ANALISIS REGRESI KOMPONEN UTAMA *ROBUST* DENGAN METODE *MINIMUM COVARIANCE DETERMINANT – LEAST TRIMMED SQUARE* (MCD-LTS)

(Skripsi)

Oleh

SISKA DIAH AYU LARASATI

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2019

ABSTRACT

ROBUST PRINCIPAL COMPONENT REGRESSION ANALYSIS USING MINIMUM COVARIANCE DETERMINANT - LEAST TRIMMED SQUARE (MCD-LTS) METHODS

By

SISKA DIAH AYU LARASATI

Principal Component Regression (PCR) is a method used to overcome multicollinearity problems by reducing the dimensions of independent variables so that new simpler variables are obtained without losing most of the information contained in independent variables. If the observation data also indicates outliers, a robust method on PCR is used by analyzing the combination of Robust Principal Component Analysis using the Minimum Covariance Determinant (MCD) method with Robust Regression Analysis using the Least Trimmed Square (LTS) method. The purpose of this study is to examine the PCR Robust analysis using the MCD-LTS method and to know the PCR Robust rigidity by looking at its sensitivity to outliers then compared to the classic PCR based on the bias and Mean Square Error (MSE) on several different sample sizes and outliers. The results of this study indicate that PCR Robust is more effective and efficient in overcoming the problem of multicollinearity and outliers.

Keywords : Principal Component Regression, Multicollinearity, Outliers, Robust

ABSTRAK

ANALISIS REGRESI KOMPONEN UTAMA *ROBUST* DENGAN METODE *MINIMUM COVARIANCE DETERMINANT – LEAST TRIMMED SQUARE* (MCD-LTS)

Oleh

SISKA DIAH AYU LARASATI

Regresi Komponen Utama (RKU) merupakan metode yang digunakan untuk mengatasi masalah multikolinearitas dengan mereduksi dimensi variabel bebas sehingga diperoleh variabel baru yang lebih sederhana tanpa kehilangan sebagian besar informasi yang terkandung pada variabel bebasnya. Apabila pada data pengamatan juga terindikasi adanya pencilan, maka digunakan metode *robust* pada RKU dengan cara melakukakan analisis kombinasi antara Analisis Komponen Utama *Robust* menggunakan metode *Minimum Covariance Determinant* (MCD) dengan Analisis Regresi *Robust* menggunakan metode *Least Trimmed Square* (LTS). Tujuan dari penelitian ini adalah mengkaji analisis RKU *Robust* dengan metode MCD-LTS serta mengetahui ketegaran RKU *Robust* dengan melihat kepekaannya terhadap pencilan kemudian dibandingkan dengan RKU klasik berdasarkan bias dan *Mean Square Error* (MSE) pada beberapa ukuran sampel dan persentase pencilan yang berbeda. Hasil dari penelitian ini menunjukkan bahwa RKU *Robust* lebih efektif dan efisien dalam mengatasi masalah multikolinearitas dan pencilan.

Kata kunci : Regresi Komponen Utama, Multikolinearitas, Pencilan, *Robust*

ANALISIS REGRESI KOMPONEN UTAMA *ROBUST* DENGAN METODE *MINIMUM COVARIANCE DETERMINANT – LEAST TRIMMED SQUARE* (MCD-LTS)

Oleh

SISKA DIAH AYU LARASATI

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA SAINS

pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2019

Judul Skripsi	: **ANALISIS REGRESI KOMPONEN UTAMA *ROBUST* DENGAN METODE *MINIMUM COVARIANCE DETERMINANT - LEAST TRIMMED SQUARE* (MCD-LTS)**
Nama Mahasiswa	: Siska Diah Ayu Larasati
Nomor Pokok Mahasiswa	: 1517031020
Program Studi	: Matematika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Dr. Khoirin Nisa, M.Si.	**Dr. Notiragayu, M.Si.**
NIP 19740726 200003 2 001	NIP 19731109 200012 2 001

2. Ketua Jurusan Matematika

Prof. Dra. Wamiliana, M.A., Ph.D.
NIP 19631108 198902 2 001

MENGESAHKAN

1. Tim Penguji

Ketua	: **Dr. Khoirin Nisa, M.Si.**	....
Sekretaris	: **Dr. Notiragayu, M.Si.**	....
Penguji Bukan Pembimbing	: **Drs. Eri Setiawan, M.Si.**	....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Drs. Suratman, M.Sc.
NIP 19640604 199003 1 002

Tanggal Lulus Ujian Skripsi : **15 Mei 2019**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama	: **Siska Diah Ayu Larasati**
Nomor Pokok Mahasiswa	: **1517031020**
Jurusan	: **Matematika**
Judul Skripsi	: **Analisis Regresi Komponen Utama *Robust* dengan Metode *Minimum Covariance Determinant – Least Trimmed Square* (MCD-LTS)**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan 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, 15 Mei 2019
Yang Menyatakan

METERAI TEMPEL
BF1F7AFF753033405
6000
ENAM RIBU RUPIAH

Siska Diah Ayu Larasati

RIWAYAT HIDUP

Penulis bernama lengkap Siska Diah Ayu Larasati, lahir di Bogor pada 20 April 1997. Penulis merupakan anak pertama dari dua bersaudara, pasangan Bapak Siswanto dan Ibu Sulastri.

Penulis menempuh pendidikan dasar di SD Negeri Citeureup 03 pada tahun 2003-2009. Kemudian melanjutkan pendidikan di SMP Negeri 1 Citeureup dan lulus pada tahun 2012. Kemudian menempuh pendidikan di SMA Al-Azhar 3 Bandar Lampung dan lulus pada tahun 2015.

Penulis melanjutkan pendidikan di perguruan tinggi dan terdaftar sebagai mahasiswa di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung (UNILA) melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) tahun 2015. Pada tahun 2016, penulis terdaftar sebagai anggota Kaderisasi dan Kepemimpinan di Himpunan Mahasiswa Jurusan Matematika (HIMATIKA). Sebagai bentuk penerapan ilmu perkuliahan, pada tahun 2018 penulis melakukan Kerja Praktik (KP) di Kantor Pelayanan Pajak Pratama Teluk Betung serta melaksanakan Kuliah Kerja Nyata Kebangsaan (KKN K) di Desa Karya Makmur, Kecamatan Labuhan Maringgai, Kabupaten Lampung Timur.

KATA INSPIRASI

"Dan Allah mengeluarkan kamu dari perut ibumu dalam keadaan tidak mengetahui sesuatu pun, dan Dia memberimu pendengaran, penglihatan, dan hati nurani agar kamu bersyukur"

(QS. An-Nahl : 78)

"Dan bahwa manusia hanya memperoleh apa yang telah diusahakannya, dan sesungguhnya usahanya itu kelak akan diperlihatkan (kepadanya), kemudian akan diberi balasan kepadanya dengan balasan yang paling sempurna"

(QS. An-Najm : 39-41)

"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 dan syukur kepada Allah SWT atas segala nikmat dan karunia-Nya. Tidak lupa shalawat beserta salam senantiasa tercurahkan kepada baginda besar Nabi Muhammad SAW yang merupakan suri tauladan terbaik bagi seluruh umat.

Kupersembahkan sebuah karya sederhana ini kepada :

Kedua Orang Tua Tercinta

Terima kasih atas limpahan kasih sayang, pengorbanan, doa, dan seluruh motivasi di setiap langkahku. Hanya atas doa dan ridho bapak mamah, Allah memudahkan tiap perjalanan hidup ini. Terimalah bukti kecil ini sebagai kado kesuksesanku dalam menggapai impian untuk menjadi seorang sarjana.

Adik Tersayang

Semoga apa yang telah kakak lakukan selalu bisa menjadi contoh dan motivasi untukmu.

Sahabat-sahabat Terbaik

Terima kasih atas kebahagiaan, keceriaan, motivasi serta seluruh kenangan indah di bangku kuliah yang kalian berikan.

