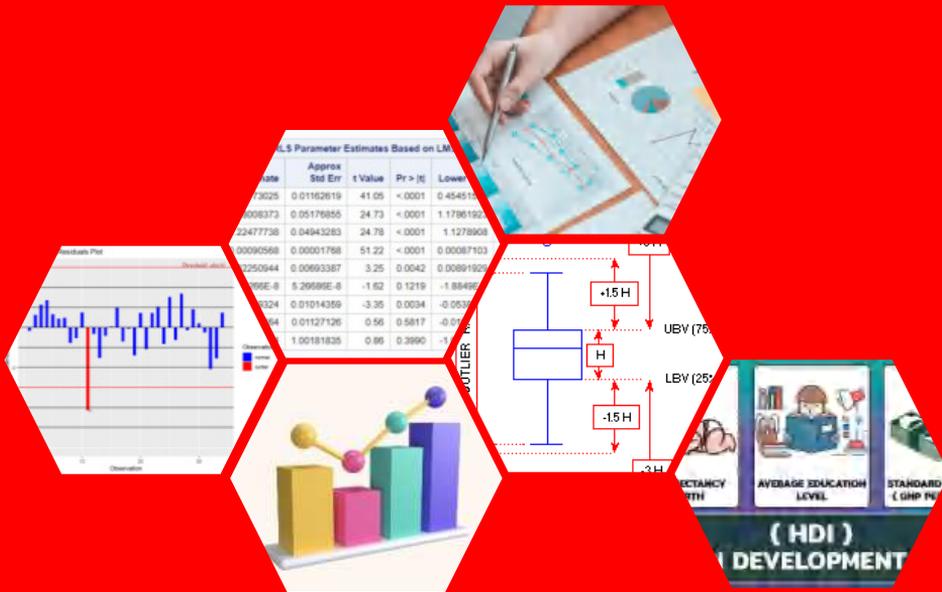


# PEMODELAN REGRESI *ROBUST* MENGGUNAKAN *LEAST MEDIAN OF SQUARES ESTIMATION* PADA INDEKS PEMBANGUNAN MANUSIA DI INDONESIA



NAHDAH AZATIL ISMAH  
H051201012



PROGRAM STUDI STATISTIKA  
DEPARTEMEN STATISTIKA  
MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2024

**PEMODELAN REGRESI *ROBUST* MENGGUNAKAN *LEAST  
MEDIAN OF SQUARES ESTIMATION* PADA INDEKS  
PEMBANGUNAN MANUSIA DI INDONESIA**

**NAHDAH AZATIL ISMAH  
H051201012**



**PROGRAM STUDI STATISTIKA  
DEPARTEMEN STATISTIKA  
MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2024**



**Optimization Software:  
[www.balesio.com](http://www.balesio.com)**

**PEMODELAN REGRESI *ROBUST* MENGGUNAKAN *LEAST  
MEDIAN OF SQUARES ESTIMATION* PADA INDEKS  
PEMBANGUNAN MANUSIA DI INDONESIA**

NAHDAH AZATIL ISMAH  
H051201012

Skripsi

Diajukan sebagai salah satu syarat untuk mencapai gelar Sarjana Statistika

Program Studi Statistika

pada

**PROGRAM STUDI STATISTIKA  
DEPARTEMEN STATISTIKA  
MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2024**



Optimization Software:  
[www.balesio.com](http://www.balesio.com)

SKRIPSI

PEMODELAN REGRESI *ROBUST* MENGGUNAKAN *LEAST  
MEDIAN OF SQUARES ESTIMATION* PADA INDEKS  
PEMBANGUNAN MANUSIA DI INDONESIA

NAHDAH AZATIL ISMAH

H051201012

Skripsi,

telah dipertahankan di hadapan Panitia Ujian Sarjana Statistika pada 24 Juni 2024  
dan dinyatakan telah memenuhi syarat kelulusan

pada

Program Studi Statistika  
Departemen Statistika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Hasanuddin  
Makassar

Mengesahkan:  
Pembimbing tugas akhir,

Mengetahui,  
Ketua Program Studi,



Optimization Software:  
[www.balesio.com](http://www.balesio.com)

Si., M.Si.  
3 1 001



Atma Islamiyati, S.Si., M.Si.  
NIP. 19770808 200501 2 002

## PERNYATAAN KEASLIAN SKRIPSI DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, skripsi berjudul "Pemodelan Regresi *Robust* Menggunakan *Least Median of Squares Estimation* pada Indeks Pembangunan Manusia di Indonesia" adalah benar karya saya dengan arahan dari pembimbing skripsi saya (Andi Kresna Jaya, S.Si., M.Si.). Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka skripsi ini. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini adalah karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut berdasarkan aturan yang berlaku.

Dengan ini saya melimpahkan hak cipta (hak ekonomis) dari karya tulis saya berupa skripsi ini kepada Universitas Hasanuddin.

Makassar, 24 Juni 2024



Nahdan Azatil Ismah  
NIM H051201012



Optimization Software:  
[www.balesio.com](http://www.balesio.com)



## UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa sallam*, yang telah membawa kita dari zaman kegelapan menuju zaman yang terang benderang. *Alhamdulillahirobbil'aalamiin*, berkat rahmat dan kemudahan yang diberikan oleh Allah *Subhanahu Wa Ta'ala*, penelitian ini dapat terlaksana dan terselesaikan atas bimbingan, diskusi dan arahan dari Bapak **Andi Kresna Jaya, S.Si., M.Si.** selaku pembimbing sekaligus penasehat akademik yang dengan penuh kesabaran telah meluangkan waktu dan pemikirannya. Ucapan terima kasih juga penulis sampaikan kepada Ibu **Prof. Dr. Dr. Georgina Maria Tinungki, M.Si.** dan Bapak **Siswanto, S.Si., M.Si.** atas saran dan kritikan yang berharga serta waktu yang telah diberikan kepada penulis. Terima kasih juga kepada pimpinan Universitas Hasanuddin, Ketua Departemen Statistika, para dosen dan staff yang telah memberikan ilmu dan fasilitas kepada penulis.

Ucapan terima kasih serta penghargaan setinggi-tingginya penulis haturkan kepada orang tua, Ayahanda **Muh. Asri Makkuraga** dan Ibunda **Nuralam** yang telah memberikan dukungan, pengorbanan, kasih sayang serta doa yang tak henti-hentinya dipanjatkan kepada penulis. Tak lupa pula kepada kakak **Anisah Ramadhani** dan adik-adik tersayang, **Nur Aliyah Asri** dan **Muh. Al Farisi**, yang selalu memotivasi penulis untuk tetap semangat. Terima kasih juga kepada keluarga besar penulis, **Nasir Sirmas, Helmi, Wahyuningsih, Muh. Nasri** dan **Haerani** yang setia mendampingi dan menyemangati penulis.

Penulis juga berterima kasih kepada teman-teman **Statistika 2020 dan Pois20n**, terkhusus **Nurhaliza Rais, Irma Mulia, Sri Rahayu Yusri, Andi Nurannisa Azzahra, Ruslinda, Fauziah Nurhidayah, Najlah Fauziah, Alisha Shafa Azzahra** dan **Azalia Filadelfia Pagalo** serta keluarga besar **Himastat FMIPA Unhas** yang telah mewarnai dunia perkuliahan penulis. Teman-teman grup **Auntie Rich, Teman Biasa, PMM 2 ITS** dan **BDK Menkeu** yang senantiasa memberikan dukungan. **Sri Rahayu** dan **Nurjihan Shahid** yang menjadi tempat berbagi cerita, serta pemilik nim **H051181510** yang tak kalah penting kehadirannya dan semua pihak yang tidak dapat disebutkan satu-persatu, semoga segala dukungan dan partisipasi yang diberikan bernilai ibadah disisi Allah *Subhanahu Wa Ta'ala*. Penulis menyadari bahwa masih banyak kekurangan, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini memberikan manfaat untuk pembaca.



