PERBANDINGAN METODE *LEAST MEDIAN SQUARE* (LMS) DAN PENDUGA-*S* DALAM ANALISIS REGRESI *ROBUST*

(Skripsi)

Oleh

NURULITA INDRIYANI

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2021

ABSTRACT

COMPARISON OF THE LEAST MEDIAN SQUARE (LMS) METHOD AND THE S-ESTIMATION IN ROBUST REGRESSION ANALYSIS

By

Nurulita Indriyani

Linear regression analysis was used to measure the effect of more than one independent variable (X) on the dependent variable (Y). Estimating the parameters of regression analysis is generally solved by the Ordinary Least Square (OLS). However, when an outlier is found in the observational data, it can cause the regression line coefficient estimator of OLS to be inaccurate. In this case we need a robust regression method to overcome the problem. Least Median of Square (LMS) and S-estimator are two methods in robust regression. The LMS itself can be used to minimize the median number of squared errors, while the S-estimator is used to minimize the scale sample on the specified residual. The purpose of this study is to empirically examine the robust LMS method and the S-estimator in estimating the regression parameters on data containing outliers through simulation data with several percentages of outliers, namely 10%, 15%, and 20% with sample sizes of 20, 50, 100 and 200, determine the best robust regression method between the LMS method and the S-estimator method by looking at the MSE value of each estimator. Based on the results of the study, it was concluded that MSE of LMS value was better than S-estimator in estimating the regression parameters containing outliers.

Keywords: OLS, LMS, S-estimator, Outlier, Robust

ABSTRAK

PERBANDINGAN METODE *LEAST MEDIAN SQUARE* (LMS) DAN PENDUGA-*S* DALAM ANALISIS REGRESI *ROBUST*

Oleh

Nurulita Indriyani

Analisis regresi linear digunakan untuk mengukur pengaruh lebih dari satu variabel bebas (*X*) terhadap variabel terikat (*Y*). Penduga parameter analisis regresi pada umumnya diselesaikan dengan Metode Kuadrat Terkecil (MKT). Namun, ketika ditemukan sebuah pencilan pada data pengamatan, dapat menyebabkan penduga koefisien garis regresi dengan MKT menjadi tidak tepat. Sehingga diperlukan metode regresi *robust* untuk mengatasi masalah tersebut. *Least Median of Square* (LMS) dan Penduga-*S* merupakan dua metode dalam regresi *robust.* LMS sendiri dapat digunakan untuk meminimumkan jumlah median kuadrat galat, sedangkan Penduga-*S* digunakan untuk dengan meminimumkan sampel skala pada residual yang ditetapkan. Tujuan dari penelitian ini adalah mengkaji secara empiris metode *robust* LMS dan Penduga-*S* dalam menduga parameter regresi pada data yang mengandung pencilan melalui data simulasi dengan beberapa persentase pencilan yaitu 10%, 15%, dan 20% dengan ukuran sampel 20, 50, 100 dan 200 serta menentukan metode regresi *robust* terbaik di antara metode LMS dan metode Penduga-*S* dengan melihat nilai MSE dari masing-masing penduga. Berdasarkan hasil penelitian diperoleh kesimpulan bahwa nilai MSE LMS lebih baik dibandingkan Penduga-*S* dalam menduga parameter regresi yang mengandung pencilan.

Kata Kunci: MKT, LMS, Penduga-*S*, Pencilan, *Robust*

PERBANDINGAN METODE *LEAST MEDIAN SQUARE* (LMS) DAN PENDUGA-*S* DALAM ANALISIS REGRESI *ROBUST*

Oleh

NURULITA INDRIYANI

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2021

Judul Skripsi : **PERBANDINGAN METODE *LEAST MEDIAN SQUARE* (LMS) DAN PENDUGA-*S* DALAM ANALISIS REGRESI *ROBUST***

Nama Mahasiswa : *Nurulita Indriyani*

Nomor Pokok Mahasiswa: 1757031019

Program Studi : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI,

1. Komisi Pembimbing

Dr. Khoirin Nisa, M.Si.
NIP. 19740726 200003 2 001

Drs. Tiryono Ruby, M.Sc., Ph.D.
NIP. 19620704 198803 1 002

2. Ketua Jurusan Matematika

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

MENGESAHKAN

1. Tim Penguji

   Ketua : **Dr. Khoirin Nisa, M.Si.** ....

   Sekretaris : **Drs. Tiryono Ruby, M.Sc., Ph.D.** ....

   Penguji
   Bukan Pembimbing : **Widiarti, S.Si., M.Si.** ....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

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

Tanggal Lulus Ujian Skripsi : **11 Oktober 2021**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama	: **Nurulita Indriyani**
Nomor Pokoknya Mahasiswa	: **1757031019**
Jurusan	: **Matematika**
Judul Skripsi	: **Perbandingan Metode *Least Median Square* (LMS) dan Penduga-*S* dalam Analisis Regresi *Robust***

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah karya penulisan ilmiah Universitas Lampung.

Bandar Lampung, 11 Oktober 2021
Penulis

Nurulita Indriyani
NPM. 1757031019

RIWAYAT HIDUP

Penulis bernama lengkap Nurulita Indriyani, lahir di Bandar Lampung pada tanggal 10 Desember 1998. Penulis merupakan anak ketiga dari tiga bersaudara yang lahir dari pasangan Bapak Malik, ST dan Ibu Ir. Heni Nugrahani.

Penulis menempuh pendidikan sekolah dasar di SD Negeri 1 Sukamaju pada tahun 2005. Kemudian melanjutkan ke jenjang sekolah menegah pertama di SMP Negeri 3 Bandar Lampung pada tahun 2011-2014, dan jenjang sekolah menengah atas di SMA YP Unila Bandar Lampung pada tahun 2014-2017.

Pada tahun 2020, Penulis melaksanakan Kuliah Kerja Nyata (KKN) di Desa Toto Multo, Kecamatan Way Bungur, Kabupaten Lampung Timur, Provinsi Lampung, sebagai bentuk pengabdian kepada masyarakat. Pada tahun 2020 penulis melaksanakan Kerja Praktek (KP) di Badan Pusat Statistik (BPS). Pengalaman organisasi penulis yaitu menjadi staf ahli bidang PPW Mahasiswa (BEM) Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung tahun 2019.

KATA INSPIRASI

"Karena sesungguhnya sesudah kesulitan itu ada kemudahan."

(Q.S. Al-Insyirah : 5-6)

"Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya."

(Q.S. Al-Baqarah : 286)

"Tetap sabar, semangat, dan tersenyum, karena kamu sedang menimba ilmu di Universitas kehidupan. Allah menaruhmu di tempatmu yang sekarang bukan karena kebetulan."

(Dahlan Iskan)

"Bersemangatlah melakukan hal yang bermanfaat untukmu dan meminta tolonglah pada Allah, serta janganlah engkau malas"

(HR. Muslim)

"Ketika kau melakukan usaha mendekati cita-citamu, di waktu yang bersamaan cita-citamu juga sedang mendekatimu. Alam semesta bekerja seperti itu "

(Fiersa Besari)

PERSEMBAHAN

Puji Sykur kepada Allah SWT, Karena atas limpahan berkah, rahmad, dan karunia-Nya skripsi ini dapat diselesaikan.