Almamater Tercinta, Universitas Lampung

SANWACANA

Puji syukur penulis ucapkan kepada Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul "Analisis Regresi Komponen Utama *Robust* dengan Metode *Minimum Covariance Determinant – Least Trimmed Square* (MCD-LTS)".

Penulis menyadari bahwa skripsi ini dapat terselesaikan dengan baik karena bimbingan, dukungan, saran, serta doa dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis menyampaikan terima kasih kepada:

1. Ibu Dr. Khoirin Nisa, M.Si., selaku dosen Pembimbing I yang telah bersedia memberikan arahan, bimbingan, kritik dan saran bagi penulis dalam menyusun skripsi.
2. Ibu Dr. Notiragayu, M.Si., selaku dosen Pembimbing II yang telah memberikan saran serta arahan bagi penulis.
3. Bapak Drs. Eri Setiawan, M.Si., selaku dosen Pembahas yang telah memberikan evaluasi, arahan dan saran sehingga terselesaikannya skripsi ini.
4. Bapak Dr. Aang Nuryaman, M.Si., selaku dosen Pembimbing Akademik yang telah membimbing penulis dalam menyelesaikan permasalahan seputar akademik.

5. Ibu Prof. Dra. Wamiliana, M.A., Ph.D., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam.
6. Bapak Drs. Suratman, M.Sc., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Seluruh dosen, staff dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
8. Orang tua tercinta dan adik tersayang, serta seluruh keluarga besar yang senantiasa memberikan kasih sayang tiada terkira, selalu menjadi penyemangat disaat lemah, selalu memotivasi penulis untuk memberikan yang terbaik, serta tiada henti berdoa untuk kesuksesan penulis.
9. Sahabat-sahabat terbaik Mahmud, Ario, Rahmad, Dony, Azam, Bagus, Nathanael, Nabila, Sandria, Almira, Caroline, Mira yang telah banyak memberikan dukungan dan kenangan indah kepada penulis.
10. Teman-teman satu bimbingan, terima kasih atas semangat dan saran selama penyelesaian skripsi.
11. HIMATIKA yang telah memberikan pengalaman berharga.
12. Teman-teman mahasiswa Jurusan Matematika angkatan 2015.

Penulis juga menyadari bahwa dalam penulisan skripsi ini masih jauh dari sempurna, sehingga informasi tambahan, saran, dan kritik untuk pengembangan lebih lanjut sangat diharapkan. Semoga skripsi ini dapat bermanfaat bagi kita.

Bandar Lampung, 15 Mei 2019

Penulis,

Siska Diah Ayu Larasati

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis regresi merupakan salah satu metode analisis data yang digunakan untuk menyelidiki hubungan antara variabel respon dengan satu atau beberapa variabel bebas. Apabila hubungan yang diselidiki terdiri dari satu variabel respon dan satu variabel bebas maka disebut dengan analisis regresi linear sederhana. Sedangkan analisis regresi linear berganda terdiri dari satu variabel respon dan lebih dari satu variabel bebas.

Permasalahan yang sering dihadapi dalam analisis regresi linear berganda yaitu adanya multikolinearitas, yang mana terjadi ketika adanya korelasi yang kuat antara variabel bebas. Hal ini dapat menyebabkan $\mathbf{X}^{\mathbf{T}}\mathbf{X}$ memiliki kondisi buruk (*ill condition*) atau hampir singular yang pada akhirnya akan menyebabkan nilai penduga ragam bagi parameter regresi menjadi lebih besar (Draper & Smith, 1992).

Salah satu metode statistik yang digunakan untuk mengatasi masalah multikolinearitas adalah regresi komponen utama (RKU). Menurut Notiragayu & Nisa (2008), pada RKU klasik terdapat dua tahap yaitu komponen utama dibentuk menggunakan vektor eigen dari matriks kovarian sampel (**S**) klasik dan diregresikan terhadap Y dengan metode kuadrat terkecil. Namun analisis seperti ini sangat sensitif terhadap pencilan *(outliers)* dan akan menghasilkan dugaan parameter yang bias akibat terpengaruh oleh data pencilan. Cara yang paling sederhana untuk mengatasi pencilan pada RKU yaitu dengan menggunakan metode *robust* pada kedua tahap.

Salah satu metode *robust* bagi matriks kovarian sampel (**S**) yaitu metode *Minimum Covariance Determinant* (MCD) dan kemudian hasil komponen-komponen utama *robust* yang terbentuk akan diregresikan dengan peubah respon menggunakan *Least Trimmed Square* (LTS). Menurut Nisa (2006), metode LTS menduga koefisien regresi dari data yang mengandung pencilan dengan meminimumkan jumlah kuadrat galat terhadap sebaran data yang sudah terpangkas *(trimmed)* yang sering juga disebut dengan istilah “sebaran terwinsorkan” *(winsorized distribution).*

Berdasarkan hal tersebut, maka pada penelitian ini akan mengkaji analisis Regresi Komponen Utama (RKU) *Robust* dengan metode *Minimum Covariance Determinant- Least Trimmed Square* (MCD-LTS) serta mengetahui ketegaran metode tersebut terhadap pencilan.

1.2 Tujuan Penelitian

Adapun tujuan penelitian ini adalah :

1. Mengkaji analisis Regresi Komponen Utama (RKU) *Robust* dengan metode *Minimum Covariance Determinant- Least Trimmed Square* (MCD-LTS).
2. Mengetahui ketegaran metode *Minimum Covariance Determinant- Least Trimmed Square* (MCD-LTS) pada analisis RKU dengan melihat kepekaannya terhadap pencilan berdasarkan bias dan *Mean Square Error* (MSE).

1.3 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah untuk menambah pengetahuan bagi penulis dan diharapkan dapat memberikan alternatif lain bagi pembaca dalam memilih metode penduga yang terbaik dalam mengkaji data yang terdapat pencilan.

II. TINJAUAN PUSTAKA

2.1 Matriks

Matriks adalah susunan bilangan atau fungsi yang diletakkan atas baris dan kolom serta diapit oleh dua kurung siku. Bilangan atau fungsi tersebut disebut entri atau elemen matriks. Elemen matriks terdiri dalam baris dan kolom. Lambang matriks dilambangan dengan huruf besar, sedangkan entri (elemen) matriks dilambangkan dengan huruf kecil.

Matriks **A** yang memiliki *m* baris dan *n* kolom dapat ditulis sebagai berikut :

$$\boldsymbol{A} = \left[a_{ij}\right] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix}$$

Dimana a_{ij} menyatakan elemen matriks **A** yang berada pada baris ke-*i* dan kolom ke-*j* (Anton, 1987).

2.1.1 Jenis-Jenis Matriks

Berikut beberapa jenis matriks, yaitu :

a. 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.

b. Matriks ortogonal

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

Matriks ortogonal dapat dituliskan sebagai:

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

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 $det(\boldsymbol{A}) = -1$

c. Matriks definit positif

Suatu matriks simetris $\boldsymbol{A}$ berorder $n \times n$ disebut matriks definit positif apabila $\boldsymbol{X^T A X} > 0$ untuk semua vektor tak nol $\boldsymbol{X}$. Suatu matriks $\boldsymbol{A}$ yang berukuran $n \times n$ dikatakan definit positif jika dan hanya jika $\boldsymbol{A^T} = \boldsymbol{A}$ dan $\boldsymbol{X^T A X} > 0$ untuk setiap vektor $\boldsymbol{X}$ tidak nol.

d. Matriks semi definit positif

Matriks $\boldsymbol{A}$ yang berukuran $n \times n$ dikatakan semi definit positif jika dan hanya jika kondisi-kondisi berikut terpenuhi :

1) $\boldsymbol{A^T} = \boldsymbol{A}$, ($\boldsymbol{A}$ matriks simetris).
2) $\boldsymbol{X^T A X} \geq 0$, untuk setiap vektor **X** tidak nol.

3) $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{A}\boldsymbol{X} = 0$, untuk paling sedikit satu vektor tidak nol.