Optimization Software:  
[www.balesio.com](http://www.balesio.com)

Makassar, 24 Juni 2024

Nahdah Azatil Ismah

## ABSTRAK

NAHDAH AZATIL ISMAH. **Pemodelan Regresi *Robust* Menggunakan *Least Median of Squares Estimation* pada Indeks Pembangunan Manusia di Indonesia** (dibimbing oleh Andi Kresna Jaya)

**Latar Belakang.** Model regresi berperan penting dalam menganalisis hubungan antara variabel bebas dengan variabel terikat. Metode yang umum digunakan untuk mengestimasi parameter regresi adalah *Ordinary Least Square* (OLS). Namun, metode ini memiliki kelemahan karena sensitif terhadap adanya pencilan, yang dapat menyebabkan bias pada estimasi koefisien regresi. Sebagai solusi, metode *robust* seperti *Least Median of Squares* (LMS) dapat digunakan. LMS meminimalkan median dari kuadrat galat, sehingga lebih tahan terhadap pencilan dengan *breakdown point* yang tinggi. **Tujuan.** Penelitian ini bertujuan untuk melakukan pemodelan regresi *robust* menggunakan estimasi LMS untuk mengidentifikasi faktor-faktor yang secara signifikan mempengaruhi Indeks Pembangunan Manusia (IPM) di Indonesia. **Metode.** Regresi *robust* digunakan untuk mengestimasi parameter regresi pada data yang mengandung pencilan. Data yang digunakan mencakup 34 provinsi di Indonesia tahun 2022 tentang IPM. **Hasil.** Hasil analisis menunjukkan bahwa model regresi *robust* LMS lebih baik dalam menangani kondisi data yang mengandung pencilan, dengan nilai  $R^2$  sebesar 99,93% dan  $R^2_{Adjusted}$  sebesar 99,84%. Model terbaik yang diperoleh adalah model dengan nilai  $C_p$  *Mallows* sebesar 5,1 yang menunjukkan kombinasi dari 5 variabel yang signifikan mempengaruhi IPM di Indonesia tahun 2022. **Kesimpulan.** Faktor-faktor yang teridentifikasi mempengaruhi IPM di Indonesia meliputi angka harapan hidup, harapan lama sekolah, rata-rata lama sekolah, pengeluaran per kapita dan angka melek huruf.

**Kata Kunci:** Analisis regresi, OLS, Pencilan, Regresi Robust, LMS



## ABSTRACT

NAHDAH AZATIL ISMAH. **Robust Regression Modeling Using Least Median of Squares Estimation on Human Development Index in Indonesia** (supervised by Andi Kresna Jaya)

**Background.** Regression models play an important role in analyzing the correlation between independent variables and dependent variables. The commonly used method to estimate regression parameters is Ordinary Least Square (OLS). However, this method has the disadvantage of being sensitive to the presence of outliers, which can cause bias in the estimation of regression coefficients. As a solution, robust methods such as Least Median of Squares (LMS) can be used. LMS minimizes the median of the squared errors, making it more resistant to outliers with high breakdown points. **Objective.** This study aims to conduct robust regression modeling using LMS estimation to identify factors that significantly affect the Human Development Index (HDI) in Indonesia. **Methods.** Robust regression is used to estimate regression parameters on data containing outliers. The data used covers 34 provinces in Indonesia in 2022 regarding HDI. **Results.** The analysis results show that the LMS robust regression model is better at handling data conditions containing outliers, with an  $R^2$  value of 99.93% and  $R^2_{Adjusted}$  of 99.84%. The best model obtained is a model with a Cp Mallows value of 5.1 which shows a combination of 5 variables that significantly affect HDI in Indonesia in 2022. **Conclusion.** The identified factors affecting HDI in Indonesia include life expectancy, expected years of schooling, average years of schooling, per capita expenditure and literacy rate.

**Keywords:** Regression analysis, OLS, Outlier, Robust Regression, LMS



## DAFTAR ISTILAH

Istilah	Arti dan Penjelasan
<i>Breakdown point</i>	Proporsi minimal dari banyaknya pencilan dibandingkan seluruh data pengamatan
<i>Error</i>	Selisih antara nilai yang diamati dan nilai yang diprediksi dalam suatu model
Intersep	Nilai di mana garis regresi memotong sumbu $y$
Kuartil	Nilai-nilai yang membagi data menjadi empat bagian yang sama besar
<i>Outlier</i>	Nilai yang berbeda secara signifikan dari data lainnya
Parameter	Konstanta yang digunakan untuk mendefinisikan karakteristik suatu populasi
Pembobot	Nilai yang berbeda pada setiap titik pengamatan untuk menaksir parameter
<i>Plot</i>	Grafik yang digunakan untuk menggambarkan visualisasi data
<i>Robust</i>	Ketahanan suatu metode terhadap penyimpangan dari asumsi atau adanya <i>outlier</i> dalam data
Sampel	Subset dari populasi yang digunakan untuk mengestimasi karakteristik populasi secara keseluruhan
<i>Standar deviation</i>	Ukuran sebaran data yang menunjukkan seberapa jauh nilai data tersebar dari nilai rata-rata
Subset	Kumpulan bagian dari himpunan yang lebih besar dan mengandung beberapa elemen (atau semua) dari himpunan asli
<i>Whisker</i>	Bagian dari <i>box plot</i> yang menunjukkan jangkauan data di luar kuartil pertama dan kuartil ketiga



## DAFTAR LAMBANG/SINGKATAN

Lambang/Singkatan	Arti dan Penjelasan
$Y$	Variabel terikat
$\beta$	Parameter regresi
$\varepsilon_i$	Residual ke- $i$
$X$	Variabel bebas
$\alpha$	Taraf signifikansi
$s^2$	Varians galat
$R^2$	Koefisien Determinasi
$n$	Jumlah pengamatan
$k$	Jumlah variabel bebas
$p$	Jumlah parameter
$v_i$	<i>Error</i> dari model residual
$\rho$	Koefisien autokorelasi
$h_{ii}$	Nilai <i>leverage value</i>
$t_i$	Nilai TRES
$\hat{\sigma}$	Sigma atau parameter skala
$M_j$	Nilai minimum dari median kuadrat galat
$\varepsilon_i^*$	Nilai <i>breakdown point</i>
$\chi^2$	Distribusi <i>chi-square</i>
$W_{LMS}$	Bobot estimasi LMS
$Q_{LMS}(\beta)$	Median kuadrat galat dari $h$ pengamatan
$Se(\hat{\beta}_j)$	Nilai standar <i>error</i> parameter regresi ke- $j$
IQR	Jangkauan antar kuartil
JKT	Jumlah Kuadrat Total
JKG	Jumlah Kuadrat Galat
KTR	Kuadrat Tengah Regresi
KTG	Kuadrat Tengah Galat
MAD	<i>Median Absolute Deviation</i>
BLUE	<i>Best Linear Unbiased Estimator</i>
OLS	<i>Ordinary Least Square</i>
LMS	<i>Least Median of Squares</i>
WLS	<i>Weighted Least Square</i>
VIF	<i>Variance Inflation Factor</i>
TRES	<i>Studentized Deleted Residual</i>
	<i>Cook's Distance</i>
	<i>The Difference In Fits Statistics</i>
	Indeks Pembangunan Manusia
	Badan Pusat Statistik



## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PENGANTAR .....	ii
HALAMAN PENGESAHAN .....	iii
PERNYATAAN KEASLIAN SKRIPSI .....	iv
UCAPAN TERIMA KASIH .....	v
ABSTRAK .....	vi
ABSTRACT .....	vii
DAFTAR ISTILAH .....	viii
DAFTAR LAMBANG/SINGKATAN .....	ix
DAFTAR ISI .....	x
DAFTAR TABEL .....	xii
DAFTAR GAMBAR .....	xiii
DAFTAR LAMPIRAN .....	xiv
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Batasan Masalah .....	2
1.3 Tujuan Penelitian .....	2
1.4 Manfaat Penelitian .....	2
1.5 Teori .....	2
1.5.1 Analisis Regresi .....	2
1.5.2 Pengujian Asumsi .....	4
1.5.3 <i>Ordinary Least Square</i> (OLS) .....	6
1.5.4 Pencilan ( <i>Outlier</i> ) .....	6
1.5.5 Regresi <i>Robust</i> .....	9
1.5.6 <i>Least Median of Squares</i> (LMS) .....	9
1.5.7 Koefisien Determinasi ( $R^2$ ) .....	11
1.5.8 Uji t .....	11
1.5.9 Uji F .....	11
1.5.10 Uji t dan F untuk Memeriksa Validitas Model Terbaik .....	12
1.5.11 Uji t dan F untuk Memeriksa Validitas Model Pembangunan Manusia (IPM) .....	12
BAB II PEMBAHASAN .....	16



2.1.	Sumber Data.....	16
2.2.	Variabel Penelitian .....	16
2.3.	Tahapan Analisis.....	16
BAB III HASIL DAN PEMBAHASAN .....		18
3.1	Estimasi Parameter Regresi <i>Robust</i> LMS.....	18
3.2	Analisis Deskriptif .....	19
3.3	Pengujian Asumsi .....	20
3.3.1.	Uji Normalitas.....	20
3.3.2.	Uji Heteroskedastisitas.....	20
3.3.3.	Uji Multikolinearitas .....	21
3.3.4.	Uji Autokorelasi .....	21
3.4	Pendugaan Parameter Regresi dengan OLS .....	22
3.5	Pendeteksian Pencilan dan Amatan Berpengaruh.....	23
3.5.1.	Pendeteksian Pencilan.....	23
3.5.2.	Pendeteksian Amatan Berpengaruh.....	24
3.5.3.	Pencilan Berpengaruh.....	25
3.6	Pendugaan Parameter Regresi <i>Robust</i> dengan LMS.....	25
3.7	Koefisien Determinasi ( $R^2$ ) .....	27
3.8	Pengujian Parameter Regresi.....	27
3.8.1.	Uji Parsial.....	27
3.8.2.	Uji Simultan .....	28
3.9	Pemilihan Model Terbaik.....	28
BAB IV KESIMPULAN DAN SARAN .....		30
4.1	Kesimpulan .....	30
4.2	Saran .....	30
DAFTAR PUSTAKA .....		31
LAMPIRAN .....		33



## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
1. Analisis Deskriptif.....	19
2. Hasil Pengujian Normalitas.....	20
3. Hasil Pengujian Heteroskedastisitas.....	21
4. Hasil Pengujian Multikolinearitas .....	21
5. Hasil Pengujian Autokorelasi.....	21
6. Hasil Pendeteksian Pencilan.....	24
7. Hasil Identifikasi Pengamatan Berpengaruh .....	25
8. Hasil Identifikasi Pencilan Berpengaruh.....	25
9. Hasil Koefisien Determinasi .....	27
10. Hasil Pengujian Parsial .....	27
11. Hasil Pengujian Simultan .....	28
12. Hasil Pemilihan Model Terbaik .....	28



**DAFTAR GAMBAR**

<b>Gambar</b>	<b>Halaman</b>
1. <i>Box Plot</i> .....	7
2. <i>Q-Q Plot</i> .....	20
3. <i>Box Plot Variabel</i> .....	24



## DAFTAR LAMPIRAN

Lampiran	Halaman
1. Data Indeks Pembangunan Manusia di Indonesia Tahun 2022 .....	34
2. Statistika Deskriptif .....	36
3. <i>Ordinary Least Square</i> (OLS) .....	37
4. Pengujian Asumsi .....	38
5. Pencilan dan Pengamatan Berpengaruh .....	40
6. Pendugaan Parameter Model Regresi <i>Robust</i> LMS .....	44
7. Pemilihan Model Terbaik .....	45
8. Pendugaan Parameter Model Regresi <i>Robust</i> LMS Terbaik.....	46



# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Analisis regresi merupakan analisis statistik yang bertujuan untuk memodelkan hubungan antara variabel bebas dengan variabel terikat (Siagian & Marpaung, 2023). Salah satu cara untuk menduga koefisien parameter regresi terbaik yaitu dengan menggunakan Metode Kuadrat Terkecil atau *Ordinary Least Square* (OLS). Penggunaan OLS harus memenuhi beberapa asumsi klasik diantaranya asumsi normalitas, homoskedastisitas, non-multikolinearitas dan non-autokorelasi. Jika semua asumsi tersebut terpenuhi, maka penduga parameter yang diperoleh bersifat *Best Linear Unbiased Estimator* (Sari dkk., 2020). Namun dalam penerapan analisis regresi memungkinkan adanya kendala dari data, salah satunya adalah terdapat data yang bersifat *outlier* (pencilan).

Pencilan (*outlier*) merupakan pengamatan yang jauh dari kelompok data yang mungkin berpengaruh besar terhadap koefisien regresi. Metode OLS mempunyai kelemahan apabila terdapat *outlier*, yang dapat mengakibatkan estimasi koefisien regresi menjadi kurang tepat (Sari dkk., 2020). Oleh karena itu, diperlukan adanya metode yang bersifat *robust* di mana nilai estimasinya tidak banyak dipengaruhi oleh perubahan kecil dalam data.

Regresi *robust* merupakan metode regresi yang digunakan ketika data pengamatan mengandung *outlier* yang mempengaruhi model. Metode ini penting untuk menganalisis data yang mengandung *outlier* agar model yang dihasilkan kekar terhadap keberadaan *outlier*. Kekekaran suatu penduga dapat diukur dari nilai *breakdown point* (Sinurat & Marpaung, 2023). Salah satu metode *robust* dengan *breakdown point* yang tinggi yaitu *Least Median of Squares*, yang bekerja dengan meminimalkan median dari kuadrat galat. Median adalah suatu ukuran pemusatan yang tidak sensitif terhadap keberadaan *outlier* (Ratri dkk., 2021). Penerapan metode ini akan digunakan dalam mengukur indikator keberhasilan pembangunan.

Salah satu indikator keberhasilan pembangunan dapat dilihat dari Indeks Pembangunan Manusia (IPM), yang bertujuan untuk meningkatkan kualitas hidup manusia sesuai dengan Undang-Undang Nomor 25 Tahun 2004 tentang Sistem Perencanaan Pembangunan Nasional (Ummah, 2024). Menurut Badan Pusat Statistik, pada tahun 2020-2023 Indonesia mengalami peningkatan nilai IPM dengan peningkatan tertinggi mencapai 0,84%. Meskipun terjadi peningkatan secara umum, masih terdapat beberapa wilayah dengan nilai IPM yang cukup rendah, yang

nya *outlier*.

elumnya oleh Febrianto dkk (2018) yang membandingkan Anggaran Pendapatan dan Belanja Daerah di Pulau Jawa, metode *Least Median of Squares* (LMS) adalah metode terbaik dan SIC. Khotimah dkk (2021) juga melakukan perbandingan produksi kelapa sawit di Indonesia tahun 2018, menghasilkan



bahwa metode LMS lebih baik dibandingkan metode MM berdasarkan nilai koefisien determinasi dan RMSE. Penelitian selanjutnya dilakukan Ratri dkk (2021) menggunakan regresi *robust* LMS pada produksi kedelai di Indonesia, menunjukkan bahwa metode ini membantu meningkatkan ketahanan model regresi dan menghasilkan estimasi yang lebih baik dengan nilai *breakdown point* yang tinggi.