Ku persembahkan karya sederhana penuh perjuangan dan kesabaran ini kepada:

Ayahanda Malik, ST dan Ibunda Ir. Heni Nugrahani

Terimakasih atas limpahan kasih sayang, pengorbanan, semangat, motivasi, serta doa dan sujud yang selalu menantikan keberhasilanku dengan sabar dan penuh pengertian. Karena atas doa dan ridho kalian, Allah memudahkan setiap perjalanan hidup ini. Terimalah bukti kecil ini sebagai kado keseriusanku untuk membalas semua pengorbanan, keikhlasan, dan jerih payah yang selama ini kalian lakukan.

Almamater yang kucintai, Universitas Lampung

SANWACANA

Penulis mengucapkan puji syukur kehadirat Allah SWT, karena dengan ridho dan karunia-Nya serta atas berkah dan rahmat-Nya sehinga penulis dapat menyelesaikan skripsi yang berjudul "Perbandingan Metode *Least Median Square* (LMS) dan Penduga-*S* dalam Analisis Regresi *Robust*". Selesainya penulisan skripsi ini adalah berkat motivasi, pengarahan serta bimbingan dari berbagai pihak. Dengan segala kerendahan dan ketulusan hati penulis ingin menyampaikan ucapan terima kasih kepada:

1. Ibu Dr. Khoirin Nisa, M.Si, selaku Pembimbing pertama atas saran, bimbingan, arahan, motivasi, dan kesabaran dalam membimbing penulis selama penelitian hingga penyelesaian skripsi dan memberi arahan kepada penulis selama menuntut ilmu di Universitas Lampung.
2. Bapak Drs. Tiryono Ruby, M.Sc., Ph.D., selaku Pembimbing kedua yang telah memberikan arahan, saran, serta dukungan bagi penulis.
3. Ibu Widiarti, S.Si., M.Si, selaku Pembahas yang telah memberikan ide, kritik dan saran sehingga terselesaikannya skripsi ini.
4. Bapak Subian Saidi, S.Si., M.Si., selaku dosen Pembimbing Akademik yang telah membimbing penulis dalam menyelesaikan permasalahan seputar akademik.
5. Bapak Dr. Aang Nuryaman, M.Si., selaku Kepala Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung.

6. Bapak Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T., selaku Dekan FMIPA Universitas Lampung.
7. Para Dosen dan Staff Jurusan Matematika FMIPA Universitas Lampung.
8. Bapak, Ibu, Kakak-kakak, dan keluarga tercinta yang selalu memberikan motivasi, semangat, dan doa yang tak terhingga kepada penulis.
9. Sahabat-sahabat penulis Rossa, Nyoman, Putri, Nupus, Sekar, dan Mutia yang telah membantu, memberikan semangat dan kecerian pada penulis.
10. Teman-temanku Nita, Dhea, Indah Suciati, Dewi, Epmi, dan Dinda yang telah memberikan keceriaan dan semangat bagi penulis.
11. Teman-teman Matematika 2017 yang selalu menjadi semangat bagi penulis.
12. Semua pihak yang terlibat dalam penyelesaian skripsi yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa masih ada kekurangan dari skripsi ini, akan tetapi besar harapan semoga skripsi ini dapat bermanfaat bagi kita semua.

Bandar Lampung, 11 Oktober 2021
Penulis

Nurulita Indriyani

DAFTAR ISI

Halaman

DAFTAR TABEL

Tabel Halaman

DAFTAR GAMBAR

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis regresi merupakan salah satu metode statistika yang digunakan untuk mencari sebuah hubungan antara dua variabel atau lebih. Variabel terikat disebut juga variabel *dependent* yaitu variabel yang keberadaannya dipengaruhi oleh variabel lainnya dan dinotasikan dengan Y. Variabel bebas disebut juga variabel *independent* yaitu variabel yang tidak dipengaruhi oleh variabel lainnya dan dinotasikan dengan X. Analisis regresi adalah metode yang sangat populer untuk mengetahui hubungan linear antara variabel respon dengan satu atau lebih peubah bebas.

Metode pendugaan parameter regresi yang paling populer adalah Metode Kuadrat Terkecil (*least square*). Metode ini menduga parameter regresi dengan cara meminimumkan jumlah kuadrat galat dan bersifat BLUE (*Best Linear Unbiased Estimasi)*. Tetapi, tak jarang ditemukannya data yang menyimpang dari sekumpulan data lainnya yang disebut pencilan (*outlier)* yang dapat berpengaruh pada model. Oleh karena itu, dibutuhkan uji kekekaran regresi (*robust*) yang bersifat tidak sensitif terhadap pelanggaran asumsi-asumsi pencilan.

Pencilan (*outlier*) merupakan data pengamatan yang mungkin adanya data terkontaminasi, yaitu adanya kesalahan pada data melakukan pengambilan sampel pada suatu populasi. Menurut Olive (2005), terkadang untuk mengatasi *outlier* seorang peneliti melakukan transformasi terhadap suatu data.

Namun seringkali asumsi tersebut tidak terpenuhi meskipun telah dilakukan transformasi yang pada akhirnya tetap membiaskan pendugaan. Oleh karena itu, ketika data mengandung pencilan diperlukan suatu metode yang resisten terhadap pencilan, yaitu metode *robust*.

Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi dari suatu residual tidak normal atau mengandung beberapa pencilan yang berpengaruh pada model. Sehingga, dapat menghasilkan model yang *robust* atau resisten terhadap pencilan. Berbagai penelitian terkait regresi robust telah dikemukakan oleh para peneliti, diantaranya yaitu oleh Herawati & Nisa (2010), Nisa (2007), dan Khotimah, *et al*. (2020).

Least Median Square (LMS) adalah salah satu metode estimasi dalam regresi *robust* yang digunakan untuk mengatasi *outlier*. Dalam metode ini, dengan meminimumkan median kuadrat sisaannya, penduga yang dihasilkan akan lebih resisten terhadap *outlier*. Suatu estimasi yang resisten adalah relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data. Penelitian ini bertujuan untuk mempelajari dan memahami tentang regresi *robust*. Selain itu tujuan dari penelitian ini adalah untuk mendeteksi *outlier* pada data yang digunakan dan menentukan persamaan regresi linear berganda yang mengandung *outlier* dengan menggunakan metode regresi *robust* LMS.

Penduga-*S* bertujuan untuk meminimumkan sampel skala-*S* pada residual yang sudah ditetapkan (Sakata & Preminger, 2007). Pendugaan parameter dengan Penduga-*S* dapat menghasilkan penduga yang bersifat *robust* terhadap *outlier* berpengaruh. Berdasarkan uraian tersebut, maka peneliti akan mengkaji secara empiris metode *robust* LMS dan Penduga-*S* dalam menduga parameter regresi pada data yang mengandung pencilan. Setelah itu, akan membandingkan metode regresi *robust* terbaik di antara metode LMS dan metode Penduga-*S*.

1.2 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dalam penulisan penelitian ini adalah:

1. Mengkaji secara empiris metode *robust Least Median Square* (LMS) dan Penduga-*S* dalam menduga parameter regresi pada data yang mengandung pencilan.
2. Menentukan metode regresi *robust* terbaik di antara metode *Least Median Square* (LMS) dan metode Penduga-*S*.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah untuk menambah wawasan dan memberi masukan kepada para peneliti serta diharapkan dapat memberikan metode alternatif lain bagi pembaca dalam melakukan menganalisa data yang mengandung pencilan.

