

SKRIPSI

**ANALISIS REGRESI ROBUST ESTIMASI M DENGAN
PEMBOBOT HUBER DAN ESTIMASI LTS DALAM
MENGATASI PENCILAN DATA**

(Studi Kasus : Faktor-faktor yang Mempengaruhi Penyakit Anemia)



DZAHABIYYAH MUTHMAINNAH

21106010080

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2025

SKRIPSI

**ANALISIS REGRESI ROBUST ESTIMASI M DENGAN
PEMBOBOT HUBER DAN ESTIMASI LTS DALAM
MENGATASI PENCILAN DATA**

(Studi Kasus : Faktor-faktor yang Mempengaruhi Penyakit Anemia)

Untuk memenuhi sebagian persyaratan

mencapai derajat Sarjana S-1

Program Studi Matematika



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

diajukan oleh

DZAHABIYYAH MUTHMAINNAH

21106010080

Kepada

PROGRAM STUDI MATEMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA

YOGYAKARTA

2025



PERSETUJUAN SEMINAR PROPOSAL

Hal : Persetujuan Seminar Proposal

Lamp : -

Kepada:

Yth. Ketua Program Studi Matematika

Fakultas Saintek UIN Sunan Kalijaga Yogyakarta

di tempat

Assalaamu' alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa proposal skripsi Saudara:

Nama : Dzahabiyyah Muthmainnah
NIM : 21106010080
Prodi / smt : Matematika / VII
Tema : *Analisis Regresi Robust Estimasi M dengan Pembobot Huber dan Estimasi LTS dalam Mengatasi Pencilan Data (Studi Kasus: Faktor-faktor Yang Mempengaruhi Penyakit Anemia)*

sudah dapat diseminarkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu' alaikum wr. wb.

Yogyakarta,

Pembimbing I

Sri Utami Zuliana, S.Si., M.Sc., Ph.D

NIP. 19741003 200003 2 002



PENGESAHAN TUGAS AKHIR

Nomor : B-1032/Un.02/DST/PP.00.9/06/2025

Tugas Akhir dengan judul : Analisis Regresi Robust Estimasi M dengan Pembobot Huber dan Estimasi LTS dalam Mengatasi Pencilan Data (Studi Kasus: Faktor-faktor Yang Mempengaruhi Penyakit Anemia)

yang dipersiapkan dan disusun oleh:

Nama : DZAHABIYYAH MUTHMAINNAH
Nomor Induk Mahasiswa : 21106010080
Telah diujikan pada : Kamis, 22 Mei 2025
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Sri Utami Zuliana, S.Si., M.Sc., Ph.D.
SIGNED

Valid ID: 684927ab0468e



Penguji I

Dr. Epha Diana Supandi, S.Si., M.Sc.
SIGNED

Valid ID: 6847adc5d4656



Penguji II

Arif Munandar, M.Sc.
SIGNED

Valid ID: 6847b3d2957a0



Yogyakarta, 22 Mei 2025
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 68492daa0be8a

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Dzahabiyyah Muthmainnah
NIM : 21106010080
Program Studi : Matematika
Fakultas : Sains dan Teknologi

Dengan ini menyatakan bahwa isi skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu Perguruan Tinggi dan sesungguhnya skripsi ini merupakan hasil pekerjaan penulis sendiri sepanjang pengetahuan penulis, bukan duplikasi atau saduran dari karya orang lain kecuali bagian tertentu yang penulis ambil sebagai bahan acuan. Apabila terbukti pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab penulis.

Yogyakarta, 6 Mei 2025



METERAI
TEMPEL
20
B1F02AMX238792280

Dzahabiyyah Muthmainnah

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

KARYA SEDERHANA INI PENULIS PERSEMBAHKAN UNTUK:

KEDUA ORANG TUA PENULIS,

ABI TERCINTA,

UMI TERCINTA,

ADIK TERSAYANG,

KELUARGA BESAR PENULIS,

SAHABAT PENULIS,

TEMAN PENULIS,

SERTA KEPADA DOSEN PEMBIMBING.

TAK LUPA JUGA KEPADA

ALMAMATER KAMPUS TERCINTA

UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA YOGYAKARTA.

"MATEMATIKA ANGKATAN 2021 TERCINTA"

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN MOTTO

"Dan apa pun yang kamu peroleh, maka itu semua adalah karunia dari Allah."

(QS. An-Nahl: 71)

"Seperti juga orang tua yang enggan memberikan pisau karena takut anaknya terluka, maka Tuhan pun punya alasan kenapa menunda semua mimpi kita. Tuhan menunggu kita dewasa dan menyaksikan apakah kita sudah terlatih dalam menjaga setiap kepercayaan."

(Kartini F.Astuti)

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PRAKATA

Puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat dan karuniannya yang tak ternilai harganya berupa keimanan, kesabaran, kekuatan dan kelancaran. Shalawat serta salam semoga selalu tercurahkan kepada Nabi Muhammad SAW sehingga penulis dapat menyelesaikan skripsi ini yang berjudul "Analisis Regresi Robust Estimasi M dengan Pembobot Huber dan Estimasi LTS dalam Mengatasi Pencilan Data (Studi Kasus: Faktor-faktor Yang Mempengaruhi Penyakit Anemia)".

