

**IMPLEMENTASI *ROBUST* PC DENGAN *WILD BOOTSTRAP* DALAM
MENGATASI *OUTLIER*, MULTIKOLINEARITAS, DAN
HETEROSKEDASTISITAS PADA DATA
HEPATITIS B KRONIS**

(Skripsi)

Oleh

JOSEPH DEDEK PARHUSIP



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026**

ABSTRACT

Robust Principal Component Regression with Wild Bootstrap for Handling Outliers, Multicollinearity, and Heteroskedasticity in Chronic Hepatitis B Data

By

Joseph Dedek Parhusip

This study aims to analyze the performance of the Robust Principal Component (RPC) method combined with Wild Bootstrap in handling outliers, multicollinearity, and heteroskedasticity in chronic hepatitis B data. The data were obtained from the World Health Organization (WHO) and consist of several epidemiological indicators. The analysis methods include Principal Component Analysis (PCA), Least Trimmed Squares (LTS), and Wild Bootstrap using Wu and Liu multipliers. The results show that the dataset contains outliers, strong multicollinearity, and heteroskedasticity. The RPC-Wild Bootstrap method produces more stable parameter estimates, with RPC Boot Wu showing lower standard error and RMSE compared to RPC Boot Liu. Therefore, the RPC-Wild Bootstrap method is effective in producing more stable and reliable parameter estimates for complex real-world data.

Keywords: Robust Principal Component, Wild Bootstrap, Outliers, Multicollinearity, Heteroskedasticity.

ABSTRAK

IMPLEMENTASI *ROBUST PC* DENGAN *WILD BOOTSTRAP* DALAM MENGATASI *OUTLIER*, MULTIKOLINEARITAS, DAN HETEROSKEDASTISITAS PADA DATA HEPATITIS B KRONIS

Oleh

Joseph Dedek Parhusip

Penelitian ini bertujuan untuk menganalisis kinerja metode *Robust Principal Component (RPC)* yang dikombinasikan dengan *Wild Bootstrap* dalam menangani pencilan (*outlier*), multikolinearitas, dan heteroskedastisitas pada data hepatitis B kronis. Data diperoleh dari *World Health Organization (WHO)* dan terdiri atas beberapa indikator epidemiologis. Metode analisis yang digunakan meliputi *Principal Component Analysis (PCA)*, *Least Trimmed Squares (LTS)*, dan *Wild Bootstrap* dengan pengali Wu dan Liu. Hasil penelitian menunjukkan bahwa dataset mengandung pencilan, multikolinearitas yang kuat, dan heteroskedastisitas. Metode *RPC Wild Bootstrap* menghasilkan estimasi parameter yang lebih stabil, dengan *RPC Boot Wu* menunjukkan nilai *standard error* dan *RMSE* yang lebih rendah dibandingkan *RPC Boot Liu*. Oleh karena itu, metode *RPC Wild Bootstrap* efektif dalam menghasilkan estimasi parameter yang lebih stabil dan andal untuk data dunia nyata yang kompleks.

Kata kunci : *Robust Principal Component, Wild Bootstrap, Pencilan (Outlier), Multikolinearitas, Heterokedastisitas.*

**IMPLEMENTASI *ROBUST* PC DENGAN *WILD BOOTSTRAP* DALAM
MENGATASI *OUTLIER*, MULTIKOLINEARITAS, DAN
HETEROSKEDASTISITAS PADA DATA
HEPATITIS B KRONIS**

Oleh
Joseph Dedek Parhusip

(Skripsi)

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

**Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026**

Judul Skripsi : **IMPLEMENTASI ROBUST PC DENGAN WILD BOOTSTRAP DALAM MENGATASI OUTLIER, MULTIKOLINEARITAS, DAN HETEROSKEDASTISITAS PADA DATA HEPATITIS B KRONIS**

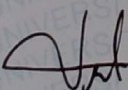
Nama Mahasiswa : **Joseph Dedek Parhusip**

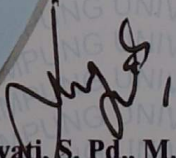
NPM : **2217031148**

Program Studi : **Matematika**

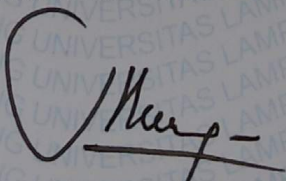
Fakultas : **Matematika dan Ilmu Pengetahuan Alam**




Drs. Nusyirwan, M. Si.
NIP.196610101992031028


Misgivati, S. Pd., M. Si.
NIP. 198509282023212032

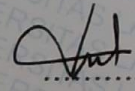
2. Ketua Jurusan Matematika


Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 197403162005011001

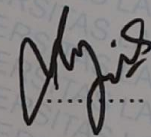
MENGESAHKAN

1. **Tim Penguji**

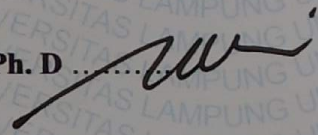
Ketua Penguji : Drs. Nusyirwan, M. Si



Sekretaris : Misgiyati, S. Pd., M. Si.



**Penguji
Bukan Pembimbing : Prof. Ir. Netti Herawati, M. Sc., Ph. D**



2. **Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam**



Dr. Ing. Heri Satria, S.Si., M.Si.

10012005011002

Tanggal Lulus Ujian Skripsi: 3 Juni 2026

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama : **Joseph Dedek Parhusip**

NPM : **2217031148**

Program Studi : **Matematika**

Judul Skripsi : **Implementasi *Robust* PC Dengan *Wild Bootstrap* dalam Mengatasi *Outlier*, Multikolinearitas, dan Heteroskedastisitas Pada Data Hepatitis B Kronis**

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

Bandar Lampung, 3 juni 2026

Penulis,



Joseph Dedek Parhusip

NPM. 2217031148

RIWAYAT HIDUP

Penulis dilahirkan di Bandar Lampung pada tanggal 24 Januari 2004. Penulis merupakan anak ke-6 dari enam bersaudara oleh pasangan Bapak Antonius Nainggolan dan Ibu Berliana Gultom.

Penulis melaksanakan pendidikan dasar di SDN 5 Sumberrejo pada tahun 2009, menyelesaikan pendidikan menengah pertama di SMP Negeri 26 Bandar Lampung pada tahun 2018, dan pendidikan menengah atas di SMA Negeri 14 Bandar Lampung pada tahun 2021.

Pada tahun 2022, penulis diterima sebagai mahasiswa Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung melalui jalur Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN). Selama menempuh pendidikan di perguruan tinggi, penulis aktif mengikuti kegiatan akademik dan organisasi kemahasiswaan, antara lain sebagai Anggota Bidang Kaderisasi dan kepemimpinan pada Himpunan Mahasiswa Matematika (HIMATIKA) FMIPA Unila tahun 2023 dan lanjut sebagai Ketua Bidang Kaderisasi dan kepemimpinan tahun 2024.

Sebagai bagian dari pemenuhan kurikulum, penulis telah melaksanakan Praktik Kerja Lapangan (PKL) di UPTD KPHK Tahura Wan Abdul Rachman pada Desember 2024 - Januari 2025 dengan judul. Selain itu, penulis juga melaksanakan Kuliah Kerja Nyata (KKN) di Kelurahan Way Lunik, Panjang, Kota Bandar Lampung pada bulan Juli - Agustus 2025 sebagai bentuk pengabdian kepada masyarakat.

MOTTO

"Dan apa saja yang kamu minta dalam doa dengan penuh kepercayaan, kamu akan menerimanya."

(Matius 21:22)

PERSEMBAHAN

Puji syukur penulis panjatkan ke hadirat Tuhan Yesus Kristus atas segala berkat, bimbingan, dan kekuatan-Nya sehingga penulis dapat menyelesaikan skripsi ini.

Saya persembahkan karya ini untuk...

Kedua orang tua tercinta, terimakasih bapak dan mamak selalu memberikan kasih sayang, dukungan, serta doa yang tiada henti. Skripsi ini saya persembahkan sebagai rasa syukur dan terimakasih atas pengorbanan yang telah diberikan kepada penulis.

Abang dan kakak yang kusayangi, Abang Yohanes dan kak Widi, Kak Tina, Abang Joshua dan Kak Bea, Abang Jonathan dan kak Rina, serta Abang Andre. Terimakasih atas segala dukungan dan kasih sayang yang kalian berikan kepada penulis.

Para dosen dan pembimbing, yang dengan sabar membimbing, mengarahkan, serta memberikan ilmu, arahan, dan kritik yang membangun. Bimbingan dan pengalaman yang diberikan menjadi bekal berharga bagi penulis, tidak hanya dalam menyelesaikan skripsi, tetapi juga dalam menapaki langkah hidup ke depan.

Tak lupa, penulis persembahkan karya ini kepada teman-teman, sahabat, dan almamater yang telah memberi dukungan moral, motivasi, dan semangat selama perjalanan kuliah.

SANWACANA

Puji dan syukur penulis panjatkan kepada Tuhan Yesus Kristus atas kasih, anugerah, dan penyertaan-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Implementasi *Robust PC* Dengan *Wild Bootstrap* Dalam Mengatasi *Outlier*, Multikolinearitas, Dan Heteroskedastisitas Pada Data Hepatitis B Kronis.”

