

**ANALISIS REGRESI BERGANDA DENGAN METODE *PARTIAL*  
*ROBUST-M* DALAM MENGATASI MASALAH MULTIKOLINEARITAS  
DAN PENCILAN**

**(Skripsi)**

**Oleh**

**TRI DEWI CAHYANI**



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

## **ABSTRACT**

### **Multiple Regression Analysis Using Partial Robust-M Method to Overcome Multicollinearity and Outliers Data**

**By**

**Tri Dewi Cahyani**

Partial Robust-M Regression (PRM) is a robust alternative approach method of Partial Least Square (PLS) regression which is used as a method of handling multicollinearity and data contained outliers in one case. The purpose of this study was to determine the performance of the Partial Robust-M method in overcoming data containing multicollinearity and outliers and to compare the estimated value with the Ordinary Least Square (OLS) method. The results of this study indicate that PRM is better at handling data containing multicollinearity and outliers than the OLS method based on the standard error and MSE values obtained. The Human Development Index data from Central Java Province in 2020 is influenced by the expected length of school, the average length of school, the number of hospitals, and the number of doctors in each district/city based on the PRM method, while the OLS method is only influenced by the school average variable.

**Keyword :** Partial Robust-M Regression (PRM), Multicollinearity, Outliers data.

## ABSTRAK

### ANALISIS REGRESI BERGANDA DENGAN METODE *PARTIAL ROBUST-M* DALAM MENGATASI MASALAH MULTIKOLINEARITAS DAN PENCILAN

Oleh

Tri Dewi Cahyani

Regresi *Partial Robust-M* (PRM) merupakan suatu metode pendekatan alternatif robust dari regresi *Partial Least Square* (PLS) yang digunakan sebagai metode penanganan multikolinieritas dan data pencilan dalam satu kasus. Tujuan dari penelitian ini adalah untuk mengetahui performa metode *Partial Robust-M* dalam mengatasi data yang mengandung multikolinearitas dan pencilan serta membandingkan nilai dugaannya dengan metode *Ordinary Least Square* (OLS). Hasil dari penelitian ini menunjukkan bahwa PRM lebih baik dalam menangani data yang mengandung multikolinearitas dan pencilan dibandingkan dengan metode OLS berdasarkan nilai *standard error* dan MSE yang didapat. Data Indeks Pembangunan Manusia (IPM) Provinsi Jawa Tengah tahun 2020 dipengaruhi oleh harapan lama sekolah, rata-rata lama sekolah, jumlah rumah sakit, dan jumlah dokter tiap kabupaten/kota berdasarkan metode PRM, sedangkan dengan metode OLS hanya dipengaruhi oleh variabel rata-rata sekolah.

**Kata Kunci:** *Partial Robust-M* (PRM), Multikolinearitas, Data pencilan.

**ANALISIS REGRESI BERGANDA DENGAN METODE *PARTIAL*  
*ROBUST-M* DALAM MENGATASI MASALAH MULTIKOLINEARITAS  
DAN PENCILAN**

Oleh

**TRI DEWI CAHYANI**

Skripsi

Sebagai Salah Satu Syarat Untuk Mencapai Gelar  
Sarjana Matematika

Pada

Jurusan Matematika  
Fakultas Matematika dan Ilmu Pengetahuan Alam



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

**Judul Skripsi : ANALISIS REGRESI BERGANDA DENGAN METODE PARTIAL ROBUST-M DALAM MENGATASI MASALAH MULTIKOLINEARITAS DAN PENCILAN**