II. TINJAUAN PUSTAKA

2.1 Matriks

Matriks merupakan susunan bilangan persegi atau persegi panjang dari bilangan-bilangan yang diatur dalam baris dan kolom. Bilangan atau fungsi tersebut disebut entri atau elemen matriks. Lambang matriks dilambangkan dengan huruf besar, sedangkan entri (elemen) matriks dilambangkan dengan huruf kecil. Matriks dapat ditulis sebagai berikut.

$$\boldsymbol{A_{mxn}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Susunan diatas disebut mariks (*mxn*), karena memiliki *m* baris dan *n* kolom. Elemen-elemen matriks berupa bilangan real (Hadley, 1992).

2.1.1 Jenis-Jenis Matriks

Menurut Trihastuti (2014), jenis-jenis matriks, yaitu:

a. Matriks Persegi

Matriks persegi adalah matriks yang memiliki jumlah baris sama dengan jumlah kolom.

b. Matriks Identitas

Matriks identitas adalah matriks yang semua elemen diagonal utamanya bernilai satu dan elemen di luar diagonal utama bernilai nol.

Contoh: $\boldsymbol{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

c. Matriks Simetris

Matriks simetris adalah matriks persegi yang elemennya simetris secara diagonal. Matriks **A** dikatakan simetris jika $a_{ij} = a_{ji}$ untuk semua *i* dan *j*, dengan a_{ij} menyatakan unsur pada baris ke *i* dan kolom ke *j*.

Matriks yang simetris dapat dikatakan pula sebagai matriks yang transposenya sama dengan dirinya sendiri.

d. Matriks Ortogonal

Matriks ortogonal merupakan matriks yang apabila dikalikan dengan matriks tranposenya menghasilkan matriks satuan (identitas).

Matriks ortogonal dapat dituliskan sebagai berikut.

$$\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^T = I$$

Sifat matriks ortogonal:

1) Invers matriks ortogonal juga matriks ortogonal
2) Hasil kali matriks-matriks ortogonal juga matriks ortogonal
3) Jika **A** matriks ortogonal, maka $\det(\boldsymbol{A}) = 1$ atau $\det(\boldsymbol{A}) = -1$

e. Matriks *Singular*

Matriks *singular* adalah matriks yang determinannya bernilai nol atau tidak memiliki invers.

f. Matriks *Nonsingular*

Matriks *nonsingular* adalah matriks yang determinannya bernilai tidak sama dengan nol atau memiliki invers.

2.1.2 *Transpose* Matriks

Menurut Hadley (1992), *transpose* dari matriks **A** adalah matriks yang dibentuk dari **A** dengan mempertukarkan baris-baris dari kolom-kolom sehingga baris *i* dari **A** menjadi kolom *i* dari matriks *transpose*. *Transpose* dinotasikan dengan $\boldsymbol{A}^T = \boldsymbol{A}^t = \boldsymbol{A}'$. Jika **A** adalah matiks *m x n*, maka $\boldsymbol{A}^T$ adalah matriks *n x m*.

2.1.3 Operasi Matriks

Berikut beberapa operasi matriks, yaitu:

a. Penjumlahan dan Pengurangan Matriks

Operasi penjumlahan matriks didefinisikan dengan memisalkan **A** = $(\boldsymbol{a}_{ij})$ dan **B** = $(\boldsymbol{b}_{ij})$ merupakan dua matriks berukuran sama *m x n* (Trihastuti, 2014). Jumlah matiks **A** dan **B** ditulis **A**+**B** adalah matriks berukuran *m x n* dengan elemennya merupakan jumlah elemen yang seletak dari kedua matriks. Berlaku pula pada pengurangan matriks. Dalam hal ini ditulis sebagai berikut.

$$\boldsymbol{A} + \boldsymbol{B} = (\boldsymbol{a}_{ij} + \boldsymbol{b}_{ij})$$

$$\boldsymbol{A} - \boldsymbol{B} = (\boldsymbol{a}_{ij} - \boldsymbol{b}_{ij})$$

b. Perkalian Skalar dengan Matriks

Menurut Trihastuti (2014), jika diketahui matriks **A** dan c merupakan bilangan. Matriks *c***A** adalah matriks yang diperoleh dengan mengalikan setiap elemen dari matriks **A** dengan *c*. Dalam hal ini ditulis sebagai berikut.

$$c\mathbf{A} = (c\boldsymbol{a}_{ij})$$

c. Determinan Matriks

Misalkan $\boldsymbol{A} = (\boldsymbol{a}_{ij})$ adalah matriks berukuran *n x n*, determinan dari $\boldsymbol{A}$ didefinisikan sebagai berikut.

$$\boldsymbol{det}(\boldsymbol{A}) = \sum \boldsymbol{a_{ij}}(-\mathbf{1})^{\boldsymbol{i+j}}\boldsymbol{M_{ij}} \tag{2.1}$$

Dengan $\boldsymbol{M}_{ij}$ adalah minor dari $\boldsymbol{a}_{ij}$ yaitu determinan sub matriks $\boldsymbol{A}$ yang diperoleh dengan cara membuang semua entri pada baris ke-*i* dan semua entri pada kolom ke-*j*.

Jika $\boldsymbol{A} = (\boldsymbol{a}_{ij})$ adalah matriks berukuran *n x n* yang mengandung sebaris bilangan nol, maka $|\boldsymbol{A}| = 0$. Jika $\boldsymbol{A} = (\boldsymbol{a}_{ij})$ matriks segitiga, maka $|\boldsymbol{A}|$ adalah hasil kali elemen-elemen diagonal utama, yaitu $\boldsymbol{A} = \boldsymbol{a}_{11}\boldsymbol{a}_{22}\boldsymbol{a}_{33}\cdots\boldsymbol{a}_{mm}$. Jika $\boldsymbol{A} = (\boldsymbol{a}_{ij})$ adalah sebarang matriks persegi, maka $|\boldsymbol{A}| = |\boldsymbol{A}^{\boldsymbol{T}}|$, dimana $\boldsymbol{A}^{\boldsymbol{T}}$ adalah transpose dari **A**. Jika $\boldsymbol{A} = (\boldsymbol{a}_{ij})$ dan jika $\boldsymbol{B} = (\boldsymbol{b}_{ij})$ adalah matriks persegi yang ordonya sama, maka $|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$.

d. Invers Matriks

Diberikan matriks bujur sangkar $\boldsymbol{A}$. Jika terdapat matriks bujur sangkar $\boldsymbol{A}^{-1}$ yang memenuhi hubungan $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$, maka $\boldsymbol{A}^{-1}$ disebut matriks *invers* atau kebalikan dari $\boldsymbol{A}$.

Menurut Hadley (1992), *invers* dari matriks $\boldsymbol{A}$ dapat dinyatakan dengan:

$$\boldsymbol{A}^{-\mathbf{1}} = \frac{\mathbf{1}}{|\boldsymbol{A}|}\boldsymbol{adj}(\boldsymbol{A}); \boldsymbol{A} \neq \mathbf{0} \tag{2.2}$$

2.2 Analisis Regresi

Analisis regresi adalah metode statistika yang digunakan untuk mengetahui hubungan linear suatu variabel tak bebas (*Y*) dengan satu atau lebih variabel bebas (*X*). Menurut Draper & Smith (1992), salah satu tujuan analisis regresi yaitu menentukan model regresi yang baik.

Sehingga model dapat digunakan untuk menerangkan dan memprediksi hal-hal yang berhubungan dengan variabel-variabel yang terlibat di dalam model regresi.

