, Maximum Likelihood

ABSTRAK

ANALISIS REGRESI MULTILEVEL DENGAN METODE *MAXIMUM LIKELIHOOD* (Studi Kasus: Kepadatan Penduduk di Provinsi Sumatera Selatan Tahun 2020)

Oleh

Siti Bunga Rohiyatun Nufus

Regresi multilevel adalah teknik statistik untuk melakukan analisis regresi pada data berstruktur hierarki. Metode estimasi parameter yang umumnya diterapkan untuk analisis regresi multilevel adalah metode *maximum likelihood*. Pada penelitian ini akan menganalisis regresi multilevel melalui metode *maximum likelihood* untuk mengetahui faktor yang mempengaruhi kepadatan penduduk di Provinsi Sumatera Selatan tahun 2020 pada level kecamatan dan kabupaten, serta mengetahui nilai keragaman yang dijelaskan masing-masing level. Berdasarkan hasil analisis diketahui bahwa model terbaik adalah model regresi multilevel dengan variabel z yaitu model dengan mengikutsertakan variabel pada level kabupaten. Faktor yang mempengaruhi kepadatan penduduk di Provinsi Sumatera Selatan adalah variabel indeks pembangunan manusia, produk domestik regional bruto per kapita, jumlah penduduk kecamatan, dan rasio jenis kelamin kecamatan. Keragaman dari kepadatan penduduk dapat dijelaskan oleh level 1 (kecamatan) sebesar 73,03%, sementara itu nilai keragaman yang dapat dijelaskan oleh level 2 (kabupaten) sebesar 24,23%.

Kata kunci: Regresi Multilevel, Kepadatan Penduduk, *Maximum Likelihood*

ANALISIS REGRESI MULTILEVEL DENGAN METODE
MAXIMUM LIKELIHOOD
(Studi Kasus: Kepadatan Penduduk di Provinsi Sumatera Selatan Tahun 2020)

Oleh

Siti Bunga Rohiyatun Nufus

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022

Judul Skripsi	:	**ANALISIS REGRESI MULTILEVEL DENGAN METODE *MAXIMUM LIKELIHOOD* (Studi Kasus: Kepadatan Penduduk di Provinsi Sumatera Selatan Tahun 2020)**
Nama Mahasiswa	:	**Siti Bunga Rohiyatun Nufus**
Nomor Pokok Mahasiswa	:	**1857031006**
Jurusan	:	**Matematika**
Fakultas	:	**Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Prof. Drs. Mustofa, M.A., Ph.D.
NIP 19570101 198403 1 020

Agus Sutrisno, S.Si., M.Si.
NIP 19700831 199903 1 001

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

Ketua	: **Prof. Drs. Mustofa, M.A., Ph.D.**	....
Sekretaris	: **Agus Sutrisno, S.Si., M.Si.**	....
Penguji Bukan Pembimbing	: **Drs. Eri Setiawan, M.Si.**	....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T.
NIP 19740705 200003 1 001

Tanggal Lulus Ujian Skripsi : **25 Juli 2022**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama Mahasiswa : **Siti Bunga Rohiyatun Nufus**

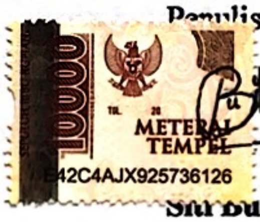
Nomor Pokok Mahasiswa : **1857031006**

Jurusan : **Matematika**

Judul Skripsi : **ANALISIS REGRESI MULTILEVEL DENGAN METODE *MAXIMUM LIKELIHOOD* (Studi Kasus: Kepadatan Penduduk di Provinsi Sumatera Selatan Tahun 2020)**

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah karya penulisan ilmiah Universitas Lampung. Apabila kemudian hari terbukti bahwa skripsi ini merupakan salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 29 Juli 2022
Penulis,



Siti Bunga Rohiyatun Nufus
NPM. 1857031006

RIWAYAT HIDUP

Penulis bernama lengkap Siti Bunga Rohiyatun Nufus, lahir di Tangerang, Provinsi Banten pada tanggal 13 November 1999. Penulis merupakan anak ketiga dari tiga bersaudara yang dilahirkan oleh pasangan Bapak Saefudin dan Ibu Komariyah.

Penulis menempuh pendidikan di Taman Kanak-kanak (TK) Inayatul Ihsan pada tahun 2004-2006. Kemudian menempuh Sekolah Dasar (SD) di SDN Buaran Asem pada tahun 2006-2012. Kemudian melanjutkan Sekolah Menengah Pertama (SMP) di MTsN 1 Tangerang pada tahun 2012-2015. Selanjutnya, melanjutkan ke jenjang sekolah Sekolah Menengah Atas (SMA) di MAN 2 Kota Serang pada tahun 2015-2018. Pada tahun 2018 penulis diterima sebagai mahasiswa S1 Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA), Universitas Lampung melalui jalur SMMPTN BARAT.

Selama menjadi mahasiswa di Universitas Lampung penulis aktif dalam berorganisasi di Himpunan Mahasiswa Jurusan Matematika (HIMATIKA) dan penulis juga aktif berorganisasi di Himpunan Mahasiswa Banten (HMB). Pada bulan Februari – Maret 2021 penulis melaksanakan Kuliah Kerja Nyata (KKN) selama 40 hari di Desa Cikoneng, Kecamatan Anyer, Kabupaten Serang, Provinsi Banten. Pada bulan Maret – Juni 2021 penulis mengikuti kegiatan Kampus Mengajar di SD Bina Mandiri Rajeg. Pada bulan Juli – Agustus 2021 penulis melaksanakan Kerja Praktik (KP) di Badan Pusat Statistik (BPS) Provinsi Banten. Pada bulan Februari – Juli 2022 penulis mengikuti kegiatan Studi Independen Bersertifikat di Zenius dengan program *Accelerated Machine Learning*.