(Anton, 1987).

e. Matriks singular dan non singular

X dikatakan mempunyai invers dilambangkan dengan $\mathbf{X}^{-1}$, jika $\mathbf{X}\mathbf{X}^{-1} = \mathbf{X}^{-1}\mathbf{X} = \mathbf{I}$. Jika suatu matriks mempunyai invers maka dikatakan matriks tersebut non singular, tetapi jika tidak mempunyai invers maka matriks tersebut singular (Usman & Warsono, 2001).

2.1.2 Operasi Matriks

Berikut beberapa operasi matriks, yaitu :

a. Jumlah unsur diagonal suatu matriks (*trace*)

Bila $\boldsymbol{A}$ adalah suatu matriks persegi dengan ukuran $n \times n$, maka jumlah unsur diagonal matriks $\boldsymbol{A}$ dilambangkan $tr(\boldsymbol{A})$, adalah

$$tr(\boldsymbol{A}) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii}$$

Lambang *tr* adalah singkatan dari *trace* dalam bahasa inggris.

b. Perkalian skalar

Jika **A** adalah suatu matriks dan *c* adalah suatu *skalar*, maka hasil kali *c***A** adalah matriks yang diperoleh dengan mengalikan setiap anggota dari **A** dan *c*. Perkalian matriks dengan skalar menghasilkan sebuah matriks baru yang elemennya adalah hasil perkalian setiap elemen matriks aslinya dengan skalar. Dalam notasi matriks, jika $\boldsymbol{A} = [a_{ij}]$, maka $c\boldsymbol{A} = c[a_{ij}]$.

c. Perkalian matriks

Jika $\boldsymbol{A}$ adalah matriks berukuran $m \times r$ dan $\boldsymbol{B}$ adalah matriks berukuran $r \times n$, maka hasil kali $\boldsymbol{AB}$ adalah matriks $\boldsymbol{C}$ berukuran $m \times n$ yang unsur-unsurnya adalah:

$$
\begin{aligned}
\boldsymbol{C}_{ij} &= Baris_i(\boldsymbol{A})Kolom_j(\boldsymbol{B}) \\
&= a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} \\
&= \sum_{k=1}^{n} a_{ik}b_{kj}
\end{aligned}
$$

Perkalian matriks hanya bisa dilakukan jika banyaknya kolom matriks yang pertama sama dengan banyaknya baris matriks yang kedua.

d. Transpose matriks

Transpose dari suatu matriks $\boldsymbol{A}$ berordo $m \times n$ yang dinotasikan $[a_{ij}]$ adalah matriks $\boldsymbol{B}$ berorde $n \times m$ yang dinotasikan $[b_{ji}]$ yang didefinisikan oleh:

$$b_{ji} = a_{ij}$$

Untuk $i = 1,2,\dots,m$ dan $j = 1,2,\dots,n$. Transpose dari matriks $\boldsymbol{A}$ dinyatakan oleh $\boldsymbol{A^T}$.

Beberapa sifat transpose matriks:

a) $(\boldsymbol{A^T})^T = \boldsymbol{A}$

b) $(\boldsymbol{A} + \boldsymbol{B})^T = \boldsymbol{A^T} + \boldsymbol{B^T}$

c) $(k\boldsymbol{A})^T = k\boldsymbol{A^T}$, dengan k sembarang skalar

d) $(\boldsymbol{AB})^T = \boldsymbol{B^T}\boldsymbol{A^T}$

e. Determinan matriks

Misalkan $\boldsymbol{A} = [a_{ij}]$ adalah matriks berukuran $n \times n$, determinan dari $\boldsymbol{A}$ didefinisikan sebagai berikut :

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

Dengan M_{ij} adalah minor dari 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 $\boldsymbol{A} = [a_{ij}]$ adalah matriks berukuran $n \times n$ yang mengandung sebaris bilangan nol, maka $|\boldsymbol{A}| = 0$. Jika $\boldsymbol{A} = [a_{ij}]$ matriks segitiga, maka $|\boldsymbol{A}|$ adalah hasil kali elemen-elemen diagonal utama, yaitu $\boldsymbol{A} = a_{11}a_{22}a_{33} \ldots a_{mm}$.
Jika $\boldsymbol{A} = [a_{ij}]$ adalah sebarang matriks persegi, maka $|\boldsymbol{A}| = |\boldsymbol{A^T}|$, dimana $\boldsymbol{A^T}$ adalah transpose dari **A**. Jika $\boldsymbol{A} = [a_{ij}]$ dan $\boldsymbol{B} = [b_{ij}]$ adalah matriks persegi yang ordonya sama, maka $|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$.

f. Invers matriks

Jika $\boldsymbol{A}$ adalah suatu matriks persegi berukuran $n \times n$ dan jika suatu matriks $\boldsymbol{B}$ yang berukuran sama $n \times n$ disebut invers (balikan) dari $\boldsymbol{A}$ jika dipenuhi $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}$, maka $\boldsymbol{A}$ bisa dibalik dan $\boldsymbol{B}$ disebut invers dari $\boldsymbol{A}$. Invers dari $\boldsymbol{A}$ dilambangkan dengan $\boldsymbol{A^{-1}}$.

Jika $\boldsymbol{A}$ dan $\boldsymbol{B}$ adalah matriks-matriks yang dapat dibalik yang ukurannya sama, maka :

1) **AB** dapat dibalik
2) $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$

Jika **A** adalah sebuah matriks yang dapat dibalik, maka

1) $\boldsymbol{A}^{-\mathbf{1}}$ dapat dibalik dan $(\boldsymbol{A}^{-\mathbf{1}})^{-1} = \boldsymbol{A}$
2) $(\boldsymbol{A}^{n})^{-1} = (\boldsymbol{A}^{-1})^{n}$ untuk $n = 0,1,2, \ldots$
3) Untuk skalar k, dimana $k \neq 0$, maka k**A** dapat dibalik dan $(k\boldsymbol{A})^{-1} = \frac{1}{k}\boldsymbol{A}^{-1}$.

(Anton, 1987).

2.1.3 Nilai Eigen dan Vektor Eigen

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

$$\boldsymbol{Ax} = \lambda\boldsymbol{x} \tag{2.1}$$

Dari pernyataan diatas, 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 $\boldsymbol{x}$ merupakan anggota dari R^{n} untuk $\boldsymbol{A}$ dan $\boldsymbol{x}$ bukan vektor nol (Anton, 1987).

2.2 Analisis Regresi

Analisis regresi adalah metode statistika yang digunakan untuk mempelajari dan mengukur hubungan yang terjadi antara dua atau lebih variabel. Dalam analisis regresi, suatu persamaan regresi akan ditentukan dan digunakan untuk menggambarkan pola atau bentuk fungsi hubungan yang terdapat antar variabel. Variabel yang akan diestimasi nilainya disebut variabel terikat (*dependent variable*) dan diplot pada sumbu-y. Sedangkan variabel bebas (*independent variable*) adalah variabel yang diasumsikan memberikan pengaruh terhadap variabel terikat dan diplot pada sumbu-x (Harinaldi, 2005).

Menurut Sembiring (2003), model regresi adalah model yang memberikan gambaran mengenai hubungan antara variabel bebas dengan variabel terikat. Secara umum bentuk persamaan model regresi adalah sebagai berikut :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i \qquad (2.2)$$

dengan :

Y_i	= variabel dependen (tak bebas)
$\beta_0, \beta_1, \beta_2, \ldots\ldots, \beta_k$	= koefisien regresi
$X_{1i}, X_{2i}, \ldots X_{ki}$	= variabel independen (bebas)
ε_i	= galat / residual

Pada Persamaan (2.1) dalam lambang matriks penulisannya menjadi :

$$\boldsymbol{Y}_{(nx1)} = \boldsymbol{X}_{(nxk)}\boldsymbol{\beta}_{(kx1)} + \boldsymbol{\varepsilon}_{(nx1)} \tag{2.3}$$

atau,

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Dalam analisis regresi, terdapat dua model regresi yaitu, regresi linear sederhana dan regresi linear berganda. Regresi linear sederhana adalah persamaan regresi yang menggambarkan hubungan antara satu peubah bebas (X) dan satu peubah tak bebas (Y), dimana hubungan keduanya dapat digambarkan sebagai suatu garis lurus. Sedangkan regresi linear berganda adalah persamaan regresi yang menggambarkan hubungan antara lebih dari satu peubah bebas (X) dan satu peubah tak bebas (Y).