Dalam proses penyusunan skripsi ini, penulis ingin menyampaikan rasa terima kasih yang tulus kepada pembimbing skripsi, Ibu Sri Utami Zuliana, S.Si., M.Sc., Ph.D., atas kesediaannya meluangkan waktu untuk membimbing, memberikan arahan, serta berbagi ilmu yang sangat bermanfaat. Bimbingan dan nasihat yang diberikan telah membantu penulis menyelesaikan skripsi ini dan menjadi bekal berharga untuk masa depan. Penulis juga memohon maaf yang sebesar-besarnya atas segala kekurangan atau kesalahan yang mungkin terjadi selama proses bimbingan. Penulis juga ingin menyampaikan ucapan terima kasih dan apresiasi yang sebesar-besarnya kepada:

1. Bapak Prof. Noorhaidi, MA., M.Phil., Ph.D. selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Ibu Prof. Dr. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
3. Ibu Dr. Epha Diana Supandi, S.Si., M.Sc., selaku Ketua Program Studi Ma-

tematika dan dosen pembimbing akademik yang telah banyak memberikan pengarahan kepada penulis dari awal perkuliahan hingga akhir.

4. Seluruh dosen program studi Matematika dan staf Fakultas Sains dan Teknologi yang senantiasa memberikan ilmu dan layanan terbaik kepada penulis dari awal hingga akhir perkuliahan.
5. Abi tercinta, terima kasih untuk do'a nya, kasih sayang, dukungan, dan kesabaran dalam menemani penulis menyusun skripsi ini. Terima kasih telah menjadi sosok abi yang luar biasa. Dedikasi ini merupakan bentuk kecil dari rasa cinta dan apresiasi yang tak akan mampu membalas semua yang telah diberikan.
6. Umi tercinta, terima kasih untuk cinta tanpa syarat yang selalu umi curahkan, terima kasih untuk perhatian kecil yang selalu umi berikan, serta doa-doa yang tak pernah henti mengalir, menjadi penyemangat dan penuntun setiap langkah saya dalam menghadapi setiap tantangan.
7. Keluarga besar, terima kasih untuk dukungan dan do'a nya sehingga saya bisa menyelesaikan skripsi ini.
8. Untuk teman-teman seperbimbingan Sindi, Laili, dan Muna yang telah menemani penulis selama proses bimbingan, terima kasih atas dukungan moral dan semangat yang kalian berikan. Setiap diskusi yang kita lakukan bersama sangat membantu dalam memahami konsep-konsep yang rumit. Penulis mengucapkan terima kasih atas bantuan dan persahabatan yang kalian curahkan selama ini.
9. Fina, Yun, Nadhifa, Ade, dan Edia, terima kasih atas semua dukungan, dan momen-momen berharga yang telah kita lalui bersama.

10. Untuk teman-teman KKN 114 Banyunganti, Kaliagung, Sentolo, terima kasih atas kenangan indah yang kita ciptakan bersama selama 45 hari pelaksanaan KKN. Kebersamaan ini bukan hanya menjadi pengalaman yang sangat berharga, tetapi juga kenangan manis yang akan selalu tersimpan dalam ingatan sepanjang hidup saya.
11. Keluarga Matematika angkatan 2021 yang telah kebersamai selama 4 tahun dibangku perkuliahan, saling mendukung, berbagi ilmu dan banyak kenangan manis yang selalu dikenang.
12. Terakhir kepada semua pihak yang tidak dapat penulis sebut satu-persatu.

Penulis sepenuhnya menyadari bahwa tidak ada sesuatu yang sempurna di dunia ini. Oleh karena itu, penulis dengan terbuka dan senang hati menerima segala bentuk saran dan kritik yang membangun terkait skripsi ini. Harapan penulis adalah semoga karya ini dapat memberikan manfaat, khususnya bagi penulis sendiri serta bagi semua pihak yang membutuhkan, sebagai referensi dalam proses penulisan. Pada akhirnya, semoga Allah SWT senantiasa melimpahkan rahmat dan hidayah-Nya kepada kita semua. *Aamiin*.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Yogyakarta, 6 Maret 2025

Dzahabiyyah Muthmainnah

DAFTAR ISI

HALAMAN JUDUL	i
PERSETUJUAN SEMINAR PROPOSAL	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTTO	vi
PRAKATA	vii
DAFTAR ISI	x
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
DAFTAR LAMBANG	xvi
INTISARI	xvii
ABSTRACT	.xviii
I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah	4
1.4. Tujuan	4
1.5. Manfaat	5
1.6. Tinjauan Pustaka	5
1.7. Sistematika Penulisan	9
II LANDASAN TEORI	10
2.1. Statistik	10

2.1.1.	Analisis Univariat	10
2.1.2.	Analisis Multivariat	11
2.2.	Variabel Acak	11
2.2.1.	Variabel Acak Diskrit	11
2.2.2.	Variabel Acak Kontinu	11
2.3.	Ekspetasi dan Variansi	12
2.3.1.	Ekspetasi Distribusi Peluang	12
2.3.2.	Variansi Distribusi Peluang	13
2.4.	Analisis Regresi Linear	13
2.5.	Metode <i>Ordinary Least Square</i> (OLS)	14
2.6.	Asumsi Klasik Analisis Regresi Linear	16
2.6.1.	Asumsi Homoskedastisitas	16
2.6.2.	Asumsi Normalitas	17
2.6.3.	Asumsi Multikolinearitas	18
2.6.4.	Asumsi Tidak Ada Autokorelasi	19
2.7.	Data Pencilan	20
2.7.1.	Metode <i>Leverage</i>	20
2.7.2.	Metode <i>DFFITS</i> (<i>Difference in fit statndarized</i>)	21
2.7.3.	Metode <i>R-student</i>	22
2.7.4.	Metode <i>Cook's Distance</i>	22
2.8.	Regresi <i>Robust</i>	23
2.9.	Anemia	24
III	METODE PENELITIAN	28
3.1.	Jenis dan Sumber Data	28
3.2.	Metode Pengumpulan Data	29
3.3.	Variabel Penelitian	29