Berdasarkan uraian sebelumnya, penelitian ini bertujuan untuk melakukan pemodelan regresi *robust* menggunakan *Least Median of Squares Estimation* pada Indeks Pembangunan Manusia di Indonesia. Penelitian ini diharapkan dapat memberikan manfaat bagi penggunaan model regresi yang tahan terhadap *outlier*, khususnya dalam bidang IPM. Selain itu, penelitian ini juga diharapkan dapat mendukung pembuatan kebijakan yang lebih tepat sasaran untuk meningkatkan kesejahteraan masyarakat Indonesia.

## 1.2 Batasan Masalah

Batasan masalah penelitian ini adalah:

1. Data yang digunakan terdiri dari 8 variabel bebas yang diduga mempengaruhi IPM pada provinsi di Indonesia tahun 2022.
2. Pemilihan model terbaik didasarkan pada kriteria *Cp Mallows*.

## 1.3 Tujuan Penelitian

Tujuan penelitian ini adalah:

1. Mendapatkan model regresi *robust* yang terbentuk menggunakan *Least Median of Squares Estimation*.
2. Memperoleh faktor-faktor apa saja yang mempengaruhi IPM di Indonesia tahun 2022.

## 1.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Menambah wawasan mengenai metode alternatif pendugaan parameter yang dapat digunakan dalam regresi *robust* pada data yang mengandung *outlier*.
2. Menjadi sarana informasi dan referensi bagi semua pihak yang tertarik dengan masalah yang dibahas dalam penelitian ini.

## 1.5 Teori

### 1.5.1. Analisis Regresi

Analisis regresi merupakan salah satu teknik dalam bidang ilmu statistik yang dapat mengetahui hubungan atau pengaruh suatu variabel bebas terhadap variabel terikat (Marcus dkk., 2012). Variabel dalam hal ini merujuk pada variabel yang sedang diamati. Terdapat 2 jenis variabel yang digunakan dalam analisis regresi yaitu variabel bebas atau variabel yang memengaruhi dan variabel terikat atau variabel yang dipengaruhi. Analisis regresi linier memiliki dua jenis yaitu regresi linier sederhana dan analisis regresi linier berganda.



### A. Analisis Regresi Linier Sederhana

Analisis regresi linier sederhana melibatkan satu variabel bebas dalam analisisnya (Astuti & Sulistiyowati, 2017). Adapun rumus umum analisis regresi linier sederhana ditunjukkan pada persamaan (1).

$$Y_i = \beta_0 + \beta_1 X_1 + \varepsilon_i; \quad i = 1, 2, \dots, n \quad (1)$$

dengan:

- $Y_i$  : nilai variabel terikat pada pengamatan ke- $i$
- $\beta_0, \beta_1$  : parameter regresi
- $\varepsilon_i$  : galat (*error*)
- $X$  : variabel bebas

### B. Analisis Regresi Linier Berganda

Analisis regresi linier berganda merupakan bentuk kompleks atau lanjutan dari analisis regresi linier sederhana, yang bertujuan untuk melihat pengaruh dua atau lebih variabel bebas terhadap satu variabel terikat apabila nilai variabel bebasnya diketahui. Selain itu, tujuan lainnya adalah untuk mengetahui arah hubungan variabel bebas dengan variabel terikat. Model ini mengasumsikan bahwa terdapat hubungan yang linier atau dalam satu garis lurus antara variabel bebas dengan masing-masing variabel terikatnya (Janie, 2012). Adapun rumus umum analisis regresi linier berganda ditunjukkan pada persamaan (2) dan (3).

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i; \quad i = 1, 2, \dots, n; \quad n \geq k + 1 \quad (2)$$

atau

$$Y_i = \beta_0 + \sum_{j=1}^k \beta_j X_{ij} + \varepsilon_i \quad (3)$$

dengan:

- $Y_i$  : nilai variabel terikat pada pengamatan ke- $i$
- $\beta_0, \beta_1, \dots, \beta_k$  : parameter regresi
- $\varepsilon_i$  : galat (*error*)
- $X_{ij}$  : nilai variabel bebas pada pengamatan ke- $i$  variabel ke- $j$
- $n$  : jumlah pengamatan
- $k$  : banyaknya variabel bebas

Jika terdapat 2 variabel bebas ( $X_1$  dan  $X_2$ ), maka persamaan (3) dapat dituliskan sebagai berikut.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} \quad (4)$$

Merujuk pada persamaan (4), ada 3 kondisi yang dapat terjadi jika koefisien arah regresi  $\beta_1$  dan  $\beta_2$  memiliki nilai sebagai berikut (Walpole., 1995).

a. Kondisi ini berarti variabel  $Y$  tidak dipengaruhi oleh variabel bebasnya

f. Kondisi ini berarti terdapat hubungan dengan arah terbalik terikat ( $Y$ ) dengan variabel bebasnya ( $X_1$  dan  $X_2$ ).



3. Nilainya positif. Kondisi ini berarti terdapat hubungan searah antara variabel terikat ( $Y$ ) dengan variabel bebasnya ( $X_1$  dan  $X_2$ ).

### 1.5.2. Pengujian Asumsi

Uji asumsi klasik merupakan syarat agar suatu model regresi dianggap model yang baik dan dapat digunakan dalam suatu analisis. Uji asumsi klasik yang dilakukan antara lain uji normalitas, uji heteroskedastisitas, uji multikolinearitas dan uji autokorelasi (Ratri dkk., 2021).

#### A. Uji Normalitas

Uji normalitas dilakukan untuk mengetahui *studentized residual* yang diteliti berdistribusi secara normal atau tidak. Pengujian asumsi normalitas dapat dilakukan melalui pendekatan grafik seperti Q-Q *plot* atau dengan uji *Shapiro-Wilk*. Hipotesis uji normalitas *Shapiro-Wilk* yaitu:

$H_0$  : galat berdistribusi normal

$H_1$  : galat tidak berdistribusi normal

Adapun statistik uji *Shapiro Wilk* ditunjukkan pada persamaan (5) berikut:

$$W = \frac{1}{(n-1)s^2} \left[ \sum_{i=1}^n a_i (\varepsilon_{n-i+1} - \varepsilon_i) \right]^2 \quad (5)$$

dengan:

$a_i$  : koefisien tabel *Shapiro-Wilks test*

$\varepsilon_{n-i+1}$  : galat ke  $n - i + 1$

$\varepsilon_i$  : galat ke- $i$

$n$  : banyaknya pengamatan

$s^2$  : varians galat

Adapun kriteria uji yaitu jika nilai  $W_{hitung}$  lebih besar dari nilai kritis pada tabel  $W_{(n,\alpha)}$  atau  $p - value > 0,05$  maka  $H_0$  diterima, yang berarti asumsi normalitas terpenuhi (Syamsudin & Wachidah, 2020).

#### B. Uji Heteroskedastisitas

Heteroskedastisitas adalah kondisi di mana varians pada galat tidak sama atau tidak identik. Sebaliknya, apabila varians dari galat bersifat seragam atau konstan maka dapat dikatakan terjadi homoskedastisitas. Salah satu uji yang dapat digunakan untuk mendeteksi heteroskedastisitas adalah uji *Breusch-Pagan* dengan hipotesis uji sebagai berikut:

$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$  (tidak terjadi heteroskedastisitas)

$1, 2, \dots, n$  (terjadi heteroskedastisitas)

akan sebagai berikut:

$$\chi_{BP}^2 = n \times R^2 \quad (6)$$

*Breusch-Pagan*

pengamatan



$R^2$  : koefisien determinasi dari regresi kuadratik residual

Adapun kriteria uji yang digunakan yaitu jika nilai  $\chi_{BP}^2 \geq \chi_{tabel}^2$  atau  $p - value < 0,05$  maka keputusan tolak  $H_0$ , artinya bahwa terjadi heteroskedastisitas dalam model regresi (Azis dkk., 2023).