Menurut Draper & Smith (1992), model regresi linear secara umum dapat ditulis sebagai berikut.

$$\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{e}\ ,\ \mathbf{e} \sim \mathrm{N}(\mathbf{0},\boldsymbol{\sigma^2 I}) \tag{2.3}$$

dengan:

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \\ \vdots \\ \boldsymbol{y_n} \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} \mathbf{1} & \boldsymbol{x_{11}} & \boldsymbol{x_{12}} & \cdots & \boldsymbol{x_{1p}} \\ \mathbf{1} & \boldsymbol{x_{21}} & \boldsymbol{x_{22}} & \cdots & \boldsymbol{x_{2p}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1} & \boldsymbol{x_{n1}} & \boldsymbol{x_{n2}} & \cdots & \boldsymbol{x_{np}} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \\ \vdots \\ \boldsymbol{\beta_p} \end{bmatrix}, \boldsymbol{e} = \begin{bmatrix} \boldsymbol{e_1} \\ \boldsymbol{e_2} \\ \vdots \\ \boldsymbol{e_n} \end{bmatrix}$$

dimana:

$\boldsymbol{Y}_{(nx1)}$: vektor variabel tak bebas

$\boldsymbol{X}_{(nx1)}$: matriks variabel bebas

$\boldsymbol{\beta}_{(nx1)}$: vektor parameter yang diduga

$\boldsymbol{e}_{(nx1)}$: vektor galat

Berdasarkan jumlah variabel bebasnya, analisis regresi linear dibagi menjadi dua, yaitu:

2.2.1 Analisis Regresi Linear Sederhana

Analisis regresi linear sederhana digunakan untuk mengetahui pengaruh antara satu variabel bebas terhadap satu variabel tak bebas. Menurut Sembiring (2003), model regresi linear sederhana dapat digunakan dalam persamaan dibawah ini.

$$Y = \beta_0 + \beta_1 X_1 + e \tag{2.4}$$

dimana:

Y	: variabel tak bebas
β_0, β_1	: parameter regresi
X_1	: variabel bebas
e	: galat

2.2.2 Analisis Regresi Linear Berganda

Analisis regresi linear berganda digunakan untuk menyelidiki hubungan di antara dua atau lebih variabel bebas X terhadap satu variabel tak bebas Y. Menurut Draper & Smith (1992), bentuk umum persamaan model regresi linear berganda dengan p jumlah variabel bebas dapat ditulis sebagai berikut.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + e \qquad (2.5)$$

dimana:

Y	: variabel tak bebas
$\beta_0, \beta_1, \beta_2, \ldots, \beta_p$	: parameter regresi
$X_1, X_2, \ldots, X_p$	: variabel bebas
e	: galat

2.3 Uji Asumsi Analisis Regresi

Menurut Imam Ghozali (2011), uji asumsi klasik terhadap model regresi linear digunakan untuk menentukan apakah model regresi itu baik atau tidak.
Tujuan pengujian asumsi klasik adalah untuk menentukan persamaan regresi yang diperoleh memiliki akurasi dalam estimasi, tidak bias, dan konsisten. Sebelum melakukan analisis regresi terlebih dahulu dilakukan pengujian asumsi.

2.3.1 Uji Normalitas

Analisis regresi mengasumsikan bahwa galat (ε_i) berdistribusi normal. Menurut Gujarati (1978), pada regresi linear klasik diasumsikan bahwa tiap ε_i berdistribusi normal dengan $\varepsilon_i \sim N(0, \sigma^2)$. Uji Normalitas dapat dilakukan dengan beberapa cara, diantaranya ialah *Normal Probability Plot*, uji *Kolmogorov-Smirnov*, dan *Anderson-Darling*. Tidak terpenuhinya normalitas, terjadi karena adanya beberapa data yang merupakan pencilan atau terdapat nilai ekstrim dalam data yang digunakan.

2.3.2 Uji Linearitas

Menurut Suliyanto (2008), diperlukan uji linearitas untuk mengetahui apakah model yang ditentukan merupakan model linear atau tidak. Uji linearitas dilakukan untuk memperoleh informasi apakah model empiris sebaiknya linear, kuadrat, atau kubik. Jika ditentukan bahwa model regresi salah, nilai prediksi nilai yang dihasilkan akan sangat menyimpang sehingga nilai prediksi akan menjadi bias.

Menurut Suliyanto (2008), uji Lagrange Multipler (LM-Test) merupakan salah satu metode yang digunakan untuk mengukur linearitas yang dikembangkan oleh Engle pada tahun 1982. Prinsip metode ini adalah membandingkan antara nilai X^2_{hitung} dengan nilai X^2_{tabel} dengan df=(n,α).

2.3.3 Uji Keberartian Simultan

Menurut Suliyanto (2008), uji keberartian simultan digunakan untuk menguji ketepatan model. Uji signifikansi simultan sering disebut dengan uji-F, digunakan untuk menguji apakah variabel bebas yang digunakan dalam model secara simultan (bersama-sama) mampu menjelaskan perubahan nilai variabel tak bebas atau tidak. Kriteria pengujian keberartian simultan yaitu jika $F_{hitung} \geq F_{tabel}$ atau nilai sig $(\hat{\alpha}) <$ 0,05 maka variabel bebas yang digunakan dalam model secara simultan (bersama sama) mampu menjelaskan perubahan nilai variabel tak bebas. Nilai F_{tabel} diperoleh dari Tabel distribusi *F* dengan df: α,(k-1),(n-k).

$$F = \frac{R^2/(k-1)}{1-R^2/(n-k)} \tag{2.6}$$

2.3.4 Uji Keberartian Parsial

Pengujian keberartian parsial dilakukan untuk menentukan keberartian masing-masing variabel bebas terhadap variabel tak bebas. Uji keberartian parsial menggunakan nilai t_{hitung}, dengan kriteria pengujian jika $t_{hitung} > t_{tabel}$ atau nilai sig $(\hat{\alpha}) < 0{,}05$ maka variabel bebas memiliki pengaruh yang berarti terhadap variabel tak bebas. Nilai t_{tabel} diperoleh dengan df: α,(n-k). Nilai t_{hitung} dapat diperoleh menggunakan rumus sebagai berikut.

$$t_{hitung} = \frac{b_j}{sb_j} \tag{2.7}$$

dimana:

b_j : koefisien regresi

sb_j : kesalahan baku koefisien regresi

2.3.5 Uji Heteroskedastisitas

Homoskedastisitas merupakan distribusi galat yang harus konstan (homogen). Jika variabel galat dalam model tidak konstan disebut dengan heteroskedastisitas.

Heteroskedastisitas disebabkan karena adanya variabel yang digunakan memiliki nilai yang sangat beragam, sehingga menghasilkan nilai galat yang tidak konstan. Untuk menguji adanya masalah heteroskedastisitas dapat dilakukan dengan menggunakan metode Glejser. Uji Glejser (Suliyanto, 2008), dilakukan dengan meregresikan semua variabel bebas terhadap nilai mutlak residualnya. Jika terdapat pengaruh variabel bebas yang signifikan terhadap nilai mutlak residualnya (sig < 0,05) maka dalam model terdapat masalah heteroskedastisitas.