3.4. Alat Pengolahan Data	29
3.5. Metode Analisis Data	29
3.6. Flowchart	31
IV PEMBAHASAN	32
4.1. Estimasi M	32
4.2. Estimasi LTS	34
4.3. Fungsi Objektif dan Fungsi Pembobot	36
4.4. <i>Breakdown Point</i>	37
4.5. Uji Kecocokan Model <i>Goodness of Fit</i>	38
4.5.1. <i>Adjusted R^2</i>	38
4.5.2. <i>Residual Standard Error (RSE)</i>	39
V STUDI KASUS	40
5.1. Studi Kasus	40
5.2. Analisis Deskriptif	40
5.3. Analisis Regresi Metode Kuadrat Terkecil (MKT)	43
5.4. Asumsi Klasik Model Analisis Regresi Linear	44
5.4.1. Asumsi Homoskedastisitas	44
5.4.2. Asumsi Normalitas	45
5.4.3. Asumsi Multikolinearitas	45
5.4.4. Asumsi Tidak Ada Autokorelasi	46
5.5. Deteksi Outlier	47
5.6. Regresi Robust Estimasi M Pembobot Huber	48
5.7. Regresi Robust Estimasi LTS	49
5.8. Pemilihan Model Estimasi Terbaik	51
VI PENUTUP	53
6.1. Kesimpulan	53

6.2. Saran	53
DAFTAR PUSTAKA	55
LAMPIRAN	59
A Data Penelitian	59
B Source Code Program R	64
C Data Outlier	68
D Biodata Penulis	69

DAFTAR TABEL

1.1	Tinjauan Pustaka	8
4.1	Breakdown Point Beberapa Estimator	38
5.1	Analisis Deskriptif	41
5.2	Hasil Estimasi Persamaan MKT	43
5.3	Nilai RSE dan \bar{R}^2 untuk MKT	44
5.4	Hasil Uji Asumsi Homoskedastisitas	45
5.5	Hasil Uji Asumsi Normalitas	45
5.6	Hasil Uji VIF	46
5.7	Hasil Uji <i>Durbin Watson</i>	46
5.8	Model Estimasi M Pembobot Huber	48
5.9	Nilai RSE dan \bar{R}^2 untuk Estimasi M	49
5.10	Model Estimasi LTS	50
5.11	Nilai RSE dan \bar{R}^2 untuk Estimasi LTS	51
5.12	Perbandingan RSE Pada Ketiga Metode	52

DAFTAR GAMBAR

3.1	Flowchart Penelitian	31
5.1	Boxplot data anemia	41



DAFTAR LAMBANG

y	: variabel respon yang akan diteliti
y_i	: variabel dependen ke- i
β_0	: konstanta
β_1	: parameter variabel independen
x	: variabel prediktor
e	: variabel error acak
β_k	: parameter variabel independen ke- k
x_{ik}	: variabel independen ke- k
h_{ii}	: leverage kasus ke- i
n	: banyaknya data
p	: banyaknya parameter
X_i	: nilai untuk kasus ke- i
\bar{X}	: mean dari X
S_x^2	: kuadrat simpangan X_i terhadap mean
e_i	: residual ke- i
k	: jumlah parameter
$f(z)$: fungsi kepadatan normal standar
R^2	: koefisien determinasi regresi

INTISARI

Analisis Regresi Robust Estimasi M dengan Pembobot Huber dan Estimasi

LTS dalam Mengatasi Pencilan Data

(Studi Kasus : Faktor-faktor yang Mempengaruhi Penyakit Anemia)

Oleh

DZAHABIYYAH MUTHMAINNAH

21106010080

Pencilan dalam data dapat menyebabkan hasil analisis regresi yang tidak akurat, terutama ketika menggunakan metode Kuadrat Terkecil (OLS). Penelitian ini bertujuan untuk menganalisis efektivitas metode regresi *robust*, khususnya Estimasi M dengan pembobot *Huber* dan Estimasi *Least Trimmed Squares* (LTS), dalam menangani pencilan pada data anemia. Anemia sendiri merupakan kondisi rendahnya kadar hemoglobin dalam darah yang dapat menyebabkan kelelahan dan penurunan daya tahan tubuh. Data dianalisis menggunakan asumsi klasik regresi linear, termasuk uji homoskedastisitas, normalitas, autokorelasi, dan multikolinearitas, sebelum dilakukan deteksi *outlier* menggunakan *leverage*, *DFBETS*, *R-student*, dan *Cook's Distance*. Studi kasus penelitian ini adalah data anemia dari rumah sakit Alok Healthcare Ltd yang berada di Dhaka, Bangladesh. Data tersebut diunggah oleh Mayen Uddin Mojumdar di Mendeley Data. Hasil penelitian menunjukkan bahwa metode Estimasi LTS memiliki nilai *Residual Standard Error* (RSE) terkecil sebesar 0,046 dan koefisien determinasi tertinggi sebesar 0,998 yang menjadikannya metode paling optimal dalam menangani pencilan dibandingkan Estimasi M dan OLS. Faktor-faktor yang mempengaruhi anemia pada anak adalah *Red Blood Cell* (RBC), *Packed Cell Volume* (PCV), *Mean Corpuscular Hemoglobin* (MCH), *Mean Corpuscular Volume* (MCV), dan *Mean Corpuscular Hemoglobin Concentration* (MCHC).

