

**SKRIPSI**

**ANALISIS DATA PANEL DENGAN MODEL REGRESI  
LOGISTIK BINER**

**Disusun dan diajukan oleh**

**FITRI**

**H051171011**



**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2021**

**ANALISIS DATA PANEL DENGAN MODEL REGRESI  
LOGISTIK BINER**

**SKRIPSI**

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

**FITRI**

**H051171011**

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN**

**MAKASSAR**

**2021**

# LEMBAR PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini :

Nama : Fitri  
NIM : H051171011  
Program Studi : Statistika  
Jenjang : Sarjana (S1)

Menyatakan dengan ini bahwa karya tulisan saya berjudul:

## **Analisis Data Panel dengan Model Regresi Logistik Biner**

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

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

**Makassar, 29 Maret 2021**



**FITRI**

**NIM. H051171011**

**ANALISIS DATA PANEL DENGAN MODEL REGRESI  
LOGISTIK BINER**

**Disetujui oleh:**

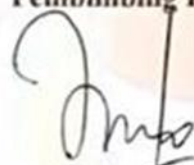
**Pembimbing Utama,**



**Dr. Anna Islamivati, S.Si., M.Si.**

**NIP. 19770808 200501 2002**

**Pembimbing Pertama,**



**Anisa, S.Si., M.Si.**

**NIP. 19730227 199802 2001**

**Ketua Departemen Statistika**



**Dr. Nurtiti Sunusi, S.Si., M.Si.**

**NIP. 19720117 199703 2002**

**Pada tanggal: 29 Maret 2021**

# LEMBAR PENGESAHAN

## ANALISIS DATA PANEL DENGAN MODEL REGRESI LOGISTIK BINER

Disusun dan diajukan oleh

**FITRI**

**H051171011**

Telah dipertahankan dihadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin pada tanggal 29 Maret 2021 dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama,



Dr. Anna Islamiyati, S.Si., M.Si.

NIP. 19770808 200501 2002

Pembimbing Pertama,



Anisa, S.Si., M.Si.

NIP. 19730227 199802 2001

Ketua Departemen Statistika


Dr. Nurhid Sunusi, S.Si., M.Si.

NIP. 19720117 199703 2002

## KATA PENGANTAR

*Assalamu'alaikum Warohmatullahi Wabarokatuh.*

*Alhamdulillah robbil'alamin*, Puji syukur kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat, nikmat, dan hidayah yang diberikan kepada penulis sehingga dapat menyelesaikan penulisan skripsi dengan judul “**Analisis Data Panel dengan Model Regresi Logistik Biner**” sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam.

Salam dan sholawat *Insyallah* senantiasa tercurah kepada **Nabi Muhammad Shallallahu'alaihi Wasallam**, sang kekasih tercinta yang telah memberikan petunjuk cinta dan kebenaran dalam kehidupan.

Dalam penyelesaian skripsi ini, penulis telah melewati perjuangan panjang dan pengorbanan yang tidak sedikit. Namun berkat rahmat dan izin-Nya serta dukungan dari berbagai pihak yang turut membantu baik moril maupun material sehingga akhirnya tugas akhir ini dapat terselesaikan. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang setinggi-tingginya dan penghargaan yang tak terhingga kepada Ayahanda **Huseng** dan Ibunda tercinta **Ida** yang telah membesarkan dan mendidik penulis dengan penuh kesabaran dan dengan limpahan cinta, kasih sayang dan doa kepada penulis yang tak pernah habis. Dan kakak tersayang **Asdianto** dan **Riska Damayanti** yang telah memberikan dukungan baik secara material maupun non material dan menjadi motivasi bagi penulis. Serta adik kecil penulis **Muh. Adzan Ashar** yang menjadi penyemangat untuk segera menyelesaikan masa studi penulis.

Ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada:

1. **Ibu Prof. Dr. Dwia Aries Tina Pulubuhu, MA**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Nurtiti Sunusi, S.Si., M.Si.**, selaku Ketua Departemen Statistika yang telah seperti orang tua sendiri. Segenap dosen pengajar dan staf

**Departemen Statistika** yang telah membekali ilmu dan kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika,

4. **Ibu Dr. Anna Islamiyati, S.Si., M.Si.**, selaku Pembimbing Utama penulis yang telah meluangkan waktu dan pemikirannya untuk memberikan arahan, pengetahuan, motivasi dan bimbingan ditengah kesibukan beliau serta menjadi tempat berkeluh kesah untuk penulis. **Ibu Anisa, S.Si., M.Si.** selaku Pembimbing Pertama sekaligus penasehat akademik penulis yang telah meluangkan waktunya ditengah kesibukan untuk memberikan arahan bagi penulis.
5. **Ibu Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.** dan **Bapak Siswanto, S.Si., M.Si.** selaku tim penguji yang telah memberikan saran dan kritikan yang membangun dalam penyempurnaan penyusunan tugas akhir ini.
6. Spesial untuk orang terdekat penulis, **Nurkamalia** yang telah menjadi sahabat bahkan saudari penulis dan senantiasa memberikan dorongan, semangat, dan motivasi dalam setiap keadaan. Terima kasih atas kebersamaan, suka dan duka dalam berjuang menjalani kehidupan dan pendidikan di Departemen Statistika. Tetap semangat, kamu pasti bisa!
7. Sahabat seperjuangan yang selalu menemaniku dibangku kuliah, the real is “**SEMANGAT BELAJAR**” **Musdalifah, S.Si, Nur Alya Tussa’ada, Siti Ihza Arsella dan Aqilah Salsabila Rahman.** Terima kasih atas kebersamaan, kebahagiaan, kesedihan, dan kebaikannya selama masa perkuliahan.
8. Spesial untuk sahabat tercinta penulis, **Nurul Wahyuni, S.Si., Riska Rasyid, Fakhriyyah Dj Junus, Nurul Annisa, Miftahul Jannah, Nurhidayatullah, Sakinah Oktoni, Nur Aprilia Dzulhijjah, Munadiah Apriliani dan Risnawati Azali** yang telah menjadi sahabat terbaik sejak awal perkuliahan dan senantiasa mendengarkan curhatan, memberikan dorongan, semangat, dan motivasi dalam setiap keadaan sehingga penulis bisa mendapatkan lebih banyak pelajaran hidup Serta sahabat terbaik sejak di bangku SMA, **Nurrifkatul Hikmayani dan Mita Ayu Lestari** yang sampai saat ini masih setia mendengarkan keluh kesah penulis.