KATA INSPIRASI

"Jika kamu berbuat baik (berarti) kamu berbuat baik untuk dirimu sendiri. Dan jika kamu berbuat jahat, maka (kerugian kejahatan) itu untuk dirimu sendiri".
(Q.S. Al-Israa: 7)

"Sesungguhnya sesudah kesulitan itu ada kemudahan. Maka apabila kamu telah selesai (dari sesuatu urusan), kerjakanlah dengan sungguh-sungguh (urusan) yang lain, dan hanya kepada Tuhanmulah hendaknya kamu berharap."
(Q.S. Al-Insyirah: 6-8)

"Belajarlah kalian ilmu untuk ketentraman dan ketenangan serta rendah hatilah pada orang yang kamu belajar darinya."
(HR. Thabrani)

"If you want to be successful, you must respect one rule – Never lie to yourself."
(Paulo Coelho)

PERSEMBAHAN

Alhamdulillahirobbil'alamin, puji syukur atas kehadirat Allah SWT yang Maha Pengasih lagi Maha Penyayang, untuk segala rahmat dan hidayah-Nya yang telah memberikan petunjuk sehingga skripsi ini dapat diselesaikan, dan selawat serta salam kepada Nabi Muahammad SAW. Saya persembahkan karya sederhana ini untuk:

Bapak Saefudin dan Umi Komariyah

Terima kasih kepada Bapak dan Umi yang selalu memberikan doa, kasih sayang, dukungan moral dan materil, motivasi, dan saran dalam setiap keputusan untuk keberhasilan dan sesuksesan penulis. Terima kasih telah memberikan arti hidup yang sangat berharga dan mengajarkan untuk selalu berbuat baik.

Kakak Fudholi dan Kakak Fikri

Terima kasih telah memberikan doa, semangat dan dukungan selama ini.

Dosen Pembimbing dan Pembahas

Terimakasih kepada dosen pembimbing dan pembahas yang sudah membantu, memberikan arahan dan ilmu yang sangat bermanfaat.

Terima kasih kepada semua orang-orang baik disekitarku yang telah memberikan doa, dukungan, keceriaan, semangat dan suka maupun duka yang telah menyertai dalam langkahku.

Almamater Tercinta, Universitas Lampung

SANWACANA

Alhamdulillah puji dan syukur kepada Allah SWT karena berkat segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Analisis Regresi Multilevel dengan Metode *Maximum Likelihood* (Studi Kasus: Kepadatan Penduduk di Provinsi Sumatera Selatan Tahun 2020)”.

Dalam proses penyusunan skripsi ini,tidak akan terselesaikan tanpa adanya pihak yang telah membantu memberikan bimbingan, saran, dan motivasi serta semangat sehingga skripsi ini dapat terselesaikan dengan baik. Untuk itu, penulis mengucapkan terimakasih kepada:

1. Bapak Prof.Drs. Mustofa Usman, M.A., Ph.D., selaku dosen pembimbing I yang telah bersedia membimbing, memberi saran dan arahan, serta memberikan waktu, motivasi dalam menyelesaikan skripsi ini.
2. Bapak Agus Sutrisno, S.Si., M.Si., selaku dosen pembimbing II yang telah memberikan bimbingan, saran dan pengarahan serta motivasi kepada penulis.
3. Bapak Drs. Eri Setiawan, M.Si., selaku dosen penguji yang telah memberikan kritik dan saran yang membangun serta evaluasi selama proses penyusunan skripsi hingga dapat menjadi lebih baik.
4. Bapak Prof.Drs. Mustofa Usman, M.A., Ph.D., selaku dosen pembimbing akademik yang telah memberi bimbingan dan arahan selama masa perkuliahan.
5. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Bapak Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

7. Seluruh dosen, staff, karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
8. Bapak dan Umi, Kakak Fudholi dan Kakak Fikri, serta seluruh keluarga besar Kakek Alm. H. Main dan keluarga besar Kakek Alm. Siam yang selalu memberikan doa, dukungan, serta memberikan semangat dan motivasi yang begitu besar kepada penulis.
9. Teman-temanku yang tersayang Nanda, Alifiah, Dila, Mega, Martha, Dalfa, Vinny, Firda, Ulil, Alfi dan Nanda Rahma yang telah menemani, memberi dukungan, memberikan kebahagiaan serta keceriaan kepada penulis.
10. Teman-teman satu bimbingan Regita, Aulia, Lauren, Risha, Muflihah, Marisa, Nurlela, dan Rika yang telah memberikan dukungan, serta bantuan kepada penulis selama menyusun skripsi ini.
11. Teman-teman Matematika 2018 dan teman kelas C yang telah bersama selama masa perkuliahan.
12. Semua pihak yang telah membantu dalam menyelesaikan skripsi ini.

Akhir kata, semoga Allah SWT senantiasa melimpahkan karunia-Nya dan membalas segala kebaikan pihak-pihak yang telah membantu penulis dalam menyusun skripsi ini. Penulis menyadari masih bahwa dalam skripsi ini masih banyak kekurangan. Oleh karena itu, kritik dan saran sangat penulis harapkan untuk menjadi bahan perbaikan kedepannya. Semoga skripsi ini dapat memberikan manfaat bagi pihak yang membutuhkan.

Bandar Lampung, Juli 2022
Penulis,

Siti Bunga Rohiyatun Nufus
NPM. 1857031006

DAFTAR ISI

Halaman

DAFTAR GAMBAR

DAFTAR TABEL

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Regresi multilevel yakni sebuah model digunakan dalam suatu data berjenjang (hierarki) (Yulian dan Pawitan, 2017). Model multilevel adalah metode statistik yang dikembangkan dari regresi sederhana. Pengembangan ini didasarkan pada fakta bahwa penelitian seringkali mengungkapkan perbedaan sumber responden yang diteliti dalam kasus yang berbeda, dan bahwa data bersifat hierarkis (struktur berjenjang) dan *cluster* (berkelompok). Goldstein (1995) mengenalkan pengembangan dari regresi reguler dalam menangani masalah yang ditimbulkan oleh data hierarkis, yakni analisis pemodelan berlevel (*Multilevel Modelling)*.

Di era modern saat ini seringkali ditemukan data berjenjang dalam penelitian survei yang satuan pengamatannya bersumber dari kelompok-kelompok. Misalnya dalam suatu data obyek yang diamati tergabung dalam kelompok dan variabel-variabelnya didefinisikan pada level berbeda. Level dalam model multilevel merupakan sebutan dari struktur data hierarki. Level terendah disebut level 1 dan level tertinggi disebut level 2 (Tantular, Aunuddin, dan Wijayanto, 2009). Metode multilevel ini biasa digunakan dalam kasus-kasus pendidikan, sosial dan kesehatan, terutama dengan data kependudukan yang terstruktur secara hierarkis.. Dalam penelitian ini, fokus penelitian adalah pada kasus-kasus sosial yang berkaitan dengan kepadatan penduduk. Kepadatan penduduk ialah rasio penduduk per kilometer persegi. Di negara berkembang yang memiliki jumlah penduduk terbanyak seperti negara Indonesia sering terjadi masalah kepadatan penduduk yang tidak merata diberbagai daerah di Indonesia termasuk di Provinsi Sumatera Selatan. Tahun 2020, kepadatan penduduk di Provinsi Sumatera Selatan mencapai 92,45 jiwa/km^2, kota Palembang sebesar 4.519,93 jiwa/km^2 menjadi kepadatan penduduk tertinggi dan 31,43 jiwa/km^2 di kabupaten Musi Rawas Utara menjadi kepadatan terendah (Badan Pusat Statistik, 2021). Kepadatan penduduk yang tidak merata dapat menimbulkan ketimpangan perekonomian dalam masyarakat. Setiap daerah memiliki permasalannya masing-masing yang berpengaruh terhadap kepadatan penduduk, sedangkan upaya pemerintah dalam mengatasi hal ini hanya memberikan kebijakan secara global. Untuk mengatasi masalah tersebut, digunakan regresi multilevel karena variabel independen yang digunakan berstruktur hierarki di level kecamatan dan kabupaten dengan estimasi parameter dengan Metode Kemungkinan Maksimum (*Maximum Likelihood Methods*).