2.3 Uji Asumsi Analisis Regresi

Menurut Imam Ghozali (2011), uji asumsi klasik terhadap model regresi linear yang dilakukan agar dapat diketahui model regresi baik atau tidak. Tujuan pengujian asumsi klasik adalah untuk memberikan kepastian bahwa persamaan regresi yang diperoleh memiliki ketepatan dalam estimasi, tidak bias, dan konsisten. Sebelum melakukan analisis regresi terlebih dahulu dilakukan pengujian asumsi.

Menurut Chatterjee & Hadi (2006), untuk melakukan analisis regresi linear harus memenuhi asumsi-asumsi berikut:

1. Linearitas

   Uji ini dilakukan untuk mengetahui apakah hubungan antara respon Y dan prediktor X membentuk hubungan linear atau tidak. Memeriksa asumsi linearitas dalam regresi sederhana mudah karena validitas asumsi ini dapat ditentukan dengan memeriksa *scatterplot* dari Y terhadap X. Namun, memeriksa linearitas dalam regresi berganda lebih sulit karena dimensi yang tinggi dari data. Ketika asumsi linearitas tidak terpenuhi maka data dianalisis dengan regresi nonlinear atau dapat dilakukan transformasi data.

2. Normalitas

   Normalitas yang dimaksudkan adalah galat yang berdistribusi normal yaitu $\varepsilon \sim N(0, \sigma^2)$. Pelanggaran terhadap kenormalan dapat terjadi karena adanya beberapa data yang merupakan pencilan atau karena terdapat nilai ekstrim dalam data yang digunakan.

3. Homogenitas

   Homoskedastisitas diartikan sebagai distribusi dari galat memiliki ragam yang konstan (homogen). Apabila varian galat dalam model tidak konstan disebut dengan heteroskedastisitas. Heteroskedastisitas disebabkan karena variabel yang digunakan memiliki nilai yang sangat beragam, sehingga menghasilkan nilai galat yang tidak konstan.

4. Tidak terjadi autokorelasi

   Autokorelasi terjadi karena galat antar pengamatan tidak saling bebas atau berkaitan satu sama lain.

5. Tidak terjadi multikorelasi

   Khusus untuk regresi linear berganda terdapat tambahan satu asumsi lagi yaitu tidak terjadi multikolinearitas. Kolinearitas dapat terjadi karena suatu faktor diukur lebih dari sekali. Kolinearitas seperti ini disebut kolinearitas sempurna, yaitu suatu peubah bebas bergantung sepenuhnya pada yang lainnya. Namun, dalam prakteknya kolinearitas sering muncul dalam bentuk tersamar sehingga sukar dikenali, dimana suatu peubah tidaklah sepenuhnya tergantung pada peubah lainnya tapi hanya sebagian. Kolinearitas tidak sempurna seperti ini sering terjadi bila dua atau lebih peubah dalam model saling berkaitan (multikolinearitas). Sehingga, yang dimaksudkan dengan multikolinearitas adalah terjadinya kolinearitas antara dua atau lebih peubah bebas dalam model.

2.4 Multikolinearitas

Istilah multikolinearitas pertama kali ditemukan oleh Ragnar Frisch pada tahun 1934 yang artinya terdapat hubungan linear diantara beberapa atau semua variabel bebas dalam model regresi. Masalah multikolinearitas hanya ditemukan dalam regresi linear berganda. Model yang baik adalah model yang bebas dari multikolinearitas. Suatu model yang bebas dari multikolinearitas adalah model yang memiliki nilai faktor *Variance Inflation Factor* (VIF) < 10. Nilai VIF > 10 mengindikasikan terdapat multikolinearitas (Myers, 1990).

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

$$\mathrm{VIF}_{(j)} = \frac{1}{(1 - R_j^2)} \quad ; j = 1,2, \dots, k \tag{2.4}$$

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

2.5 Matriks Kovarian

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

Menurut Krzanowski (1998), matriks kovarian didefinisikan sebagai $\boldsymbol{\Sigma} = E\left[\left(\boldsymbol{X} - E(\boldsymbol{X})\right)\left(\boldsymbol{X} - E(\boldsymbol{X})\right)^T\right]$ yang merupakan pengembangan dari nilai $cov(x_i, x_j) = E\left[(x_i - E(x_i))\left(x_j - E(x_j)\right)\right]$ dan $var\ (x_i) = E[(x_i - E(x_i))]^2$.

Dengan *k* peubah, x_1, x_2, …, x_k , ada *k* varian dan $\frac{k(k-1)}{2}$ kovarian. Secara umum, perhitungan ini dihasilkan dari suatu $k \times k$ matriks simetris $\boldsymbol{\Sigma}$ yaitu:

$$\boldsymbol{\Sigma} = E\left[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T\right] \tag{2.5}$$

$$= E\left(\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_k - \mu_k \end{bmatrix} \begin{bmatrix} X_1 - \mu_1 & X_2 - \mu_2 & \dots & X_k - \mu_k \end{bmatrix}\right)$$

$$= \begin{bmatrix} E(X_1-\mu_1)^2 & E(X_1-\mu_1)(X_2-\mu_2) & \dots & E(X_1-\mu_1)(X_k-\mu_k) \\ E(X_2-\mu_2)(X_1-\mu_1) & E(X_2-\mu_2)^2 & \dots & \mathrm{E}(X_2-\mu_2)(X_k-\mu_k) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_k-\mu_k)(X_1-\mu_1) & E(X_k-\mu_k)(X_2-\mu_2) & \dots & E(X_k-\mu_k)^2 \end{bmatrix}$$

$$= \begin{bmatrix} var(x_1) & cov(x_1,x_2) & \dots & cov(x_1,x_k) \\ cov(x_2,x_1) & var(x_2) & \dots & cov(2,x_k) \\ \vdots & \vdots & \ddots & \vdots \\ cov(x_k,x_1) & cov(x_k,x_2) & \dots & var(x_k) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{bmatrix}$$

Matriks ini biasanya disebut matriks varian kovarian. $\boldsymbol{\Sigma}$ merupakan matriks simetrik karena $\mathrm{cov}(x_i, x_j) = \mathrm{cov}(x_j, x_i)$ atau $\sigma_{ii} = \sigma_{ij}$. Matriks $\boldsymbol{\Sigma}$ diduga oleh matriks **S**.

S adalah penduga matriks varian kovarian kelompok ke-*i* yang didefinisikan oleh:

$$\boldsymbol{S} = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{x_i} - \overline{\boldsymbol{x}})(\boldsymbol{x_i} - \overline{\boldsymbol{x}})^T \qquad (2.6)$$

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

2.6 Metode Kuadrat Terkecil

Menurut Sembiring (2003), Metode kuadrat terkecil adalah metode yang digunakan untuk mengestimasi nilai parameter regresi dengan meminimumkan jumlah kuadrat galat. Estimasi koefisien regresi dengan metode kuadrat terkecil yaitu dengan meminimumkan fungsi :

$$J = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \qquad (2.7)$$

Dalam persamaan (2.7), x_i dan y_i bilangan yang berasal dari pengamatan, sedangkan β_0 dan β_1 berubah bila garis regresinya berubah. Untuk mendapatkan nilai β_0 dan β_1, maka J diturunkan terhadap β_0 dan β_1 kemudian menyamakannya dengan nol. Jadi, bila J diturunkan terhadap β_0 , diperoleh :

$$\frac{\partial J}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0$$

atau,

$$\sum_{i=1}^{n} y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} x_i = 0 \tag{2.8}$$

selanjutnya, J diturunkan terhadap β_1 dan samakan dengan nol,

$$\frac{\partial J}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i = 0$$

atau,

$$\sum_{i=1}^{n} y_i x_i - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 = 0 \tag{2.9}$$