### C. Uji Multikolinearitas

Multikolinearitas merupakan suatu kondisi adanya hubungan linier sempurna atau mendekati antar variabel bebas dalam model regresi. Jika dalam model regresi terjadi multikolinearitas, maka dapat menyebabkan pemakaian yang kurang tepat karena taksiran regresi menjadi tidak stabil dan variabel koefisien regresi menjadi sangat besar. Uji Multikolinearitas dapat diketahui melalui nilai *Variance Inflation Factor* (VIF) yang dinyatakan pada persamaan (7).

$$VIF = \frac{1}{1 - R_j^2} \quad (7)$$

dengan:

$VIF$  : nilai *Variance Inflation Factor*

$j$  : banyaknya variabel bebas

$R_j^2$  : koefisien determinasi variabel bebas ke- $k$

Adapun kriteria uji yang digunakan yaitu jika nilai  $VIF > 10$  pada masing-masing variabel bebas maka dapat disimpulkan bahwa model regresi mengalami multikolinearitas atau dapat dikatakan bahwa terdapat hubungan linier antara variabel bebas dengan variabel bebas lainnya (Azis dkk., 2023).

### D. Uji Autokorelasi

Uji autokorelasi dilakukan untuk mengetahui ada tidaknya masalah autokorelasi dalam model regresi. Autokorelasi terjadi karena adanya galat yang berhubungan antar pengamatan tidak berkaitan satu sama lain atau saling bebas. Salah satu uji yang dapat digunakan yaitu uji *Breusch-Godfrey* atau *Lagrange-Multiplier*. Uji *Breusch-Godfrey* dilakukan dengan meregresikan variabel *error*  $\varepsilon_i$  menggunakan *autoregressive model* orde  $p$  yang ditunjukkan pada persamaan (8).

$$\varepsilon_i = \rho_1 \varepsilon_{i-1} + \rho_2 \varepsilon_{i-2} + \dots + \rho_p \varepsilon_{i-p} + v_i \quad (8)$$

dengan:

$\varepsilon_i$  : residual atau *error* dari model regresi

$\varepsilon_{i-1}$  : residual atau *error* sebelumnya

$\rho$  : koefisien autokorelasi

$v_i$  : *error* dari model residual ( $\varepsilon_i$ )

Hipotesis uji yang digunakan adalah sebagai berikut:

$\rho_p = 0$  (tidak terjadi autokorelasi antar galat)

$\rho_1, \rho_2, \dots, \rho_p$  (terjadi autokorelasi antar galat)

dapat digunakan rumus berikut:

$$\rho = \frac{\sum \varepsilon_i \varepsilon_{i-1}}{\sum \varepsilon_i^2} \quad (9)$$



Hasil yang diperoleh dari uji *Breusch-Godfrey* disebut  $\chi^2_{hitung}$  sehingga uji ini membandingkan nilai  $\chi^2_{tabel}$  dengan  $\chi^2_{hitung}$ . Adapun kaidah keputusan pada uji *Breusch-Godfrey* adalah jika nilai  $\chi^2_{hitung} < \chi^2_{tabel}$  atau  $p - value > 0,05$  maka tidak terdapat gejala autokorelasi, begitupun sebaliknya (Huseynli, 2022).

### 1.5.3. Ordinary Least Square (OLS)

Metode Kuadrat Terkecil atau *Ordinary Least Square* merupakan salah satu metode yang digunakan untuk menduga parameter dalam model regresi. Metode ini memerlukan beberapa uji asumsi yang harus dipenuhi oleh komponen  $\varepsilon$ , yaitu kenormalan, kehomogenan ragam, keacakan atau tidak adanya autokorelasi (Auqino dkk., 2019). OLS bertujuan untuk meminimalkan Jumlah Kuadrat Galat (JKG). Adapun persamaannya dapat ditulis dalam bentuk matriks sebagai berikut:

$$Y = X\beta + \varepsilon \quad (10)$$

dengan:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Keterangan:

$Y$  : vektor variabel terikat ( $n \times 1$ )

$X$  : matriks dari variabel bebas ( $n \times (k + 1)$ )

$\beta$  : vektor parameter model regresi ( $(k + 1) \times 1$ )

$\varepsilon$  : vektor galat ( $n \times 1$ )

Dalam menentukan penduga  $Y$  yaitu  $\hat{Y} = X\hat{\beta}$ , diperlukan nilai penduga untuk parameter  $\beta$  yaitu  $\hat{\beta}$  dengan parameter  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$  tidak diketahui dan perlu ditentukan nilai dugaannya dengan meminimumkan JKG. Nilai pendugaan parameter model regresi dinyatakan pada persamaan (11).

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (11)$$

### 1.5.4. Pencilan (*Outlier*)

Pencilan (*outlier*) merupakan pengamatan yang jauh dari kelompok data yang mungkin berpengaruh besar terhadap koefisien regresi. Keberadaan *outlier* dalam suatu analisis regresi dapat menimbulkan masalah yang akan mempengaruhi keakuratan hasil analisis, sehingga asumsi yang dihasilkan menjadi tidak tepat atau estimasi menjadi bias. Beberapa hal yang dapat menjadi penyebab terjadinya *outlier* seperti kesalahan dalam memasukkan data, gagal dalam spesifikasi atau terdapat *outlier* yang sebenarnya. Menghilangkan *outlier* bukanlah pilihan yang tepat dilakukan karena *outlier* justru memberikan informasi penting yang tidak dapat diabaikan.

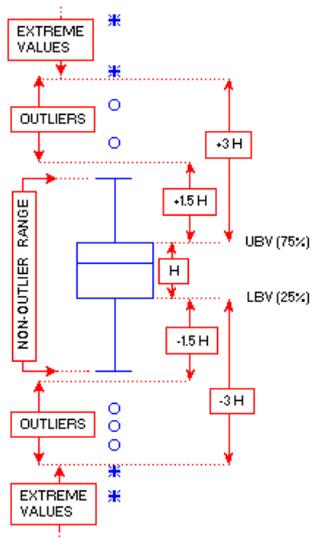


## A. Pendeteksian Pencilan

Setelah semua asumsi yang ada terpenuhi maka salah satu hal yang harus diperhatikan adalah dengan deteksi adanya pencilan dan pengaruh pengamatan pada data. Melakukan pendeteksian terhadap *outlier* sangat penting sebelum menerapkan metode-metode statistika pada data (Derquenne, 2023). Pendeteksian pencilan dapat dilakukan dengan metode grafis, *leverage value* atau perhitungan statistik. Beberapa diantaranya yaitu:

### 1. *Box Plot*

Metode *box plot* diperkenalkan pertama kali oleh John Tukey (1977) yang digunakan untuk memeriksa informasi dari suatu sampel. Selain itu, *box plot* juga digunakan untuk menampilkan data ekstrem dan menunjukkan pengamatan yang dianggap atau diperkirakan sebagai *outlier*. Terdapat beberapa informasi yang dapat ditampilkan dalam *box plot* seperti nilai minimum dan maksimum data, median, kuartil bawah ( $Q_1$ ) dan kuartil atas ( $Q_3$ ) serta besar selisih antara ( $Q_1$ ) dan ( $Q_3$ ) adalah jangkauan antar kuartil (*IQR*).