2.3.6 Uji Autokorelasi

Uji autokorelasi bertujuan untuk mengetahui ada tidaknya masalah autokorelasi dalam persamaan regresi. Autokorelasi terjadi karena adanya galat yang berhubungan antar pengamatan tidak berkaitan satu sama lain atau saling bebas. Menurut Gujarati (Suliyanto, 2008) ada beberapa cara untuk mendeteksinya, salah satunya yaitu Uji *Durbin Watson* (Uji *DW*). Rumus yang digunakan untuk Uji *DW* adalah sebagai berikut.

$$DW = \frac{\sum (e - e_{t-1})^2}{\sum e} \quad (2.8)$$

dimana:

DW : nilai *Durbin-Watson Test*

e : nilai residual

e_{t-1} : nilai residual satu baris/periode sebelumnya

Adapun kaidah keputusan dalam uji DW adalah:

1. Jika nilai DW terletak antara batas atas (*upper bound*) d_u dan $4 - d_u$, maka tidak terjadi autokorelasi.

2. Jika nilai DW lebih rendah daripada batas bawah (*lower bound*) d_l, maka ada autokorelasi positif.
3. Jika nilai DW terletak antara batas $4 - d_l$ dan 4, maka ada autokorelasi negatif.

2.3.7 Uji Multikolinearitas

Uji multikolinearitas bertujuan untuk menentukan bahwa tidak adanya hubungan yang kuat antar variabel bebas. Multikolinearitas terjadi karena adanya korelasi yang cukup tinggi diantara variabel bebas (Montgomery & Peck, 1992). VIF (*Variance Inflation Factor*) merupakan metode untuk mengetahui ada tidaknya multikolinearitas antar variabel. Dengan rumus sebagai berikut:

$$VIF = \frac{1}{1-R_m^2} \tag{2.9}$$

Dengan $m = 1,2, \dots, p$ dan p adalah banyaknya variabel bebas. R_m^2 merupakan koefisien determinasi yang dihasilkan regresi variabel bebas X_m dengan variabel bebas lainnya. Nilai VIF semakin besar jika terdapat korelasi yang semakin besar diantara variabel bebas. Jika nilai VIF lebih dari 10, maka terjadi multikolinearitas yang kuat.

2.4 Metode Kuadrat Terkecil (MKT)

Metode kuadrat terkecil atau yang biasa disebut dengan *Ordinary Least Square* (OLS) merupakan metode pendugaan dengan meminimumkan jumlah kuadrat galat. Dibandingkan dengan metode lainnya, metode ini telah banyak digunakan dalam pembentukan model regresi atau penduga parameter regresi.

Menurut Sembiring (2003), pendugaan parameter regresi pada *p* data pengamatan MKT diperoleh dengan fungsi meminimumkan berikut.

$$J = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \ldots - \beta_p X_{ip})^2 \quad (2.10)$$

Pada persamaan (2.10) nilai *X* dan *Y* berasal dari pengamatan. Jika J berubah diturunkan terhadap β_0,β_1,β_2,…,β_p kemudian menyamakannya dengan nol, sehingga diperoleh (Sembiring, 2003):

$$\frac{\partial J}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_p x_{ip}) = 0, \quad (2.11)$$

$$\frac{\partial J}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_p X_{ip}) x_{i1} = 0, \quad (2.12)$$

$$\vdots$$

$$\frac{\partial J}{\partial \beta_p} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_p x_{ip}) x_{ip} = 0, \quad (2.13)$$

Nilai β_0,β_1,β_2,…,β_p pada Persamaan (2.11), (2.12), dan (2.13) diganti dengan masing-masing penduganya, yaitu $\hat{\beta}_0$,$\hat{\beta}_1$,$\hat{\beta}_2$,…,$\hat{\beta}_p$. Sehingga dapat disederhanakan menjadi:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} + \cdots + \hat{\beta}_p \sum_{i=1}^{n} x_{ip} = \sum_{i=1}^{n} y_i \quad (2.14)$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_1 \sum_{i=1}^{n} {x_{i1}}^2 + \hat{\beta}_2 \sum_{i=1}^{n} x_{i1} x_{i2} + \cdots + \hat{\beta}_p \sum_{i=1}^{n} x_{i1} x_{ip} = \sum_{i=1}^{n} x_{i1} y_i \quad (2.15)$$

$$\vdots$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{ip} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} x_{ip} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} x_{ip} + \cdots + \hat{\beta}_p \sum_{i=1}^{n} {x_{ip}}^2 = \sum_{i=1}^{n} x_{ip} y_i \quad (2.16)$$

Apabila dinyatakan dalam bentuk matriks, persamaan normal (2.14), (2.15), (2.16) menjadi:

$$\boldsymbol{X^T X \widehat{\beta}} = \boldsymbol{X^T Y} \quad (2.17)$$

dengan:

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_p \end{bmatrix}$$

$$\boldsymbol{X^T X} = \begin{bmatrix} n & \sum_{i=1}^{n} x_{i1} & \cdots & \sum_{i=1}^{n} x_{ip} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}{}^2 & \cdots & \sum_{i=1}^{n} x_{i1}x_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_{ip} & \sum_{i=1}^{n} x_{i1}x_{ip} & \cdots & \sum_{i=1}^{n} x_{ip}{}^2 \end{bmatrix}$$

$$\boldsymbol{X^T Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_1 \\ \sum_{i=1}^{n} x_{i1}y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ip}y_i \end{bmatrix}$$

Untuk menyelesaikan persamaan (2.17) kalikan kedua sisinya dengan invers dari $\boldsymbol{X^T X}$. Sehingga diperoleh:

$$(\boldsymbol{X^T X})^{-1}\boldsymbol{X^T X}\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T Y}$$

$$\boldsymbol{I}\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T Y}$$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T Y} \tag{2.18}$$

2.5 Regresi *Robust*

Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi galat tidak normal atau terdapat beberapa pencilan yang mempengaruhi model. Dengan menggunakan metode ini, dapat dihasilkan model *robust* atau resisten terhadap pencilan. Estimasi resisten adalah relatif tidak terpengaruh oleh perubahan besar di sebagian kecil data atau perubahan kecil di sebagian besar data.

Prosedur *robust* ditujukan untuk mengakomodasi adanya keanehan data, sekaligus meniadakan identifikasi adanya data pencilan dan juga bersifat otomatis dalam menanggulangi data pencilan. Regresi *robust* terdiri dari 5 metode estimasi, yaitu M-estimator, *Least Median Square* (LMS), *Least Trimmed Square* (LTS), S-estimator, dan MM-estimator.

2.6 Pencilan (*Outlier*)

Menurut Rousseeuw, *et al*. (1987), pencilan (*outlier*) adalah data yang tidak mengikuti pola umum pada model regresi yang dihasilkan, atau tidak mengikuti model data secara keseluruhan. Apabila terdapat masalah berkaitan dengan pencilan maka diperlukan alat diagnosis yang dapat mengidentifikasi masalah pencilan. Salah satunya adalah dengan menyisihkan *outlier* dari kumpulan data dan kemudian menganalisis data tanpa *outlier*.