Kata Kunci: Regresi Robust, Estimasi M, Estimasi LTS, Pencilan, Anemia

ABSTRACT

Robust Regression Analysis Using M-Estimation with Huber Weights and

LTS Estimation to Handle Data Outliers

(Case Study: Factors Affecting Anemia Disease)

By

DZAHABIYYAH MUTHMAINNAH

21106010080

Outliers in data can lead to inaccurate regression analysis results, especially when using the *Ordinary Least Squares* (OLS) method. This study aims to analyze the effectiveness of robust regression methods, specifically M-estimation with *Huber* weights and the *Least Trimmed Squares* (LTS) estimation, in handling outliers in anemia data. Anemia itself is a condition characterized by low levels of hemoglobin in the blood, which can lead to fatigue and reduced immune function. The data were analyzed using classical linear regression assumptions, including tests for homoscedasticity, normality, autocorrelation, and multicollinearity, before *outlier* detection was conducted using *leverage*, *DFFITS*, *R-student*, and *Cook's Distance*. The case study for this research is anemia data from Alok Healthcare Ltd hospital located in Dhaka, Bangladesh. The data were uploaded by Mayen Uddin Mojumdar on Mendeley Data. The results showed that the LTS estimation method had the smallest *Residual Standard Error* (RSE) of 0.046 and the highest coefficient of determination of 0.998, making it the most optimal method for handling outliers compared to M-estimation and OLS. The factors influencing pediatric anemia include Red Blood Cell (RBC), Packed Cell Volume (PCV), Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Volume (MCV), and Mean Corpuscular Hemoglobin Concentration (MCHC).

Keywords: Robust Regression, M-Estimation, LTS Estimation, Outliers, Anemia

BAB I

PENDAHULUAN

1.1. Latar Belakang

Analisis regresi linier pertama kali diperkenalkan oleh Sir Francis Galton (1822-1911). Menurut Montgomery et al. (2021), analisis regresi linier adalah analisis terhadap hubungan satu variabel tak bebas (Y) dengan satu atau lebih variabel bebas (X). Estimasi parameter biasanya diselesaikan dengan Metode Kuadrat Terkecil (MKT). Metode Kuadrat Terkecil (MKT) adalah metode yang digunakan pada regresi untuk mencari estimasi parameter dengan meminimalkan jumlah kesalahan kuadrat antara nilai yang sebenarnya dan nilai yang diprediksi oleh model. Namun, apabila terdapat pencilan atau *outlier* pada data, maka estimasi koefisien garis regresi dengan Metode Kuadrat Terkecil menjadi tidak tepat.

Pencilan data (*outlier*) adalah nilai-nilai yang signifikansi berbeda dari sebagian besar data lainnya dalam suatu dataset dan dapat mempengaruhi hasil analisis. Pencilan dapat disebabkan oleh kesalahan pengukuran, kesalahan input data, atau fenomena alami yang memerlukan perhatian khusus dalam analisis statistik. Salah satu cara dalam mengatasi pencilan data dengan melakukan Analisis Regresi *Robust*.

Metode *robust* pertama kali diperkenalkan oleh Peter J. Huber (1964), merupakan metode regresi yang digunakan ketika distribusi residual tidak normal atau ada beberapa *outlier* yang berpengaruh pada model (Ryan (2008)). Menurut Chen (2002), model *robust* memiliki empat metode estimasi parameter, yaitu estimasi M ,

estimasi S , estimasi LTS , dan estimasi MM . Metode estimasi M merupakan salah satu metode *robust* yang paling populer dan mudah dalam pengaplikasiannya dimana estimasi M memiliki nilai *breakdown point* sebesar 50%, artinya lebih toleran terhadap outlier sehingga dapat memberikan hasil analisis lebih stabil. Estimasi M juga menggunakan Metode Kuadrat Terkecil (MKT) untuk menduga parameter regresi dengan meminimumkan jumlah kuadrat residu sehingga dapat memberikan estimasi parameter regresi yang lebih akurat. Sedangkan, metode estimasi LTS dirancang untuk mengurangi pengaruh pencilan pada estimasi parameter. Dalam metode ini, sejumlah residu terkecil dari model regresi yang dihasilkan akan dipangkas untuk menghindari pengaruh data yang tidak representatif.