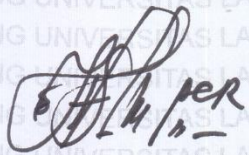
**Nama Mahasiswa : Tri Dewi Cahyani**

**Nomor Pokok Mahasiswa : 1717031085**


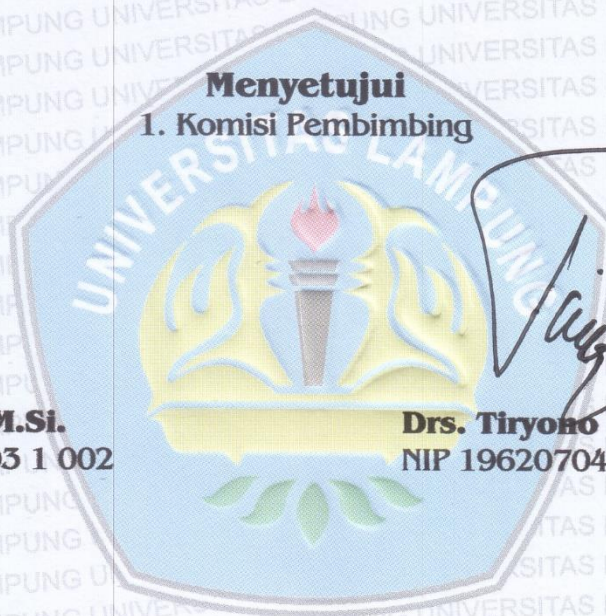
**Jurusan : Matematika**

**Fakultas : Matematika dan Ilmu Pengetahuan Alam**

**Menyetujui**  
**1. Komisi Pembimbing**

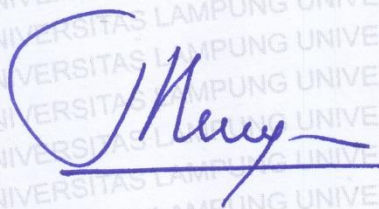


**Drs. Eri Setiawan, M.Si.**  
**NIP 19581101 198803 1 002**



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

**2. Mengetahui**  
**Ketua Jurusan Matematika**

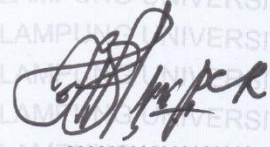


**Dr. Aang Nuryaman, S.Si., M.Si.**  
**NIP 19740316 200501 1 001**

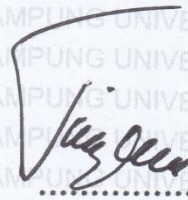
**MENGESAHKAN**

**1. Tim Peguji**

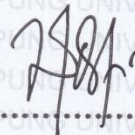
**Ketua : Drs. Eri Setiawan, M.Si.**



**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. Supto Dwi Yuwono, S.Si., M.T.**  
**NIP 19740705 200003 1 001**

**Tanggal Lulus Ujian Skripsi : 05 Agustus 2021**

## PERNYATAAN SKRIPSI MAHASISWA

Saya yang bertanda tangan di bawah ini:

Nama : Tri Dewi Cahyani

Nomor Pokok Mahasiswa : 1717031085

Jurusan : Matematika

Judul : **ANALISIS REGRESI BERGANDA DENGAN  
METODE *PARTIAL ROBUST*-M DALAM  
MENGATASI MASALAH MULTIKOLINEARITAS  
DAN PENCILAN**

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

Bandar Lampung, 05 Agustus 2021

Penulis,



**Tri Dewi Cahyani**

NPM 1717031085

## **RIWAYAT HIDUP**

Penulis bernama lengkap Tri Dewi Cahyani lahir di Bandar Lampung pada tanggal 30 Juni 1999. Penulis merupakan anak ketiga dari empat bersaudara dari pasangan Bapak Muhammad Saman dan Ibu Erni.

Penulis telah menempuh Pendidikan di Taman Kanak-Kanak (TK) Aisyiah Bustaful Anfal pada tahun 2004-2005, melanjutkan ke Sekolah Dasar (SD) Negeri 1 Sidodadi selama 6 tahun 2005-2011, melanjutkan ke Sekolah Menengah Pertama (SMP) Negeri 22 Bandar Lampung 2011-2014, dan melanjutkan ke Sekolah Menengah Atas di Madrasah Aliyah Negeri (MAN) 1 Bandar Lampung 2014-2017.

Pada tahun 2017 penulis terdaftar sebagai mahasiswa di jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung melalui jalur Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN). Selama menjadi mahasiswa penulis cukup aktif mengikuti organisasi kampus antara lain, menjadi staff Menteri Dalam Negeri BEM U KBM pada tahun 2018, serta menjadi anggota Divisi Komisi IV DPM U pada tahun 2019.

Pada awal tahun 2020 penulis melaksanakan KKN (Kuliah Kerja Nyata) di Desa Pasar Batang, Kecamatan Penawar Aji, Kabupaten Tulang Bawang selama 40 hari sebagai bentuk pengabdian mahasiswa terhadap Tri Dharma Perguruan Tinggi. Kemudian pada pertengahan tahun 2020 penulis melaksanakan Kuliah Praktik di PT. Telkom Witel Lampung.