Dalam proses penyusunan skripsi ini, banyak pihak yang telah membantu memberikan bimbingan, dukungan, arahan, motivasi serta saran sehingga skripsi ini dapat terselesaikan. Oleh karena itu, dalam kesempatan ini penulis mengucapkan terimakasih kepada:

1. Bapak Drs. Nusyirwan, M. Si., selaku Pembimbing 1 yang telah banyak meluangkan waktunya untuk memberikan arahan, bimbingan, motivasi, saran serta dukungan kepada penulis sehingga dapat menyelesaikan skripsi ini.
2. Ibu Misgiyati, S. Pd., M. Si., selaku Dosen Pembimbing II yang telah memberikan saran, masukan, dan koreksi yang sangat berarti selama proses penyusunan skripsi.
3. Prof. Ir. Netti Herawati, M. Sc., Ph. D., selaku Dosen Pembahas yang telah memberikan kritik dan saran yang membangun demi kesempurnaan skripsi ini.
4. Seluruh dosen, staf, dan karyawan Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung atas bantuan dan pelayanan yang diberikan selama penulis menempuh pendidikan

5. Kedua orang tua tercinta, Bapak Antonius Nainggolan dan Ibu Berliana Gultom, yang senantiasa memberikan kasih sayang, doa yang tiada henti, nasihat, perhatian, dukungan moral maupun materil, serta semangat dan motivasi kepada penulis dalam setiap langkah kehidupan.
6. Abang dan kakak tercinta, Abang Yohanes dan kak Widi, Kak Tina, Abang Joshua dan Kak Bea, Abang Jonathan dan kak Rina, serta Abang Andre beserta keluarga besar, yang selalu memberikan doa, dukungan, perhatian, nasihat, motivasi, serta semangat kepada penulis selama menjalani perkuliahan hingga penyusunan skripsi ini.
7. Kepada Vania Yosinka Putri S.Si., atas doa, perhatian, dan dukungan kepada saya selama proses penyusunan skripsi ini.
8. Pimpinan Himatika 2024 yang selalu memberikan dukungan, doa, serta menemani penulis selama masa perkuliahan baik secara langsung ataupun tidak langsung.
9. Kaderisasi dan Kepemimpinan 2024 yang memberikan semangat kepada penulis, selama proses penyusunan skripsi.
10. Margo, Udin, Rafael, Alfredo, dan Fajar, terima kasih atas saran, masukan, dan kebersamaan yang telah diberikan kepada penulis sejak awal masuk di Jurusan Matematika.
11. Penulis mengucapkan terima kasih kepada Arifin, Insan, Zul, dan Djoe atas kebersamaan serta dukungan yang telah diberikan sejak awal proses penulisan skripsi hingga saat ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna dan masih banyak kekurangan, sehingga penulis mengharapkan kritik dan saran dari kalian. Semoga skripsi ini dapat berguna dan bermanfaat bagi kita semua. Amin.

Bandar Lampung, 3 juni 2026

Penulis

Joseph Dedek Parhusip

DAFTAR ISI

	Halaman
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
I. PENDAHULUAN	2
1.1 Latar Belakang dan Masalah	2
1.2 Tujuan Penelitian.....	2
1.3 Manfaat Penelitian.....	2
II. TINJAUAN PUSTAKA	5
2.1 Matriks.....	5
2.1.1 Nilai Eigen dan Vektor Eigen	5
2.2 Analisis Regresi Linear	5
2.3 Uji Asumsi Analisis Regresi	6
2.4 <i>Outlier</i>	7
2.5 Multikolinearitas.....	8
2.6 Heteroskedastisitas	9
2.7 Matriks Kovarian.....	10
2.8 <i>Ordinary Least Squares</i>	11
2.9 Analisis Komponen Utama (AKU/PCA)	12
2.10 <i>Principal Component Regression</i> (PCR).....	15
2.11 <i>Least Trimmed Squares</i>	16
2.12 <i>Wild bootstrap</i>	19
2.13 <i>Robust Principal Component</i> (RPC)	20
2.14 <i>Robust Principal Component with Wild Bootstrap</i>	20
2.14.1 Alur RPC dengan LTS, dan <i>Wild Bootstrap</i>	21
2.14.2 Residual dan Skala <i>Robust LTS</i>	22
2.14.3 Bobot <i>Tukey Bisquare</i>	24
2.14.4 Residual <i>Wild Bootstrap</i>	25
2.14.5 Evaluasi Metode <i>Bootstrap</i>	27
2.15 Hepatitis B	29

III. METODOLOGI PENELITIAN	32
3.1 Tempat dan Waktu Penelitian	32
3.2 Data Penelitian.....	32
3.3 Metode Penelitian.....	32
IV. HASIL DAN PEMBAHASAN.....	35
4.1 Deskriptif Data	35
4.2 Pemeriksaan Asumsi	36
4.2.1 Deteksi <i>Outlier</i>	36
4.2.2 Uji Multikolinearitas	38
4.2.3 Uji Heterokedastisitas	39
4.3 Pembentukan Komponen Utama.....	40
4.4 Pemodelan Regresi <i>Robust</i> LTS.....	41
4.5 Pembobotan <i>Robust</i> (MAD & <i>Tukey</i>)	42
4.6 <i>Robust Principal Component with Wild Bootstrap</i>	45
V. KESIMPULAN DAN SARAN	47
5.1 Kesimpulan.....	47
5.2 Saran.....	47
DAFTAR PUSTAKA.....	49
LAMPIRAN.....	52

DAFTAR TABEL

Tabel	Halaman
1. Statistik Deskriptif	35
2. Negara-negara yang terdeteksi sebagai <i>outlier</i>	37
3. Nilai VIF	38
4. Hasil Uji <i>Glejser</i>	39
5. Hasil Nilai Eigen.....	40
6. Estimasi Parameter Regresi <i>Robust</i>	41
7. Pembobotan Residual <i>Robust</i>	43
8. Residual Berbobot LTS.....	44
9. Hasil Estimasi Parameter <i>Robust Wild Bootstrap</i>	45

DAFTAR GAMBAR

Gambar	Halaman
Gambar 1. Boxplot <i>outlier</i>	37

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Robust Principal Component (Robust PC) merupakan pengembangan dari *Principal Component Analysis (PCA)* yang dirancang untuk lebih tahan terhadap pengaruh *outlier*, sehingga mampu memberikan representasi variabilitas data yang lebih stabil ketika data mengandung nilai-nilai ekstrim atau observasi menyimpang (Hubert, *et al.*, 2005). *Robust PC* berfungsi untuk mengurangi dampak *outlier* dan multikolinearitas dengan mentransformasi variabel menggunakan *Principal Component Analysis*. Sehingga proses ini mempertahankan informasi dalam data namun mengurangi pengaruh titik-titik ekstrim yang dapat merusak model regresi dapat diminimalisir secara signifikan (Hubert, *et al.*, 2005).

Wild Bootstrap adalah teknik resampling yang secara khusus dikembangkan untuk menangani kondisi heteroskedastisitas dan *error* yang tidak normal sehingga berbeda dengan *bootstrap* klasik yang mengandalkan pengambilan sampel ulang residual secara langsung (Efron, 1979). *Wild bootstrap* menghasilkan sampel-sampel baru dengan memodifikasi residual menggunakan distribusi pembobot tertentu sehingga struktur varians yang tidak homogen dapat dipertahankan (Wu, 1986). Oleh karena itu, kombinasi kedua metode ini banyak memberikan analisis yang lebih stabil dan mampu menghasilkan hasil statistik yang lebih andal ketika data dunia tidak memenuhi asumsi dasar regresi klasik, sehingga meningkatkan hasil analisis statistik secara menyeluruh (Davidson & Flachaire, 2008).

Data hepatitis B kronis memiliki karakteristik yang kompleks, seperti cakupan imunisasi Hep B3, angka infeksi baru hepatitis B, jumlah orang yang hidup dengan hepatitis kronis, serta cakupan Hep B *Birth dose* (BD). Seringkali data tersebut memiliki variasi yang sangat besar antar negara, variasi ini terjadi karena perbedaan akses layanan kesehatan, kebijakan imunisasi, kondisi ekonomi masyarakat, serta kualitas program pencegahan (WHO, 2023). Data semacam ini rentan mengandung *outlier*, karena setiap negara memiliki masalah tersendiri misalnya dengan wabah lokal yang menyebabkan lonjakan kasus secara drastis, atau negara dengan program imunisasi yang hampir sempurna sehingga nilai kasus tertentu cenderung rendah.

Multikolinearitas juga sangat mungkin terjadi karena faktor-faktor tersebut saling berkaitan, misalnya rendahnya cakupan imunisasi berkaitan dengan infeksi baru dan meningkatnya jumlah orang yang hidup dengan hepatitis B kronis. Keberagaman varian antar negara menyebabkan heteroskedastisitas yang kuat. Oleh karena itu, metode seperti RPC Boot Wu dan RPC Boot Liu diperlukan untuk menghasilkan estimasi yang lebih stabil dan tahan terhadap *outlier* serta heteroskedastisitas (Wu, 1986).

Sebelumnya, penelitian oleh Rasheed, Adnan, Saffari, dan Atiyaye (2022) menunjukkan bahwa penggabungan *Robust Principal Component Regression* dengan prosedur *Wild Bootstrap* menghasilkan dua metode baru, yaitu *RPC Boot Wu* dan *RPC Boot Liu*. Hasil penelitian tersebut menunjukkan bahwa pendekatan metode ini mampu meningkatkan performa estimasi regresi secara signifikan terutama menangani data dengan *outlier*, multikolinearitas, dan heteroskedastisitas secara bersamaan, sehingga sangat direkomendasikan untuk analisis regresi pada data dunia nyata yang kompleks dan tidak ideal.

Meskipun metode *RPC Boot* telah terbukti unggul secara teoritis dalam menghasilkan estimasi yang lebih stabil pada data yang mengandung *outlier*, multikolinearitas, dan heteroskedastisitas (Hubert, *et al.*, 2005). Namun penerapannya pada data hepatitis B kronis masih sangat terbatas. Selain itu belum

banyak studi yang memanfaatkan kedua *wild bootstrap* yaitu (Wu dan Liu) secara langsung menganalisis dinamika hepatitis B. Padahal data hepatitis B cenderung memiliki heteroskedastisitas, mengandung *outlier*, serta memiliki korelasi tinggi antar variabel. Dengan demikian menerapkan RPC Boot Wu dan RPC Boot Liu pada data hepatitis B kronis untuk memperoleh estimasi yang lebih baik dan lebih akurat.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut :

1. Menganalisis kinerja dari metode *Robust PC* dan *Wild Bootstrap* untuk mengatasi masalah *outlier*, multikolinearitas, dan heteroskedastisitas.
2. Membandingkan hasil estimasi parameter dari dua metode dalam konteks studi kasus populasi terinfeksi hepatitis B kronis.
3. Menemukan model yang paling tepat digunakan dalam situasi data yang tidak memenuhi asumsi regresi linear.