Menurut Paludi (2009), keberadaan data *outlier* akan mengganggu dalam proses analisis data dan harus dihindari dari beberapa hal. Dalam kaitannya dalam analisis regresi, pencilan dapat menyebabkan hal-hal berikut:

1. Residual yang besar dari model yang terbentuk $E(\varepsilon) \neq 0$
2. Varians pada data tersebut menjadi lebih besar
3. Taksiran interval memiliki rentang yang lebar

Adapun beberapa tipe *outlier* yang memperngaruhi model regresi menurut Verardi & Croux (2008), yaitu antara lain:

1. Pencilan Vertikal (*Vertical Outlier*) adalah pencilan yang disebabkan oleh peubah respon. Titik pencilan ini memencil karena koordinat y yang ekstrim, sehingga y_i menjauh, tetapi x_i cocok dengan garis linear. Data yang mengandung *Vertical Outlier* akan menyebabkan pendugaan parameternya menjadi tidak efisien, sehingga harus diatasi dengan metode *robust*.
2. Titik Pengaruh Baik (*Good Leverage Points*) adalah pencilan yang disebabkan oleh peubah bebas. Titik pencilan ini memencil pada koordinat x dan terletak dekat dengan garis regresi.

3. Titik Pengaruh Buruk (*Bad Leverage Points*) adalah pencilan yang disebabkan oleh peubah respon dan peubah bebas. Titik pencilan yang memencil pada kedua koordinat (Y dan X) dan terletak jauh dari garis regresi sebenarnya.

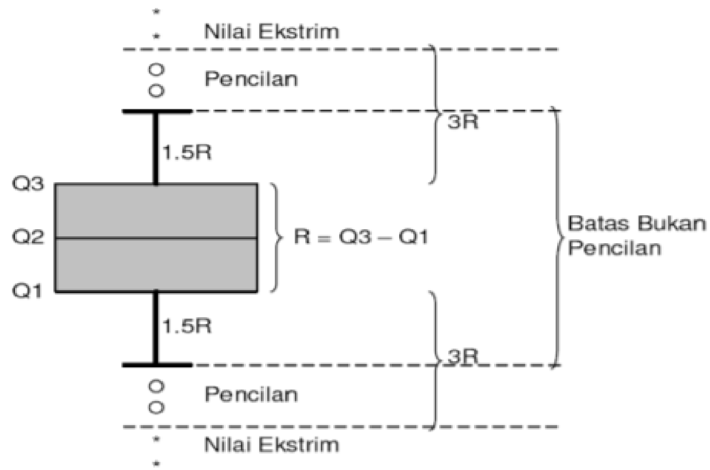
2.7 Deteksi *Outlier*

Ketika peneliti telah mendeteksi *outlier*, maka yang dilakukan yaitu dengan melihat kemungkinan bahwa *outlier* merupakan data yang terkontaminasi. Data *outlier* harus dilihat terhadap posisi dan sebaran data yang lainnya sehingga akan dievaluasi apakah data *outlier* tersebut perlu dihilangkan atau tidak. Adapun berbagai macam metode pendeteksian data *outlier* yang berpengaruh dalam koefisien regresi diantaranya adalah metode *boxplot*, *Leverage Value*, *Standardized Residual*, dan *Cook's Distance*. Dalam penelitian ini hanya dibahas metode *boxplot* dan *Cook's Distance* dalam pendeteksian *outlier*.

2.7.1 Metode *Boxplot*

Metode *boxplot* merupakan suatu metode yang menggunakan nilai kuartil dan jangkauan untuk mendeteksi *outlier*. Kuartil 1, 2, dan 3 akan membagi urutan data menjadi empat bagian. Tingkat jangkauan (IQR, *Interquartile Range*) didefinisikan sebagai selisih antara kuartil 1 dengan kuartil 3, atau $IQR = Q_3 - Q_1$.

Data dapat diketahui *outlier* dengan nilai yang kurang dari 1,5*IQR terhadap kuartil 1 dan nilai yang lebih dari dari 1,5*IQR terhadap kuartil 3.



Gambar 1. Skema Identifikasi *Outlier* Menggunakan *Boxplot*.

2.7.2 Metode *Cook's Distance*

Cook's Distance merupakan salah satu ukuran untuk mendeteksi *outlier* dalam data.

$$D_i = \frac{{e_i}^2 h_{ii}}{pMSE(1-h_{ii})^2} \tag{2.19}$$

Dengan e_i adalah residual ke-i, MSE adalah rata-rata jumlah kuadrat residual, h_{ii} adalah nilai *leverage* untuk kasus ke-i , dan p banyaknya variabel independen ditambah konstan. Nilai *leverage* adalah elemen-elemen diagonal dari matriks *H*.

$$H = X_i\left({X_i}^T X_i\right)^{-1}{X_i}^T \tag{2.20}$$

suatu data yang mempunyai nilai $D_i > \frac{4}{n}$ disebut *outlier*.

2.8 *Least Median Squares* (LMS)

Metode *Least Median Squares* **(**LMS) merupakan salah satu metode penduga regresi yang melakukan perhitungan dengan menghilangkan pengaruh-pengaruh dari galat. Dengan menggunakan penduga yang dihasilkan akan lebih baik dalam menghadapi pencilan. Jika pada metode kuadrat terkecil hal yang perlu dilakukan adalah meminimumkan jumlah kuadrat sisaan ($\sum_{i=1}^{n} e_i{}^2$), maka pada LMS hal yang perlu dilakukan adalah mencari nilai median kuadrat sisaan pada setiap iterasi, yaitu:

$$M_j = med\ e_i{}^2 \tag{2.21}$$

sehingga terbentuk $M_1, M_2, \ldots, M_s$ yang merupakan median kuadrat sisaan dari setiap h_i pengamatan. Untuk mendapatkan nilai M_1, dicari himpunan bagian data sejumlah h_i pengamatan, yaitu:

$$h_i = \left[\frac{n}{2}\right] + \left[\frac{p+1}{2}\right] \tag{2.22}$$

Dengan *n* adalah banyaknya data, dan *p* adalah banyaknya parameter. Dalam perhitungan nilai h_i harus selalu dalam bentuk bilangan bulat. Oleh karena itu, jika nilai h_i bukan dalam bentuk bilangan bulat maka dilakukan pembulatan ke atas. Demikian seterusnya sampai iterasi berakhir pada iterasi ke-*I* yaitu saat h_i=h_{i+1}. Setelah itu dicari nilai minimum dari $M_1, M_2, \ldots, M_s$.

Berdasarkan Rousseeuw (1987), bobot w_{ii} dirumuskan dengan ketentuan sebagai berikut.

$$\boldsymbol{W}_{ii} = \begin{cases} 1, & \text{jika } \left|\frac{e_i}{\hat{\sigma}}\right| \leq 2{,}5 \\ 0, & lainnya \end{cases}$$

dengan

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

Setelah bobot w_{ii} dihitung dapat dibentuk matriks W sebagai berikut:

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{in} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix}$$

Dengan entri matriks $\boldsymbol{w}_{ij}$ = 0, dengan $i \neq j$.

Setelah entri matriks **W**, maka penaksir parameter regresi LMS dapat dihitung dengan menggunakan rumus

$$\widehat{\boldsymbol{\beta}}_{LMS} = (\boldsymbol{X^T W X})^{-\mathbf{1}}(\boldsymbol{X^T W Y})$$

(Parmikanti, dkk., 2013).

2.9 Penduga-*S*

Metode penduga-*S* pertama kali dikembangkan oleh Rousseeuw & Yohai (1984), dimana metode ini merupakan keluarga *high breakdown point* yaitu ukuran umum proporsi dari data pencilan yang dapat ditangani sebelum pengamatan tersebut mempengaruhi model prediksi. Disebut penduga-*S* karena mengestimasi berdasarkan skala, dengan skala yang digunakan adalah standar deviasi residual. Metode penduga-*S* mempunyai kelebihan yaitu bisa digunakan untuk pencilan dengan proporsi hingga 50% serta digunakan ketika variabel *dependen* dan variabel *independen* terdapat pencilan. Metode ini menggunakan nilai pembobot dengan fungsi *Tukey's biweight* dengan nilai *breakdown value*= 0,5 dimana nilai konstanta c = 1,547. Menggunakan fungsi pembobotan *Tukey's biweight* karena iterasi yang digunakan lebih sedikit dibandingkan fungsi pembobot yang lain (Salibian & Yohai, 2006).