## **KATA INSPIRASI**

*“Karena sesungguhnya sesudah kesulitan itu ada kemudahan, sesungguhnya sesudah kesulitan itu ada kemudahan”.*

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

*“Sesungguhnya Allah tidak akan mengubah keadaan suatu kaum hingga mereka merubah keadaan yang ada pada diri mereka sendiri”.*

(QS. Ar Rad : 11)

*“If you never try, You’ll never know”.*

(John Barrow)

*“It always seem imposible until it’s done”.*

(Nelson Mandela)

## **PERSEMBAHAN**

Dengan mengucapkan puji dan syukur kehadirat Allah SWT yang telah memberikan petunjuk dan kemudahan untuk menyelesaikan studiku, kupersembahkan karyaku yang sederhana ini kepada

### **Mama dan Papa Tercinta**

Terimakasih kepada mama dan papa yang selalu memberikan semangat, doa, mendukung apapun yang dilakukan penulis, yang tak pernah kenal lelah dan selalu berusaha memberikan yang terbaik.

### **Ayukku Sari, Ayukku Maya, Adikku Rani, Raihan, dan seluruh keluarga besar**

Terimakasih atas semangat yang selalu kalian berikan juga doa, yang selalu ada dimasa senang dan sulitku, memberikan keceriaan dalam hidup ini.

### **Sahabat-sahabatku**

Terimakasih kepada para teman dan sahabat yang selalu ada disisiku dalam menjalani semua proses baik suka dan duka selama ini. Terimakasih atas semangat, doa, motivasi, serta pembelajaran yang diberikan.

## SANWACANA

Alhamdulillah Robbil' alamin, puji dan syukur Penulis ucapkan kepada Allah SWT, yang selalu melimpahkan rahmat dan kasih sayang-Nya, sehingga Penulis dapat menyelesaikan skripsi ini. Sholawat serta salam senantiasa tetap tercurah kepada Nabi Muhammad SAW, tuntunan dan tauladan utama bagi seluruh umat manusia. Skripsi dengan judul “Analisis Regresi Berganda Dengan Metode *Partial Robust-M* Untuk Mengatasi Multikolinearitas dan Pencilan” adalah salah satu syarat untuk memperoleh gelar Sarjana Matematika di Universitas Lampung.

Dalam menyelesaikan skripsi ini, banyak pihak yang telah membantu Penulis dalam memberikan bimbingan, dorongan, dan saran-saran. Sehingga dengan segala ketulusan dan kerendahan hati pada kesempatan ini Penulis mengucapkan terimakasih yang sebesar-besarnya kepada:

1. Bapak Drs. Eri Setiawan M.Si., selaku dosen pembimbing utama yang senantiasa memberikan bimbingan serta masukan sehingga penulis dapat menyelesaikan perkuliahan skripsi ini.
2. Bapak Drs. Tiryono Ruby, M.Sc., Ph.D., selaku dosen pembimbing kedua yang telah memberikan masukan dan saran dalam penyelesaian skripsi.
3. Ibu Widiarti, S.Si., M.Si., selaku pembahas dan penguji yang telah memberikan evaluasi, arahan, dan saran kepada penulis dalam penyelesaian skripsi.
4. Ibu Dr. Notiragayu, S.Si., M.Si., selaku dosen pembimbing akademik.
5. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam.
6. Bapak Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Seluruh dosen, staff dan karyawan Jurusan Matematika Universitas Lampung.

8. Mama dan Papa tercinta yang tak pernah lelah memberikan doa serta dukungan kepada penulis sehingga penulis dapat menyelesaikannya dengan baik.
9. Ayukku Sari, Ayukku Maya, Mas Ruli dan Raihan yang selalu memberikan semangat, motivasi, dorongan, dukungan dan juga doa kepada penulis.
10. Adikku Rani Nirmala yang juga menjadi sahabat bagi penulis, terimakasih atas semangat, dukungan, bantuan, terimakasih untuk selalu ada. Semoga adikku bisa menjadi lebih baik lagi.
11. Sahabat-sahabatku tersayang Nauli, Zahra, Innas, Sabila dan Febri yang selalu ada untuk penulis, memberikan semangat, motivasi dan tempat penulis berbagi cerita.
12. Sahabat-sahabatku semasa perkuliahan yaitu Indah, Dhea, Yustika, Epmi, Nita, Inas, Atina yang telah mewarnai masa-masa perkuliahan ini, yang selalu mendukung penulis, terimakasih atas semangat, doa, dan kenangan yang telah kita buat selama masa perkuliahan ini.
13. Teman-teman Matematika kelas C yaitu Ananto, Deo, Obit, Reza, Restu, Shintia, Nadhira, Vina, Dindha, Della, Maul, Indri, Putri, Nyoman, Rossa, Jeje, Horas dan teman-teman yang lain yang telah memberikan pengalaman berharga dan atas kebersamaan serta canda tawanya.
14. Teman-teman seperbimbingan Yulica, Widya, Umroh, Ayu dan teman-teman yang lain yang saling membantu penulis dalam mengerjakan skripsi.
15. Teman-teman mahasiswa jurusan Matematika 2017.
16. Teman-teman maupun dari pihak lain yang membantu dalam pengerjaan skripsi.