Nilai β_0 dan β_1 pada persamaan (2.8) dan (2.9) diganti dengan masing-masing taksirannya, b_0 dan b_1. Persamaan (2.8) dan (2.9) kemudian menjadi suatu sistem persamaan linear disebut persamaan normal.

$$\sum_{i=1}^{n} b_0 + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{2.10}$$

$$b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i \tag{2.11}$$

dengan $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ dan $\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$, maka persamaan (2.10) menjadi :

$$b_0 = \frac{\sum_{i=1}^{n} y_i}{n} - b_1 \frac{\sum_{i=1}^{n} x_i}{n}$$

$$= \overline{y} - b_1 \overline{x}$$

Persamaan (2.11) menjadi :

$$\sum_{i=1}^{n} y_i x_i - \left\{ \frac{\sum_{i=1}^{n} y_i}{n} - b_1 \left(\frac{\sum_{i=1}^{n} x_i}{n} \right) \right\} \left(\sum_{i=1}^{n} x_i \right) - b_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} y_i x_i - \frac{(\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i)}{n} - b_1 \left\{ \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n} \right\} = 0$$

Jadi,

$$b_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{(\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}$$

Rumus ini dengan mudah dapat disederhanakan menjadi:

$$b_1 = \frac{\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \qquad (2.12)$$

Seperti halnya dengan regresi linear sederhana, untuk penaksir dari nilai β_0, β_1, ..., β_k diganti dengan $b_0, b_1, \dots, b_k$. Menurut metode kuadrat terkecil penaksir tersebut dapat diperoleh dengan meminimumkan jumlah kuadrat galat.

$$J = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik})^2 \qquad (2.13)$$

Maka turunkan J secara parsial terhadap β_0, β_1, ..., β_k dan samakan dengan nol. Sehingga diperoleh,

$$\frac{\partial J}{\partial \beta_0} = -2\sum(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik}) = 0$$

$$\frac{\partial J}{\partial \beta_1} = -2\sum(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik})\, x_{i1} = 0$$

$$\frac{\partial J}{\partial \beta_2} = -2\sum(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik})\, x_{i2} = 0$$

$$\vdots$$

$$\frac{\partial J}{\partial \beta_k} = -2\sum(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik})\, x_{ik} = 0$$

Setelah disusun kembali dan ganti semua parameter dengan penaksirnya, sistem persamaan ini dapat ditulis sebagai :

$$\begin{aligned}
\sum y_i &= nb_0 + b_1 \sum x_{i1} + b_2 \sum x_{i2} + \cdots + b_k \sum x_{ik} \\
\sum y_i x_{i1} &= b_0 \sum x_{i1} + b_1 \sum x_{i1}^2 + b_2 \sum x_{i2} x_{i1} + \cdots + b_k \sum x_{ik} x_{i1} \\
\sum y_i x_{i2} &= b_0 \sum x_{i2} + b_1 \sum x_{i1} x_{i2} + b_2 \sum x_{i2}^2 + \cdots + b_k \sum x_{ik} x_{i2} \\
&\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (2.14) \\
\sum y_i x_{ik} &= b_0 \sum x_{ik} + b_1 \sum x_{i1} x_{ik} + b_2 \sum x_{i2} x_{ik} + \cdots + b_k \sum x_{ik}^2
\end{aligned}$$

Persamaan (2.14) disebut sebagai persamaan normal. Jika ditulis dalam lambang matriks maka bentuknya menjadi:

$$(\boldsymbol{X^T X})\boldsymbol{b} = \boldsymbol{X^T Y} \tag{2.15}$$

atau,

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i2} & \dots & \sum_{i=1}^{n} X_{ik} \\ \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i1}^2 & \sum_{i=1}^{n} X_{i2}X_{i1} & \dots & \sum_{i=1}^{n} X_{ik}X_{i1} \\ \sum_{i=1}^{n} X_{i2} & \sum_{i=1}^{n} X_{i1}X_{i2} & \sum_{i=1}^{n} X_{i2}^2 & \dots & \sum_{i=1}^{n} X_{ik}X_{i2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} X_{ik} & \sum_{i=1}^{n} X_{i1}X_{ik} & \sum_{i=1}^{n} X_{i2}X_{ik} & \dots & \sum_{i=1}^{n} X_{ik}^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & \dots & X_{n1} \\ X_{12} & X_{22} & \dots & X_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{ik} & X_{2k} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}$$

Jika $\boldsymbol{X^T X}$ tidak singular maka untuk estimasi kuadrat terkecil dari **b** pada persamaan (2.15) yaitu dikalikan dengan invers $\boldsymbol{X^T X}$**.** Jadi,

$$(\boldsymbol{X^T X})^{-\boldsymbol{1}}\boldsymbol{X^T X}\,\mathbf{b} = (\boldsymbol{X^T X})^{-\boldsymbol{1}}\boldsymbol{X^T Y}$$

$$\mathbf{b} = (\boldsymbol{X^T X})^{-\boldsymbol{1}}\boldsymbol{X^T Y} \tag{2.16}$$

Sifat-sifat penduga metode kuadrat terkecil yaitu sebagai berikut:

1. $\hat{\boldsymbol{\beta}}$ linear

   $\hat{\boldsymbol{\beta}}$ linear jika $\hat{\boldsymbol{\beta}}$ merupakan fungsi linear dari $\boldsymbol{\beta}$.

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\boldsymbol{X^T X})^{-\boldsymbol{1}}\boldsymbol{X^T Y} \\ &= (\boldsymbol{X^T X})^{-\boldsymbol{1}}\boldsymbol{X^T}(\boldsymbol{X\beta} + \boldsymbol{\varepsilon}) \\ &= (\boldsymbol{X^T X})^{-\boldsymbol{1}}\boldsymbol{X^T X\beta} + (\boldsymbol{X^T X})^{-\boldsymbol{1}}\boldsymbol{X^T \varepsilon} \\ &= \boldsymbol{I\beta} + (\boldsymbol{X^T X})^{-\boldsymbol{1}}\boldsymbol{X^T \varepsilon} \end{aligned}$$

2. $\widehat{\boldsymbol{\beta}}$ tak bias

$\widehat{\boldsymbol{\beta}}$ adalah penduga tak bias jika $\widehat{\boldsymbol{\beta}}$ linear jika $E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$

$$
\begin{aligned}
E(\widehat{\boldsymbol{\beta}}) &= E((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T Y}) \\
&= E((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T X\beta} + \boldsymbol{\varepsilon}) \\
&= E((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T X\beta} + (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T \varepsilon}) \\
&= ((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T X\beta} + (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T} E(\boldsymbol{\varepsilon}) \\
&= ((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T X\beta}) \\
&= \boldsymbol{\beta}
\end{aligned}
$$

3. $\widehat{\boldsymbol{\beta}}$ memiliki variansi minimum

$$
\begin{aligned}
var(\widehat{\boldsymbol{\beta}}) &= E\left[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\boldsymbol{T}}\right] \\
&= E[((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T Y} - \boldsymbol{\beta})((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T Y} - \boldsymbol{\beta})^{\boldsymbol{T}}] \\
&= E[((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T}(\boldsymbol{X\beta} + \boldsymbol{\varepsilon}) - \boldsymbol{\beta})((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T}(\boldsymbol{X\beta} + \boldsymbol{\varepsilon}) - \boldsymbol{\beta})^{\boldsymbol{T}}] \\
&= E[((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T X\beta} + (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T \varepsilon}) - \boldsymbol{\beta})\ ((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T X\beta} + \\
&\quad (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T \varepsilon}) - \boldsymbol{\beta})^{\boldsymbol{T}}] \\
&= E[(\boldsymbol{\beta} + (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T \varepsilon} - \boldsymbol{\beta})\ (\boldsymbol{\beta} + (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T \varepsilon} - \boldsymbol{\beta})^{\boldsymbol{T}}] \\
&= E[((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T \varepsilon})\ ((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T \varepsilon})^{\boldsymbol{T}}] \\
&= E[((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T \varepsilon^T \varepsilon}\ \boldsymbol{X}\ (\boldsymbol{X^T X})^{-1}] \\
&= E[((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T E}(\boldsymbol{\varepsilon^T \varepsilon})\ \boldsymbol{X}\ (\boldsymbol{X^T X})^{-1}] \\
&= (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T X}\ (\boldsymbol{X^T X})^{-1}\ E(\boldsymbol{\varepsilon\varepsilon^T}) \\
&= (\boldsymbol{X^T X})^{-1}\sigma^2
\end{aligned}
$$

$var(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X^T X})^{-1}\sigma^2$ merupakan varians terkecil dari semua penaksir linear tak bias

Selama asumsi-asumsi regresi terpenuhi, maka dugaan metode kuadrat terkecil bersifat linear, tak bias, dan mempunyai varians minimum ini bersifat BLUE (*Best Linier Unbiased Estimator*).