9. Teman-teman **Statistika 2017**, terima kasih atas kebersamaan, suka dan duka selama menjalani pendidikan di Departemen Statistika.
10. Keluarga besar **DISKRIT 2017**, terima kasih telah memberikan pelajaran yang berharga dan arti kebersamaan selama ini kepada penulis. Pengalaman yang berharga telah penulis dapatkan dari teman-teman selama berproses.
11. **Keluarga Mahasiswa FMIPA Unhas** terkhusus anggota keluarga **Himatika FMIPA Unhas** dan **Himastat FMIPA Unhas**, terima kasih atas ilmu yang mungkin tidak bisa didapatkan di proses perkuliahan dan telah menjadi keluarga selama penulis kuliah di Universitas Hasanuddin.
12. Teman-teman **KKN Bone 3 Gelombang 104**, terkhusus **Ishar** yang merupakan teman seperjuangan dari SMA terima kasih telah menjadi teman sekaligus saudara bagi penulis.
13. Kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih setinggi-tingginya untuk segala dukungan dan partisipasi yang diberikan kepada penulis semoga bernilai ibadah di sisi **Allah SWT**.

Penulis berharap skripsi ini dapat memberikan tambahan pengetahuan baru bagi para pembelajar statistika. Penulis menyadari bahwa dalam penulisan tugas akhir ini masih banyak terdapat kekurangan. Oleh karena itu, dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga dapat bermanfaat bagi pihak-pihak yang berkepentingan. *Aamiin Yaa Rabbal Alamin.*

Makassar, 19 Maret 2021

  
Fitri



**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR  
UNTUK KEPENTINGAN AKADEMIS**

---

---

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Fitri  
NIM : H051171011  
Program Studi : Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam  
Jenis Karya : Skripsi

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non eksklusif (*Non-exclusive Royalty-Free Right*)** atas tugas akhir saya yang berjudul:

**“Analisis Data Panel dengan Model Regresi Logistik Biner”**

beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada tanggal 29 Maret 2021

Yang menyatakan,

  
Fitri

## ABSTRAK

Model regresi logistik biner dapat digunakan pada data panel dengan respon kategorik yang mengalami pengukuran berulang berdasarkan waktu. Penelitian ini bertujuan untuk mengetahui faktor-faktor yang berpengaruh terhadap Indeks Pembangunan Manusia di Provinsi Sulawesi Selatan pada tahun 2015-2019. Data dianalisis melalui regresi logistik biner dengan pendekatan model efek tetap melalui *Conditional Maximum Likelihood Estimation* (CMLE) untuk data panel. Hasil penelitian ini menunjukkan bahwa variabel yang berpengaruh signifikan yaitu angka harapan hidup ( $X_1$ ), harapan lama sekolah ( $X_2$ ) dan rata-rata lama sekolah ( $X_3$ ). Diperoleh nilai probabilitas kabupaten/kota yang memiliki indeks pembangunan manusia menengah rendah dan menengah tinggi dengan ketepatan klasifikasi 56,25%.

**Keywords:** Indeks Pembangunan Manusia, Angka Harapan Hidup, Harapan Lama Sekolah, Rata-rata Lama Sekolah, *Conditional Maximum Likelihood Estimation*, Model Efek Tetap.

## ABSTRACT

Binary logistic regression models can be used on panel data with categorical responses that experience repeated measurements based on time. This study aims to determine the factors that influence the Human Development Index in South Sulawesi Province in 2015-2019. Data were analyzed through binary logistic regression with fixed effect model approach through Conditional Maximum Likelihood Estimation (CMLE) for panel data. The results of this study indicate that the variables that have a significant effect are life expectancy ( $X_1$ ), school length expectancy ( $X_2$ ) and the average length of schooling ( $X_3$ ). Obtained the probability value of districts/cities that have a medium low and medium high human development index with a classification accuracy of 56.25%.

**Keywords:** Human Development Index, Life Expectancy Rate, Expectation of Old School Years, Average Length of Schooling, Conditional Maximum Likelihood Estimation, Fixed Effects Model.

## DAFTAR ISI

HALAMAN SAMPUL .....	i
HALAMAN JUDUL.....	ii
LEMBAR PERNYATAAN KEASLIAN .....	iii
LEMBAR PERSETUJUAN PEMBIMBING .....	iv
LEMBAR PENGESAHAN .....	v
KATA PENGANTAR .....	vi
PERSETUJUAN PUBLIKASI KARYA ILMIAH.....	ix
ABSTRAK .....	x
ABSTRACT.....	xi
DAFTAR ISI.....	xii
DAFTAR GAMBAR .....	xiv
DAFTAR TABEL.....	xv
DAFTAR LAMPIRAN.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Regresi Logistik Biner .....	5
2.2 Regresi Data Panel.....	7
2.3 Regresi Logistik Biner Data Panel Efek Tetap .....	9
2.4 Estimasi Parameter Model Regresi logistik Panel .....	10
2.5 Pengujian Panel <i>Unit Root</i> .....	11
2.6 Multikolinearitas .....	11
2.7 Pemilihan Model.....	12
2.8 Pengujian Model Regresi Logistik Biner Data Panel Efek Tetap.....	13
2.9 Kriteria Pemilihan Model Terbaik.....	14
2.9.1 <i>Akaike's Information Criterion (AIC)</i> .....	14

2.9.2 Ketepatan Klasifikasi .....	14
2.10 Indeks Pembangunan Manusia (IPM).....	15
<b>BAB III METODOLOGI PENELITIAN.....</b>	<b>17</b>
3.1 Jenis dan Sumber Data.....	17
3.2 Identifikasi Variabel.....	17
3.3 Metode Analisis .....	20
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>22</b>
4.1 Estimasi Parameter Model Regresi Logistik Biner Data Panel Efek Tetap yang Bersesuaian dengan Data Indeks Pembangunan Manusia di Provinsi Sulawesi Selatan Tahun 2015-2019.....	22
4.2 Penerapan Model Logistik Biner Data Panel Efek Tetap Pada Data Indeks Pembangunan Manusia Data Indeks Pembangunan Manusia di Provinsi Sulawesi Selatan Tahun 2015-2019 .....	27
4.2.1 Uji Panel Unit Root.....	28
4.2.2 Pengujian Multikolinearitas .....	28
4.2.3 Pemilihan Model Regresi Logistik Biner Data Panel .....	28
4.2.4 Estimasi Parameter Model Regresi Logistik Biner Data Panel Efek Tetap .....	30
4.2.5 Uji Signifikansi Parameter .....	30
<b>BAB V KESIMPULAN DAN SARAN.....</b>	<b>40</b>
5.1 Kesimpulan .....	40
5.2 Saran .....	40
<b>DAFTAR PUSTAKA .....</b>	<b>41</b>
<b>LAMPIRAN.....</b>	<b>45</b>