1.3 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan beberapa manfaat, yaitu:

1. Menyediakan pendekatan analisis statistik yang lebih handal melalui penerapan metode *Robust Principal Component (PC)* dan *Wild Bootstrap* dalam studi kasus epidemiologi.
2. Menghasilkan estimasi model yang lebih stabil dan minim bias karena metode yang digunakan mampu mengatasi permasalahan *outlier*, multikolinearitas, dan heteroskedastisitas .
3. Memberikan pemahaman penulis dan pembaca mengenai analisis faktor-faktor yang memengaruhi kasus hepatitis B kronis di Asia.

II. TINJAUAN PUSTAKA

2.1 Matriks

Matriks adalah susunan bilangan yang terdiri atas baris dan kolom. Sebuah matriks berukuran $n \times m$ memiliki n baris dan m kolom, dan setiap elemennya biasanya dituliskan sebagai a_{ij} , yaitu elemen pada baris ke- i dan kolom ke- j .

Bentuk umum sebuah matriks dapat dituliskan sebagai barisan nilai yang diatur dalam n baris dan m kolom. Matriks \mathbf{A} dengan $n \times m$ dapat ditulis:

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}$$

Kita juga bisa menulis matriks \mathbf{A} di atas sebagai berikut:

$$\mathbf{A} = (a_{ij}) \tag{2.1}$$

Notasi $\mathbf{A} = (a_{ij})$ dipakai untuk menyatakan bahwa seluruh elemen matriks berada pada posisi indeks baris i dan kolom j (Anton, 1987).

2.1.1 Nilai Eigen dan Vektor Eigen

Nilai eigen dan vektor eigen merupakan konsep dasar yang membantu memahami pola utama yang terdapat dalam suatu data. Dalam analisis multivariat seperti *Principal Component Analysis* (PCA), nilai eigen digunakan untuk menunjukkan seberapa besar variasi data yang dapat dijelaskan oleh setiap komponen, sedangkan vektor eigen menggambarkan arah pola utama yang terbentuk dari data tersebut.

Misalkan A adalah matriks berordo $n \times n$, vektor $x \in R^n$, dan $x \neq 0$, disebut vektor eigen, jika terdapat bilangan riil λ yang disebut nilai eigen, sehingga memenuhi persamaan:

$$Ax = \lambda x \quad (2.2).$$

Dari pernyataan di atas, dapat diketahui persyaratan-persyaratan untuk nilai eigen maupun vektor eigen. Nilai eigen λ merupakan bilangan riil, yang berarti dapat bernilai nol, negatif dan juga positif. Sedangkan vektor eigen x merupakan anggota dari R^n untuk A dan x bukan vektor nol (Anton, 1987).

2.2 Analisis Regresi Linear

Analisis regresi merupakan suatu kajian yang bertujuan untuk melihat hubungan antara satu variabel yang diterangkan (*the explained variable*) dan satu atau lebih variabel yang menerangkan (*the explanatory variable*). Jika variabel bebas yang digunakan hanya satu, maka metode tersebut disebut sebagai regresi sederhana, sedangkan apabila jumlah variabel bebas lebih dari satu, maka analisis regresinya disebut regresi linear berganda. Analisis regresi banyak digunakan dalam penelitian ilmiah untuk menilai pengaruh dan prediksi variabel bebas terhadap

variabel terikat, serta menentukan keberhasilan atau tidaknya hipotesis penelitian (Sembiring, 2003).

Proses analisis regresi melibatkan beragam uji statistika, antara lain uji signifikansi koefisien (uji-t), uji simultan (uji-F), analisis varians (ANOVA), serta pengujian hipotesis. Hasil utama dari analisis regresi berupa suatu persamaan regresi yang digunakan untuk memprediksi variabel yang dipengaruhi oleh variabel lainnya. Regresi merupakan suatu alat ukur yang digunakan untuk mengukur ada tidaknya korelasi antar variabel. Secara umum, bentuk persamaan model regresi linear berganda dapat dinyatakan sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i \quad (2.3)$$

di mana:

- Y_i : variabel dependen,
- $\beta_0, \beta_1, \beta_2, \dots, \beta_k$: koefisien regresi,
- $X_{1i}, X_{2i}, \dots, X_{ki}$: variabel independent,
- ε_i : residual.

2.3 Uji Asumsi Analisis Regresi

Uji asumsi dalam analisis regresi dilakukan untuk bertujuan untuk memastikan bahwa model regresi yang diperoleh memenuhi kriteria statistik sehingga hasil estimasi menjadi tidak bias, stabil, dan dapat digunakan untuk pengambilan keputusan ilmiah. Beberapa penelitian menegaskan bahwa estimasi regresi harus melalui uji asumsi klasik agar model yang dihasilkan valid dan dapat dipercaya. Pelanggaran asumsi menyebabkan parameter regresi tidak efisien dan data pengujian tidak akurat (Iheaka, 2025).

Uji asumsi klasik meliputi uji linearitas, normalitas, homoskedastisitas, autokorelasi, dan multikolinearitas. Model regresi dikatakan baik apabila sebuah asumsi terpenuhi sehingga regresi dapat menghasilkan *Best Linear Unbiased Estimator (BLUE)*. Menurut Chatterjee & Hadi (2006), untuk melakukan analisis regresi linear harus memenuhi asumsi-asumsi berikut:

2.4 *Outlier*

Secara umum, *outlier* dapat diartikan data yang tidak mengikuti pola umum pada model atau data yang keluar dari model dan tidak berada dalam daerah selang kepercayaan (Sembiring, 2003). Menurut Weissberg (1985), jika terdapat masalah yang berkaitan dengan *outlier* maka diperlukan alat diagnosis yang dapat mengidentifikasi masalah *outlier*. Salah satunya dengan menyisihkan *outlier* dari kelompok data kemudian menganalisis data tanpa *outlier*. Keberadaan data *outlier* akan mengganggu dalam proses analisis data dan harus dihindari dalam banyak hal. Dalam kaitannya dengan analisis regresi, *outlier* dapat menyebabkan hal-hal berikut :

1. Residual yang besar dari model yang terbentuk.
2. Varians pada data tersebut menjadi lebih besar.
3. Taksiran interval memiliki rentang yang besar.

Dalam penelitian ini, deteksi *outlier* dilakukan menggunakan boxplot. Boxplot merupakan metode visual yang digunakan untuk menggambarkan penyebaran data serta mengidentifikasi adanya nilai yang menyimpang dari distribusi data.

Selain menggunakan boxplot, deteksi observasi yang memiliki pengaruh besar terhadap model regresi juga dilakukan menggunakan metode *Difference in Fits (DFFITS)*. Metode ini digunakan untuk mengidentifikasi observasi yang memiliki pengaruh signifikan terhadap nilai prediksi dalam model regresi. Nilai DFFITS dihitung menggunakan rumus sebagai berikut:

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{i(i)}}{s_i \sqrt{h_{ii}}} \quad (2.4)$$

dimana:

- \hat{y}_i : nilai prediksi dengan seluruh data,
- $\hat{y}_{i(i)}$: nilai prediksi tanpa observasi ke-i,
- $s_{(i)}$: standar *error* tanpa observasi ke-i,
- $\sqrt{h_{ii}}$: leverage observasi ke-i.

Suatu observasi dikatakan memiliki pengaruh terhadap model regresi apabila memenuhi kriteria berikut:

$$|DFFITS| > 2 \sqrt{\frac{k}{n}} \quad (2.5)$$

dengan k merupakan jumlah parameter dalam model regresi dan n merupakan jumlah observasi, Jika nilai absolut DFFITS lebih besar dari batas tersebut, maka observasi tersebut dikategorikan sebagai observasi berpengaruh (*influential observation*) terhadap model regresi.

2.5 Multikolinearitas

Multikolinearitas adalah kondisi ketika terdapat hubungan linear antara beberapa atau seluruh variabel bebas dalam model regresi. Masalah ini hanya muncul pada regresi linear berganda. Model regresi linear berganda yang baik adalah model yang bebas dari multikolinearitas. Suatu model dikatakan bebas dari multikolinearitas apabila memiliki nilai *Variance Inflation Factor* (VIF) < 10 . Sebaliknya nilai VIF > 10 mengindikasikan adanya multikolinearitas dalam model (Myers, 1990).

Menurut Montgomery & Runger (2011), multikolinearitas dapat dideteksi menggunakan nilai *Variance Inflation Factor* (VIF). Nilai VIF dapat dicari menggunakan rumus sebagai berikut:

$$VIF_{(j)} = \frac{1}{(1 - R_j^2)} \quad ; \quad j = 1, 2, \dots, k \quad (2.6)$$

dengan R_j^2 merupakan koefisien determinasi yang didapat dari variabel bebas X_j yang diregresikan dengan variabel bebas lainnya pada model dugaan regresi.