2.7 Analisis Komponen Utama

Menurut Jollife (2002), mendefinisikan bahwa 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 negative) antar peubah, makin baik pula hasil yang diperoleh dari AKU (Mattjik & Sumertajaya, 2011).

Dalam analisis komponen utama misalkan $X_1,\ X_2,\ \ldots\ ,\ X_k$ memiliki sebaran peubah ganda dengan vektor rataan $\boldsymbol{\mu}$ dan matriks kovarian $\boldsymbol{\Sigma}$ serta vektor ciri $\boldsymbol{a_k}$. Komponen utama (Q) seperti telah dijelaskan diatas merupakan kombinasi linier dari *k* peubah asal, atau dapat ditulis :

$$\boldsymbol{Q} = \boldsymbol{AX} \qquad (2.17)$$

dimana:

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

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

$$Q_1 = a_{11}X_1 + a_{21}X_2 + \cdots + a_{k1}X_k = \boldsymbol{a_1^T X}$$

Komponen utama kedua dapat dituliskan sebagai:

$$Q_2 = a_{12}X_1 + a_{22}X_2 + \cdots + a_{k2}X_k = \boldsymbol{a_2^T X}$$

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

$$Q_j = a_{1j}X_1 + a_{2j}X_2 + \cdots + a_{kj}X_k = \boldsymbol{a_j^T X} \qquad (2.18)$$

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.

2.8 Regresi Komponen Utama

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 MKT menjadi gagal terpenuhi dan salah satu cara membebaskan korelasi antar peubah bebasnya dengan regresi komponen utama.

Menurut Notiragayu & Nisa (2008), regresi komponen utama merupakan regresi dari peubah tak bebas terhadap komponen-komponen utama yang tidak saling berkorelasi, dimana setiap komponen utama merupakan kombinasi linear dari semua peubah bebas yang telah dispesifikasikan sejak awal.

Bentuk persamaan regresi dalam bentuk peubah asal X dapat ditulis sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon \qquad (2.19)$$

Dengan Y merupakan peubah tak bebas, X_i peubah bebas ke-i *(i = 1, 2, ..., k)*, β_i adalah parameter-parameter regresi, dan ε merupakan galat.

Peubah baru sebagai komponen utama (Q) adalah hasil transformasi dari peubah asal (X) yang modelnya dalam bentuk matriks pada Persamaan (2.17), dan komponen utama ke-j ditulis pada persamaan (2.18) dimana vektor pembobot $\boldsymbol{a}_j^T$ diperoleh dengan memaksimumkan keragaman komponen utama ke-j, yaitu $\boldsymbol{S}_{Q_j}^2 = \boldsymbol{a}_j^T \boldsymbol{S} \boldsymbol{a}_j$ dengan kendala $\boldsymbol{a}_j^T \boldsymbol{a}_j = \mathbf{1}$ serta $\boldsymbol{a}_h^T \boldsymbol{a}_j = \mathbf{0},$ untuk $h \neq j$. Vektor pembobot $\boldsymbol{a}_j^T$ diperoleh dari matriks kovarian $\boldsymbol{\Sigma}$ yang diduga dengan matriks $\boldsymbol{S}$, yaitu $\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x_i} - \overline{\boldsymbol{x}})(\boldsymbol{x_i} - \overline{\boldsymbol{x}})^T.$ Vektor $\boldsymbol{a}_j^T$ yang memenuhi kendala diatas adalah vektor eigen dari matriks kovarian $\boldsymbol{\Sigma}$. Model regresi komponen utama dapat ditulis sebagai berikut:

$$Y = \beta_0 + \beta_1 Q_1 + \beta_2 Q_2 + \ldots + \beta_j Q_j + \varepsilon \text{ , dengan } j \leq k \qquad (2.20)$$

2.9 Pencilan *(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 pencilan akan mengganggu dalam proses analisis data dan harus dihindari dalam banyak hal. Dalam kaitannya dengan analisis regresi, pencilan dapat menyebabkan hal-hal berikut :

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

2.10 *Minimum Covariance Determinant* (MCD)

Minimum Covariance Determinant (MCD) merupakan metode penduga parameter dengan meminimumkan determinan matriks kovarians. Metode MCD menggunakan vektor rata-rata dan matriks kovarians yang didapat dari penduga MCD untuk menentukan bobot dari setiap data, sehingga akan didapat penduga parameter model MCD. Prinsip MCD adalah mencari subsampel H yang berukuran h dari keseluruhan n amatan dengan $h \leq n$ yang matriks kovariansnya memiliki determinan terkecil di antara semua kombinasi kemungkinan data (Kumalasari *et al.*, 2017).

Penduga MCD adalah pasangan $(\boldsymbol{\bar{X}}, \boldsymbol{S})$, dimana $\boldsymbol{\bar{X}}$ adalah vektor rata-rata dan $\boldsymbol{S}$ adalah matriks kovariansi yang meminimumkan nilai determinan $\boldsymbol{S}$ pada subsampel yang berisikan tepat sebanyak *h* anggota dari *n* pengamatan, dimana nilai standar dari *h*=[(n+k+1)/2]. Pada populasi dengan jumlah pengamatan yang kecil, penduga MCD dapat dengan cepat dihitung dan ditemukan, tetapi jika jumlah pengamatan besar, maka akan banyak sekali kombinasi subsampel dari *H* yang harus ditemukan dan penghitungan pun akan cukup memakan waktu. Dalam mengatasi keterbatasan ini, maka digunakan suatu algoritma baru untuk metode MCD yang dinamakan dengan metode *fast*-MCD (Rousseeuw & Van Driessen, 1999).

Menurut Dayanti *et al.* (2016), langkah-langkah penduga MCD dalam menduga nilai koefisien regresi dengan menggunakan *fast*-MCD adalah sebagai berikut.

1. Mengambil subsampel pertama dari data pengamatan secara acak, misalkan subsampel tersebut H_1 dengan jumlah elemen sebanyak *h*.
2. Menghitung vektor rata-rata $\boldsymbol{\bar{X}}_{MCD}$ pada H_1 yang selanjutnya dapat disebut $\boldsymbol{\bar{X}}_1$ dengan persamaan berikut.

$$\boldsymbol{\bar{X}}_{MCD} = \frac{1}{h} \sum_{i \epsilon H} \boldsymbol{x_i} \tag{2.21}$$

3. Menghitung matriks kovarians $\boldsymbol{S}_{MCD}$ pada H_1 yang selanjutnya dapat disebut $\boldsymbol{S}_1$ dengan persamaan berikut.

$$\boldsymbol{S}_{MCD} = \frac{1}{h} \sum_{i \epsilon H} [\boldsymbol{x_i} - \boldsymbol{\bar{X}}_{MCD}][\boldsymbol{x_i} - \boldsymbol{\bar{X}}_{MCD}]^T \tag{2.22}$$

4. Menghitung determinan $\boldsymbol{S}_1$. Jika *det*($\boldsymbol{S}_1$) = 0, maka berhenti. Namun apabila *det*($\boldsymbol{S}_1$) ≠ 0, maka dilanjutkan dengan menghitung jarak *robust* (*RD*) untuk setiap pengamatan menggunakan persamaan berikut.

$$RD_i = \sqrt{(\boldsymbol{x}_i - \overline{\boldsymbol{X}}_{MCD})^T \boldsymbol{S}_{MCD}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{X}}_{MCD})} \tag{2.23}$$

5. Mengurutkan nilai *RD* mulai dari yang terkecil hingga terbesar.
6. Mengambil elemen sejumlah h pengamatan dengan jarak terkecil berdasarkan pada langkah 5 untuk menjadi himpunan bagian H_2, kemudian ulangi langkah 2 sampai langkah 5 sehingga ditemukan himpunan bagian yang konvergen det ($\boldsymbol{S}_{i+1}$) = det ($\boldsymbol{S}_{\mathbf{i}}$).
7. Memilih subsampel H yang memiliki nilai determinan matriks kovarians terkecil. Kemudian dari subsampel yang terpilih akan dicari nilai $\overline{\boldsymbol{X}}_{MCD}$ dan $\boldsymbol{S}_{MCD}$.