## DAFTAR GAMBAR

Gambar 4.1 Grafik Hubungan antara Probabilitas $P(Y=1)$ , $P(Y=0)$ dan Angka Harapan Hidup .....	36
Gambar 4.2 Grafik Hubungan antara Probabilitas $P(Y=1)$ , $P(Y=0)$ dan Harapan Lama Sekolah.....	37
Gambar 4.3 Grafik Hubungan antara Probabilitas $P(Y=1)$ , $P(Y=0)$ dan Rata-rata Lama Sekolah.....	37
Gambar 4.4 Grafik Prediksi Probabilitas Indeks Pembangunan Manusia Kabupaten Selayar, Gowa, Pangkep, Soppeng, Sidrap, Enrekang, Luwu, Toraja Utara, Makassar, Parepare dan Palopo Tahun 2015-2019.....	38

## DAFTAR TABEL

Tabel 2.1 Struktur Data Panel Secara Umum.....	9
Tabel 2.2 Ketepatan Klasifikasi .....	15
Tabel 4.1 Hasil Uji Panel <i>Unit Root</i> Variabel Bebas .....	28
Tabel 4.2 Nilai VIF .....	28
Tabel 4.3 Uji <i>Hausman</i> .....	29
Tabel 4.4 Hasil Estimasi Model Regresi Logistik Biner dengan Data Panel Efek Tetap .....	30
Tabel 4.5 Pengujian Serentak .....	30
Tabel 4.6 Pengujian Parameter Secara Parsial .....	31
Tabel 4.7 Pengujian Parameter Secara Parsial Berdasarkan Variabel yang Signifikan.....	31
Tabel 4.8 Perbandingan <i>Akaike's Information Criterion</i> (AIC) .....	32
Tabel 4.9 Pengelompokkan Indeks Pembangunan Manusia di 24 Kabupaten/Kota .....	39
Tabel 4.10 Ketepatan Klasifikasi Indeks Pembangunan Manusia.....	39

## DAFTAR LAMPIRAN

Lampiran 1. Data Penelitian Variabel Y (IPM) .....	45
Lampiran 2. Data Variabel $X_1$ (AHH).....	46
Lampiran 3. Data Variabel $X_2$ (HLS).....	47
Lampiran 4. Data Variabel $X_3$ (RLS) .....	48
Lampiran 5. Data Variabel $X_4$ (PRP) .....	49
Lampiran 6. Probabilitas Variabel yang Mempengaruhi Indeks Pembangunan Manusia di Provinsi Sulawesi Selatan Tahun 2018.....	50
Lampiran 7. Nilai Indeks Pembangunan Manusia Kabupaten/Kota di Provinsi Sulawesi Selatan tahun 2015-2019 .....	51
Lampiran 8. Nilai Probabilitas dan Prediksi Indeks Pembangunan Manusia .....	52



# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Analisis regresi merupakan suatu metode statistika yang dapat menjelaskan hubungan antara variabel terikat dan variabel bebas. Pada umumnya analisis regresi digunakan untuk menganalisis data dengan variabel terikat kuantitatif (kontinu) yang berdistribusi normal. Akan tetapi dalam prakteknya seringkali dijumpai variabel terikat kualitatif (kategorik), misalnya dalam bidang pendidikan, sosial, ekonomi, dan kesehatan. Apabila data pada variabel terikat adalah kategori, model regresi yang mampu menyelesaikannya adalah model regresi logistik. Model regresi logistik adalah salah satu model yang digunakan untuk mencari hubungan antara variabel terikat kategorik dengan satu atau lebih variabel bebas yang kontinu ataupun kategorik.

Regresi logistik biner merupakan satu teknik analisis statistika yang digunakan untuk menganalisis hubungan antara satu atau lebih variabel bebas dengan variabel terikat yang bersifat biner (dikotomis). Beberapa penelitian yang mengkaji mengenai model regresi logistik biner diantaranya Kotimah (2014) menggunakan model regresi logistik biner stratifikasi pada partisipasi ekonomi perempuan di Provinsi Jawa. Wardhani, ddk (2015) menggunakan model regresi logistik biner dan model logistik untuk menganalisis keputusan konsumen memilih bahan bakar minyak. Marlina (2017) menggunakan penerapan regresi logistik pada model *problem based learning* berbantu *software cabri 3D* dan Khaerin (2018) menggunakan penerapan model regresi logistik biner pada motif faktor kewirausahaan. Penelitian-penelitian tersebut masih sebatas penggunaan data *cross section*.

Dalam penelitian ini model regresi logistik biner dikembangkan dengan menggunakan data panel yang merupakan data pengamatan dilakukan secara berulang terhadap subjek dan variabel yang sama. Data panel merupakan penggabungan antara data *cross section* dan data *time series*. Baltagi (2015) menjelaskan bahwa kelebihan data panel adalah lebih komprehensif karena mengandung unsur waktu sehingga jumlah data akan meningkat dan dapat

meningkatkan efisiensi dalam penaksiran parameternya. Terdapat kelebihan lain dari data panel seperti yang dijelaskan oleh Hsio (2014) dalam Basuki (2015) bahwa penggunaan panel data dalam penelitian memiliki beberapa keuntungan dibandingkan menggunakan data jenis *cross section* maupun *time series* saja. Pertama, dapat memberikan peneliti jumlah pengamatan yang besar, meningkatkan *degree of freedom* (derajat kebebasan), data memiliki variabilitas yang besar dan mengurangi kolinieritas antara variabel penjelas, dapat menghasilkan estimasi yang efisien. Kedua, data panel dapat memberikan informasi lebih banyak yang tidak dapat diberikan hanya oleh data *cross section* atau *time series* saja. Ketiga, data panel dapat memberikan penyelesaian yang lebih baik dalam inferensi perubahan dinamis dibandingkan data *cross section*.