Penelitian analisis regresi multilevel dengan metode *maximum likelihood* pernah dilakukan oleh Tantular, Aunuddin dan Wijayanto pada tahun 2009 dalam pemilihan model regresi multilevel terbaik dengan variabel dependen pada penelitian ini adalah nilai ujian akhir semester pada mata kuliah Analisis Statistika mahasiswa Pascasarjana Institut Pertanian Bogor dan variabel independen pada level 1 adalah nilai ujian pertama mahasiswa, pada level 2 adalah nilai ujian pertama tingkat program studi, dan pada level 3 adalah nilai ujian pertama pada tingkat kelas.

Oleh sebab itu, penelitian ini menggunakan analisis regresi multilevel menggunakan metode *maximum likelihood* untuk menjelaskan faktor-faktor yang akan mempengaruhi kepadatan penduduk di Provinsi Sumatera Selatan pada tahun 2020.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Menentukan model regresi multilevel terhadap faktor-faktor yang akan mempengaruhi kepadatan penduduk di Provinsi Sumatera Selatan tahun 2020.

2. Menguraikan keragaman pada tingkat kecamatan dan kabupaten untuk mengetahui besarnya pengaruh faktor-faktor pada tiap level terhadap kepadatan penduduk di Provinsi Sumatera Selatan tahun 2020.

1.3 Manfaat penelitian

Penelitian diharapkan memberikan manfaat:

1. Untuk mengetahui model regresi multilevel terhadap faktor-faktor yang mempengaruhi kepadatan penduduk di Provinsi Sumatera Selatan tahun 2020.
2. Untuk mengetahui besarnya pengaruh faktor-faktor pada tiap level yang dijelaskan oleh keragaman pada tingkat kecamatan dan kabupaten terhadap kepadatan penduduk di provinsi Sumatera Selatan tahun 2020.

II. TINJAUAN PUSTAKA

2.1 Data Hierarki

Data hierarki merupakan data yang terdiri dari item-item pengamatan yang disarangkan atau dikelompokkan ke dalam item-item yang tingkatnya lebih tinggi. Data hierarki juga dikenal sebagai data multilevel atau bersarang (Tantular, 2009). Suatu studi menemukan data populasi dengan data hierarkis atau berjenjang. Data terstruktur secara hierarki adalah data yang muncul karena tingkat yang lebih rendah bersarang di dalam tingkat yang lebih tinggi, dikumpulkan dalam satu set populasi hierarkis di mana individu dapat menentukan variabel dari setiap tingkat.

Hox (1995), menganalisis data hierarki menggunakan metode regresi linier berganda menghadirkan beberapa permasalahan, antara lain:

1. Apabila analisis dilakukan di tingkat tertinggi, tingkat informasi terendah akan hilang. Alhasil akan mengurangi kemampuan pengujian statistik di tingkat ini, sebab kehilangan banyak informasi pada tingkat terendah.
2. Apabila analisis dilakukan di tingkat terendah, pengelompokan data akan terabaikan. Model regresi dibangun dari semua observasi minimal tingkat. Masalah yang muncul ialah multikolinearitas, sehingga model yang dihasilkan tidak sesuai.

2.2 Analisis Regresi Linear

Menurut Sembiring (1995), analisis regresi ialah metode statistik untuk membangun model untuk menentukan hubungan kausal antara dua variabel atau lebih. Model ini adalah fungsi dari variabel-variabel tersebut dan digunakan untuk memahami, menjelaskan, dan memprediksi perilaku sistem yang diamati. Model regresi memiliki tiga tujuan. Artinya, tujuan pertama menjelaskan pola hubungan sebab-akibat yang terdapat antara variabel dependen dan independen, tujuan kedua menjelaskan kontribusi masing-masing variabel independen terhadap dependen. Tujuan ketiga adalah memprediksi nilai variabel dependen untuk nilai tertentu dari variabel independen tertentu (Gujarati, 2006).

Analisis regresi yang hanya berisi satu variabel independen (bebas) dan satu variabel dependen (terikat) disebut analisis regresi linier sederhana. Persamaan umum untuk model regresi linier sederhana adalah:

$$Y = \alpha + \beta X + \varepsilon \tag{2.1}$$

dengan,

Y = variabel terikat (dependen)

X = variabel bebas (independen)

α = *intercept* (kostanta regresi)

β = *slop* (koefisien regresi)

ε = galat

Analisis regresi di sisi lain, yang melibatkan beberapa variabel independen seta satu variabel dependen disebut analisis regresi linier berganda. Rumus umum untuk regresi berganda adalah:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_n X_{ni} + \varepsilon_i,\ i = \text{indeks } (i = 1, 2, \ldots, \text{n}) \tag{2.2}$$

dengan,

Y = variabel terikat (dependen)

X = variabel bebas (independen)

α = *intercept* (kostanta regresi)

β = *slop* (koefisien regresi)

ε = galat

2.3 Model Linear Campuran

Model linier campuran (*linear mixed models*) adalah model dalam suatu persamaan terdiri dari faktor tetap (*fixed effect*) dan faktor acak (*random effect*). Variabel bebas berpengaruh tetap apabila koefisien regresinya sama untuk semua anggota sampel, dan variabel bebas berpengaruh acak apabila ada perbedaan pada nilai koefisien regresinya antara kelompok dua ataupun lebih anggota sampel (Harlan, 2016). Menurut Bryck dan Raudenbush (1987), model linier campuran umumnya mencakup tiga hal:

1. Efek acak, disebabkan oleh pengaruh variabel yang nilainya diperoleh dari sampel acak.
2. Efek hierarki, disebabkan oleh pengaruh variabel yang diukur di tingkat berbeda.
3. Pengukuran berulang, mengacu pada pengamatan sebelumnya.