**Gambar 1.** *Box Plot*. Sumber: [www.docs.tibco.com](http://www.docs.tibco.com)

Dalam *box plot* berbagai informasi tentang distribusi data dapat dilihat. Garis di dalam *box plot* menunjukkan median (nilai tengah). *Whisker* di atas dan di bawah persegi panjang menunjukkan lokasi minimum dan maksimum data. Kuartil bawah ( $Q_1$ ) membagi data menjadi 25% terbawah, sedangkan kuartil atas ( $Q_3$ ) membagi data menjadi 25% teratas.

Adapun rumus untuk  $Q_1$  dan  $Q_3$  dapat dilihat pada (13).

$$Q_1 = \text{data ke } - \frac{n}{4} \quad (12)$$

$$Q_3 = \text{data ke } - \frac{3n}{4} \quad (13)$$



Garis tengah yang membentang dari kuartil bawah ( $Q_1$ ) ke kuartil atas ( $Q_3$ ) dikenal sebagai *Interquartile Range* (IQR), yang dapat ditulis dengan  $IQR = Q_3 - Q_1$ . Untuk mendeteksi *outlier* (nilai yang berbeda secara signifikan dari data lainnya), data di luar rentang  $1,5IQR$  di atas atau di bawah kuartil dianggap sebagai dugaan *outlier*. Sedangkan data di luar rentang  $3IQR$  di atas atau di bawah kuartil dianggap sebagai nilai ekstrim (Walpole., 1995).

## 2. Leverage Value

Metode yang digunakan dalam mengidentifikasi adanya pencilan terhadap variabel  $X$  adalah nilai pengaruh (*leverage value*). Nilai pengaruh ( $h_{ii}$ ) dari pengamatan ( $X_i, Y$ ) menunjukkan besarnya peranan  $Y$  terhadap  $\hat{Y}$  yang didefinisikan seperti pada persamaan (14).

$$h_{ii} = X_i^T (X^T X)^{-1} X_i \quad (14)$$

dengan  $i = 1, 2, \dots, n$  dan  $X_i^T = [X_{i1}, X_{i2}, \dots, X_{ik}]$  adalah vektor baris yang berisi nilai-nilai dari  $k$  variabel bebas pada pengamatan ke- $i$ . Nilai  $h_{ii}$  berada diantara  $0 \leq h_{ii} \leq 1$ . Jika  $h_{ii}$  lebih besar dari  $\frac{2p}{n}$ , dengan  $p = k + 1$  maka pengamatan ke- $i$  dikatakan pencilan terhadap  $X$  (Rahmatika dkk., 2014).

## 3. Studentized Deleted Residual

Menurut Draper dan Smith (1998) metode yang digunakan dalam mengidentifikasi pencilan terhadap variabel  $Y$  adalah *Studentized Deleted Residual* (TRES) yang didefinisikan seperti pada persamaan (15).

$$t_i = \varepsilon_i \left[ \frac{n - k}{JKG(1 - h_{ii}) - \varepsilon_i^2} \right]^{\frac{1}{2}} \quad (15)$$

dengan:

$\varepsilon_i$  :  $Y_i - \hat{Y}_i$

$n$  : banyaknya pengamatan

$k$  : banyaknya variabel bebas

$JKG$  : jumlah kuadrat galat

Hipotesis yang digunakan adalah sebagai berikut:

$H_0$  : pengamatan ke- $i$  bukan pencilan

$H_1$  : pengamatan ke- $i$  merupakan pencilan

Adapun kriteria uji yang melandasi keputusan yaitu tolak  $H_0$  jika nilai  $|t_i| \leq t_{(\frac{\alpha}{2}, n-k-1)}$ , dan terima  $H_0$  jika nilai  $|t_i| > t_{(\frac{\alpha}{2}, n-k-1)}$  dengan  $t_{\frac{\alpha}{2}}$  adalah distribusi *t-student*

(Rahmatika dkk., 2014).



## Amatan Berpengaruh

Berpengaruh adalah pengamatan yang baik secara individu atau beberapa pengamatan lainnya, memiliki pengaruh yang jauh berbagai nilai estimasi dibandingkan pengaruh sebagian besar. Oleh karena itu, perlu dilakukan deteksi pengamatan

berpengaruh pada data dan menilai dampaknya terhadap model. Beberapa metode pendeteksian amatan berpengaruh antara lain:

1. **Cook's Distance** digunakan untuk mengidentifikasi apakah suatu pengamatan berpengaruh terhadap semua estimasi koefisien regresi. Perhitungan *Cook's Distance* dapat dinyatakan secara umum dengan persamaan (16) berikut:

$$CD_i = \sqrt{\frac{1}{p} t_i^2 \frac{h_{ii}}{1 - h_{ii}}} \quad (16)$$

Suatu data pengamatan dikatakan berpengaruh terhadap koefisien regresi jika nilai  $CD_i > F_{\alpha; p, n - (k+1)}$ .

2. **The Difference In Fits Statistics (DFFITS)** digunakan untuk mengetahui apakah suatu pengamatan ke- $i$  berpengaruh terhadap model regresi yang ditinjau dari nilai dugaan  $\hat{Y}$ -nya. Perhitungan DFFITS dapat dinyatakan secara umum dengan (Rousseeuw & Wiley, 1987):

$$DFFITS_i = t_i \sqrt{\frac{h_{ii}}{1 - h_{ii}}} \quad (17)$$

Suatu pengamatan ke- $i$  dikatakan berpengaruh terhadap nilai dugaan  $\hat{Y}$ -nya apabila memiliki nilai  $|DFFITS_i| > 2 \sqrt{\left(\frac{p}{n}\right)}$ .

### 1.5.5. Regresi Robust

Regresi *robust* merupakan solusi alternatif untuk mengestimasi parameter regresi ketika terdapat *outlier* yang mempengaruhi model. Metode ini penting untuk menganalisis data yang dipengaruhi oleh *outlier* dengan hasil model yang kekar terhadap *outlier*. Dalam hal ini berarti tidak dipengaruhi oleh perubahan besar pada sebagian kecil data atau perubahan kecil pada sebagian besar data. Pada umumnya, regresi *robust* akan ditambahkan pembobot yang disimbolkan sebagai  $W$  seperti pada persamaan (18).

$$\hat{\beta}_{robust} = (X^T W X)^{-1} X^T W Y \quad (18)$$

dengan  $\hat{\beta}_{robust}$  adalah koefisien regresi *robust* berukuran  $(k + 1) \times 1$ .

Pada regresi *robust* data diberi bobot yang berbeda dalam mengestimasi parameter dengan bobot yang lebih kecil diberikan pada data yang terdeteksi sebagai *outlier*, sedangkan pada metode kuadrat terkecil data diberi bobot yang sama. Karena bobot yang diberikan berbeda maka pada pendugaan regresi *robust* *weighted least square* (Husain & Jamaluddin, 2023).



Optimization Software:  
[www.balesio.com](http://www.balesio.com)

### of Squares (LMS)

regresi *robust* yang dapat digunakan untuk mengatasi adanya estimasi *Least Median of Squares* (LMS). LMS merupakan parameter regresi *robust* yang memiliki nilai *breakdown point*

yang tinggi dengan cara meminimalkan median dari kuadrat galat. Penduga LMS didefinisikan pada persamaan (19) berikut:

$$\hat{\beta}_{LMS} = \arg \min Q_{LMS}(\beta) \quad (19)$$

dengan  $Q_{LMS}(\beta) = med \varepsilon_h^2$  adalah median kuadrat galat dari  $h$  pengamatan yang telah diurutkan yaitu  $(\varepsilon^2)_{1:n} \leq (\varepsilon^2)_{2:n} \leq \dots \leq (\varepsilon^2)_{n:n}$ .