Meskipun Metode Kuadrat Terkecil (MKT) dapat mengatasi pencilan (*outliers*) dalam data, namun sering kali tidak akurat. Oleh karena itu, masalah pencilan ini harus diatasi. Masalah ini penting untuk diteliti, sebab pencilan dalam analisis data dapat merusak interpretasi hasil penelitian. Banyak penelitian sebelumnya telah mencoba menyelesaikan masalah ini, namun pada penelitian ini, penulis akan lebih berfokus pada penerapan metode *robust* dengan pendekatan Estimasi M dan Estimasi LTS yang masih belum banyak dibahas. Alasan mengapa penulis memilih Estimasi M dengan pembobot *Huber* dan Estimasi LTS sebab merupakan solusi yang tepat untuk mengatasi masalah pencilan dalam data. Hal ini karena kedua metode lebih unggul daripada Metode Kuadrat Terkecil (MKT) sebab memiliki *breakdown point* 50%, yang berarti bisa memberikan hasil estimasi yang stabil dan akurat. Metode ini dirancang untuk lebih toleran terhadap pencilan dan memberikan estimasi yang lebih akurat dibandingkan dengan metode regresi biasa.

Hal ini mendorong penulis untuk menganalisis pencilan data menggunakan metode estimasi M dengan pembobot *Huber* dan metode estimasi LTS sebagai pen-

dekatan untuk mengatasi pencilan dalam data. Estimasi M dan Estimasi LTS adalah teknik *robust* yang mampu meminimalisir dampak pencilan pada analisis statistik, sedangkan pembobot *Huber* adalah fungsi bobot yang dirancang khusus untuk mengurangi pengaruh pencilan dengan efisien dengan mengkombinasikan kuadrat terkecil dan absolut terkecil untuk memberikan solusi yang *robust* terhadap *outlier*. Studi kasus dalam penelitian ini berfokus pada faktor-faktor yang mempengaruhi penyakit anemia.

Anemia adalah kondisi medis yang ditandai dengan rendahnya kadar hemoglobin dalam darah, yang dapat menyebabkan berbagai gejala seperti kelelahan, pucat, dan penurunan daya tahan tubuh. Berdasarkan data dari WHO, anemia mempengaruhi sekitar 40% anak usia 6–59 bulan dan 30% wanita usia 15–49 tahun.

Dalam penelitian ini, dilakukan analisis terhadap faktor-faktor yang mempengaruhi anemia dengan menggunakan metode regresi *robust*. Metode ini digunakan karena data yang dianalisis mengandung pencilan (*outlier*), yang dapat mengganggu hasil analisis jika hanya menggunakan metode regresi biasa. Dataset yang digunakan dalam studi ini berasal dari Mendeley Data dan mencakup variabel seperti kadar hemoglobin (Hb), *packed cell volume* (PCV), jumlah sel darah merah (RBC), serta indeks eritrosit seperti MCV, MCH, dan MCHC.

Tujuan dari penelitian ini adalah untuk membandingkan metode regresi *robust* estimasi M dan estimasi LTS dalam mendeteksi serta mengatasi pencilan data, sekaligus mengevaluasi faktor-faktor utama yang berkontribusi terhadap anemia. Hasil penelitian ini diharapkan dapat memberikan pendekatan yang lebih akurat dalam analisis data kesehatan serta mendukung pengembangan metode diagnosis anemia yang lebih efisien.

1.2. Rumusan Masalah

Berdasarkan pada latar belakang yang telah diuraikan, maka rumusan masalah penelitian adalah sebagai berikut:

1. Bagaimana performa metode regresi robust estimasi M dengan pembobot *Huber* dan estimasi *LTS* dalam mendeteksi serta mengatasi pencilan data?
2. Dari kedua metode tersebut, metode manakah yang menghasilkan model terbaik dalam menangani data dengan pencilan?
3. Apa saja faktor-faktor yang berkontribusi terhadap munculnya pencilan dalam dataset yang digunakan dalam penelitian ini?

1.3. Batasan Masalah

Agar fokus pada sasaran yang diharapkan maka perlu dilakukan pembatasan masalah dalam penelitian ini sebagai berikut:

1. Menganalisis *outlier* pada data dengan menggunakan Metode Regresi *Robust* estimasi M dan estimasi *LTS*.
2. Data yang digunakan adalah data anemia, dengan variabel independen yaitu RBC (X_1), PCV (X_2), MCV (X_3), MCH(X_4), MCHC (X_5). Sedangkan untuk variabel dependen adalah HGB (Y).
3. *Software* yang digunakan adalah R Studio versi R.4.4.2.

1.4. Tujuan

Berdasarkan pada latar belakang yang telah diuraikan, maka tujuan penelitian adalah sebagai berikut:

1. Membandingkan dan menganalisis pencilan data menggunakan metode regresi *robust* estimasi M dengan pembobot *Huber* dan estimasi *LTS* sebagai pendekatan untuk mengatasi pencilan dalam dataset.
2. Mengevaluasi efektivitas metode regresi *robust* parameter estimasi M dengan pembobot *Huber* dan parameter estimasi *LTS* dalam mengidentifikasi dan menangani pencilan data.
3. Untuk mengetahui faktor-faktor apa saja yang mempengaruhi dan untuk memperoleh hasil yang lebih stabil dan akurat pada kasus data *outlier*.

1.5. Manfaat

Manfaat dari penelitian ini adalah memberikan pemahaman yang lebih mendalam mengenai metode regresi *robust*, khususnya Estimasi M dan Estimasi *LTS*, dalam menangani pencilan data di bidang kesehatan. Serta membantu dalam pemilihan model terbaik untuk menganalisis data dengan pencilan, sehingga hasil penelitian lebih valid dan dapat digunakan untuk pengambilan keputusan yang lebih baik.