Rencher dan Schaalje (2007), menyatakan persamaan untuk model linier campuran memiliki bentuk sederhana sebagai berikut:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z_1 u_1} + \boldsymbol{Z_2 u_2} + \cdots + \boldsymbol{Z_i u_i} + \boldsymbol{\varepsilon} \qquad (2.3)$$

dengan,

$\boldsymbol{Y}$ = vektor respon $(n \times 1)$

$\boldsymbol{X}$ = matriks prediktor efek tetap $(n \times p)$

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

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

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

$\boldsymbol{\varepsilon}$ = vektor galat $(r_i \times 1)$

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

$E(\boldsymbol{u_i}) = 0$ dan $Cov(\boldsymbol{u_i}) = \sigma^2 \boldsymbol{I_{ri}}$ untuk $i = 1,2,\dots,m$

$Cov\left(\boldsymbol{u_i}, \boldsymbol{u_j}\right) = O$ untuk $i \neq j$, dimana O adalah $r_i \; x \; r_j$

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

Maka,

$$\begin{aligned} E(\boldsymbol{Y}) &= E(\boldsymbol{X\beta} + \textstyle\sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \boldsymbol{Z_i u_i} + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{X\beta} + E(\textstyle\sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \boldsymbol{Z_i u_i} + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{X\beta} + \textstyle\sum_{i=1}^{m} \boldsymbol{Z_i} E(\boldsymbol{u_i}) + E(\boldsymbol{\varepsilon}) \\ &= \boldsymbol{X\beta}. \end{aligned}$$

$$\begin{aligned} Cov(\boldsymbol{Y}) &= Cov(\boldsymbol{X\beta} + \textstyle\sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \boldsymbol{Z_i u_i} + \boldsymbol{\varepsilon}) \\ &= Cov(\textstyle\sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \boldsymbol{Z_i u_i} + \boldsymbol{\varepsilon}) \\ &= \textstyle\sum_{i=1}^{m} Cov\,(\boldsymbol{Z_i u_i}) + Cov(\boldsymbol{\varepsilon}) + \sum_{i \neq j} Cov\left(\boldsymbol{Z_i u_i}, \boldsymbol{Z_j u_j}\right) + \\ &\quad \textstyle\sum_{i=1}^{m} Cov\,(\boldsymbol{Z_i u_i}, \boldsymbol{\varepsilon}) + \sum_{i=1}^{m} Cov\,(\boldsymbol{\varepsilon}, \boldsymbol{Z_i u_i}) \\ &= \textstyle\sum_{i=1}^{m} \boldsymbol{Z_i} Cov\,(\boldsymbol{u_i}) \boldsymbol{Z_i'} + Cov(\boldsymbol{\varepsilon}) + \sum_{i \neq j} \boldsymbol{Z_i} Cov\left(\boldsymbol{u_i}, \boldsymbol{u_j}\right) \boldsymbol{Z_j'} + \\ &\quad \textstyle\sum_{i=1}^{m} Z_i Cov\,(\boldsymbol{u_i}, \boldsymbol{\varepsilon}) + \sum_{i=1}^{m} Cov\,(\boldsymbol{\varepsilon}, \boldsymbol{u_i}) \boldsymbol{Z_i'} \\ &= \textstyle\sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \boldsymbol{\sigma_i^2 Z_i Z_i'} + \boldsymbol{\sigma^2 I_n}. \end{aligned}$$

2.4 Analisis Regresi Multilevel

Analisis regresi multilevel adalah metode statistik yang berguna dalam memperkirakan hubungan antar variabel prediktor dan respons di model regresi dengan menggunakan setiap kumpulan data dalam suatu kelompok. Analisis regresi multilevel ialah metode statistik yang dikembangkan dari regresi klasik. Data dalam analisis regresi multilevel adalah data hierarkis. Artinya, data yang terdiri dari item-item observasi yang bersarang atau dikelompokkan menjadi item-item yang tingkatnya lebih tinggi. Data hierarki juga dikenal sebagai data multi-level. Ringdal (1992) menyatakan bahwa meskipun data yang diperoleh berisi informasi ini, analisis awalnya digunakan terlepas dari informasi tentang milik individu terhadap lingkungan. Hal tersebut mengakibatkan tidak puas dengan hasil analisis karena tidak ada lagi kesimpulan konkret yang dapat ditarik untuk setiap tingkat.

Goldstein (1995) mengenalkan pengembangan regresi reguler dalam menangani masalah yang ditimbulkan oleh data yang terstruktur secara hierarkis: analisis *Multilevel Modelling.* Analisis regresi multilevel dicirikan oleh adanya tingkat data di mana data tingkat rendah hadir dalam data tingkat tinggi. Variabel respon diukur pada tingkat terendah dan prediktor dapat diukur pada setiap tingkat data dalam analisis regresi multilevel (Hox, 2010).

Menurut Hox (2010), model regresi multilevel dibagi menjadi dua bentuk dasar:

1. Model multilevel dasar *random intercept*

   Model ini mengasumsikan bahwa *intercept* dimodelkan sebagai efek acak dari variabel tingkat 2, dan bahwa *intercept* setiap kelompok berbeda tetapi gradiennya sama, sehingga efek masing-masing prediktor pada variabel respons berbeda.
2. Model multilevel dasar *random slope*

   Model ini ialah model dimana koefisien prediktor dimodelkan pada level yang lebih rendah sebagai efek acak dari variabel level 2, dan setiap kelompok diasumsikan memiliki gradien yang berbeda, dengan mempertimbangkan efek prediktor terhadap variabel respon bervariasi dari kelompok ke kelompok.

Model regresi multilevel yang paling sederhana yakni model dua tingkat, dimana tingkat kesatu ialah data individu dan tingkat kedua ialah data kelompok (West *et,al*., 2007). Representasi persamaan regresi multilevel dengan 2 level sebagai berikut:

a. Model level 1

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

$$Y_{ij} = \beta_{0j} + \sum_{P=1}^{P} \beta_{pj}X_{pij} + \varepsilon_{ij} \tag{2.4}$$

dengan,

Y_{ij} = variabel respon ke-i dalam kelompok ke-j pada level 1

β_{0j} = *intercept* pada level 1

β_{pj} = koefisien regresi variabel bebas ke-p pada kelompok ke-j pada level 1

X_{pij} = variabel prediktor ke-p pada level 1

ε_{ij} = galat pada level 1, diasumsikan berdistribusi $N(0, \sigma_\varepsilon^2)$