Data panel juga memungkinkan memiliki respon kategori sehingga dapat dimodelkan dengan pendekatan regresi logistik seperti pada penelitian Jayanti (2017) yang menggunakan penerapan regresi logistik data panel pada evaluasi kebijakan *inflation targeting framework* di pulau Jawa periode 2006-2015. Model regresi logistik biner data panel disusun berdasarkan beberapa pendekatan model yaitu model efek umum, model efek tetap, dan model efek acak. Model regresi logistik data panel sangat rentan terhadap bias variabel karena model regresi logistik data panel sulit dimodelkan. Model efek tetap memberikan solusi untuk masalah ini dengan menghilangkan parameter efek individu (Allison, 2009). Model efek tetap membantu mencapai kesimpulan menggunakan variansi dalam kelompok daripada hanya mengeksploitasi variansi antar kelompok. Dalam model regresi data panel efek tetap, transformasi harus dilakukan untuk mendapatkan estimator yang konsisten. Akan tetapi, cara yang sama tidak dapat diterapkan pada analisis data panel model regresi logistik (Baltagi, 2005). Sehingga diperlukan cara yang lain untuk memperoleh estimator yang konsisten. Dengan demikian diperkenalkan *conditional maximum likelihood estimation* untuk data panel respon biner pada model logistik efek tetap yaitu fungsi likelihood bersyarat untuk mendapatkan estimator yang konsisten (Kurniawan, 2019).

Berdasarkan uraian diatas, maka penulis menyusunnya dalam sebuah penelitian dengan judul “**Analisis data panel dengan model regresi logistik biner**”. Model ini diaplikasikan pada data indeks pembangunan manusia (IPM)

yang diamati tiap kabupaten/kota di Provinsi Sulawesi Selatan tahun 2015-2019. Indeks pembangunan manusia adalah suatu tolak ukur angka kesejahteraan suatu daerah atau negara yang dilihat berdasarkan tiga dimensi yaitu indeks pendidikan, indeks kesehatan dan indeks pengeluaran. Indeks pendidikan ini memiliki beberapa indikator yang digunakan, salah satunya yaitu rata-rata lama sekolah dan harapan lama sekolah. Indeks kesehatan memiliki indikator yaitu angka harapan hidup. Selain itu, indeks pengeluaran memiliki indikator yaitu pengeluaran per kapita (BPS, 2015). Indikator ini dapat digunakan sebagai contoh penerapan dalam analisis data panel dengan model regresi logistik biner.

## 1.2 Rumusan Masalah

Adapun rumusan masalah dari penelitian ini adalah sebagai berikut:

1. Bagaimana estimasi parameter model regresi logistik biner data panel dengan *conditional maksimum likelihood estimation* yang bersesuaian dengan data indeks pembangunan manusia di Provinsi Sulawesi Selatan tahun 2015-2019?
2. Bagaimana pengaruh rata-rata lama sekolah, angka harapan hidup, harapan lama sekolah dan pengeluaran per kapita terhadap indeks pembangunan manusia di Provinsi Sulawesi Selatan tahun 2015-2019 dengan *conditional maksimum likelihood estimation*?

## 1.3 Batasan Masalah

Penelitian ini menggunakan estimasi parameter model efek tetap dengan *conditional maksimum likelihood estimation* pada model regresi logistik data panel yang diaplikasikan pada data rata-rata lama sekolah, angka harapan hidup, harapan lama sekolah dan pengeluaran per kapita terhadap indeks pembangunan manusia di Provinsi Sulawesi Selatan tahun 2015-2019 yang diperinci berdasarkan 24 kabupaten/kota.

## 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah ditentukan, maka tujuan penulisan ini sebagai berikut:

1. Memperoleh hasil estimasi parameter model regresi logistik biner data panel dengan *conditional maksimum likelihood estimation* yang bersesuaian dengan

data indeks pembangunan manusia di Provinsi Sulawesi Selatan tahun 2015-2019.

2. Memperoleh hasil yang berpengaruh signifikan terhadap indeks pembangunan manusia di Provinsi Sulawesi Selatan tahun 2015-2019.

### **1.5 Manfaat Penelitian**

Penelitian ini diharapkan bermanfaat untuk berbagai pihak baik peneliti, mahasiswa, dan instansi. Manfaat yang diharapkan yaitu:

1. Menambah wawasan dan pengetahuan mengenai estimasi parameter model regresi logistik biner data panel dengan *conditional maksimum likelihood estimation*.
2. Sebagai bahan rujukan dan pengembangan pembelajaran statistika tentang model regresi logistik biner data panel.

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Regresi Logistik Biner

Regresi logistik merupakan regresi global. Regresi logistik biner adalah analisis yang digunakan untuk data respon berkategori atau hanya memiliki dua kemungkinan nilai/hasil (*dikotomus*), misalnya ya/tidak, sukses/gagal. Regresi logistik termasuk dalam model linier diperumum atau *Generalized Linier Models* (GLM). Model linier diperumum merupakan pengembangan dari model linier klasik yang mengasumsikan variabel terikat tidak harus mengikuti distribusi normal, tetapi harus termasuk dalam distribusi keluarga eksponensial (Pradita, 2011). Hasil pengamatan variabel acak  $Y$  terikat hanya mempunyai dua kategori yaitu 0 dan 1, sehingga mengikuti distribusi *Bernoulli* dengan distribusi peluang sebagai berikut:

$$P(Y = y) = \pi(x_i)^{y_i}[1 - \pi(x)]^{1-y_i}; y_i = 0,1$$

Pada model regresi logistik, nilai variabel terikat dinyatakan dengan  $y = \pi(x) + \varepsilon$  dengan  $\varepsilon$ , yang merupakan nilai eror, diasumsikan memiliki dua kemungkinan nilai. Jika  $y = 1$  maka  $\varepsilon = 1 - \pi(x)$  dengan peluang  $\pi(x)$ , dan jika  $y = 0$  maka  $\varepsilon = -\pi(x)$  dengan peluang  $1 - \pi(x)$  (Hosmer, 2013).

Menurut Agresti (2002), fungsi distribusi peluang untuk  $Y$  dengan parameter  $(x)$  dapat dilihat pada persamaan berikut:

$$f(y_i; \beta) = \pi(x_i)^{y_i}[1 - \pi(x)]^{1-y_i}; y_i = 0,1$$

Menurut Hosmer & Lemashow (2000) model regresi logistik biner dapat dilihat pada persamaan berikut:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_j x_j + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_j x_j + \dots + \beta_k x_k)}$$

$$\pi(x) = \frac{\exp(\beta X)}{1 + \exp(\beta X)}$$

dengan  $\beta_0$  merupakan konstanta dan  $\beta_k$  koefisien regresi logistik untuk variabel bebas ke- $k$ .