8. Dengan menggunakan $\overline{\boldsymbol{X}}_{MCD}$ dan $\boldsymbol{S}_{MCD}$ tersebut, maka selanjutnya akan dihitung jarak *robust* kuadrat dengan persamaan sebagai berikut.

$$RD_i^2 = (\boldsymbol{x}_i - \overline{\boldsymbol{X}}_{MCD})^T \boldsymbol{S}_{MCD}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{X}}_{MCD}) \tag{2.24}$$

9. Menentukan pencilan data dengan kriteria sebagai berikut.

$$RD_i^2 > X_{p;\alpha}^2$$

2.11 *Least Trimmed Square* (LTS)

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 pencilan yang mengakibatkan estimasi koefisien regresi menjadi bias. Tujuan utama dari regresi *robust* adalah untuk memberikan hasil yang resisten terhadap adanya pencilan. Dalam keadaan ini regresi robust merupakan metode terbaik yang tahan terhadap pencilan (Chen, 2002).

Prosedur *robust* ditunjukkan untuk memberikan dugaan yang lebih tepat dan cepat terhadap data yang melanggar asumsi dengan cara meniadakan identifikasi adanya data pencilan, serta bersifat otomatis dalam menanggulangi data pencilan. 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 pencilan adalah metode penduga *Least Trimmed Square* (LTS). Metode ini merupakan salah satu metode pendugaan parameter pada regresi *robust* yang kekar terhadap keberadaan pencilan. Metode LTS mempunyai prinsip pendugaan parameter yang sama dengan Metode Kuadrat Terkecil, 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 pencilan. (Rousseeuw & Hubert, 1997).

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

$$\hat{\beta} = \arg\min_{\beta}\left(\sum_{i=1}^{h} e_i^2\right) = \arg\min_{\beta}\left(\sum_{i=1}^{h}(y_i - \hat{y}_i)^2\right),\ \frac{(3n+k+1)}{4} \leq h \leq n \quad (2.25)$$

Solusi $\hat{\beta}$ pada Persamaan (2.25) dapat diperoleh dengan menggunakan turunan/differensial seperti pada penyelesaian pendugaan metode OLS, hanya pada LTS persamaan tersebut dihitung pada subhimpunan H terbaik dilakukan dengan menggunakan algoritma resampling dari seluruh kemungkinan subhimpunan yang dapat dibentuk yaitu sebanyak $\binom{n}{h}$. Subhimpunan H yang diperoleh merupakan sebaran data yang sudah terpangkas (Chen, 2002).

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 tesebut, 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 yang terbesar, yaitu $|e_1| \le |e_2| \le \cdots \le |e_n|$.
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 subsampel 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 h pengamatan untuk mendapatkan galat terkecil. Langkah-langkah *C-Step* adalah sebagai berikut; misalnya telah diberikan sub himpunan H berukuran h dari sampel berukuran n, selanjutnya:

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, $e_1 \leq e_2 \leq \cdots \leq e_n$.
4. Ambil H sebanyak h pengamatan yang memiliki nilai mutlak galat terkecil.

Pengulang *C-Step* akan menghasilkan proses iterasi.

2.12 Ukuran Kebaikan Penduga

2.12.1 Bias

Menurut Wulandari *et al.* (2010), 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 yang tanpa pencilan, dibagi dengan banyaknya perulangan. Semakin kecil nilai bias, maka hasil pendugaan suatu parameter semakin baik.

Bias dinotasikan sebagai berikut.

$$Bias(\hat{\beta}) = \frac{1}{m}\sum_{i=1}^{m}\left|\hat{\beta}_i^{(s)} - \hat{\beta}^{(0)}\right| \quad (2.26)$$

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

2.12.2 *Mean Square Error* (MSE)

Mean Square Error (MSE) adalah salah satu pengukuran kesalahan yang popular dan mudah digunakan. 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:

$$MSE(\hat{\beta}) = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{\beta}_i^{(s)} - \hat{\beta}^{(0)}\right)^2 \quad (2.27)$$

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

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun akademik 2018/2019, bertempat di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data simulasi. Pada data simulasi akan dibangkitkan data dengan variabel bebas sebanyak k=6 dengan ukuran sampel (n = 20, 60, 100, 200) serta 5 jenis persentase pencilan (5%, 10% 15%, 20%, 25%) dan diulang sebanyak 1000 kali. Kemudian pencilan dibangkitkan dari distribusi normal dengan mean 50 dan simpangan baku 0,05 yaitu $\varepsilon^{*} \sim N\left(50, (0.05)^2\right)$.

Untuk mendapatkan data kolinearitas pada setiap himpunan data, X_{ik} dibangkitkan menggunakan simulasi Monte Carlo berdasarkan McDonald & Galarneau (1975) dengan persamaan sebagai berikut:

$$X_{ij} = (1-\rho^2)^{1/2} x_{ij} + \rho x_{i(k+1)} \qquad ;\ i = 1, 2, \dots, n \qquad j = 1, 2, \dots, k \tag{3.1}$$

dengan $x_{i1}, x_{i2}, \dots, x_{i(k+1)}$ merupakan data yang dibangkitkan berdistribusi normal N(0,1) dan ρ ditentukan sehingga korelasi antarvariabel bebas diberikan oleh ρ^2. Dua himpunan dari variabel yang saling berkorelasi dalam penelitian ini dibuat berdasarkan nilai $\rho = 0.99$.

Variabel terikat (Y) untuk setiap k variabel bebas diperoleh berdasarkan model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ dimana β adalah $\beta_{i,j} = 1$; untuk i = j, dan 0 selainnya. Selanjutnya ε dibangkitkan berdasarkan distribusi normal $N(0,1)$ sehingga Y merupakan kombinasi linear dari k variabel bebas ditambah galat seperti persamaan berikut:

$$Y = 1 + X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + N(0,1)$$

3.3 Metode Penelitian

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*) SAS 9.4. Adapun tahapan yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Melakukan simulasi data.
2. Melakukan regresi komponen utama klasik dengan langkah sebagai berikut:
   a. Menghitung matriks kovarian dari variabel asal (X).
   b. Menghitung nilai eigen λ_i dan vektor eigen a dari matriks kovarian.
   c. Menghitung skor komponen utama Q.

d. Memilih komponen utama yang memiliki nilai eigen lebih dari 1.

e. Menghitung nilai duga koefisien regresi komponen utama berdasarkan komponen yang terpilih dengan Metode Kuadrat Terkecil (MKT) simpan sebagai $\widehat{\beta}^{(0)}$.