Yang didefinisikan:

$$\hat{\beta}_s = arg\ \min_{\beta} \hat{\sigma}_s(e_1, e_2, \dots, e_n) \tag{2.23}$$

Dengan menentukan nilai estimator skala *robust* ($\hat{\sigma}_s$) yang minimum dan memenuhi

$$min \sum_{i=1}^{n} \rho \left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}_s} \right) \quad (2.24)$$

dengan

$$\hat{\sigma}_s = \sqrt{\frac{1}{nK} \sum_{i=1}^{n} w_i {e_i}^2} \quad (2.25)$$

ρ merupakan fungsi pembobot *Tukey's biweight*, dengan

K=0,199, $w_i = w_\sigma(u_i) = \frac{\rho(u_i)}{{u_i}^2}$, dan dipilih estimasi awal

$$\hat{\sigma}_s = \frac{median|e_i - median(e_i)|}{0,6745} \quad (2.26)$$

Penyelesaian persamaan (2.6) dengan cara mencari turunannya terhadap $\hat{\beta}$ sehingga diperoleh:

$$\sum_{i=1}^{n} \rho' \left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta}{\hat{\sigma}_s} \right) = 0 \qquad j = 0,1,\ldots,k$$

$$\sum_{i=1}^{n} x_{ij} \psi \left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta}{\hat{\sigma}_s} \right) = 0 \qquad j = 0,1,\ldots,k \quad (2.27)$$

ψ disebut fungsi pengaruh yang merupakan turunan dari ρ ($\rho' = \psi$), turunan dari fungsi ρ adalah

$$\psi(u_i) = \rho'(u_i) = \begin{cases} u_i \left(1 - \left(\frac{u_i}{c} \right)^2 \right)^2 & , \ |u_i| \leq c \\ 0 & , \ |u_i| > c \end{cases} \quad (2.28)$$

Dengan w_i merupakan fungsi pembobot IRLS

$$W_i(u_i) = \frac{\psi(u_i)}{c} = \begin{cases} \frac{u_i \left(1 - \left(\frac{u_i}{c} \right)^2 \right)^2}{u_i} & , \ |u_i| \leq c \\ 0 & , \ |u_i| > c \end{cases} \quad (2.29)$$

$$W_i(u_i) = \begin{cases} \left(1 - \left(\frac{u_i}{c}\right)^2\right)^2 & , \ |u_i| \leq c \\ 0 & , \ |u_i| > c \end{cases} \tag{2.30}$$

dengan $u_i = \frac{e_i}{\hat{\sigma}_s}$ dan $c = 1{,}547$. Persamaan (2.27) dapat diselesaikan dengan IRLS sehingga mencapai konvergen.

2.10 Bias dan *Mean Square Error* (MSE)

2.10.1 Bias

Bias penduga dari suatu parameter pada simulasi data didefinisikan sebagai jumlah selisih dari penduga parameter pada data yang terdapat pencilan dengan penduga parameter pada data tanpa pencilan, dan dibagi dengan banyaknya pengulangan. Semakin kecil nilai bias, maka hasil pendugaan suatu parameter semakin baik. Bias dinotasikan sebagai berikut:

$$Bias((\hat{\beta}) = \frac{1}{m}\sum_{i=1}^{m}\left|\hat{\beta}^{(s)} - \hat{\beta}^{(0)}\right| \tag{2.31}$$

Dengan $\hat{\beta}^{(0)}$ adalah koefisien regresi untuk data tanpa pencilan, dan $\hat{\beta}^{(s)}$ adalah koefisien regresi untuk data yang telah diberi kontaminasi pencilan, sedangkan *m* adalah banyaknya pengulangan.

2.10.2 *Mean Square Error* (MSE)

Menurut Wulandari, *et al*. (2010), MSE penduga pada simulasi data adalah jumlah selisih kuadrat dari penduga parameter pada data yang terdapat pencilan dengan penduga parameter pada data yang tanpa pencilan, dibagi dengan banyaknya pengulangan. MSE dinotasikan sebagai berikut.

$$MSE(\hat{\beta}) = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{\beta}^{(s)} - \hat{\beta}^{(0)}\right)^2 \tag{2.32}$$

Dengan $\hat{\beta}^{(0)}$ adalah koefisien regresi untuk data tanpa pencilan. Dan $\hat{\beta}^{(s)}$ koefisien regresi untuk data yang telah diberi kontaminasi pencilan, sedangkan m adalah banyaknya pengulangan. Kebaikan suatu penduga dapat dilihat dari tingkat kesalahannya, semakin kecil tingkat kesalahan semakin baik pendugaannya.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat

Penelitian ini dilakukan pada semester ganjil tahun akademik 2021/2022 di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data simulasi. Data akan dibangkitkan yang terdiri dari galat dan variabel bebas (X_1 dan X_2) yang akan digunakan untuk menentukan variabel tak bebas (Y). Data akan dibangkitkan dengan berukuran 20, 50, 100, dan 200. Data galat dan variabel bebas dibangkitkan berdistribusi N (0,1). Pencilan akan dibangkitkan dengan persentase 10%, 15%, dan 20% dari jumlah sampel.

3.3 Metode Analisis

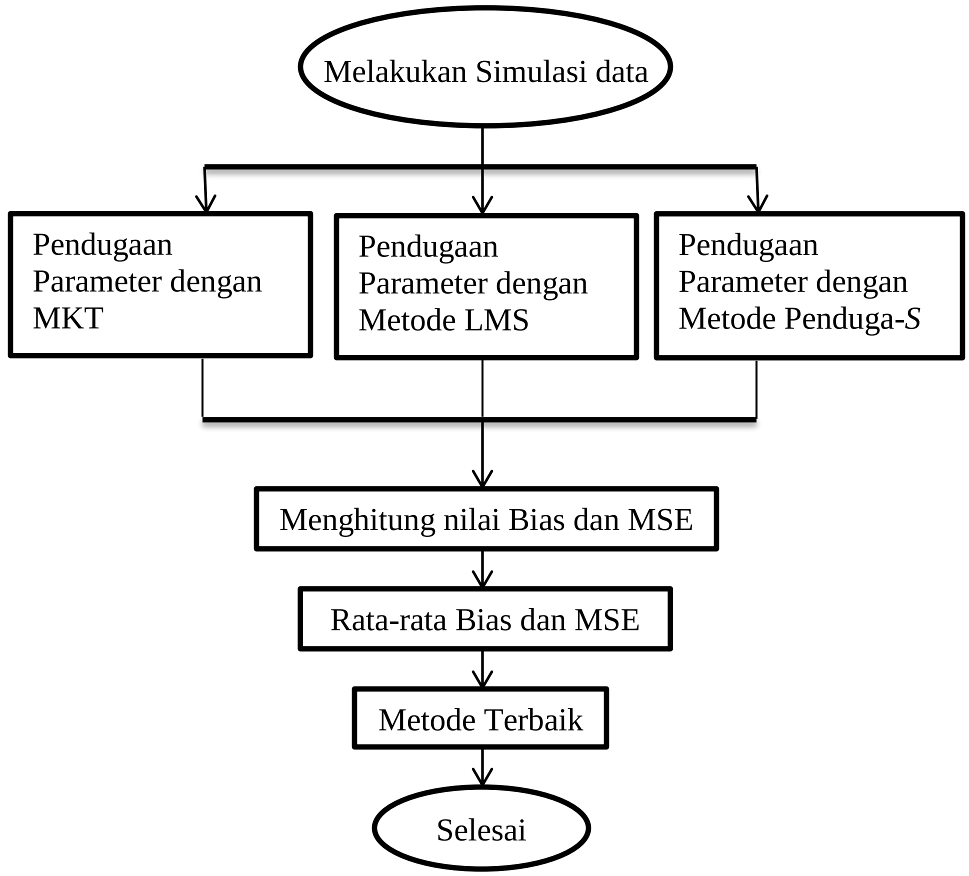
Penelitian ini dilakukan dengan menggunakan studi literatur secara sistematis yang diperoleh dari buku-buku teks penunjang, jurnal dan juga media lain seperti internet.