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{12} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \mathbf{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

Fungsi  $\pi(x)$  merupakan fungsi non linear sehingga perlu dilakukan transformasi logit untuk memperoleh fungsi linear sehingga hubungan antara variabel bebas dan terikat dapat dilihat, yaitu:

$$y = g(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_j + \dots + \beta_k x_k$$

maka model regresi logistik dapat dituliskan dalam bentuk persamaan berikut:

$$\pi(x) = \frac{\exp g(x)}{1 + \exp g(x)} \quad (2.1)$$

Selanjutnya estimasi parameter regresi logistik dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Prinsip dari MLE adalah menemukan penduga  $\beta$  yang memaksimalkan fungsi *likelihood* disamakan dengan nol. Langkah pertama yang akan dilakukan yaitu membentuk fungsi *likelihood*nya dapat dilihat pada persamaan sebagai berikut:

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (2.2)$$

untuk mempermudah perhitungan, maka fungsi *likelihood* Persamaan (2.2) dimaksimalkan dalam bentuk  $\ln l(\beta)$  dapat dituliskan sebagai berikut:

$$\begin{aligned} \ln l(\beta) &= \ln \left( \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \right) \\ \ln l(\beta) &= \sum_{i=1}^n [y_i \ln \pi(x_i) + (1 - y_i) \ln(1 - \pi(x_i))] \end{aligned} \quad (2.3)$$

Dengan mensubstitusikan  $\pi(x_i)$  pada Persamaan (2.1) ke Persamaan (2.3) maka:

$$\begin{aligned} \ln l(\beta) &= \sum_{i=1}^n \left[ y_i \ln \left( \frac{\exp g(x)}{1 + \exp g(x)} \right) + (1 - y_i) \ln \left( 1 - \frac{\exp g(x)}{1 + \exp g(x)} \right) \right] \\ \ln l(\beta) &= \sum_{i=1}^n [y_i g(x) - \ln(1 + \exp g(x))] \\ \ln l(\beta) &= \sum_{i=1}^n [y_i (\beta_0 + \beta_1 x_1 + \dots + \beta_k x_{ik}) - \ln(1 + \exp \beta_0 + \beta_1 x_1 + \dots + \beta_k x_{ik})] \\ \ln l(\beta) &= \sum_{j=0}^k (\sum_{i=1}^n y_i x_{ij}) \beta_j - \sum_{i=1}^n \ln \{ 1 + \exp(\beta_0 + \sum_{j=1}^k \beta_j x_{ij}) \} \end{aligned} \quad (2.4)$$

Selanjutnya, bentuk  $\ln$  dari fungsi *likelihood* pada Persamaan (2.4) diturunkan terhadap  $\beta_j$  untuk mendapat  $\beta$  dan hasil dari turunan tersebut disamakan dengan nol, sehingga dapat dinyatakan dalam persamaan:

$$\begin{aligned} \frac{\partial \ln l(\beta)}{\partial \beta_j} &= 0 \\ \frac{\partial \ln l(\beta)}{\partial \beta_j} &= \frac{\partial \sum_{j=0}^k (\sum_{i=1}^n y_i x_{ij}) \beta_j - \sum_{i=1}^n \ln \{ 1 + \exp(\beta_0 + \sum_{j=1}^k \beta_j x_{ij}) \}}{\partial \beta_j} = 0 \\ &= \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n x_{ij} \pi(x_i) = 0 \end{aligned}$$

$$= \sum_{i=1}^n x_{ij} \left( y_i - \frac{\exp(\beta_0 + \sum_{j=1}^k \beta_j x_{ij})}{1 + \exp(\beta_0 + \sum_{j=1}^k \beta_j x_{ij})} \right) \quad (2.5)$$

dari turunan pertama fungsi *ln likelihood* akan diperoleh turunan kedua fungsi *ln likelihood* pada persamaan berikut:

$$\begin{aligned} \frac{\partial^2 \ln l(\beta)}{\partial \beta_j^2} &= 0 \\ \frac{\partial^2 \ln l(\beta)}{\partial \beta_j^2} &= \sum_{i=1}^n x_{ij} \left( - \left\{ \frac{x_{ij} (\exp(\beta_0 + \sum_{j=1}^k \beta_j x_{ij}))}{(1 + \exp(\beta_0 + \sum_{j=1}^k \beta_j x_{ij}))^2} \right\} \right) \\ \frac{\partial^2 \ln l(\beta)}{\partial \beta_k^2} &= - \sum_{i=1}^n x_{ij} x_{ij} \pi(x_i) (1 - \pi(x_i)) \end{aligned} \quad (2.6)$$

Persamaan (2.5) dan (2.6) merupakan persamaan yang tidak linear dalam parameter ( $\beta$ ) sehingga digunakan metode *fisher Scoring* untuk mendapatkan nilai  $\beta$  (Agresti, 2002). Untuk memudahkan perhitungan, digunakan metode iterasi dengan program komputer yang akan berhenti ketika diperoleh nilai konvergen pada iterasi yang merupakan estimator yang memaksimalkan fungsi *likelihood*-nya (Hosmer, 2013).

## 2.2 Regresi Data Panel

Data panel merupakan data yang menggambarkan observasi individu (*cross section*) dan data deret waktu (*time series*) dalam suatu priode penelitian. Beberapa istilah yang digunakan untuk data panel yaitu data gabungan atau *pooled data* (menggabungkan observasi *time series* dan *cross section*), data mikropanel, data longitudinal (penelitian dari waktu ke waktu terhadap sebuah variabel atau kelompok subjek), analisis sejarah kejadian, dan analisis kohort. Walaupun memiliki banyak istilah, pada dasarnya data panel merupakan pergerakan waktu ke waktu dari unit *cross sectional*.

Dalam studi ekonometrik, variabel bebas pada model regresi data panel tidak selalu bersifat kontinu, tetapi juga dapat berupa variabel diskrit. Literatur terbaru, terutama dalam bidang analisis data panel, telah menghasilkan sejumlah teknik baru. Penggunaan data panel telah diperluas kedalam bentuk model non linear. Pada penelitian ini, digunakan analisis data panel efek tetap dengan model regresi logistik biner.