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

j = indeks pada level 2 $(1,2, \dots, m)$

P = banyak variabel prediktor pada level 1 $(1,2, \dots, p)$

b. Model level 2

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

$$\beta_{0j} = \gamma_{00} + \sum_{Q=1}^{Q} \gamma_{0q} Z_{qj} + u_{0j} \tag{2.5}$$

dengan,

β_{0j} = *intercept* pada level 1

γ_{00} = *intercept* pada level 2

γ_{0q} = koefisien regresi untuk variabel bebas ke-q pada level 2

Z_{qj} = variabel bebas ke-q untuk kelompok ke-j pada level 2

u_{0j} = galat pada level 2, diasumsikan berdistribusi $N(0, \sigma_\varepsilon^2)$

j = indeks pada level 2 $(1,2, \dots, m)$

Q = banyak variabel prediktor pada level 2 $(1,2, \dots, q)$

Model koefisien acak digunakan untuk mengetahui ada tidaknya perbedaan *slope* pada tiap level, sehingga model ini digunakan untuk mengetahui selisih kemiringan antar level. Berdasarkan Persamaan (2.5), diperoleh model koefisien acak untuk level 2 adalah:

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_{1j} + \cdots + \gamma_{1q} Z_{qj} + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21} Z_{2j} + \cdots + \gamma_{2q} Z_{qj} + u_{2j}$$

$$\vdots$$

$$\beta_{pj} = \gamma_{p0} + \gamma_{p1} Z_{pj} + \cdots + \gamma_{pq} Z_{qj} + u_{pj}$$

$$\beta_{pj} = \gamma_{p0} + \sum_{Q=1}^{Q} \gamma_{pq} Z_{qj} + u_{pj} \tag{2.6}$$

dengan,

β_{pj} = koefisien regresi variabel bebas ke-p dalam kelompok ke-j pada level 1

γ_{p0} = *intercept* pada level 2

γ_{pq} = koefisien regresi ke-p untuk variabel bebas ke-q pada level 2

Z_{qj} = variabel bebas ke-q dalam kelompok ke-j pada level 2

u_{pj} = galat pada level 2, diasumsikan berdistribusi $N(0, \sigma_\varepsilon^2)$

j = indeks pada level 2 $(1,2, \dots, m)$

P = banyak variabel prediktor pada level 1 $(1,2, \dots, p)$

Q = banyak variabel prediktor pada level 2 $(1,2, \dots, q)$

Model campuran atau disebut model regresi untuk kedua level didapat dengan cara mendistribusikan pada Persamaan (2.5) dan (2.6) ke (2.4). Oleh karena itu, didapatkan model regresi multilevel adalah sebagai berikut:

$$
\begin{aligned}
Y_{ij} &= \beta_{0j} + \textstyle\sum_{P=1}^{P} \beta_{pj} X_{pij} + \varepsilon_{ij} \\
&= \gamma_{00} + \textstyle\sum_{Q=1}^{Q} \gamma_{0q} Z_{qj} + u_{0j} + \sum_{P=1}^{P} \left(\gamma_{p0} + \sum_{Q=1}^{Q} \gamma_{pq} Z_{qj} + u_{pj} \right) X_{pij} + \varepsilon_{ij} \\
&= \gamma_{00} + \textstyle\sum_{Q=1}^{Q} \gamma_{0q} Z_{qj} + \sum_{P=1}^{P} \gamma_{p0} X_{pij} + \sum_{P=1}^{P} \sum_{Q=1}^{Q} \gamma_{pq} Z_{qj} X_{pij} + \\
&\quad \textstyle\sum_{P=1}^{P} u_{pj} X_{pij} + u_{0j} + \varepsilon_{ij} \qquad (2.7)
\end{aligned}
$$

dengan,

i = indeks pada level 1 yang berada dalam indeks level 2

j = indeks pada level 2 $(1,2, \dots, m)$

P = banyak variabel prediktor pada level 1 $(1,2, \dots, p)$

Q = banyak variabel prediktor pada level 2 $(1,2, \dots, q)$

2.5 Uji Asumsi

Uji asumsi klasik regresi linear merupakan sebutan dari beberapa asumsi yang harus terpenuhi di dalam menggunakan analisis regresi linear. Dalam melakukan pendugaan parameter tentu saja tidak lepas dari kesalahan, baik kecil maupun besar.

Ketika beberapa asumsi terpenuhi maka kesalahan pendugaan merupakan yang terkecil (Kurniawan, 2008).

2.4.1 Uji Normalitas

Uji normalitas digunakan untuk mengetahui apakah data berdistribusi normal, dapat menggunakan uji *Kolmogorov-Smirnov* untuk menghitung uji normalitas. Uji *Kolmogorov-Smirnov* dapat membandingkan dua distribusi data, yakni distribusi hipotesis dan teramati. Jika distribusi yang diamati ada kesamaan dengan distribusi hipotetis, kita dapat menyimpulkan bahwa data yang teramati berdistribusi secara normal (Kurniawan, 2008). Selain *uji Kolmogorov-Smirnov*, dapat melakukan dengan memeriksa plot probabilitas normal. Jika plot probabilitas normal ditunjukan adanya titik-titik yang menyebar di sekitar diagonal serta distribusi mengikuti arah diagonal, maka model garis regresi data berdistribusi normal.

2.4.2 Uji Multikolinearitas

Uji multikolinearitas berguna menguji korelasi antar variabel prediktor pada model regresi (Gujarati, 2004). Uji statistik yang digunakan deteksi multikolinearitas yakni dengan nilai *Vaiance Inflation Factor* (VIF). Asumsi non-multikolinearitas tidak terpenuhi jika nilai VIF lebih dari 10. Untuk mendapatkan nilai *Vaiance Inflation Factor* (VIF) digunakan rumus sebagai berikut:

$$VIF = \frac{1}{1-{R_j}^2} \qquad (2.8)$$

dengan,

${R_j}^2$ = nilai koefisien determinasi X_j sebagai respon dengan $(p-1)$ prediktor lain.

2.6 Metode Penduga Parameter Regresi Multilevel

Pada dasarnya, pendugaan parameter dengan regresi multilevel ini memiliki kesamaan manfaat dengan regresi linier biasa yakni mencari nilai parameter yang dipakai pada proses regresi. Metode estimasi parameter yang paling umum digunakan dalam analisis regresi multilevel adalah metode *Maximum Likelihood Estimation* (MLE). *Maximum Likelihood Estimation* memperkirakan parameter dengan memaksimalkan fungsi *likelihood* (Hox, 2010). Berikut merupakan metode *Maximum Likelihood* untuk estimasi parameter.