Kemudian melakukan simulasi sebagai aplikasi untuk menerapkan teori yang telah didapat. Untuk mempermudah perhitungan digunakan perangkat lunak (*software*) *R* 4.1.0.

Metode yang digunakan dalam penelitian ini adalah simulasi data dengan tahapan sebagai berikut:

1. Membangkitkan galat serta variabel bebas X_1 dan X_2 masing-masing berdistribusi N(0,1).
2. Mengkontaminasikan galat dari distribusi N(0,1) dengan pencilan berdistribusi N (50,1). Pencilan yang diberikan yaitu sebesar 10%, 15%, dan 20% dari jumlah sampel.
3. Menentukan nilai variabel tak bebas $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$. Dengan menetapkan $\beta_0 = \beta_1 = \beta_2 = 1$.
4. Mengulangi langkah a sampai dengan c sebanyak 1000 kali.
5. Menduga parameter regresi dengan MKT.
6. Menduga parameter regresi *robust* dengan metode LMS.
7. Menduga parameter regresi *robust* dengan metode Penduga-*S*.
8. Menghitung nilai bias dan MSE pada masing-masing metode untuk semua persentase pencilan.

Kerangka



Gambar 2. Diagram Alur Langkah-langkah Penelitian dalam Simulasi Data.

V. KESIMPULAN

Berdasarkan hasil dan pembahasan, dapat diperoleh kesimpulan sebagai berikut:

1. Berdasarkan hasil simulasi telah diperoleh bahwa nilai bias dan MSE dari Metode LMS lebih kecil dibandingkan dengan Metode Penduga-*S* untuk semua ukuran sampel, baik pada data dengan perasentase pencilan 10%, 15% maupun 20%.
2. Dengan demikian, dapat disimpulkan bahwa Metode LMS lebih baik dibandingkan Metode Penduga-*S* dalam menduga parameter regresi pada yang mengandung pencilan.

.

DAFTAR PUSTAKA

Anton, H. 1987. *Aljabar Linear Elementer*. Diterjemahkan oleh Pantur Silaban, Ph.D. & Drs. I Nyoman Susila, M.Sc. Erlangga, Jakarta.

Draper, N. & Smith, H. 1992. *Analisis Regresi Terapan*. Ed. Ke-2. Diterjemahkan oleh Ir. Bambang Sumantri. Gramedia, Jakarta.

Ghozali, I. 2011. *Aplikasi Analisis Multivariate dengan Program SPSS*. Badan Penerbit Universitas Diponegoro, Semarang.

Gujarati, D.N. 2004. *Basic Econometrics*. 4th Edition. The McGraw-Hill Companies, New York.

Hadley, G. 1992. *Aljabar Linear*. Diterjemahkan oleh Naipospos & Noenick Soemartoyo. Erlangga, Jakarta.

Herawati, N. & Nisa, K. 2010. *Analisis regresi robust menggunakan metode penduga -MM*. Prosiding Seminar Nasional Sains dan Teknologi-III Published by LPPM Unila 18-19 Oktober 2010. ISSN 978 – 979 - 8510 - 20 – 5.

Khotimah, K., Sadik, K., & Rizki, A. 2020. *Study of Robust Regression Modeling Using MM-Estimator and Least Median Squares*. Proceedings of the 1st International Conference on Statistics and Analytics, ICSA 2019, 2-3 August 2019, Bogor, Indonesia.

Montgomery, D.C. & Peck E.A. 1992. *Introduction to Linear Regression Analisis*. 2th Edition. John Wiley & Sons, New York.

Nisa, K. 2007. *Pembandingan Empiris Tiga Metode Regresi Robust (An Empirical Comparison Of Three Robust Regression Methods)*. Prosiding Seminar Hasil Penelitian & Pengabdian Kepada Masyarakat Universitas Lampung Bandar Lampung September Universitas Lampung. ISSN ISBN 978-979-15535-1-3.

Olive, D.J. 2005. *Applied Robust Statistics*. Southern Illinois University, Carbondale.

Parmikanti, K., Rusyaman, E., & Suryamah, E. 2013. Model Regresi Kandungan Batubara Menggunakan Metode Least Median Of Squares. *Prosiding Seminar Nasional Sains dan Teknologi Nuklir PTNBR,* Batan Bandung.

Rousseeuw, P.J. & Yohai, V. J. 1984. *Robust Regression by Mean of S-Estimation*. Berlin, New York.

Rousseeuw, P.J. & Leroy, A.M . 1987. *Robust Regression and Outlier Detection*. John Wiley & Sons, Canada.

Sakata, S. & Preminger, A. 2007. A Model Selection method for S-estimation. *The Econometrics Journal*. **10**(2): 294-319.

Salibian-Barrera, M. & Yohai, V. J. 2006. A Fast Algoritm for S-Regression Estimates. *Journal of Computational and Graphical Statistics*. **15**(2): 414-427.

Sembiring, R.K. 2003. *Analisis Regresi.* Ed. Ke-2. Penerbit ITB, Bandung.

Setiawan & Kusrini, D.E. 2010. *Ekonometrika*. C.V. Andi Offset, Yogyakarta.

Sungkawa, I. 2009. Pendeteksian Pencilan dan Residual pada Regresi Linear. http://www.litbang.pertanian.go.id/warta-ip/pdf-file/2.iwa_ipvol18-02-2009. Diakses pada 10 November 2019.

Suliyanto. 2008. *Teknik Proyeksi Bisnis*. C.V. Andi Offset, Yogyakarta.

Suliyanto. 2011. *Ekonometrika Terapan*. C.V. Andi Offset, Yogyakarta.

Trihastuti, R.E. 2014. Bentuk Normal Jordan untuk Menentukan Invers Moore Penrose. Skripsi. Program Studi Pendidikan Matematika FKIP Universitas Muhammadiyah Purwokerto, Purwokerto.

Verardi, V. & Croux, C. 2008. Robust Statistics in Stata. *Departement of Decision Sciences and Management,* Katholieke Universitiet Leuven.

Wulandari, S., Salam, N., & Anggraini, D. 2010. Perbandingan Metode Robust MCD-LMS, MCD-LTS, MVE-LMS, dan MVE-LTS dalam Analisis Regresi Komponen Utama. *Jurnal Matematika Murni dan Terapan*. **4**(1):57-64.