3. Mencatat nilai duga β pada regresi komponen utama klasik.
4. Membangkitkan sebuah matriks noise dari distribusi $N(0,(0.01)^2)$.
5. Membangkitkan matriks pencilan dari distribusi $N(50,(0.05)^2)$ sehingga diperoleh matriks kontaminasi. Elemen dari matriks kontaminasi adalah nol kecuali untuk beberapa elemen yang dijadikan pencilan.
6. Menambahkan matriks noise dan matriks kontaminasi pada data simulasi. Sehingga diperoleh matriks untuk data yang telah terkontaminasi pencilan (XYk).
7. Melakukan regresi komponen utama klasik seperti langkah 2 pada data yang telah terkontaminasi pencilan.
8. Mencatat nilai duga β pada regresi komponen utama klasik di langkah 7.
9. Melakukan regresi komponen utama *robust* dengan langkah sebagai berikut:
   a. Meghitung matriks kovarian *Minimum Covariance Determinant* (MCD) dari data yang terkontaminasi pencilan.
   b. Menghitung nilai eigen λ_i dan vektor eigen *a* dari matriks kovarian MCD.
   c. Menghitung skor komponen utama *Q*.
   d. Memilih komponen utama yang memiliki nilai eigen lebih dari 1.
   e. Menghitung nilai duga koefisien regresi komponen utama berdasarkan komponen yang terpilih dengan *Least Trimmed Square* (LTS).

10. Mencatat nilai duga β pada regresi komponen utama *robust*.
11. Ulangi langkah 4 sampai 10 sebanyak 1000 kali untuk seluruh jumlah data.
12. Menghitung nilai bias dan *Mean Square Error* (MSE) dari RKU klasik dan RKU *robust* dengan menggunakan rumus sebagai berikut:

$$\text{Bias}\left(\hat{\beta}_j\right) = \frac{1}{m}\sum\nolimits_{i=1}^{m}\left|\hat{\beta}_{ij}^{(s)} - \hat{\beta}_j^{(0)}\right|$$

$$\text{MSE}\left(\hat{\beta}_j\right) = \frac{1}{m}\sum\nolimits_{i=1}^{m}\left(\hat{\beta}_{ij}^{(s)} - \hat{\beta}_j^{(0)}\right)^2$$

dimana,

j = 0, 1, ..., k

m = 1000

13. Menganalisa dan membandingkan hasil perhitungan RKU klasik dengan RKU *robust* berdasarkan nilai bias dan *Mean Square Error* (MSE) dari dugaan koefisien regresi yang dihasilkan.

V. KESIMPULAN

Berdasarkan hasil dan pembahasan, maka diperoleh kesimpulan sebagai berikut:

1. Analisis Komponen Utama *Robust* dapat dilakukan dengan menggunakan matriks kovarian *robust Minimum Covariance Determinant* $(\boldsymbol{\Sigma}_{\boldsymbol{MCD}})$ yang dalam hal ini diperoleh:

   a. Penduga bagi matriks kovarian *Minimum Covariance Determinant* (MCD) yaitu $\boldsymbol{S}_{MCD} = \frac{1}{h}\sum_{i\in H}[\boldsymbol{x_i} - \overline{\boldsymbol{X}}_{MCD}][\boldsymbol{x_i} - \overline{\boldsymbol{X}}_{MCD}]^T$.

   b. Penduga nilai koefisien Regresi Komponen Utama *Robust* diperoleh sebagai berikut:

   $$\hat{\beta}_0 = \frac{\sum_{i=1}^{h} y_i - \beta_1 \sum_{i=1}^{h} Q_i}{h} \quad , \quad \hat{\beta}_1 = \frac{h\sum_{i=1}^{h} y_i Q_i - \sum_{i=1}^{h} y_i \sum_{i=1}^{h} Q_i}{h\sum_{i=1}^{h} {Q_i}^2 - \sum_{i=1}^{h}(Q_i)^2}$$

   dimana,

   Q = Komponen Utama

   h = Banyaknya anggota dalam subhimpunan terbaik dari data yang terpangkas.

2. Pada hasil simulasi diperoleh:

   a. Metode Regresi Komponen Utama *Robust* lebih efektif dan efisien dalam mengatasi masalah multikolinearitas dan pencilan dibandingkan RKU klasik.

b. Pada satu ukuran sampel dengan presentase pencilan 5%-25%, nilai rata-rata bias dan MSE metode RKU klasik selalu naik secara konstan disetiap penambahan presentase pencilan sedangkan metode RKU *robust* diperoleh nilai rata-rata bias yang lebih kecil dibandingkan RKU klasik. Hal ini menunjukkan bahwa metode RKU *robust* merupakan metode yang kekar terhadap pencilan, sedangkan metode RKU klasik sangat sensitif terhadap adanya pencilan.

DAFTAR PUSTAKA

Anton, H. 1987. *Aljabar Linier Elementer*. Erlangga, Jakarta.

Chatterjee, S. & Hadi, A.S. 2006. *Regression Analysis by Example*. 4thEdition. Jhon Wiley and Sons Inc., New Jersey.

Chen, C. 2002. *Robust Regression and Outlier Detection with the ROBUSTREG procedure*. SAS Institue : Cary, NC. 265-27.

Dayanti, N.P., Suciptawati, N.L., & Susilawati, M. 2016. Penerapan Bootstrap dalam Metode Minimum Covariance Determinant (MCD) dan Least Median Square (LMS) pada Analisis Regresi Linear Berganda. *E-Jurnal Matematika*, **5**(1), 22-26.

Draper, N.R. & Smith, H. 1992. *Analisis Regresi Terapan*. Ed. Ke-2. Diterjemahkan oleh Bamang Sumantri. Gramedia Pustaka Umum, Jakarta.

Gaspersz, V. 1991. *Ekonometri Terapan*. Ed. Ke-2. Tarsito, Bandung.

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

Harinaldi. 2005. *Prinsip-Prinsip Statistik untuk Teknik dan Sains*. Erlangga, Jakarta.

Jollife, I.T. 2002. *Principal Component Analysis.* Second Edition. Springer-Verlag. Inc, New York.

Krzanowski, W. 1998. *An Introduction to Statistical Modelling*. Wiley-Blackwell, New Jersey.

Kumalasari, N., Suciptawati, N., & Susilawati, M. 2017. Perbandingan Metode MCD-Bootstrap dan LAD-Bootstrap dalam Mengatasi Pengaruh Pencilan pada Analisis Regresi Linear Berganda. *E-Jurnal Matematika*. **6**(1): 47-55.

Mattjik, A. A., & Sumertajaya, I. M. 2011. *Sidik Peubah Ganda dengan Menggunakan SAS*. IPB PRESS, Bogor.

McDonald, G.C. & Galarneau, D.I. 1975. Amonte Carlo Evaluation of Some Ridge-Type Estimators. *J. Amer. Statist. Asoc.* **70**:407-416.

Montgomery, D.C. & Runger, G.C. 2011. *Applied Statistics and Probability for Engineers*. Ed. ke-5. John Wiley & Sons, New York.

Morrison, D.F. 1967. *Multivariate Statistical Methods.* McGraw-Hill, New York.

Myers, R.H. 1990. *Clasical and Modern Regression with Application.* PWSKENT publishing Company, Boston.

Nisa, K. 2006. Analisis Regresi Robust Menggunakan Metode *Least Trimmed Square* untuk Data Mengandung Pencilan. *Jurnal Ilmiah MIPA.* **IX**(2):1-9.

Notiragayu & Nisa, K. 2008. Analisis Regresi Komponen Utama *Robust* untuk Data Mengandung Pencilan. *Jurnal Sains MIPA.* **14**(1):45-50.

Raykov, T. & Marcoulides, G.A. 2008. *An Introduction to Applied Multivariat Analysis*. Taylor and Fracis Group, New York.

Rousseeuw, P.J. & Van Driessen, K. 1999. A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*. **41**(3), 212-223.

Roousseeuw, P.J. & Hubert, M. 1997. Robust Regression with Continuous and Binary Regressors. *Journal of Statistical Planning an Inference.* **57**:153-163.

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

Usman, M. & Warsono. 2001. *Teori Model Linier & Aplikasinya*. CV. Darmajaya, Bandar Lampung.

Weisberg, S. 1985. *Applied Linear Regression.* Wiley, New York.

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.