Persamaan model linier campuran yaitu:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon} \qquad (2.9)$$

dimana $E(\boldsymbol{\varepsilon}) = 0, cov(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{I_n} = \boldsymbol{\Sigma_i}$ dan $E(u_i) = 0, cov(u_i) = \sigma^2\boldsymbol{I_{ri}} = \boldsymbol{D}$

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

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{D} & 0 & \cdots & 0 \\ 0 & \boldsymbol{D} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{D} \end{bmatrix} \text{ dan } \boldsymbol{R} = \begin{bmatrix} \boldsymbol{\Sigma_1} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Sigma_2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\Sigma_n} \end{bmatrix}$$

dengan,

$\boldsymbol{Y}$ = vektor respon $(n \times 1)$

$\boldsymbol{X}$ = matriks desain efek tetap $(n \times p)$

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

$\boldsymbol{Z}$ = matriks desain efek acak $(n \times q)$

$\boldsymbol{u}$ = vektor parameter efek acak $(q \times 1)$

$\boldsymbol{\varepsilon}$ = vektor galat $(n \times 1)$

$\boldsymbol{G}$ = matriks kovarians dari efek acak u

$\boldsymbol{R}$ = matriks kovarians dari vektor galat ε

Misalkan:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}^* \tag{2.10}$$

dengan,

$$\begin{aligned}\boldsymbol{\varepsilon}^* &= \boldsymbol{Zu} + \boldsymbol{\varepsilon} \\ &= (\boldsymbol{Z} \quad \boldsymbol{I}_{n\times n})\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{\varepsilon}\end{pmatrix} \\ &= \boldsymbol{A}\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{\varepsilon}\end{pmatrix}\end{aligned}$$

$$\boldsymbol{\varepsilon}^* \sim N_n(0, \boldsymbol{V})$$

dimana:

$$\begin{aligned}\boldsymbol{V} &= \boldsymbol{A}\begin{pmatrix}\boldsymbol{G} & 0\\ 0 & \boldsymbol{R}\end{pmatrix}\boldsymbol{A}^{\boldsymbol{t}} \\ &= (\boldsymbol{Z} \quad \boldsymbol{I}_{n\times n})\begin{pmatrix}\boldsymbol{G} & 0\\ 0 & \boldsymbol{R}\end{pmatrix}\begin{pmatrix}\boldsymbol{Z}^{\boldsymbol{t}}\\ \boldsymbol{I}_{n\times n}\end{pmatrix} \\ &= (\boldsymbol{ZG} \quad \boldsymbol{R})\begin{pmatrix}\boldsymbol{Z}^{\boldsymbol{t}}\\ \boldsymbol{I}_{n\times n}\end{pmatrix} \\ &= \boldsymbol{ZGZ}^{\boldsymbol{t}} + \boldsymbol{R}\end{aligned}$$

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

Vektor parameter kovarian θ diduga $\hat{\theta}_{ML}$ dengan pemaksimalan kegunaan *likelihood*.

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}$$

dengan $\begin{bmatrix}u_i\\ \boldsymbol{\varepsilon}\end{bmatrix} \sim N_{mq+n}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\boldsymbol{G} & 0_{mq\times n}\\ 0_{n\times mq} & \boldsymbol{R}\end{bmatrix}\right)$

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

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

Fungsi *likelihood*:

$$\begin{aligned}L &= \prod_{i=1}^{n} f(y_i|\theta) \\ &= \prod_{i=1}^{n} f(y_i, \mu, \sigma^2) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2}\sum\left(\frac{y_i-\mu}{\sigma}\right)^2 \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^{2n}} \exp -\frac{1}{2}\sum\left(\frac{y_i-\mu}{\sigma}\right)^2\end{aligned}$$

$$\ln L = \ln\left[\frac{1}{(2\pi)^{\frac{n}{2}}\sigma^{2n}} \exp -\frac{1}{2}\sum\left(\frac{y_i-\mu}{\sigma}\right)^2\right]$$

$$= \ln \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^{2n}} + \ln\left(\exp - \frac{1}{2}\sum(\frac{y_i-\mu}{\sigma})^2\right)$$

$$= \ln \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^{2n}} - \frac{1}{2}\sum(\frac{y_i-\mu}{\sigma})^2$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}V(\theta)} \exp - \frac{1}{2}(\varepsilon - \mu_\varepsilon)^T V(\theta)^{-1}(\varepsilon - \mu)$$

dengan fungsi *likelihood* variabel acak galat (ε) diberikan oleh:

$$L = \prod_{i=1}^{n} f(\varepsilon, 0, V(\theta))$$

$$= \prod_{i=1}^{n} \frac{1}{(2\pi)^{\frac{n}{2}}V(\theta)} \exp - \frac{1}{2}(\varepsilon - \mu_\varepsilon)^T V(\theta)^{-1}(\varepsilon - \mu)$$

$$L = \prod_{i=1}^{n} \frac{1}{(2\pi)^{\frac{n}{2}}V(\theta)} \exp - \frac{1}{2}(Y - X\beta)^t V(\theta)^{-1}(Y - X\beta)$$

$$L(\beta, \sigma^2) = -\frac{1}{2}\{\ln|\boldsymbol{V}(\theta)| + (\boldsymbol{Y} - \boldsymbol{X\beta})^{\boldsymbol{t}}\boldsymbol{V}(\theta)^{-\boldsymbol{1}}(\boldsymbol{Y} - \boldsymbol{X\beta})\} \tag{2.11}$$

dengan memaksimumkan log *likelihood* untuk θ terhadap β maka didapatkan

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = (\boldsymbol{X^t V}(\theta)^{-\boldsymbol{1}}\boldsymbol{X})^{-\boldsymbol{1}}\boldsymbol{X^t V}(\boldsymbol{\theta})^{-\boldsymbol{1}}\boldsymbol{Y} \tag{2.12}$$

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

$$Cov(\boldsymbol{Y}, \boldsymbol{u}) = Cov(\boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}, \boldsymbol{u})$$

$$= Cov(\boldsymbol{X\beta}, \boldsymbol{u}) + Cov(\boldsymbol{Zu}, \boldsymbol{u}) + Cov(\boldsymbol{\varepsilon}, \boldsymbol{u})$$

$$= 0 + \boldsymbol{ZG} + 0$$

$$= \boldsymbol{ZG}$$

$$Cov(\boldsymbol{Y}, \boldsymbol{u}) = \boldsymbol{ZG} \tag{2.13}$$

$$E(\boldsymbol{u}|\boldsymbol{Y}) = E(\boldsymbol{u}) + cov(\boldsymbol{u}, \boldsymbol{Y})[cov(\boldsymbol{Y})]^{-1}[\boldsymbol{Y} - E(\boldsymbol{Y})]$$

$$= 0 + \boldsymbol{GZ^t V^{-1}}(\boldsymbol{Y} - \boldsymbol{X\beta})$$

$$= \boldsymbol{GZ^t V^{-1}}(\boldsymbol{Y} - \boldsymbol{X\beta})$$

$$E(\boldsymbol{u}|\boldsymbol{Y}) = \boldsymbol{GZ^t V^{-1}}(\boldsymbol{Y} - \boldsymbol{X\beta}) \tag{2.14}$$