1.6. Tinjauan Pustaka

Rohmah et al. (2020) membahas tentang perbandingan metode Regresi *Robust* estimasi M dan estimasi *LTS* dalam menganalisis jumlah kasus tuberkulosis di Indonesia. Penelitian ini bertujuan untuk menentukan metode estimasi yang lebih baik dalam menganalisis data jumlah kasus tuberkulosis, sehingga dapat memberikan pemahaman yang lebih baik tentang faktor-faktor yang mempengaruhi penyakit tersebut. Variabel yang digunakan dalam penelitian ini adalah jumlah kasus tuberkulosis, persentase tempat pengelolaan makanan, jumlah kasus HIV, persentase

penduduk, dan jumlah tenaga kesehatan masyarakat. Metode yang digunakan pada penelitian ini adalah estimasi M dan estimasi LTS dengan Metode Kuadrat Terkecil (MKT). Hasil penelitian menunjukkan bahwa metode estimasi LTS lebih baik dibandingkan dengan metode estimasi M dalam mengestimasi parameter pada data jumlah kasus tuberkulosis di Indonesia tahun 2019.

Pradewi & Sudarno (2012) membahas tentang estimasi metode Regresi *Robust* menggunakan estimasi M dengan fungsi pembobot *Huber* dan *Bisquare Tukey* pada data ketahanan pangan di Jawa Tengah tahun 2007. Penelitian ini berfokus pada pengujian ketidakpuhan asumsi klasik dalam analisis regresi dan penanganan pencilan dalam data. Tujuan dilakukannya penelitian ini untuk mendapatkan model terbaik pada data ketahanan pangan yang mengandung pencilan dan mendeteksi pencilan dalam data, serta menguji ketidakpuhan asumsi klasik dalam analisis regresi. Variabel yang digunakan pada penelitian ini adalah rasio ketersediaan beras, stok beras, luas panen padi, rata-rata produksi, harga beras, dan jumlah konsumsi. Hasil penelitian menunjukkan bahwa estimasi M dengan pembobot *Bisquare Tukey* lebih baik dalam mengatasi pencilan data pada data ketahanan pangan di Jawa Tengah tahun 2007 dengan pencilan data yang terdeteksi di Kabupaten Cilacap, Kabupaten Sragen, dan Kota Semarang.

Akolo & Nadjamuddin (2022) membahas tentang analisis Regresi *Robust* estimasi LTS dan estimasi *Maximum Likelihood* pada pemodelan Indeks Pembangunan Manusia (IPM) di Sulawesi, serta faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia (IPM). Tujuan dilakukannya penelitian ini adalah untuk mengetahui faktor-faktor yang mempengaruhi IPM di Sulawesi dan menentukan metode terbaik dalam pemodelan IPM. Variabel yang digunakan pada penelitian ini adalah umur harapan hidup, harapan lama sekolah, rata-rata lama sekolah, penge-

luaran riil per kapita, dan IPM. Metode yang digunakan pada penelitian ini adalah estimasi LTS dan estimasi *Maximum Likelihood* dengan pembobot *Huber* dan *Tukey Bisquare*. Hasil penelitian menunjukkan bahwa metode Regresi *Robust* dengan estimasi LTS sebagai model terbaik.

Matdoan (2020) membahas tentang penggunaan metode Regresi *Quantil* dan Regresi *Robust Least Trimmed Square (LTS)* untuk menganalisis faktor-faktor yang mempengaruhi penyebaran penyakit malaria di Indonesia. Selain itu, metode *Ordinary Least Square (OLS)* juga digunakan sebagai pembanding. Tujuan utama dari penelitian ini adalah untuk membandingkan metode Regresi *Quantil* dan Regresi *Robust LTS* dalam mengidentifikasi faktor-faktor yang berkontribusi terhadap penyebaran malaria serta menentukan model regresi terbaik yang dapat menjelaskan fenomena tersebut. Variabel yang digunakan adalah *Annual Parasite Incidence (API)* malaria, persentase akses sanitasi layak, persentase rumah layak huni, persentase penduduk miskin, rasio tenaga dokter, persentase akses air bersih, dan rasio jumlah puskesmas. Hasil penelitian menunjukkan bahwa model terbaik diperoleh dengan Regresi *Quantil* yang lebih tinggi dibandingkan metode *OLS* dan Regresi *Robust LTS*.

Atamia et al. (2021) membahas tentang penggunaan analisis regresi *robust* dengan dua metode estimasi, yaitu estimasi *M* dan estimasi *S*, yang keduanya menggunakan pembobot *Huber* untuk mengatasi *outlier* dalam data. Tujuan penelitian ini adalah untuk menentukan metode terbaik dalam memodelkan Indeks Pembangunan Manusia (IPM) di Indonesia, dengan membandingkan nilai *Adjusted R-Square* tertinggi dan *Mean Square Error (MSE)* terkecil. Dalam penelitian ini, IPM dijadikan sebagai variabel terikat, sementara variabel bebas terdiri dari angka harapan hidup saat lahir (X_1), rata-rata lama sekolah (X_2), dan pengeluaran per kapita (X_3). Hasil

analisis menunjukkan adanya outlier pada observasi ke-11, 14, 18, dan 34 dalam data IPM Indonesia tahun 2017. Setelah dilakukan pemodelan, diperoleh kesimpulan bahwa estimasi M lebih efektif dibandingkan estimasi S , karena memiliki *Adjusted R-Square* lebih tinggi dan MSE lebih kecil.