Seperti pada Persamaan (2.2), Green (2012) menyatakan model logistik biner data panel ditulis sebagai berikut:

$$y_{it} = \beta_0 + \sum_{k=1}^K \beta_k X_{kit} + \varepsilon_{it} \quad i = 1, 2, \dots, N \quad t = 1, 2, \dots, T$$

Keterangan :

$y_{it}$  : Variabel terikat untuk unit individu ke- $i$  dan waktu ke- $t$

$\beta_0$  : Konstanta intersep

$\beta_k$  : Koefisien regresi

$X_{kit}$  : Variabel bebas ke- $k$  untuk unit individu ke- $i$  dan waktu ke- $t$

$\varepsilon_{it}$  : Error regresi untuk unit individu ke- $i$  dan periode waktu ke- $t$

Bentuk matriks dituliskan sebagai berikut:

$$Y = X \beta + \varepsilon$$

Keterangan:

$$Y_{NT \times 1} = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{NT} \end{bmatrix}$$

$$X_{NT \times (K+1)} = \begin{bmatrix} 1 & X_{111} & X_{211} & \cdots & X_{K1T} \\ 1 & X_{121} & X_{221} & \cdots & X_{K2T} \\ 1 & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1NT} & X_{2NT} & \cdots & X_{KNT} \end{bmatrix}$$

$$\beta_{(K+1)} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}, \varepsilon_{NT \times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{NT} \end{bmatrix}$$

Pada model efek tetap, model regresi logistik biner data panel menjadi:

$$y_{it} = \beta x_{it} + \varepsilon_{it} + \alpha_i; i = 1, \dots, n; t = 1, \dots, T \quad (2.7)$$

Keterangan:

$y_{it}$  : Variabel terikat untuk unit individu ke- $i$  dan waktu ke- $t$

$\beta$  : Koefisien regresi

$\alpha_i$  : Efek individu yang berbeda-beda untuk setiap individu ke- $i$

$x_{it}$  : Variabel bebas ke- $k$  untuk unit individu ke- $i$  dan waktu ke- $t$

$\varepsilon_{it}$  : Error regresi untuk unit individu ke- $i$  dan periode waktu ke- $t$

Adapun struktur dari data panel secara umum dapat dilihat pada Tabel 2.1 sebagai berikut:



**Tabel 2.1** Struktur Data Panel Secara Umum

Individu (i)	Tahun (t)	Variabel Terikat (Y <sub>it</sub> )	Variabel Bebas (X <sub>1</sub> )	Variabel Bebas (X <sub>2</sub> )	...	Variabel Bebas (X <sub>k</sub> )
1	1	Y <sub>11</sub>	X <sub>111</sub>	X <sub>211</sub>	...	X <sub>K11</sub>
	2	Y <sub>12</sub>	X <sub>112</sub>	X <sub>212</sub>	...	X <sub>K12</sub>
	⋮	⋮	⋮	⋮	...	⋮
	T	Y <sub>1T</sub>	X <sub>11T</sub>	X <sub>21T</sub>	...	X <sub>K1T</sub>
2	1	Y <sub>21</sub>	X <sub>121</sub>	X <sub>221</sub>	...	X <sub>K21</sub>
	2	Y <sub>22</sub>	X <sub>122</sub>	X <sub>222</sub>	...	X <sub>K22</sub>
	⋮	⋮	⋮	⋮	...	⋮
	T	Y <sub>2T</sub>	X <sub>12T</sub>	X <sub>22T</sub>	...	X <sub>K2T</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮
N	1	Y <sub>N1</sub>	X <sub>1N1</sub>	X <sub>2N1</sub>	...	X <sub>KN1</sub>
	2	Y <sub>N2</sub>	X <sub>1N2</sub>	X <sub>222</sub>	...	X <sub>KN2</sub>
	⋮	⋮	⋮	⋮	...	⋮
	T	Y <sub>NT</sub>	X <sub>1NT</sub>	X <sub>2NT</sub>	...	X <sub>KNT</sub>

**2.3 Regresi Logistik Biner Data Panel Efek Tetap**

Jika efek spesifik individu  $\alpha_i$  diasumsikan tetap, maka  $\alpha_i$  dan  $\beta$  merupakan parameter yang tidak diketahui diestimasi untuk model:

$$y_{it} = \beta x_{it} + \varepsilon_{it} + \alpha_i; i = 1, \dots, n; t = 1, \dots, T$$

Model logistik biner data panel efek tetap juga dapat dinyatakan sebagai berikut:

$$\Pr(y_{it} = 1) = F(\beta x_{it} + \alpha_i)$$

Jika  $z = \beta x_{it} + \varepsilon_{it} + \alpha_i$  dengan  $\alpha_i = 0$  dan  $\varepsilon_{it} = 0$  maka model logistik panel adalah

$$P(Y = 1) = \pi(z)$$

$$P(Y = 0) = 1 - \pi(z)$$

Untuk model regresi data panel linear, ketika  $T$  tetap, hanya  $\beta$  yang dapat diperkirakan secara konsisten dengan terlebih dahulu menghilangkan  $\alpha_i$  menggunakan *within transformation*. Hal ini dimungkinkan untuk kasus linear karena MLE dari  $\beta$  dan  $\alpha_i$  independen asimtotik (Hasio, 2003). Untuk model pilihan biner kualitatif dengan model tetap  $T$ , ini tidak mungkin seperti yang ditunjukkan oleh Chambelain.

#### 2.4 Estimasi Parameter Model Regresi logistik Panel

Model data panel efek tetap dapat dinyatakan sebagai berikut:

$$y_{it} = \beta x_{it} + \varepsilon_{it} + \alpha_i; i = 1, \dots, n; t = 1, \dots, T$$

Ketika  $T$  mendekati tak hingga, estimator MLE menjadi konsisten. Namun, biasanya  $T$  kecil untuk data panel. Dengan demikian, estimasi  $\alpha_i$  dengan menggunakan estimator yang memiliki sifat sampel besar, seperti MLE, akan menjadi *meaningless* (tidak ada artinya). Permasalahan ini sering disebut dengan *incidental parameter problem* (Neyman, 1948).

Berbeda dengan kasus model peluang linear yang dapat menghasilkan estimator konsisten untuk  $\beta$  saat  $N$  menuju tak hingga melalui bentuk *first difference*, pada model non linear fungsi sederhana untuk  $\beta$  tidak mudah ditemukan. Pada kasus efek tetap model logistik, batas peluang MLE tidak mudah diketahui. Hal ini dapat diselesaikan melalui *sufficient* statistik  $\tau_i = \sum_{t=1}^T y_{it}$  untuk parameter insidental  $\alpha_i$  yang tidak berkorelasi dengan parameter struktural  $\beta$ , dengan peluang bersyarat sebagai berikut (Wooldridge, 2010):

$$P(y_i | \sum_{t=1}^T y_{it}) = \frac{\exp(\beta \sum_{t=1}^T x_{it} y_{it})}{\sum_{d \in B_i} \exp(\beta \sum_{t=1}^T x_{it} d_{it})} \quad (2.8)$$

dengan  $B_i = \{d = (d_{i1}, d_{i2}, \dots, d_{iT}) | d_{it} = 0 \text{ atau } 1 \text{ dan } \sum_{t=1}^T d_{it} = \sum_{t=1}^T y_{it}\}$ .