2.6 Heteroskedastisitas

Menurut Ima (2025) heteroskedastisitas adalah kondisi dalam model regresi di mana varians kesalahan (*error*) tidak konstan pada seluruh rentang nilai variabel independen. Dalam model regresi linear klasik, salah satu asumsi dasar adalah bahwa varians *error* bersifat homogen (homoskedastik), yang berarti kesalahan model memiliki distribusi dengan varians yang sama pada seluruh nilai prediktor. Heteroskedastisitas bisa menjadi masalah yang penting karena dapat memengaruhi keakuratan estimasi koefisien dan mengurangi kemampuan model dalam memprediksi. Heteroskedastisitas dapat disebabkan oleh berbagai faktor. Beberapa penyebab umum heteroskedastisitas dalam model regresi antara lain:

1. Variabel yang tidak disertakan dalam model.
2. Penyimpangan dari linearitas.
3. Skala variabel dependen.
4. Efek sub kelompok atau kategori.
5. Kesalahan pengukuran.

Untuk mengetahui keberadaan heteroskedastisitas dalam model regresi dapat dilakukan dengan menggunakan uji *Glejser*. Pengujian ini dilakukan dengan

meregresikan nilai absolut residual terhadap variabel independen dalam model regresi.

Hipotesis yang digunakan adalah sebagai berikut:

a) Hipotesis

H_0 : Tidak terdapat heteroskedastisitas dalam model regresi

H_1 : Terdapat heteroskedastisitas dalam model regresi

b) Taraf signifikansi: $\alpha = 5\% = 0.05$

c) Statistik uji:

$$\text{Nilai absolut residual} = |e_i|$$

di mana:

$|e_i|$: nilai absolut residual ke-i

e_i : residual ke-i dari model regresi

d) Kriteria keputusan:

Tolak H_0 jika terdapat paling sedikit satu variabel independen yang memiliki nilai ($p\text{-value}$) $< \alpha$

Tidak tolak jika seluruh variabel independen memiliki nilai ($p\text{-value}$) $\geq \alpha$

e) Kesimpulan.

2.7 Matriks Kovarian

Menurut Raykov & Marcoulides (2008), matriks varian kovarian merupakan suatu matriks simetris yang berisi varian pada diagonal utamanya dan kovarian pada elemen lainnya. Koefisien varian menggambarkan sebuah indeks dari hubungan linear antara dua peubah penjelas.

Menurut Krzanowski (1998), matriks kovarian didefinisikan sebagai berikut:

$$\Sigma = E \left[(X - E(X))(X - E(X))^T \right] \quad (2.7)$$

dengan k peubah x_1, x_2, \dots, x_k , ada k varian dan $\frac{k(k-1)}{2}$ kovarian. Matriks ini biasanya disebut matriks varian kovarian Σ merupakan matriks simetris karena $cov(x_i x_j) = cov(x_j x_i)$ atau $\sigma_{ii} = \sigma_{ij}$. Matriks Σ diduga oleh matriks S . S adalah penduga matriks varian kovarian kelompok ke- i yang didefinisikan oleh:

$$S = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T \quad (2.8)$$

dengan $x_i^T = [x_{i1}, x_{i2}, \dots, x_{ip}]$ adalah vektor pengamatan untuk i pengamatan diagonal utama dari matriks S berisi varian dari peubah lainnya.

2.8 Ordinary Least Squares

Ordinary least squares (OLS) adalah metode estimasi paling umum dalam regresi linear yang bertujuan mencari garis regresi yang meminimumkan jumlah kuadrat selisih antara nilai observasi dan nilai prediksi model. Dengan kata lain, OLS menentukan nilai parameter (koefisien regresi) sehingga *sum of squared error* (SSE) menjadi sekecil mungkin (Wooldridge, 2020).

OLS bekerja berdasarkan asumsi klasik seperti linearitas, homoskedastisitas, dan tidak adanya autokorelasi. Jika asumsi terpenuhi, estimator OLS bersifat *BLUE* (*Best Linear Unbiased Estimator*) sesuai teorema Gauss-Markov (Greene, 2018), namun pelanggaran asumsi seperti heteroskedastisitas atau *outlier* dapat menyebabkan estimasi menjadi tidak efisien (Amin, *et al.*, 2021).

Secara matematis misalkan model regresi linear:

$$y = \beta_0 + \beta_1 X + \varepsilon \quad (2.9)$$

maka estimator OLS diberikan oleh:

$$\hat{\beta} = (X'X)^{-1}X'y \quad (2.10)$$

rumus ini diperoleh dari proses meminimalisasi jumlah kuadrat residual:

$$\min_{\beta} \sum_{i=1}^n (y_i - X_i\beta)^2 \quad (2.11).$$

OLS tetap menjadi metode dasar dalam analisis regresi karena kesederhanaannya serta sifat teoritisnya, meskipun berbagai metode *robust* atau regularisasi sering digunakan ketika asumsi klasik tidak terpenuhi (Amin, *et al.*, 2021).

2.9 Analisis Komponen Utama (AKU/PCA)

Menurut Jolliffe (2002) ide sentral dari analisis komponen utama adalah untuk mereduksi dimensi dari peubah asal sehingga diperoleh peubah baru yang disebut komponen utama. Komponen tersebut tidak saling berkorelasi dan tetap mempertahankan sebagian besar informasi yang terkandung pada peubah asalnya. Hal ini dicapai dengan melakukan transformasi dari segugus data besar menjadi segugus data baru, yaitu komponen utama (KU), yang tidak saling berkorelasi dan disusun sedemikian sehingga untuk k KU pertama mewakili keragaman terbesar dari keseluruhan peubah aslinya.

Komponen utama merupakan kombinasi linear dari peubah yang diamati, informasi yang terkandung pada komponen utama merupakan gabungan dari semua peubah dengan bobot tertentu. Kombinasi linear yang dipilih merupakan kombinasi linear dengan ragam paling besar yang memuat informasi paling banyak. Antar komponen utama bersifat ortogonal, tidak berkorelasi dan informasinya tidak tumpang tindih. Hasil dari prosedur ini nantinya digunakan pada analisis lebih lanjut, seperti analisis pengelompokan dan regresi komponen utama.

AKU tidak selalu bermanfaat digunakan untuk mereduksi banyaknya peubah asal menjadi beberapa peubah baru yang dapat menjelaskan dengan baik keragaman data asal. Bila tidak ada korelasi antara peubah asal, AKU tidak akan memberikan hasil yang diinginkan, karena peubah baru yang diperoleh hanyalah peubah asal yang ditata berdasarkan besar keragamannya. Makin erat korelasi (baik positif maupun negatif) antar peubah, makin baik pula hasil yang diperoleh dari AKU (Mattjik & Sumertajaya, 2011).

Sebelum dilakukan pembentukan komponen utama, data terlebih dahulu dipusatkan (*centering*) dan diskalakan (*scaling*) untuk menghilangkan pengaruh perbedaan satuan pengukuran antar variabel. Proses standardisasi ini bertujuan agar setiap variabel memiliki kontribusi yang sebanding dalam analisis komponen utama. Proses pemusatan dan penskalaan data dilakukan dengan persamaan sebagai berikut:

$$X^* = \frac{x_i - \bar{x}_i}{\sqrt{\sum(x_i - \bar{x}_i)^2}} \quad (2.12)$$

di mana:

X^* : standardisasi dari data ke-i,

x_i : nilai data observasi ke-i,

\bar{x}_i : nilai rata-rata dari data x.

Data yang telah distandardisasi selanjutnya digunakan untuk membentuk matriks korelasi. Matriks korelasi ini menjadi dasar dalam menentukan struktur hubungan antar variabel serta digunakan dalam perhitungan nilai eigen dan vektor eigen pada analisis komponen utama, yang dirumuskan sebagai berikut:

$$\mathbf{R} = X^{*'} X^* \quad (2.13).$$

Dalam analisis komponen utama misalkan X_1, X_2, \dots, X_k , memiliki sebaran

peubah ganda dengan vektor rata-rata $\boldsymbol{\mu}$ dan matriks kovarian $\boldsymbol{\Sigma}$ serta vektor ciri \mathbf{a}_k . Komponen utama (Q) seperti telah dijelaskan di atas, merupakan kombinasi linear dari k peubah asal, atau dapat ditulis:

$$\mathbf{Q} = \mathbf{A}\mathbf{X} \quad (2.14)$$

$$\begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_j \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} \\ a_{21} & a_{22} & \cdots & a_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kj} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{bmatrix}$$

Komponen utama pertama adalah kombinasi linear terbobot variabel asal yang dapat menerangkan keragaman terbesar (Gasperz, 1991).

Komponen utama pertama dapat dituliskan sebagai:

$$Q_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1k}X_k = \mathbf{a}_1^T \mathbf{X}.$$

Komponen utama kedua dapat dituliskan sebagai:

$$Q_2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2k}X_k = \mathbf{a}_2^T \mathbf{X}.$$

Secara umum komponen utama ke- j dapat dituliskan sebagai berikut:

$$Q_j = a_{j1}X_1 + a_{j2}X_2 + \cdots + a_{jk}X_k = \mathbf{a}_j^T \mathbf{X} \quad (2.15).$$

Secara matriks, pembentukan komponen utama dapat pula dipandang sebagai hasil proyeksi data hasil standardisasi ke dalam ruang vektor eigen, sehingga skor komponen utama diperoleh melalui persamaan:

$$\mathbf{Z} = \mathbf{X}^* \mathbf{V} \quad (2.16)$$

dengan \mathbf{V} merupakan matriks vektor eigen dari matriks korelasi.

Menurut D.F Marrison tahun 1976 dalam edisi kedua *Multivariat Statistical Methods*, komponen-komponen dapat dihitung sampai sejumlah tertentu proporsi keragaman ($> 75\%$) yang telah terjelaskan (Draper & Smith, 1992). Selain itu, kontribusi keragaman dari setiap komponen utama ke-k terhadap total keragaman dapat juga dilihat nilai eigen yang lebih dari satu.