Tabel 1.1 Tinjauan Pustaka

No	Penulis	Metode Penelitian	Perbedaan	Persamaan
1	Rohmah, D., Susanti, Y., & Zukhronah, E. (2020)	Regresi <i>robust</i> estimasi M dan estimasi LTS dalam menganalisis jumlah kasus tuberkulosis di Indonesia	Studi kasus yang dibahas berbeda dan ada menambahkan <i>OLS</i> sebagai pembanding	Membandingkan estimasi M dan estimasi LTS dengan tujuan menangani <i>outlier</i>
2	Pradewi, E. D. & Sudarno, S. (2012)	Regresi <i>robust</i> estimasi M dengan pembobot <i>Huber</i> dan <i>Bisquare Tukey</i> pada data ketahanan pangan di Jawa Tengah	Tidak menggunakan pembobot <i>Bisquare Tukey</i> dan data yang digunakan berbeda (bukan data pangan)	Menggunakan estimasi M dengan pembobot <i>Huber</i>
3	Akolo, I. R. & Nadjamuddin, A. (2022).	Regresi <i>robust</i> estimasi LTS dan estimasi <i>Maximum Likelihood</i> pada pemodelan IPM di Sulawesi	Metode yang digunakan untuk perbandingan berbeda dan data yang digunakan adalah sosial ekonomi agregat	Bertujuan untuk mencari model terbaik dalam pemodelan studi kasus
4	Matdoan, M. Y. (2020)	Regresi <i>Quantil</i> dan Regresi <i>Robust Least Trimmed Square (LTS)</i> untuk menganalisis faktor-faktor yang mempengaruhi penyebaran penyakit malaria di Indonesia	Mengutamakan perbandingan model untuk sebaran ekstrem dan menambahkan regresi <i>quantil</i>	Membandingkan estimasi LTS dengan metode lain
5	Atamia, N. A., Susanti, Y., & Handajani, S. S. (2021)	Regresi <i>Robust</i> estimasi M dan estimasi S untuk mengatasi <i>outlier</i> dalam data	Estimasi yang digunakan berbeda	Menggunakan estimasi M dengan pembobot <i>Huber</i>

1.7. Sistematika Penulisan

Sistematika pada penelitian ini yaitu:

BAB 1 PENDAHULUAN

Pada bab ini berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, tinjauan pustaka, dan sistematika penulisan.

BAB II DASAR TEORI

Pada bab ini menjelaskan terkait landasan teori yang menjadi acuan, berisi tentang konsep, hipotesis dan rancangan penelitian.

BAB III METODE PENELITIAN

Pada bab ini membahas terkait metode penelitian yang akan digunakan dalam pemecahan masalah yang berisi tentang pendekatan penelitian, sumber data, metode pengolahan data dan diagram analisis data.

BAB IV PEMBAHASAN

Pada bab ini membahas penjelasan tentang Uji Kecocokan Model *Goodness of Fit*, analisis regresi *robust* Estimasi M, Estimasi LTS, Fungsi Objektif dan Fungsi Pembobot, serta *breakdown point*.

BAB V STUDI KASUS

Pada bab ini akan dilakukan proses analisis data yang kemudian hasilnya diinterpretasikan secara terperinci.

BAB VI PENUTUP

Pada bab ini berisi kesimpulan dan saran dari penelitian yang dilakukan oleh penulis.

BAB VI

PENUTUP

6.1. Kesimpulan

Setelah melakukan analisis dan evaluasi terhadap data anemia menggunakan metode regresi *robust* Estimasi M dengan pembobot Huber dan Estimasi LTS, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Estimasi M dengan pembobot Huber dan Estimasi LTS menunjukkan kemampuan yang baik dalam mendeteksi dan mengatasi pencilan data. Estimasi LTS terbukti lebih unggul dalam dataset dengan banyak pencilan, sedangkan Estimasi M cocok untuk dataset dengan jumlah pencilan moderat.
2. Estimasi LTS dipilih sebagai metode terbaik berdasarkan nilai RSE (*Residual Standard Error*) yang lebih rendah dan nilai R^2 yang lebih tinggi dibandingkan MKT dan Estimasi M. Hal ini menunjukkan bahwa Estimasi LTS mampu menjelaskan variasi data dengan lebih baik.
3. Variabel seperti RBC, PCV, MCV, MCH, dan MCHC memiliki pengaruh signifikan terhadap kadar hemoglobin (Hb). Analisis regresi *robust* berhasil mengidentifikasi hubungan antara faktor-faktor ini dengan risiko anemia.

6.2. Saran

Berdasarkan hasil penelitian ini, disarankan untuk penelitian selanjutnya agar mencoba metode regresi *robust* lainnya seperti Estimasi S atau Estimasi MM

untuk membandingkan performanya dengan Estimasi M dan LTS sehingga dapat memberikan wawasan lebih lanjut tentang metode mana yang paling optimal dalam berbagai kondisi dataset. Selain itu, sebaiknya ditambahkan variabel-variabel lain yang relevan dengan anemia, seperti status gizi, jenis kelamin, usia, dan riwayat penyakit tertentu, untuk meningkatkan akurasi model prediksi. Di samping itu, temuan dari penelitian ini juga dapat digunakan sebagai dasar untuk mengembangkan model diagnosis anemia yang lebih efisien dan akurat, di mana faktor-faktor seperti RBC, PCV, MCV, MCH, dan MCHC dapat dijadikan indikator utama dalam deteksi dini anemia.