Selanjutnya, fungsi *log likelihood conditional*-nya adalah:

$$\ln L = \sum_i^N \ln[\exp(\beta \sum_{t=1}^T x_{it} y_{it})] / \sum_{d \in B_i} \exp(\beta \sum_{t=1}^T x_{it} d_{it}) \quad (2.9)$$

Fungsi pada Persamaan (2.9) disebut *conditional likelihood* yang sudah tidak mengandung parameter insidental  $\alpha_i$ . Memaksimalkan Persamaan (2.9) dapat dilakukan dengan metode numerik yaitu metode *fisher scoring* (Agresti, 2002). *Fisher scoring* menurut Smyth (2002) adalah salah satu bentuk pengembangan

dari metode *Newton-Raphson* dengan mengganti matriks hessian dengan matriks informasi. Adapun bentuk penduga parameter dengan metode *fisher scoring* adalah sebagai berikut:

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + \mathbf{I}(\boldsymbol{\beta}^{(t)})^{-1} \mathbf{s}(\boldsymbol{\beta}^{(t)})$$

dengan,  $\mathbf{I} = -E\left(\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \mid \mathbf{d}\right) = -\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$  dan  $\mathbf{s}(\boldsymbol{\beta}^{(t)}) = \frac{\partial \ln L}{\partial \boldsymbol{\beta}}$

$\boldsymbol{\beta}^{(t+1)}$  dan  $\boldsymbol{\beta}^{(t)}$  : Vektor untuk  $\boldsymbol{\beta}$  pada iterasi ke-  $t$  dan ke-  $t+1$

$\mathbf{I}^{(t)}$  : Matriks informasi yang berisi negatif ekspektasi dari turunan kedua *ln-likelihood* terhadap  $\boldsymbol{\beta}^{(t)}$

$\mathbf{s}(\boldsymbol{\beta}^{(t)})$  : Vektor gradien dari turunan pertama *ln-likelihood* terhadap  $(t)$

Iterasi tersebut berhenti pada saat keadaan konvergen, yaitu saat:  $|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}| \leq \epsilon$  dengan  $\epsilon$  adalah konstanta positif kecil yaitu  $\epsilon = 0,001$

## 2.5 Pengujian Panel *Unit Root*

Regresi data panel merupakan metode analisis yang menggunakan data dengan *series* yang menyimpan permasalahan *unit root* yang menyebabkan data tidak stasioner. Baltagi (2011) menyatakan bahwa data dinyatakan bersifat stasioner apabila memiliki rata-rata dan varians konstan yang tidak dipengaruhi oleh waktu. Apabila menggunakan data yang tidak stasioner, regresi yang dihasilkan akan bersifat *spurious* (lancung). Regresi lancung adalah situasi ketika hasil regresi akan terlihat baik dengan nilai koefisien determinasi ( $R^2$ ) tinggi dan uji hipotesis signifikan tetapi sebenarnya hubungan antarvariabel dalam model tidak memiliki arti (Enders, 2004). Oleh karena itu, pengujian stasioneritas penting dilakukan dalam data panel yang disebut dengan pengujian *panel unit root*.

## 2.6 Multikolinearitas

Multikolinearitas adalah suatu kondisi terjadi korelasi antara satu variabel bebas dengan variabel bebas yang lainnya sehingga menyebabkan parameter dalam model regresi yang dihasilkan memiliki sisaan yang sangat besar. Salah satu cara untuk mengecek multikolinearitas yaitu dengan melihat nilai *Variance Inflation Factor (VIF)*. Nilai *Variance Inflation Factor (VIF)* dapat digunakan sebagai kriteria untuk mendeteksi kasus multikolinearitas pada model regresi yang

memiliki lebih dari satu variabel bebas. Nilai  $VIF$  dapat dinyatakan dalam persamaan berikut: (Yan dan Su, 2009).

$$VIF_k = \frac{1}{1-R_k^2}$$

dengan  $R_k^2$  merupakan koefisien determinasi antara  $x_k$  dengan variabel bebas lainnya pada persamaan regresi dan apabila nilai  $VIF > 10$ , maka dapat diindikasikan bahwa terdapat kasus multikolinearitas yang serius (Gujarati, 2004).

## 2.7 Pemilihan Model

Pemilihan model antara efek tetap atau efek acak dilakukan melalui uji *Hausman*. Pada kasus regresi logistik biner data panel, uji *Hausman* dilakukan untuk membandingkan estimator Gauss Hermite quadrature untuk model efek acak dengan Chamberlain's *conditional maximum likelihood estimator* (CMLE) untuk model efek tetap. Uji *Hausman* adalah pengujian statistik untuk memilih model efek tetap atau model efek acak yang paling tepat digunakan. Dengan asumsi hipotesis nol yaitu kedua estimator tersebut konsisten, namun CLME tidak efisien. Sebaliknya, dengan asumsi hipotesis alternatif benar, Gauss Hermite quadrature tidak konsisten, sedangkan CMLE konsisten dan efisien. Uji *Hausman* didasarkan pada statistik bervariasi antar unit waktu di dalam model.

$H_0$  : Korelasi  $(X_{it}, \varepsilon_{it}) = 0$  (model mengikuti efek acak)

$H_1$  : Korelasi  $(X_{it}, \varepsilon_{it}) \neq 0$  (model mengikuti efek tetap)

Statistik uji yang digunakan adalah uji *chi-squared* berdasarkan kriteria *Wald*, yaitu :

$$W = \{(q)^T [Var(q)]^{-1} (q)\} \quad (2.10)$$

Keterangan:

$q$  :  $\hat{\beta}_{CML} - \hat{\beta}_{GH}$

$\hat{\beta}_{CML}$  : Vektor estimasi koefisien regresi model efek tetap

$\hat{\beta}_{GH}$  : Vektor estimasi koefisien regresi model efek acak

Jika nilai  $W > \chi_{(a,K)}^2$  atau nilai *p-value* kurang dari taraf signifikansi yang ditentukan, maka tolak  $H_0$  sehingga model yang terpilih model efek tetap (Greene, 2002).