Prinsip dasar metode regresi *robust* estimasi LMS adalah mencocokkan sebagian besar data setelah *outlier* teridentifikasi sebagai titik yang tidak berhubungan dengan data. Pada pendugaan OLS yang perlu dilakukan adalah meminimalkan kuadrat galat ( $\sum_{i=1}^n \varepsilon_i^2$ ), sedangkan pada *robust* estimasi LMS yang diminimalkan adalah median dari kuadrat galat yang didefinisikan dengan:

$$M_j = \min\{med \varepsilon_i^2\} = \min\{M_1, M_2, \dots, M_r\} \quad (20)$$

dengan  $\varepsilon_i^2$  adalah kuadrat galat dari hasil estimasi OLS, yang dinyatakan dengan:

$$\varepsilon_i = (Y_i - \beta_0 - \beta_1 X_1 - \dots - \beta_k X_k) \quad (21)$$

Menurut Rousseeuw & Leroy (2005) untuk memperoleh nilai  $M_1$  dapat dicari dengan himpunan bagian dari matriks  $X$  sebanyak  $h$  pengamatan, yang dinyatakan pada persamaan (22).

$$h = \left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{p+1}{2} \right\rfloor \quad (22)$$

dengan banyaknya data dinotasikan  $n$  dan  $p$  banyaknya parameter.

Nilai  $h$  harus selalu berbentuk bilangan bulat. Apabila nilai  $h$  tidak dalam bentuk bilangan bulat maka dapat dilakukan pembulatan ke atas (Ratri dkk., 2021). Nilai  $h$  akan berpengaruh pada *breakdown point* yang dinyatakan dengan:

$$\varepsilon_i^* = \frac{n-h+1}{n} \quad (23)$$

Estimasi LMS memberikan pembobot  $W_{ii}$  pada data, yang membuat pencilan tidak berpengaruh terhadap model parameter regresi. Bobot  $W_{ii}$  ditentukan berdasarkan asumsi *robust standar deviation* yang diperoleh berdasarkan hasil perhitungan  $M_j$  dan  $\hat{\sigma}$ , yang dirumuskan sebagai berikut:

$$W_{ii_{LMS}} = \begin{cases} 1, & \left| \frac{\varepsilon_i}{\hat{\sigma}} \right| \leq 2,5 \\ 0, & \text{lainnya} \end{cases} \quad (24)$$

dengan:

$$\hat{\sigma} = 1,4826 \left( 1 + \frac{5}{(n-p)} \right) \sqrt{M_j} \quad (25)$$

Setelah dihitung nilai dari bobot  $W_{ii}$ , dapat dibentuk matriks  $W$  seperti persamaan (26) berikut:

$$W = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{pmatrix} \quad (26)$$

$w_{ij} = 0$  dan  $i \neq j$ .

Setelah matriks  $W$  terbentuk, maka penduga parameter regresi *robust* LMS dinyatakan dengan persamaan (27) (Khotimah dkk., 2021).



$$\hat{\beta}_{LMS} = (X^T W X)^{-1} X^T W Y \quad (27)$$

dengan:

$Y$  : vektor variabel terikat berukuran  $n \times 1$

$X$  : matriks dari variabel bebas berukuran  $n \times (k + 1)$

$\hat{\beta}_{LMS}$  : vektor koefisien regresi berukuran  $(k + 1) \times 1$

### 1.5.7. Koefisien Determinasi ( $R^2$ )

Koefisien determinasi ( $R^2$ ) merupakan alat untuk mengukur seberapa baik model dapat menjelaskan variasi dalam variabel terikat berdasarkan variabel-variabel bebas. Nilai koefisien determinasi berada diantara 0 hingga 1. Nilai yang mendekati 1 berarti variabel-variabel bebas hampir memberikan informasi yang dijelaskan untuk memprediksi variabel terikat. Dengan kata lain, model dengan nilai koefisien determinasi tertinggi lebih mampu menjelaskan keadaan sebenarnya dengan baik. Untuk menghitung nilai  $R^2$  regresi *robust* digunakan rumus berikut (Rousseeuw & Leroy, 2005):

$$R^2 = 1 - \left( \frac{\text{med}|\hat{\epsilon}_i|}{\text{mad}(Y_i)} \right)^2 = 1 - \left( \frac{\text{med}|\hat{\epsilon}_i|}{\text{med}(Y_i - \text{med } Y_i')} \right)^2 \quad (28)$$

### 1.5.8. Pengujian Parameter

Pengujian parameter bertujuan untuk melihat apakah koefisien regresi yang diperoleh signifikan atau tidak. Koefisien regresi dapat dikatakan signifikan apabila secara statistik bernilai sama dengan nol. Apabila koefisien regresi yang diperoleh tidak sama dengan nol, maka dapat dikatakan bahwa variabel bebas tidak memiliki bukti yang cukup untuk dikatakan berpengaruh terhadap variabel terikat. Oleh karena itu, semua nilai koefisien regresi harus diuji dengan menggunakan 2 uji yakni uji parsial dan uji simultan.

#### A. Uji Parsial

Uji parsial digunakan untuk mengetahui signifikansi variabel bebas secara individu terhadap variabel terikat dalam suatu model regresi. Uji parsial yang sering digunakan adalah uji t parsial. Adapun hipotesis yang digunakan dalam uji t adalah sebagai berikut:

$$H_0 : \beta_j = 0 ; j = 1, 2, \dots, k$$

$$H_1 : \beta_j \neq 0$$

Statistik uji yang digunakan yaitu:

$$t_{hitung} = \frac{\hat{\beta}_j}{Se(\hat{\beta}_j)} \quad (29)$$

dan nilai standar galat dari  $\hat{\beta}_j$ , yang didefinisikan sebagai:

$$Se(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)} \quad (30)$$



Adapun kriteria pengambilan keputusan uji t parsial yaitu jika  $t_{hitung} > t_{tabel}$  atau  $p - value < 0,05$  maka  $H_0$  ditolak yang artinya variabel bebas secara parsial berpengaruh terhadap variabel terikat (Atamia dkk., 2021).

## B. Uji Simultan

Uji simultan atau uji F bertujuan untuk mengetahui apakah parameter model regresi yang terbentuk sudah signifikan atau belum secara serentak. Adapun hipotesis untuk uji simultan yaitu sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \text{Minimal ada } \beta_j \neq 0 ; j = 1, 2, \dots, k$$

Statistik uji yang digunakan:

$$F_{hitung} = \frac{KTR}{KTG} \quad (31)$$

Adapun kriteria pengambilan keputusan dari uji simultan yaitu jika  $F_{hitung} > F_{tabel}$  atau  $p - value < 0,05$  maka  $H_0$  ditolak yang artinya secara bersama-sama variabel bebas berpengaruh signifikan terhadap variabel terikat (Atamia dkk., 2021).

### 1.5.9. Pemilihan Model Terbaik

Sebuah metode statistik yang dapat digunakan untuk memperoleh kriteria model terbaik adalah statistik Cp. Dengan kriteria statistik adalah nilai minimum Cp dianggap sebagai model terbaik. Kriteria lain berdasarkan statistik Cp *Mallows* adalah model dengan nilai Cp paling dekat dengan jumlah variabel dalam model. Nilai Cp *Mallows* dapat dihitung dengan persamaan (32) berikut:

$$C_p = \frac{JKG_p}{s^2} - (n - 2p) \quad (32)$$

dengan  $p$  adalah jumlah variabel dalam model regresi,  $JKG$  adalah jumlah kuadrat galat untuk setiap model, dan  $s^2$  merupakan kuadrat rata-rata galat dari semua variabel (Saputra & Radam, 2023).