2.10 *Principal Component Regression (PCR)*

Analisis regresi komponen utama merupakan suatu analisis kombinasi antara analisis regresi dan analisis komponen utama. Analisis regresi komponen utama ditetapkan bila dalam pembentukan model pendugaan peubah bebas yang digunakan banyak dan terdapat hubungan yang erat antar peubah bebasnya. Adanya korelasi antar peubah bebas menyebabkan salah satu asumsi dasar analisis regresi linear berganda dalam metode kuadrat terkecil menjadi gagal terpenuhi dan salah satu cara membebaskan korelasi antar peubah bebasnya dengan regresi komponen utama.

Menurut Notiragayu & Nisa (2008), Regresi Komponen Utama (RKU) adalah satu diantara metode yang bisa dipergunakan dalam mengatasi masalah multikolinearitas dengan dua tahapan, tahap pertama melakukan analisis komponen utama terhadap peubah bebas X yang kemudian untuk tahap selanjutnya ialah meregresikan komponen-komponen utama tersebut dengan peubah tak bebas Y. Sesudah melakukan analisis komponen utama tahapan berikutnya yaitu meregresikan skor komponen utama yang telah terpilih dengan variabel terikat Y dengan metode *Ordinary Least Squares* (OLS). Sehingga, persamaan regresi komponen utama dapat dituliskan sebagai berikut:

$$Y_i = a_0 + a_1U_{1i} + a_2U_{2i} + \dots + a_nU_{mi} + \gamma_i \quad (2.17)$$

di mana:

- Y_i : variabel terikat,
- U_{ji} : variabel bebas, yaitu komponen utama terpilih,
- a_0 : konstanta (intersep),
- a_j : koefisien regresi,
- γ_i : *error* random,
- n : banyaknya pengamatan,
- m : banyaknya komponen utama yang terpilih.

Apabila dimuat dalam bentuk matriks sebagai berikut:

$$\mathbf{y} = \mathbf{U}\boldsymbol{\alpha} + \boldsymbol{\gamma} \quad (2.18)$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & U_{11} & U_{21} & \cdots & U_{m1} \\ 1 & U_{12} & U_{22} & \cdots & U_{m2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & U_{1n} & U_{2n} & \cdots & U_{mn} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} + \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix}$$

di mana:

- \mathbf{y} : vektor $n \times 1$ dari variabel respon,
- \mathbf{U} : matriks $n \times q$ dari variabel prediktor (komponen utama terpilih),
- $\boldsymbol{\alpha}$: vektor $q \times 1$ dari koefisien regresi, dengan $q = m+1$,
- $\boldsymbol{\gamma}$: vektor $n \times 1$ dari *error* random.

2.11 Least Trimmed Squares

Regresi *robust* merupakan metode yang digunakan ketika residual berdistribusi tidak normal. Ketika melakukan uji asumsi untuk model regresi sering ditemui bahwa asumsi regresi dilanggar, transformasi data tidak menghilangkan pengaruh dari *outlier* yang mengakibatkan estimasi koefisien regresi menjadi bias. Oleh karena itu regresi *robust* dikembangkan untuk menghasilkan estimasi yang tetap stabil terhadap observasi ekstrem. Metode ini memberikan alternatif yang lebih

andal dibandingkan regresi klasik ketika data mengandung *outlier*, karena mampu menahan pengaruh observasi yang menyimpang secara signifikan.

Prosedur *robust* ditunjukkan untuk memberikan dugaan yang lebih tepat dan cepat terhadap data yang melanggar asumsi dengan cara meniadakan identifikasi adanya data *outlier*, serta bersifat otomatis dalam menanggulangi data *outlier*. Beberapa metode dalam regresi *robust* diantaranya adalah *Least Trimmed Square* (LTS), *Least Mean Square* (LMS), *Theil-Sein*, *MM estimator*, *S estimator*, dan *M estimator* (penduga M).

Salah satu metode pendugaan parameter model regresi terhadap data yang mengandung *outlier* adalah metode penduga *Least Trimmed Square* (LTS). Metode ini merupakan salah satu metode pendugaan parameter pada regresi *robust* yang kuat terhadap keberadaan *outlier*. Metode LTS mempunyai prinsip pendugaan parameter yang sama dengan OLS, yaitu meminimumkan jumlah kuadrat galat. Hanya saja pada metode LTS, jumlah kuadrat galat yang diminimumkan adalah jumlah kuadrat galat dari h pengamatan yang dianggap bukan *outlier*. (Rousseeuw & Hubert, 1997).

Metode LTS menduga koefisien regresi dengan menggunakan metode OLS terhadap sub himpunan data berukuran h , yaitu:

$$\hat{\beta} = \arg \min_{\beta} \left(\sum_{i=1}^h e_i^2 \right) = \arg \min_{\beta} \left(\sum_{i=1}^h (y_i - \hat{y}_i)^2 \right), \frac{(3n + k + 1)}{4} \leq h \leq n \quad (2.19)$$

Solusi $\hat{\beta}$ pada persamaan (2.16) diperoleh melalui prosedur pendugaan parameter regresi. Pada metode *Least Trimmed Squares* (LTS) pendugaan tersebut tidak dilakukan menggunakan seluruh pengamatan, melainkan hanya berdasarkan sub himpunan H terbaik. Sub himpunan H dipilih dengan menggunakan algoritma resampling dari seluruh kemungkinan sub himpunan yang dapat dibentuk, yaitu

sebanyak $\binom{n}{h}$ sub himpunan data yang terpilih merupakan data yang telah terpangkas, sehingga pengaruh pencilan dapat diminimalkan.

Menurut Nisa (2006), algoritma resampling untuk pendugaan LTS dapat dijelaskan secara garis besar sebagai berikut:

1. Ambil secara acak k pengamatan, dengan k adalah banyaknya variabel bebas.
2. Hitung koefisien regresi berdasarkan k pengamatan tersebut, misalkan diperoleh persamaan regresi $\hat{Y}_k = f(x)$
3. Hitung galat dari n data berdasarkan persamaan regresi $\hat{Y}_k = f(x)$
4. Urutkan galat dari yang terkecil hingga terbesar, yaitu $|e|_1 \leq |e|_2 \leq \dots \leq |e|_n$
5. Pilih h_1 pengamatan pertama yang memiliki nilai mutlak galat terkecil himpunan pengamatan yang terpilih disebut sebagai subsampel pertama (H_1)
6. Lakukan *C-step* (akan dijelaskan lebih lanjut) terhadap H_1 sebanyak 2 kali dan didapatkan H_1^* .
7. Ulangi langkah 1-6 sebanyak $\binom{n}{k}$ kali. Dari $\binom{n}{k}$ hasil yang diperoleh, pilih 10 subsampel H terbaik, yaitu yang memiliki jumlah kuadrat galat terkecil.
8. Lakukan *C-Step* terhadap 10 sub sampel tersebut sampai konvergen dan dari 10 subsampel tersebut, pilih 1 subsampel terbaik yang akan digunakan sebagai acuan untuk perhitungan dugaan koefisien regresi.

C-Step merupakan langkah pemusatan (*Concentration Step*) terhadap pengamatan untuk memperoleh galat (*error*) terkecil. Langkah-langkah *C-Step* adalah sebagai berikut. Misalkan telah diberikan suatu sub himpunan H berukuran h dari sampel berukuran n , maka:

1. Hitung koefisien regresi dari H .
2. Hitung galat dengan menggunakan koefisien regresi H dan sampel n .
3. Urutkan galat dari yang terkecil sampai yang terbesar, $|e|_1 \leq |e|_2 \leq \dots \leq |e|_n$.
4. Ambil H sebanyak h pengamatan yang memiliki nilai mutlak galat terkecil. Pengulang *C-Step* akan menghasilkan proses iterasi.

2.12 *Wild bootstrap*

Wild bootstrap adalah metode resampling statistika yang dimodifikasi dari teknik *bootstrap* standar, terutama digunakan saat ada masalah heteroskedastisitas (varians *error* tidak konstan), sehingga mengambil sampel ulang residu (*error*) dari model regresi dengan mempertahankan variabel prediktor (X) pada nilai aslinya, untuk menghasilkan distribusi statistik uji yang lebih akurat untuk pengujian hipotesis atau interval kepercayaan. Metode ini mengatasi keterbatasan *bootstrap* tradisional yang mengasumsikan homoskedastisitas.

Metode *bootstrap* pertama kali diperkenalkan oleh Efron pada tahun 1979. Nama *bootstrap* sendiri diambil dari sebuah frase “*Pull up by your own bootstraps*” yang artinya adalah bergantunglah pada sumbermu sendiri. Dalam hal ini, metode *bootstrap* bergantung pada sampel yang merupakan satu-satunya sumber yang dimiliki oleh seorang peneliti.

Metode *bootstrap* digunakan untuk mengestimasi koefisien dari suatu persamaan regresi dengan melakukan penyampelan ulang dari sampel yang sudah ada. Dalam konteks regresi, metode *bootstrap* memiliki 3 cara (Shao dan Tu, 1995):

1. *Bootstrap* berdasarkan residual (*bootstrap based on residual*) atau dikenal dengan *resample fixed-x* atau model residual.
2. *Bootstrap* pasangan (*paired bootstrap*) atau dikenal dengan *resample random-x* atau model korelasi.
3. *Bootstrap* eksternal (*external bootstrap*).

Statistik yang akan digunakan sebagai parameter tingkat akurasi adalah bias, *standard error*, dan *root mean square error*. Untuk analisis data digunakan *software* R dalam melakukan *bootstrap*.

2.13 Robust Principal Component (RPC)

Robust principal component (RPC) merupakan suatu pendekatan regresi yang dikembangkan untuk menangani permasalahan multikolinearitas dan keberadaan pengamatan-pengamatan ekstrim atau *outlier* dalam data. Metode ini mengintegrasikan dua teknik statistik yaitu, *principal component analysis* (PCA) untuk mengatasi masalah terhadap multikolinearitas, dan *least trimmed squares* (LTS) sebagai *estimator robust* yang memiliki kemampuan untuk mengatasi *outlier*. Kombinasi kedua metode tersebut menghasilkan *estimator* regresi yang lebih stabil, efisien, dan tahan terhadap data yang tidak memenuhi asumsi klasik dibandingkan metode regresi klasik seperti *ordinary least squares* (OLS).