West, et. al., (2007) menyatakan efek tetap β dan efek acak u diduga oleh

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^t}\widehat{\boldsymbol{V}}^{-\boldsymbol{1}}\boldsymbol{X})^{-\boldsymbol{1}}\boldsymbol{X^t}\widehat{\boldsymbol{V}}^{-\boldsymbol{1}}\boldsymbol{Y} \tag{2.15}$$

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

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

Metode *Maximum Likelihood* terdapat dua keunggulan dibandingkan dengan REML. Artinya, relatif sederhana dalam hal perhitungan, dan sementara umumnya digunakan untuk memperkirakan efek tetap, lebih tepat digunakan untuk efek acak untuk REML. Namun, perbedaan antara hasil kedua metode tersebut relatif kecil, jika sampelnya besar, perbedaan antara hasil kedua metode tersebut dapat terabaikan (Hox, 1995).

2.7 Pengujian Hipotesis

Pengujian hipotesis atau pengujian estimasi parameter parsial dilakukan pada level 1 (kecamatan) dan level 2 (kabupaten) dengan menggunakan uji statistik Wald berdasarkan hipotesis sebagai berikut (Jones and Steenbergen, 2002):

a. Parameter pada level 1

$H_0: \beta_{pj} = 0$

$H_1: \beta_{pj} \neq 0$

Dengan p menyatakan banyaknya variabel prediktor untuk level 1. Statistik uji *Wald* adalah sebagai berikut:

$$t = \frac{\hat{\beta}_{pj}}{SE(\hat{\beta}_{pj})} \tag{2.17}$$

dimana:

$\hat{\beta}_{pj}$ = penduga variabel bebas ke-p dalam kelompok ke-j pada level 1

$SE(\hat{\beta}_{pj})$ = galat baku penduga variabel bebas ke-p dalam kelompok ke-j pada level 1.

b. Parameter pada level 2

$H_0: \gamma_{qj} = 0$

$H_1: \gamma_{qj} \neq 0$

Dengan q menyatakan banyaknya variabel prediktor untuk level 2. Statistik uji *Wald* adalah sebagai berikut:

$$t = \frac{\hat{\gamma}_{qj}}{SE(\hat{\gamma}_{qj})} \tag{2.18}$$

dimana:

$\hat{\gamma}_{qj}$ = penduga variabel bebas ke-q dalam kelompok ke-j pada level 2

$SE(\hat{\gamma}_{qj})$ = galat baku penduga variabel bebas ke-q dalam kelompok ke-j pada level 2.

Uji *Wald* menuruti distribusi *t-student*, maka derajat bebas (db) untuk level 1 adalah $i - p - 1$, sedangkan db untuk level 2 adalah $j - q - 1$.

2.8 Pemilihan Model Terbaik

Pemilihan model terbaik untuk regresi multilevel dengan memeriksa nilai deviasi. Deviasi ialah ukuran yang dipakai dalam menetapkan kesesuaian suatu model. Tantular (2009), menyatakan dalam memilih model regresi multi-level terbaik, bisa memakai uji rasio *likelihood*, yang dikenal sebagai distribusi penyimpangan, yang merupakan ukuran kesesuaian model. Perhitungan pengujian ini yaitu selisih (diff) nilai simpangan antara kedua model dan dapat dituliskan:

$$diff = -2\ln\left(\frac{\lambda_0}{\lambda_1}\right) \qquad (2.19)$$

dengan,

λ_0 = nilai deviasi *null* model

λ_1 = nilai deviasi *full* model

Hox (2010) menyatakan bahwa semakin kecil nilai pada deviasi model maka semakin baik model itu, serta jika nilainya $diff > \chi^2{}_{(a,db)}$ maka H_0 akan di tolak. Dimana db adalah selisih jumlah parameter pada kedua model. Kita dapat menyimpulkan efek acak adalah signifikan, yang berarti ada variasi atau keragaman signifikan pada variabel terikat antar kelompok.

2.9 Koefisien Korelasi Interklas

Kistiana, dkk. (2020) menyatakan korelasi adalah ukuran kedekatan hubungan antara dua variabel. Koefisien korelasi interklas digunakan dalam pengukuran variabilitas respon (keragaman) yang dijelaskan oleh perbedaan karakteristik antar kelompok atau menguji korelasi unit-unit di kelompok sama (Hox, 2010). Model regresi multilevel dianggap independen antara pengamatan. West, *et.al*., (2007) menyatakan independensi antar observasi diukur dengan nilai koefisien korelasi interklas. Dalam analisis regresi dua tingkat, nilai koefisien korelasi antar kelas menentukan nilai korelasi untuk setiap unit level 2. Model yang digunakan untuk merepresentasikan koefisien korelasi antar kelas adalah model tanpa variabel bebas pada setiap levelnya (*null model*). Nilai Koefisien Korelasi Antar Kelas atau *Interclass Correlation Coefficient* (ICC) dihitung dengan menggunakan rumus:

$$\rho = \frac{\sigma_{eij}^2}{\sigma_{eij}^2 + \sigma_{u0j}^2}, 0 \leq \rho \leq 1 \tag{2.20}$$

dengan,

ρ = koefisien korelasi interklas

σ_{u0j}^2 = ragam galat di level 1

σ_{eij}^2 = ragam galat di level 2.