DAFTAR PUSTAKA

- Ahmed, M. H., Ghatge, M. S., & Safo, M. K. (2020). Hemoglobin: structure, function and allostery. *Vertebrate and invertebrate respiratory proteins, lipoproteins and other body fluid proteins*, pages 345–382.
- Akolo, I. R. & Nadjamuddin, A. (2022). Analisis regresi robust estimasi least trimmed square dan estimasi maximum likelihood pada pemodelan ipm di pulau sulawesi. *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi*, 10(2):211–221.
- Atamia, N. A., Susanti, Y., & Handajani, S. S. (2021). Perbandingan analisis regresi robust estimasi-s dan estimasi-m dengan pembobot huber dalam mengatasi outlier. In *PRISMA, Prosiding Seminar Nasional Matematika*, volume 4, pages 673–679.
- Chen, C. (2002). Robust regression and outlier detection with the robustreg procedure. In *Proceedings of the Proceedings of the Twenty-Seventh Annual SAS Users Group International Conference*. SAS Institute.
- Contesa, A. Y., Wathan, F. M., & Yunola, S. (2022). Hubungan pengetahuan, lama menstruasi dan status gizi dengan kejadian anemia pada mahasiswi kebidanan reguler di universitas kader bangsa palembang tahun 2022. *Jurnal Doppler*, 6(1):88–96.
- Damodar, N. G. (2021). Basic econometrics.
- Draper, N. (1998). *Applied regression analysis*. McGraw-Hill. Inc.

- Ghozali, I. (2005). Analisis multivariate dengan program spss. *Semarang: Badan Penerbit Universitas Diponegoro*, 12(2):33–56.
- Ghozali, I. (2006). *Aplikasi analisis multivariate dengan program SPSS*. Badan Penerbit Universitas Diponegoro.
- Hagul, P. et al. (1989). Penentuan variabel penelitian dan hubungan antar variabel “metode penelitian dan survey”. *Jakarta: LP3ES*.
- Huber, P. J. (1973). Robust regression: asymptotics, conjectures and monte carlo. *The annals of statistics*, pages 799–821.
- Huber, P. J. (1992). Robust estimation of a location parameter. In *Breakthroughs in statistics: Methodology and distribution*, pages 492–518. Springer.
- Kurniawati, L. D. (2011). Kekekaran regresi linier ganda dengan estimasi mm (method of moment) dalam mengatasi pencilan. *Skripsi. Yogyakarta: Universitas Negeri Yogyakarta*.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models*. McGraw-hill.
- Matdoan, M. Y. (2020). Estimasi parameter regresi quantil dengan regresi robust least trimmed square (lts)(studi kasus: Faktor-faktor yang mempengaruhi penyebaran penyakit malaria di indonesia). *Jurnal Riset dan Aplikasi Matematika (JRAM)*, 4(1):50–62.
- Mojumdar, M. U. (2024). Pediatric anemia dataset: Hematological indicators and diagnostic classification. <https://data.mendeley.com/datasets/y7v7ff3wpj/1>.

- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.
- Ortiz-Prado, E., Dunn, J. F., Vasconez, J., Castillo, D., & Viscor, G. (2019). Partial pressure of oxygen in the human body: a general review. *American journal of blood research*, 9(1):1.
- Pradewi, E. D. & Sudarno, S. (2012). Kajian estimasi-m irls menggunakan fungsi pembobot huber dan bisquare tukey pada data ketahanan pangan di jawa tengah. *Media Statistika*, 5(1):1–10.
- Rohmah, D., Susanti, Y., & Zukhronah, E. (2020). Perbandingan model regresi robust estimasi m dan estimasi least trimmed squares (lts) pada jumlah kasus tuberkulosis di indonesia. *Kontinu: Jurnal Penelitian Didaktik Matematika*, 4(2):136–146.
- Rousseeuw, P. & Yohai, V. (1984). Robust regression by means of s-estimators. In *Robust and Nonlinear Time Series Analysis: Proceedings of a Workshop Organized by the Sonderforschungsbereich 123 "Stochastische Mathematische Modelle", Heidelberg 1983*, pages 256–272. Springer.
- Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American statistical association*, 79(388):871–880.
- Rousseeuw, P. J. & Leroy, A. M. (2005). *Robust regression and outlier detection*. John wiley & sons.
- Ryan, T. P. (2008). *Modern regression methods*, volume 655. John Wiley & Sons.
- Sembiring, R. (2003). Analisis regresi edisi kedua. *Bandung: ITB*.

Setiawan, A. (2014). Segmentasi citra sel darah merah berdasarkan morfologi sel untuk mendeteksi anemia defisiensi besi.

Walpole, R. E. & Myers, R. H. (1995). Ilmu peluang dan statistika untuk insinyur dan ilmuwan. *Bandung: Itb*, pages 149–150.

Wijaya, S. (2009). Taksiran parameter pada model regresi robust dengan menggunakan fungsi huber. *Universitas Indonesia*.

World Health Organization (2023). *Accelerating anaemia reduction: a comprehensive framework for action*. World Health Organization.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of statistics*, pages 642–656.