Gabungan PCA dan LTS memungkinkan RPC menangani dua masalah utama dalam regresi secara simultan, yaitu multikolinearitas dan *outlier*. Beberapa penelitian menunjukkan bahwa *estimator* RPC memberikan hasil yang lebih stabil dibandingkan OLS maupun *robust estimator* lain ketika data tidak memenuhi asumsi klasik (Rasheed, *et al.*, 2022)

Meskipun RPC efektif mengatasi *outlier* dan multikolinearitas, namun metode ini memiliki kekurangan, khususnya dalam konteks inferensi statistik yaitu tidak mampu mengatasi heteroskedastisitas yang menyebabkan estimasi varians parameter menjadi tidak valid, dan kurang baik ketika multikolinearitas, *outlier*, dan heteroskedastisitas terjadi secara simultan (Wu, 1986).

2.14 Robust Principal Component with Wild Bootstrap

Metode *robust principal component with wild bootstrap* merupakan pendekatan estimasi regresi yang dirancang untuk mengatasi tiga masalah klasik dalam analisis data, yaitu multikolinearitas, *outlier*, dan heteroskedastisitas secara simultan. Metode ini penggabungan dari metode *robust principal component*

analysis (Robust PCA) untuk menstabilkan struktur kovarian pada data yang mengandung *outlier*. *Least trimmed squares (LTS)* untuk menghasilkan residual dan koefisien regresi yang tahan terhadap pengaruh *outlier*, serta *wild bootstrap* untuk menghasilkan varians distribusi estimasi yang valid ketika varians *error* yang tidak konstan. Penggabungan ketiga komponen ini menjadi RPC-WB sebagai metode yang unggul dalam kondisi data kompleks, terbukti dari penelitian (Rasheed, *et al.*, 2022) yang menunjukkan bahwa RPC-WB menghasilkan bias, RMSE, dan *standrad error*, yang lebih kecil dibandingkan dengan metode *bootstrap non-robust* atau *robust* biasa.

2.14.1 Alur RPC dengan LTS, dan *Wild Bootstrap*

Alur metode *robust principal component (RPC)* yang dikombinasikan dengan *least trimmed squares (LTS)* dan *wild bootstrap* terdiri dari tiga tahap utama yang saling berkaitan. Tahap pertama adalah pembentukan *robust principal component* yaitu proses transformasi variabel prediktor ke dalam komponen utama menggunakan *robust covariance estimator* agar komponen utama tidak dipengaruhi *outlier* serta mampu mengurangi multikolinearitas. Pendekatan *robust PCA* ini terbukti efektif menghasilkan struktur data yang stabil dalam keberadaan *outlier*, sebagaimana dijelaskan Roy (2024).

Tahap kedua adalah estimasi regresi menggunakan LTS, yaitu teknik regresi *robust* yang hanya meminimalkan kuadrat residual dari subset observasi terbaik sehingga *outlier* tidak memengaruhi koefisien regresi. Estimasi LTS memberikan residual *robust* yang kemudian distandardisasi menggunakan skala *robust* seperti *median absolute deviation (MAD)* dan diberi bobot melalui fungsi *tukey bisquare*, yang dirancang untuk menekan pengaruh residual besar secara efektif. Tahap selanjutnya tahap *wild bootstrap*, yaitu teknik resampling residual tersebut menggunakan *wild multipliers*. Pendekatan *wild bootstrap* ini secara empiris memberikan estimasi varians dan *standard error* yang lebih reliabel ketika varians

error tidak homogen, sebagaimana ditunjukkan (Rasheed, *et al.*, 2022). Dalam analisis *robust* linear model dengan kombinasi RPC dengan *wild bootstrap*. Proses ini kemudian diulang sehingga menghasilkan distribusi *bootstrap* untuk menghitung bias, varians, dan *standard error* estimator.

2.14.2 Residual dan Skala *Robust* LTS

Residual dalam regresi *robust* berbasis *Least Trimmed Squares* (LTS) merupakan selisih antara nilai aktual dan nilai prediksi yang dihitung dari subset observasi dengan kontribusi kuadrat residual terkecil. Karena LTS hanya menggunakan sebagian data yang bersih untuk membentuk estimasi, dan residual yang dihasilkan lebih tahan terhadap pengaruh *outlier* dibanding residual OLS, sehingga mampu menggambarkan pola *error* yang lebih representatif. LTS memiliki *high-breakdown point*, sehingga residual tetap stabil meskipun sebagian observasi adalah *outlier*, menjadikannya komponen fundamental dalam metode RPC-LTS sebelum dilakukan proses *wild bootstrap* (Rasheed, *et al.*, 2022).

Setelah residual LTS diperoleh, langkah berikutnya adalah menghitung skala residual *robust*, yaitu ukuran penyebaran residual yang tidak sensitif terhadap *outlier*. Salah satu ukuran skala *robust* yang paling banyak digunakan adalah *Median Absolute Deviation* (MAD) dan variannya *Normalized MAD* (NMAD). Skala *robust* digunakan untuk menstandarisasi residual agar nilai residual ekstrem tidak mendominasi proses pembobotan maupun resampling pada tahap *wild bootstrap*. Ukuran MAD diterima luas dalam statistik modern karena sifatnya yang efisien, stabil terhadap *outlier*, dan mudah diimplementasikan, sebagaimana dijelaskan dalam studi *robust multivariate* terbaru oleh (Roy, 2024).

Secara matematis, residual dari model *Least Trimmed Squares* (LTS) didefinisikan sebagai selisih antara nilai pengamatan dan nilai pendugaan model, yaitu

$$\varepsilon_i = y_i - \hat{y}_i \quad (2.20)$$

di mana :

- ε_i : residual pengamatan ke- i ,
- y_i : nilai respon aktual pada pengamatan ke- i ,
- \hat{y}_i : nilai respon hasil pendugaan model LTS pada pengamatan ke- i .

Standardisasi residual secara *robust* digunakan ukuran *Median Absolute Deviation* (MAD), yang dirumuskan sebagai berikut:

$$MAD = \frac{1}{0.6745} \text{median}|\varepsilon_i - \text{median}(\varepsilon_i)| \quad (2.21).$$

Residual yang telah diskalakan secara *robust* dinyatakan sebagai:

$$r_i = \frac{\varepsilon_i}{MAD} \quad (2.22)$$

di mana:

- MAD : ukuran skala residual *robust*,
- ε_i : residual ke- i ,
- $\text{median}(\varepsilon_i)$: nilai median dari seluruh residual,
- r_i : residual terskalakan secara *robust*.

Residual terskalakan ini selanjutnya digunakan sebagai dasar dalam proses pembobotan menggunakan fungsi *Tukey bisquare* serta dalam pembentukan residual pada prosedur *wild bootstrap*, sebagaimana digunakan dalam metode RPC-LTS-Wild Bootstrap (Rasheed, et al., 2022).

Penggunaan skala *robust* ini terbukti meningkatkan stabilitas dan akurasi model ketika digabungkan dengan *robust* PCA dan *wild bootstrap*, sebagaimana ditunjukkan oleh hasil evaluasi model *robust* dan *bootstrap* dalam penelitian (Kalantan, *et al.*, 2025). Dengan demikian, residual dan skala *robust* LTS merupakan dua komponen penting yang berperan dalam mempersiapkan residual *robust* sebelum memasuki tahap pembobotan dan resampling pada metode RPC-LTS-*Wild Bootstrap*.

2.14.3 Bobot Tukey Bisquare

Bobot *Tukey bisquare* merupakan salah satu fungsi pembobot (*weight function*) yang digunakan dalam regresi *robust* untuk mengurangi pengaruh observasi dengan residual besar terhadap estimasi parameter. Fungsi ini bekerja dengan memberikan bobot penuh pada residual kecil, bobot menurun pada residual menengah, dan bobot nol pada residual yang dianggap *outlier*.

Secara matematis, fungsi bobot Tukey bisquare didefinisikan sebagai berikut:

$$w_i = \begin{cases} \left[1 - \left(\frac{r_i}{c}\right)^2\right]^2 & , |r_i| \leq c \\ 0 & , |r_i| > c \end{cases} \quad (2.23)$$

di mana:

- w_i : bobot pengamatan ke- i ,
- r_i : residual terskalakan secara *robust*,
- c : konstanta tuning, dengan nilai umum $c = 4,685$.

Dengan demikian, observasi ekstrem tidak memengaruhi proses estimasi, sementara observasi yang konsisten dengan pola data umum tetap dipertahankan kontribusinya. Penggunaan bobot *Tukey bisquare* sangat relevan dalam metode LTS dan RPC-LTS-*Wild Bootstrap* karena residual yang telah dihitung dan

distandardisasi akan diproses lebih lanjut untuk memperkuat ketahanan model terhadap *outlier* fungsional maupun struktural.

Dalam implementasinya, fungsi bobot *Tukey bisquare* didefinisikan melalui fungsi kerugian (ρ -function) yang memotong kontribusi residual setelah mencapai ambang tertentu (*cut-off point*), sehingga menjadikannya salah satu teknik *M-estimation* yang paling efektif dalam konteks *robust regression* modern.

Penelitian terbaru mendukung efektivitas pendekatan *trimming* dan pembobotan *Tukey*, terutama dalam model *robust* multivariat dan regresi ketika data memiliki *noise* tinggi atau *outlier* yang signifikan, sebagaimana ditunjukkan oleh (Roy, 2024) dalam analisis *robust* PCA berbasis *density power divergence*.