### 1.5.10. Indeks Pembangunan Manusia (IPM)

Indeks Pembangunan Manusia digunakan untuk mengukur sejauh mana tingkat keberhasilan kualitas hidup manusia. IPM juga berfungsi untuk mengukur hasil pembangunan manusia berdasarkan sejumlah komponen dasar kualitas hidup (Ningrum dkk., 2020). Adapun cara yang dapat digunakan untuk menentukan nilai IPM adalah sebagai berikut:

$$IPM = \frac{D_{UPHS} + D_{Pengetahuan} + D_{SHL}}{3} \quad (33)$$



IPM mencoba memberikan gambaran yang lebih lengkap tentang kemajuan suatu masyarakat dengan memperhitungkan aspek-aspek kesehatan, pendidikan dan standar hidup. IPM dibangun melalui pendekatan tiga dimensi dasar di mana setiap dimensi diwakili oleh suatu indikator.

1. Umur Panjang dan Hidup Sehat. Dinyatakan dengan harapan hidup yang mengukur rata-rata tahun yang diharapkan untuk hidup pada saat lahir. Ini mencerminkan kualitas dan panjang umur penduduk. Berikut cara untuk menghitung nilai dimensi umur panjang dan hidup sehat:

$$D_{UPHS} = \frac{AHH - AHH_{min}}{AHH_{max} - AHH_{min}} \quad (34)$$

dengan:

$D_{UPHS}$  : Dimensi Umur Panjang dan Hidup Sehat

$AHH$  : Angka Harapan Hidup

2. Pengetahuan. Diukur dengan dua indikator utama, yaitu harapan lama sekolah (rata-rata tahun sekolah yang diharapkan) dan rata-rata tahun sekolah (rata-rata tahun yang telah ditempuh oleh penduduk berusia 25 tahun ke atas). Cara menghitung nilai dimensi pengetahuan ditunjukkan pada persamaan (35).

$$D_{Pengetahuan} = \frac{I_{HLS} + I_{RLS}}{2} \quad (35)$$

dengan:

$D_{Pengetahuan}$  : Dimensi Pengetahuan

$I_{HLS}$  : Indikator Harapan Lama Sekolah

$I_{RLS}$  : Indikator Rata-rata Lama Sekolah

3. Standar Hidup Layak. Diukur dengan pendapatan riil per kapita, yang mencerminkan daya beli penduduk. Berikut cara untuk menghitung nilai dimensi standar hidup layak:

$$D_{SHL} = \frac{\ln(Pengeluaran) - \ln(pengeluaran_{min})}{\ln(Pengeluaran_{max}) - \ln(Pengeluaran_{min})} \quad (36)$$

dengan:

$D_{SHL}$  : Dimensi Standar Hidup Layak

$Pengeluaran$  : Pengeluaran per Kapita disesuaikan

IPM memberikan gambaran lebih lengkap tentang keadaan kesejahteraan masyarakat dibandingkan dengan indikator ekonomi tunggal seperti Produk Domestik Bruto (PDB). Oleh karena itu, IPM sering digunakan sebagai alat untuk memantau proses pembangunan manusia di tingkat nasional dan internasional, juga sebagai dasar perencanaan kebijakan pembangunan yang lebih berorientasi pada manusia (Ummah, 2024). Adapun faktor-faktor yang mungkin mempengaruhi IPM adalah sebagai berikut.

### Standar Hidup

Menurut Statistik (2022), Angka Harapan Hidup merupakan perkiraan rata-rata tahun yang akan dijalani oleh seseorang selama hidup. Indikator ini digunakan untuk mengevaluasi tingkat perkembangan ekonomi suatu negara.



Negara dengan angka harapan hidup yang tinggi biasanya termasuk dalam kelompok negara maju yang memiliki perekonomian yang besar.

### **B. Harapan Lama Sekolah**

Menurut Badan Pusat Statistik (2022), Harapan Lama Sekolah didefinisikan sebagai jumlah tahun pendidikan yang diharapkan akan dirasakan oleh anak-anak pada usia tertentu di masa depan. Indikator ini dapat digunakan untuk melihat kondisi pembangunan pada sistem pendidikan di berbagai jenjang yang ditunjukkan dalam bentuk lamanya pendidikan yang diharapkan dapat dicapai. Jika harapan lama sekolah tinggi, berarti anak-anak di wilayah tersebut diperkirakan akan mendapatkan pendidikan yang lebih lama dan lebih baik.

### **C. Rata-rata Lama Sekolah**

Menurut Badan Pusat Statistik (2022), Rata-rata Lama Sekolah menggambarkan jumlah tahun yang digunakan oleh masyarakat usia 25 tahun ke atas dalam menyelesaikan pendidikan formal. Rata-rata lama sekolah dapat digunakan untuk mengetahui kualitas pendidikan masyarakat dalam suatu wilayah. Jika rata-rata lama sekolah tinggi, berarti lebih banyak orang telah mencapai tingkat pendidikan yang lebih tinggi. Semakin tinggi rata-rata lama sekolah, maka semakin baik tingkat pendidikan yang dicapai oleh masyarakat. Hal ini memiliki dampak positif terhadap kesejahteraan ekonomi negara dalam jangka panjang.

### **D. Pengeluaran per Kapita**

Pengeluaran per Kapita biasa juga disebut dengan daya beli adalah kemampuan masyarakat dalam membelanjakan uang, baik itu dalam bentuk barang ataupun jasa. Hal yang diharapkan dari perhitungan daya beli ini adalah untuk melihat gambaran tingkat kesejahteraan masyarakat sebagai efek dari membaiknya perekonomian suatu negara (Kharazishvili dkk., 2020).

### **E. Angka Melek Huruf**

Menurut Kemenkes RI (2023), Angka Melek Huruf merupakan persentase penduduk berusia 15 tahun ke atas yang dapat membaca, menulis dan memahami kalimat sederhana dalam kehidupan sehari-hari. Indikator ini menunjukkan seberapa banyak penduduk di suatu wilayah yang memiliki kemampuan dasar untuk memperluas akses informasi, meningkatkan pengetahuan dan keterampilan sehingga dapat meningkatkan kualitas hidup diri, keluarga maupun negara. Menurut UNDP, semakin tinggi angka melek huruf suatu negara maka semakin baik kualitas Sumber Daya Manusia negara tersebut (Mursyidah dkk., 2022).



tingkat pendidikan, keterampilan yang dibutuhkan, tingkat risiko pekerjaan dan faktor lainnya (Nurfiat & Rustariyuni, 2018).

### **G. Angka Partisipasi Sekolah**

Angka Partisipasi Sekolah merupakan perbandingan jumlah anak pada kelompok usia dan jenjang pendidikan tertentu yang sedang menempuh pendidikan formal. Indikator ini digunakan untuk melihat seberapa banyak anak-anak di suatu wilayah yang berpartisipasi dalam pendidikan. Angka Partisipasi sekolah juga dapat menjadi tolok ukur dalam menentukan pemerataan pendidikan. Semakin tinggi nilai angka partisipasi sekolah menunjukkan semakin besar partisipasi masyarakat dalam pendidikan pada usia sekolah tertentu.

### **H. Angka Partisipasi Murni**

Menurut Kemenkes RI (2023), Angka Partisipasi Murni merupakan perbandingan jumlah penduduk kelompok usia tertentu yang sedang menempuh pendidikan tepat pada tingkat pendidikan yang sesuai dengan usia mereka. Indikator ini menunjukkan seberapa tepat penduduk pada usia tertentu berpartisipasi dalam pendidikan yang sesuai dengan usia mereka. Angka partisipasi murni bertujuan untuk mengukur ketepatan usia penduduk dalam berpartisipasi untuk menempuh suatu jenjang pendidikan tertentu.