## 2.8 Pengujian Model Regresi Logistik Biner Data Panel Efek Tetap

Pengujian parameter model regresi logistik biner data panel efek tetap bertujuan untuk menguji signifikansi dari pengaruh variabel bebas terhadap variabel terikat dalam model. Ada dua pengujian yang dilakukan, yaitu uji serentak dan uji parsial. Untuk menguji pengaruh koefisien  $\beta$  secara keseluruhan dalam model dilakukan uji serentak, sedangkan untuk menguji apakah suatu variabel bebas berpengaruh signifikan untuk dimasukkan dalam model atau tidak dilakukan uji parsial (Hosmer, dkk, 2013). Pengujian secara serentak menggunakan uji rasio *likelihood* dengan hipotesisnya sebagai berikut:

$H_0: \beta_1 = \dots = \beta_k = 0$  (tidak terdapat variabel bebas yang berpengaruh terhadap variabel terikat)

$H_1: \text{minimal terdapat satu } \beta_j \neq 0$  (minimal terdapat satu variabel bebas yang berpengaruh terhadap variabel terikat)

Statistik uji yang digunakan:

$$G = -2 \ln \left[ \frac{\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_{it}} (1-\hat{\pi}_i)^{1-y_{it}}} \right] \quad (2.11)$$

Keterangan:

$n_0$  : Jumlah data yang berkategori nol

$n_1$  : Jumlah data yang berkategori satu

$n$  : Jumlah data

Statistik uji  $G$  mengikuti sebaran distribusi *Chi-squared* ( $\chi^2$ ) dengan derajat bebas  $k$ , di mana  $H_0$  akan ditolak jika nilai  $G > \chi^2_{(a;p)}$  dengan  $p$  adalah jumlah variabel bebas dan nilai  $\chi^2_{(a;p)}$  dapat diperoleh dari tabel *Chi-squared* yang berarti variabel bebas secara bersama-sama mempengaruhi variabel terikat.

Selanjutnya dilakukan pengujian secara parsial menggunakan *Wald test* dengan hipotesisnya sebagai berikut:

$H_0 = \beta_j = 0$

$H_1 = \text{paling tidak terdapat satu } \beta_j \neq 0, j = 1, 2, \dots, p$

Statistik uji *Wald*:

$$W = \left[ \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right]^2$$

dengan  $\hat{\beta}_j$  adalah penaksir  $\beta_j$  dan  $SE(\hat{\beta}_j)$  adalah penaksir error  $\beta_j$ ,  $W$  adalah nilai uji *Wald* mengikuti sebaran  $\chi^2$  dengan derajat bebas satu.  $H_0$  ditolak jika  $W > \chi^2_{(1,p)}$  atau  $p\text{-value} > \alpha$ , sehingga dapat disimpulkan bahwa variabel bebas berpengaruh pada variabel terikat.

## 2.9 Kriteria Pemilihan Model Terbaik

Kriteria pemilihan model terbaik pada analisis regresi logistik biner data panel digunakan untuk mendapatkan model yang terbaik yang mampu menjelaskan pola hubungan antara variabel bebas dengan variabel terikat, serta model yang terbaik yang mampu secara akurat melakukan estimasi. Untuk mendapatkan model terbaik yang mampu menjelaskan faktor-faktor yang berpengaruh terhadap variabel terikat dapat digunakan kriteria AIC. Sedangkan untuk mendapatkan model terbaik yang mampu memprediksi secara akurat dapat digunakan nilai ketepatan klasifikasi.

### 2.9.1 Akaike's Information Criterion (AIC)

Pemilihan model terbaik berdasarkan kriteria AIC dilakukan dengan memilih model yang menghasilkan nilai AIC terkecil. Hal tersebut dikarenakan besarnya nilai AIC sejalan dengan nilai devians dari model. Semakin kecil nilai devians maka akan semakin kecil pula tingkat kesalahan yang dihasilkan sehingga model yang diperoleh akan semakin tepat. Adapun untuk mendapatkan nilai AIC dapat digunakan rumus sebagai berikut:

$$AIC = -2\log l(\beta_k) + 2p$$

Keterangan:

$p$  : Jumlah parameter

$l(\beta_k)$  : Maksimum *likelihood* dari model

### 2.9.2 Ketepatan Klasifikasi

Salah satu ukuran untuk pemilihan model terbaik yang dapat digunakan pada pemodelan statistik yang melibatkan variabel terikat kualitatif adalah ketepatan klasifikasi (Ratnasari, 2012). Ketepatan klasifikasi dapat digunakan dalam suatu evaluasi model. Menurut Johnson & Wichern (1992), evaluasi ketepatan klasifikasi adalah suatu evaluasi yang melihat probabilitas kesalahan klasifikasi yang dilakukan oleh suatu fungsi klasifikasi. Nilai ketepatan klasifikasi tersebut dapat diperoleh dengan membandingkan nilai prediksi yang benar dari

model dengan nilai observasi yang sebenarnya. Adapun tabel ketepatan klasifikasi yang biasa digunakan pada model regresi dengan variabel terikat yang bersifat kategori disajikan pada Tabel 2.2.

**Tabel 2.2** Ketepatan Klasifikasi

Hasil Observasi	Prediksi	
	$y_1$	$y_2$
$y_1$	$n_{11}$	$n_{12}$
$y_2$	$n_{21}$	$n_{22}$

Keterangan:

$y_i$  : Variabel terikat,  $i = 1, 2, \dots, n$

$n_{ij}$  : Jumlah subjek dari  $y_i$  yang tepat diklasifikasikan sebagai  $y_j$  ( $i = j$ )

$n_{ij}$  : Jumlah subjek dari  $y_i$  yang salah diklasifikasikan sebagai  $y_j$  ( $i \neq j$ )

Nilai ketepatan klasifikasi (akurasi) dapat dihitung menggunakan persamaan sebagai berikut:

$$\text{Akurasi} = \frac{n_{11} + n_{22}}{n} \times 100\% \quad (2.12)$$

### 2.10 Indeks Pembangunan Manusia (IPM)

Indeks pembangunan manusia (IPM) merupakan indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia. IPM menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan serta pendidikan. Bagi Indonesia, IPM merupakan data strategis karena selain sebagai ukuran kinerja pemerintah, IPM juga digunakan sebagai salah satu alokator penentuan Dana Alokasi Umum (DAU) serta IPM dapat menentukan peringkat atau level pembangunan suatu wilayah.

Indeks pembangunan manusia (IPM) merupakan indikator yang digunakan untuk melihat perkembangan manusia, terdapat dua aspek yang perlu diperhatikan, yaitu kecepatan dan status pencapaian. IPM digunakan untuk mengklasifikasikan negara maju, negara berkembang atau negara terbelakang dan juga untuk mengukur pengaruh dari kebijaksanaan ekonomi terhadap kualitas hidup. Perkembangan angka IPM memberikan indikasi peningkatan atau penurunan kinerja pembangunan manusia. IPM disusun dari tiga dimensi dasar

yaitu umur panjang dan hidup sehat (*a long and a healthy life*), pengetahuan (*knowledge*) dan standar hidup layak (*decent standard of living*) (BPS, 2017).