2.10 Keragaman Model

Besarnya variasi pada variabel respon dijelaskan oleh model tertentu disebut koefisien determinasi. Dalam model multilevel, koefisien determinasi yang ditentukan lebih besar dari 1. Koefisien determinasi merupakan persentase varians dijelaskan pada setiap tingkat respon (Hox, 2010). Koefisien determinasi perlu memperkirakan rasio error spread terhadap total spread. Koefisien determinasi untuk level 1 dirumuskan sebagai berikut:

$${R_1}^2 = 1 - \frac{\alpha_{up}^2}{\alpha_{u0}^2} \tag{2.21}$$

dengan,

$R_1{}^2$ = koefisien determinasi pertama pada level 1

α_{up}^2 = ragam galat level 1 dengan p variabel prediktor

α_{u0}^2 = penduga ragam galat level 1 tanpa variabel prediktor

Koefisien determinasi kedua untuk level 2 dirumuskan sebagai berikut:

$$R_2{}^2 = 1 - \frac{\alpha_{eq}^2}{\alpha_{e0}^2} \tag{2.22}$$

dengan,

$R_2{}^2$ = koefisien determinasi kedua pada level 2

α_{eq}^2 = ragam galat level 2 dengan q variabel prediktor

α_{e0}^2 = penduga ragam galat level 2 tanpa variabel prediktor

Menurut Kurniawan (2010), kisaran nilai koefisien determinasi (R^2) antara 0 hingga 1. Apabila nilai R^2 dikalikan 100% menunjukkan variasi tanggapan atau persentase informasi yang diberikan oleh model regresi yang dihasilkan. Semakin besar nilai R^2 yang diperoleh maka perolehan model regresi semakin baik.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian dilaksanakan pada semester genap 2021/2022 di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA), Universitas Lampung.

3.2 Data Penelitian

Penelitian menggunakan data sekunder berupa kepadatan penduduk di Provinsi Sumatera Selatan tahun 2020 yang didapat dari website https://sumsel.bps.go.id/ yang merupakan data unggahan oleh Badan Pusat Statistik Provinsi Sumatera Selatan. Unit analisis Level 1 mencakup 204 kecamatan di Sumatera Selatan, dan unit analisis Level 2 mencakup 13 kabupaten di Sumatera Selatan. Variabel penelitian ini secara singkat pada Tabel 1.

Tabel 1. Variabel penelitian

Variabel	Keterangan
Y	Kepadatan Penduduk
Level 1 (Kecamatan)	
X_1	Jumlah Penduduk Kecamatan
X_2	Rasio Jenis Kelamin Kecamatan
X_3	Jumlah Sarana Sekolah Kecamatan
Level 2 (Kabupaten)	

Z_1	Indeks Pembangunan Manusia
Z_2	Produk Domestik Regional Bruto Per Kapita

3.3 Metode Penelitian

Proses perhitungan dalam penelitian dengan menggunakan software SAS 9.4. Berkaitan dengan tujuan yang ingin di raih, langkah-langkah pada penelitian ini:

1. Melakukan analisis deskriptif terhadap masing-masing variabel di Level 1 (Kecamatan) serta Level 2 (Kabupaten).
2. Melakukan uji asumsi data.
    a. Uji normalitas. Apabila data tidak normal, dilakukan transformasi data dengan logaritma.
    b. Uji multikolinearitas
3. Memodelkan menggunakan analisis regresi multilevel dalam menentukan faktor-faktor di tingkat kecamatan dan kabupaten yang akan mempengaruhi kepadatan penduduk Sumatera Selatan pada tahun 2020 menggunakan metode *Maximum Likelihood.*
4. Membandingkan nilai deviasi dan *diff* antara model yang dihasilkan untuk memperoleh model terbaik.
5. Menghitung nilai koefisien korelasi *intraclass*.
6. Mendeskripsikan keanekaragaman yang diterangkan oleh kepadatan penduduk Level 1 (Kecamatan) serta Level 2 (Kabupaten) di Sumatera Selatan tahun 2020.

V. KESIMPULAN

Dari penelitian ini, dapat ditarik kesimpulan faktor-faktor yang mempengaruhi kepadatan penduduk di Provinsi Sumatera Selatan tahun 2020 yaitu variabel jumlah penduduk kecamatan, rasio jenis kelamin kecamatan, indeks pembangunan manusia, serta produk domestik regional bruto per kapita. Model regresi diperoleh dari hasil analisis regresi multilevel:

Kepadatan penduduk = -1,6481 + 12,8921 indeks pembangunan manusia – 0,07028 produk domestik regional bruto per kapita + 0,5752 jumlah penduduk kecamatan – 9,4219 rasio jenis kelamin kecamatan.

Nilai keragaman yang dijelaskan level 1 (kecamatan) yaitu sebesar 73,03%, sedangkan nilai keragaman yang dijelaskan level 2 (kabupaten) yaitu sebesar 24,23%. Hal tersebut menunjukkan faktor pada level kecamatan memberikan pengaruh sebesar 73,03% terhadap kepadatan penduduk di Provinsi Sumatera Selatan dan faktor pada level kabupaten memberikan pengaruh sebesar 24,23%. Nilai koefisien korelasi interklas yang diperoleh sebesar 0,818 yang berarti tanpa adanya pertimbangan faktor-faktor lain yang dapat berpengaruh kepadatan penduduk, sebesar 81,8% proporsi varians adanya perbedaan karakteristik antar kabupaten di Provinsi Sumatera Selatan.

DAFTAR PUSTAKA

Badan Pusat Statistik. 2021. *Provinsi Sumatera Selatan Dalam Angka 2021*. Badan Pusat Statistik Provinsi Sumatera Selatan, Palembang.

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. 1995. *Multilevel Statistical Models*. Edward Arnold, London.

Gujarati, D. N. 2004. Basic Econometric. 4th Edition. Institute of Education Multilevel Models Project, London.

Gujarati, D. 2006. *Dasar-dasar Ekonometrika*. Erlangga, Jakarta.

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 Apllications*. 2nd Edition. Routledge, New York.

Jones, B. S. and Steenbergen, M. R. 2002. Modelling Data Structures. *American Journal of Political Science*. **46**(1): 2018-237.

Kistiana, S., Nasution, L. S., dan Naibaho, M. M. P. 2020. Faktor Kontekstual dan Individual Terhadap Jumlah Anak Lahir Hidup: Sebuah Analisis Multilevel. *Jurnal Kependudukan Indonesia*. **15**(1): 33-38.

Kurniawan, D. 2008. *Regresi Linier*. R Development Core Team, Austria.

Rencher, A. C. and Schaalje, G. B. 2007. *Linear Models In Statistics*. 2nd Edition. John Wiley & Sons, Inc., New Jersey.

Ringdal, K. 1992. Methods for Multilevel Analysis. *Acta Sosiologica*. **35**: 235-243.

Sembiring, R. K. 1995. *Analisis Regresi*. ITB, Bandung.

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

Tantular, B., Aunuddin, dan Wijayanto, H. 2009. Pemilihan Model Regresi Linier Multilevel Terbaik. *Forum Statistika dan Komputasi*. **14**(2): 1-7.

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

Yulian, E. dan Pawitan, G. 2017. Pemodelan Status Usaha (Pengusaha dan Pekerja/Karyawan) Menggunakan Regresi Logistik Multilevel. *Jurnal Matematika MANTIK.* **3**(1): 32-40.