Selain itu, bobot *Tukey* terbukti memberikan performa stabil ketika digunakan bersama residual *robust* LTS dan *wild bootstrap* pada regresi linear dengan kondisi heteroskedastik, berdasarkan temuan (Rasheed, *et al.*, 2022). Dengan demikian, fungsi bobot *Tukey bisquare* merupakan komponen penting dalam alur RPC-LTS-Wild Bootstrap karena memperkuat ketahanan model sebelum proses resampling residual dilakukan.

2.14.4 Residual Wild Bootstrap

Residual *wild bootstrap* merupakan teknik *resampling* yang digunakan untuk menghasilkan estimasi varians dan *standard error* yang lebih andal ketika model regresi mengalami heteroskedastisitas, yaitu kondisi ketika varians *error* tidak konstan di seluruh pengamatan. Dalam konteks metode RPC-LTS-Wild Bootstrap, residual yang digunakan untuk proses *resampling* merupakan residual *robust* LTS yang telah distandardisasi menggunakan skala *robust*, sehingga *outlier* tidak memengaruhi distribusi residual yang akan di-bootstrap.

Prosedur *wild bootstrap* bekerja dengan mengalikan residual *robust* tersebut dengan *wild multipliers*, yaitu variabel acak dengan nilai rata-rata nol dan varians satu. Pengali acak tersebut berfungsi mempertahankan pola heteroskedastisitas dalam residual sehingga distribusi *bootstrap* mencerminkan struktur *error* sebenarnya pada data.

Dalam pendekatan *fixed-x wild bootstrap*, sampel *bootstrap* dibentuk dengan mempertahankan nilai peubah penjelas dan membangkitkan nilai respon semu dengan menambahkan residual *robust* yang telah dimodifikasi. Nilai respon *bootstrap* diperoleh melalui persamaan berikut:

$$y_i^{*b} = f(x_i, \hat{\beta}_{LTS}) + t_i^* \frac{\hat{\varepsilon}_i}{\sqrt{1 - h_{ii}}} \quad (2.24)$$

di mana:

- y_i^{*b} : nilai respon *bootstrap* pada pengamatan ke- i
- $f(x_i, \hat{\beta}_{LTS})$: nilai prediksi model regresi LTS
- $\hat{\varepsilon}_i$: residual berbobot LTS
- t_i^* : *wild multiplier* dengan nilai harapan nol dan varians satu
- h_{ii} : *leverage* pengamatan ke- i

Leverage pengamatan ke- i didefinisikan sebagai:

$$h_{ii} = x_i'(X'X)^{-1}x_i \quad (2.25)$$

Nilai *leverage* digunakan untuk mengurangi pengaruh pengamatan dengan *leverage* tinggi dalam pembentukan residual *bootstrap*, sehingga proses resampling menjadi lebih stabil terhadap observasi ekstrem.

Untuk menggantikan ukuran rata-rata dan simpangan baku yang tidak *robust*, residual berbobot LTS dinormalisasi menggunakan pendekatan *Normalized Median Absolute Deviation* (NMAD), yang dirumuskan sebagai berikut:

$$NMAD = \text{median}|\hat{\varepsilon}_i^{WLTS} - \text{median}(\hat{\varepsilon}^{WLTS})| \quad (2.26)$$

Residual ternormalisasi selanjutnya dinyatakan sebagai:

$$R_{ai} = \frac{\hat{\varepsilon}_i^{WLTS} - \text{median}(\hat{\varepsilon}^{WLTS})}{NMAD_{norm}(\hat{\varepsilon}^{WLTS})} \quad (2.27)$$

di mana:

- R_{ai} : residual *robust* ternormalisasi,
- $\hat{\varepsilon}_i^{WLTS}$: residual berbobot LTS,
- $NMAD$: ukuran skala *robust* residual berbobot.

Setelah sampel *bootstrap* terbentuk, model diestimasi ulang menggunakan LTS untuk memperoleh estimasi parameter *bootstrap*. Proses ini diulang ratusan hingga ribuan kali untuk menghasilkan distribusi *bootstrap* yang selanjutnya digunakan menghitung *standard error*, *bias*, dan *root mean square error*.

2.14.5 Evaluasi Metode *Bootstrap*

Untuk mengevaluasi kinerja *robust wild bootstrap* yang digunakan dalam penelitian ini, dilakukan perhitungan beberapa ukuran evaluasi, yaitu *bias*, *root mean square error* (RMSE), dan galat baku (*standard error*). Metode estimasi dikatakan baik apabila menghasilkan nilai *bias*, RMSE, dan galat baku yang kecil. Perhitungan ketiga ukuran evaluasi tersebut dilakukan dengan menggunakan rumus-rumus matematika dalam analisis statistik.

Least trimmed squares digunakan sebagai estimator awal karena kemampuannya dalam mengurangi pengaruh *outlier* pada hasil estimasi parameter. Penerapan prosedur *robust wild bootstrap* dengan pembobot Wu atau pembobot Liu yang dikombinasikan dengan estimator *Least Trimmed Squares* yang diharapkan menghasilkan estimasi bias, RMSE, dan galat baku yang kecil. Rumus-rumus yang digunakan dalam perhitungan ukuran evaluasi tersebut disajikan sebagai berikut:

1. Estimator *bootstrap* dari parameter regresi diperoleh sebagai rata-rata dari seluruh estimator hasil pengulangan *bootstrap*, yaitu:

$$\hat{\beta}_{LTS}^* = \frac{1}{k} \sum_{b=1}^k \hat{\beta}_{LTS}^{*b} \quad (2.28).$$

2. Estimasi bias *bootstrap* didefinisikan sebagai selisih antara estimator *bootstrap* dan estimator awal, yaitu:

$$\text{Bias}(\hat{\beta}_{LTS}) = \hat{\beta}_{LTS}^* - \hat{\beta}_{LTS} \quad (2.29).$$

3. Galat baku *bootstrap* dihitung berdasarkan simpangan baku dari estimator *bootstrap* sebagai berikut:

$$SE(\hat{\beta}_{bLTS}^*) = \sqrt{\frac{\sum_{b=1}^k \hat{\beta}_{(bLTS)}^* - (\hat{\beta}_{bLTS}^*)^2}{k}} \quad (2.30).$$

4. Varians *bootstrap* dari estimator regresi diberikan oleh:

$$S_{\hat{\beta}_{bLTS}^*}^2 = \frac{1}{k-1} \sum (\hat{\beta}_{LTS}^{*b} - \hat{\beta}_{(bLTS)}^*)^2 \quad (2.31).$$

5. Kovarians estimator diperoleh dari elemen diagonal matriks kovarians, yang dirumuskan sebagai:

$$cov(\hat{\beta}_{bLTS}) = \frac{1}{k-1} \sum_{b=1}^k (\hat{\beta}_{LTS}^{*b} - \hat{\beta}_{(bLTS)}^*)(\hat{\beta}_{LTS}^{*b} - \hat{\beta}_{(bLTS)}^*)' \quad (2.32).$$

6. Ukuran kesalahan estimator diukur menggunakan *robust mean squared error*, yang didefinisikan sebagai berikut:

$$RMSE(\hat{\beta}_{bLTS}) = (bias)^2 + var(\hat{\beta}_{bLTS}) \quad (2.33).$$

7. *Robust root mean squared error* didefinisikan sebagai akar kuadrat dari *robust mean squared error*, yaitu:

$$RRMSE = \sqrt{(bias)^2 + var(\hat{\beta}_{bLTS})} \quad (2.34)$$

di mana:

- $\hat{\beta}$: Estimasi koefisien regresi,
- k : Jumlah pengulangan *bootstrap*,
- b : Indeks pengulangan *bootstrap*.

2.15 Hepatitis B

Hepatitis B merupakan penyakit infeksi yang disebabkan oleh *Hepatitis B Virus* (HBV) yang menyerang hati dan dapat berkembang menjadi hepatitis akut maupun kronis. Infeksi kronis berpotensi menyebabkan komplikasi berat seperti sirosis dan karsinoma hepatoseluler. Menurut *World Health Organization* (WHO, 2023), lebih dari 296 juta orang di dunia hidup dengan infeksi hepatitis B kronis, sehingga penyakit ini menjadi salah satu masalah kesehatan global dengan beban

ekonomi dan mortalitas yang tinggi. Dalam konteks penelitian ini, populasi terinfeksi hepatitis B kronis, yaitu jumlah individu yang mengalami infeksi HBV menetap lebih dari enam bulan. Tingginya jumlah penderita hepatitis B kronis sangat dipengaruhi oleh cakupan vaksinasi dan tingkat penularan di suatu populasi (WHO, 2024).

Variabel imunisasi Hep B3 merupakan persentase anak yang menerima tiga dosis lengkap vaksin hepatitis B. WHO (2023) menyatakan bahwa vaksin Hep B3 memberikan perlindungan lebih dari 90% terhadap risiko infeksi HBV dan telah menjadi indikator utama keberhasilan program imunisasi nasional. Wilayah dengan cakupan Hep B3 tinggi cenderung memiliki *trend* infeksi HBV yang lebih rendah. Selanjutnya, jumlah infeksi baru hepatitis B menggambarkan angka penularan HBV yang masih terjadi. Infeksi baru dapat disebabkan oleh transmisi vertikal dari ibu ke bayi, praktik medis yang tidak steril, perilaku berisiko, maupun kurangnya cakupan imunisasi dasar.

World Health Organization menekankan bahwa tingginya kasus infeksi baru biasanya berkaitan dengan rendahnya pemberian dosis vaksin pada periode bayi dan anak (WHO, 2023). Variabel orang yang hidup dengan hepatitis B kronis mengukur jumlah individu dalam suatu populasi yang memiliki infeksi kronis aktif pada periode tertentu. Variabel ini penting untuk memetakan beban penyakit jangka panjang serta kebutuhan intervensi kesehatan seperti terapi antiviral. *Centers for Disease Control and Prevention* menjelaskan bahwa individu dengan infeksi kronis menjadi populasi yang berpotensi menularkan virus dan memengaruhi angka infeksi baru (CDC, 2024).

Variabel jumlah kematian akibat hepatitis B kronis mengukur total kematian yang disebabkan oleh komplikasi infeksi HBV kronis dalam suatu populasi pada periode tertentu. Variabel ini mencerminkan tingkat keparahan beban penyakit serta efektivitas deteksi dini, cakupan pengobatan antivirus, dan kualitas layanan kesehatan di suatu wilayah.

Adapun Hep B *Birth Dose* (BD) merupakan vaksin hepatitis B dosis lahir yang diberikan dalam 24 jam pertama kehidupan. Vaksin BD sangat penting untuk mencegah transmisi vertikal, sebab tanpa dosis ini, bayi yang terpapar virus dari ibu memiliki risiko mencapai 90% untuk berkembang menjadi hepatitis kronis (WHO, 2023).

III. METODOLOGI PENELITIAN

3.1 Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan pada semester ganjil tahun ajaran 2025/2026 bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data penelitian ini merupakan data sekunder yang diperoleh dari WHO dan terdiri atas beberapa indikator epidemiologi hepatitis B. Variabel yang digunakan meliputi persentase kasus hepatitis B kronis pada populasi umum (Y), persentase cakupan imunisasi Hep B3 pada anak usia 1 tahun (X_1), persentase cakupan vaksinasi hepatitis B dosis lahir (X_2), jumlah infeksi baru hepatitis B (X_3), jumlah orang yang hidup dengan hepatitis B kronis (X_4), serta jumlah kematian akibat hepatitis B kronis (X_5). Data tersebut digunakan untuk menganalisis faktor-faktor yang memengaruhi prevalensi hepatitis B kronis.

3.3 Metode Penelitian

Penelitian ini menggunakan perbandingan metode pendekatan estimasi regresi, yaitu *Robust Principal Component (Robust PC)* dan *Wild Bootstrap*, dalam

mengatasi permasalahan *outlier*, multikolinearitas, dan heteroskedastisitas.

Dengan perhitungan penelitian menggunakan *software R-studio* 4.2.0.

Langkah-langkah dalam penelitian ini adalah sebagai berikut:

1. Mempersiapkan data riil yang diperoleh dari *website* WHO.
2. Melakukan pemeriksaan asumsi klasik pada data riil yaitu deteksi *outlier*, uji multikolinearitas, dan uji heterokedastisitas.
3. Melakukan pemusatan (*centering*) dan penskalaan (*scaling*) terhadap data pengamatan.
4. Membentuk matriks korelasi untuk data yang telah dipusatkan dan diskalakan.
5. Menghitung nilai eigen dan vektor eigen dari matriks korelasi.
6. Membentuk komponen utama dengan memproyeksikan data hasil standardisasi ke dalam ruang vektor eigen.
7. Mengevaluasi nilai eigen untuk menentukan kontribusi informasi dari setiap komponen utama, termasuk komponen dengan nilai eigen terkecil.
8. Membangun model regresi dengan metode *least trimmed squares* (LTS) pada data asli untuk memperoleh estimasi parameter regresi.
9. Residual dari model LTS dihitung, lalu menentukan bobot awal untuk setiap pengamatan berdasarkan kebalikan dari nilai absolut residual hasil pendugaan.
10. Residual diskalakan menggunakan pendekatan *Median Absolute Deviation* (MAD).
11. Menentukan bobot akhir menggunakan fungsi pembobot *Tukey bisquare* sebagai bagian dari prosedur M-estimator.
12. Menghitung W_I^{LTS} (residual LTS berbobot) dengan mengkombinasikan bobot awal dan bobot *Tukey bisquare*.
13. Membentuk sampel *bootstrap* dengan pendekatan *fixed- x*, serta di normalisasi secara *robust* menggunakan *Normalized Median Absolute Deviation* (NMAD).
14. Melakukan pendugaan ulang model regresi LTS berbobot pada nilai *bootstrap* untuk memperoleh estimator parameter *bootstrap*.

15. Proses *bootstrap* diulang sebanyak 1000 kali untuk memperoleh distribusi empiris dari estimator parameter regresi.
16. Melakukan evaluasi kinerja metode *robust principal component with wild bootstrap* dengan menghitung nilai bias, *Root Mean Square Error* (RMSE), dan galat baku (*standard error*). Estimator *Least Trimmed Squares* (LTS) digunakan sebagai penduga awal dalam proses estimasi.

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan, penelitian ini menunjukkan bahwa :

1. Metode *Robust* PC dengan *Wild Bootstrap* mampu mengatasi pengaruh *outlier*, multikolinearitas, dan heteroskedastisitas serta menghasilkan estimasi yang lebih stabil.
2. Hasil perbandingan menunjukkan bahwa metode RPC-Boot-Wu memberikan nilai standar *error* dan RMSE yang lebih kecil dibandingkan dengan RPC-Boot-Liu.
3. Secara keseluruhan, metode *robust* terbukti efektif dan layak digunakan pada data yang tidak memenuhi asumsi regresi linear klasik, terutama pada studi epidemiologi.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran untuk pengembangan penelitian selanjutnya. Penelitian ini masih terbatas pada data hepatitis B kronis dengan penggunaan metode *Robust Principal Component* dan *Wild Bootstrap* menggunakan pengali Wu dan Liu

Oleh karena itu, beberapa saran yang dapat di berikan adalah:

1. penelitian selanjutnya disarankan menggunakan jumlah data yang lebih besar atau mencakup wilayah yang lebih luas agar hasil estimasi model menjadi lebih stabil.
2. Penelitian berikutnya dapat membandingkan metode *Robust Principal Component* dengan metode *robust* lainnya.
3. Penelitian selanjutnya juga dapat mengembangkan metode ini pada data kesehatan lainnya atau data yang memiliki karakteristik kompleks serupa dengan data hepatitis B kronis.

DAFTAR PUSTAKA

- Anton, H. 1987. *Aljabar Linear Elementer*. Erlangga, Jakarta.
- Amin, M., Saeed, A., & Shahbaz, M. 2021. Linear regression and ordinary least squares: A review. *Journal of Applied Probability and Statistics*. **16**(3): 45–62.
- CDC. 2024. *Hepatitis B Information*. Centers for Disease Control and Prevention, United States of America.
- Chatterjee, S. & Hadi, A.S. 2006. *Regression Analysis by Example*. Ed. ke-4. Jhon Wiley and Sons Inc, New Jersey.
- Davidson, R., & Flachaire, E. 2008. The wild bootstrap, tamed at last. *Journal of Econometrics*. **146**(1): 162–169.
- Draper, N.R. & Smith, H. 1992. *Analisis Regresi Terapan*. Ed. Ke-2. Diterjemahkan oleh Bamang Sumantri. Gramedia Pustaka Umum, Jakarta.
- Efron, B. 1979. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*. **7**(1): 1–26.
- Hubert, M., Rousseeuw, P. J., dan Vanden Branden, K. 2005. ROBPCA: A new approach to robust principal component analysis. *Technometrics*. **47**(1): 64–79.
- Ima. 2025. Uji heterokedastisitas dalam regresi: pengertian dan solusinya. <https://tesis.id/blog/uji-heteroskedastisitas-dalam-regresi-pengertian-dan-solusinya/>. Diakses pada 9 Januari 2026.

- Jolliffe, I. T. 2002. *Principal Component Analysis*. Ed. ke-2. Springer-Verlag, New York.
- Kalantan, Z. I., Alharbi, L. S., Al-Zahrani, M. H., dan Binhimd, S. M. S. 2025. Robust dimensionality reduction: A bootstrap-based evaluation of PCA with applications in nutritional and environmental sciences. *Contemporary Mathematics*. **6**(1): 923-942.
- Montgomery, D. C., & Runger, G. C. 2011. *Applied Statistics and Probability for Engineers*. Ed. Ke 5. John Wiley & Sons, New York.
- Mattjik, A. A., & Sumertajaya, I. M. 2011. *Sidik Peubah Ganda dengan Menggunakan SAS*. IPB PRESS, Bogor.
- Nisa, K. 2006. Analisis regresi robust menggunakan metode least trimmed square untuk data mengandung pencilan. *Jurnal Ilmiah MIPA*. **9**(2): 1–9.
- Notiragayu dan Nisa, K. 2008. Analisis regresi komponen utama robust untuk data mengandung pencilan. *Jurnal Sains MIPA*. **14**(1): 45–50.
- Rasheed, B. A., Adnan, R., Saffari, S. E., & Atiyaye, D. M. 2022. Robust PC with wild bootstrap estimation of linear model in the presence of outliers, multicollinearity and heteroskedasticity error variance. *International Journal of Statistics and Applied Mathematics*. **7**(3): 85–93.
- Raykov, T. & Marcoulides, G.A. 2008. *An Introduction to Applied Multivariate Analysis*. Ed. ke-1. Taylor and Francis Group, New York.
- Roy, S. 2024. Robust principal component analysis using density power divergence. *Journal of Machine Learning Research*. **25**(109): 1–34.
- Roosseeuw, P.J. & Hubert, M. 1997. Robust Regression with Continuous and Binary Regressors. *Journal of Statistical Planning and Inference*. **57**:153-163.
- Shao, J. dan Tu, D. 1995. *The Jackknife and Bootstrap*. Ed. Ke-1. Springer-Verlag, New York.

Wooldridge, J. M. 2020. *Introductory Econometrics: A Modern Approach* Ed. Ke-7. Cengage Learning, New York.

World Health Organization 2023. *Hepatitis B: Key facts*. WHO, Geneva.

World Health Organization 2024. *Global Health Observatory (GHO) Data Repository: Hepatitis Indicators*. WHO, Geneva.

Wu, C. F. J. 1986. Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*. **14**(4): 1261–1295.