Penulis juga menyadari bahwa dalam penulisan skripsi ini masih banyak terdapat kekurangan. Oleh karena itu, penulis mengharapkan saran dan kritik yang membangun guna penelitian selanjutnya agar lebih baik.

Bandar Lampung, 05 Agustus 2021  
Penulis,

**Tri Dewi Cahyani**

## DAFTAR ISI

Halaman

<b>DAFTAR TABEL</b> .....	xv
---------------------------	----

<b>DAFTAR GAMBAR</b> .....	xvii
----------------------------	------

### I. PENDAHULUAN

1.1 Latar Belakang dan Masalah.....	1
1.2 Tujuan Penelitian .....	3
1.3 Manfaat Penelitian .....	3

### II. TINJAUAN PUSTAKA

2.1 Analisis Regresi .....	4
2.2 Analisis Regresi Linear Berganda .....	5
2.3 Asumsi Analisis Regresi .....	6
2.4 <i>Ordinary Least Square</i> (OLS) .....	6
2.5 Multikolinearitas .....	9
2.6 Pencilan.....	10
2.7 <i>Partial Least Square</i> (PLS).....	10
2.8 Regresi <i>Robust</i> .....	12
2.9 Regresi <i>Robust-M</i> .....	12
2.10 <i>Partial Robust-M</i> (PRM) .....	14
2.11 <i>Mean Square Error</i> (MSE) .....	16

### III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian .....	17
3.2 Data Penelitian .....	17
3.3 Metodologi Penelitian .....	19

### IV. HASIL DAN PEMBAHASAN

4.1 Hasil Analisis Pada Data Simulasi.....	22
4.1.1 Hasil Simulasi untuk Kelompok Data dengan n = 40.....	22
4.1.2 Hasil Simulasi untuk Kelompok Data dengan n=60.....	31
4.1.3 Hasil Simulasi untuk Kelompok Data dengan n=100.....	40

4.2 Perbandingan Nilai <i>Mean Square Error</i> (MSE) Data Simulasi .....	50
4.3 Hasil Analisis pada Data Indeks Pembangunan Manusia Provinsi Jawa Tengah.....	51
4.3.1 Estimasi Parameter Menggunakan Metode <i>Ordinary Least Square</i> (OLS) .....	51
4.3.2 Uji Asumsi Regresi .....	52
4.3.3 Deteksi Pencilan.....	54
4.3.4 Estimasi Parameter dengan Metode <i>Partial Robust-M</i> (PRM).....	55
4.3.5 Uji Signifikansi Parameter .....	57

## **V. KESIMPULAN**

## **DAFTAR PUSTAKA**

## **LAMPIRAN**

## DAFTAR TABEL

Tabel	Halaman
1. Simulasi Monte Carlo .....	18
2. Simulasi Pencilan .....	19
3. Korelasi antar variabel bebas dengan $n = 40$ .....	22
4. VIF (Variance Inflation Factor) dengan $n = 40$ .....	23
5. Nilai VIF pada $n=40$ dan 5% pencilan dengan metode PRM .....	24
6. Rata-rata $\beta$ , <i>Standard Error</i> (SE), MSE pada OLS dan PRM pada $n=40$ dan 5% pencilan dengan 1000 kali pengulangan.....	25
7. Nilai VIF pada $n=40$ dan 10% pencilan dengan metode PRM .....	27
8. Rata-rata $\beta$ , <i>Standard Error</i> (SE), MSE pada OLS dan PRM pada $n=40$ dan 10% pencilan dengan 1000 kali pengulangan.....	27
9. Nilai VIF pada $n=40$ dan 20% pencilan dengan metode PRM .....	30
10. Rata-rata $\beta$ , <i>Standard Error</i> (SE), MSE pada OLS dan PRM pada $n=40$ dan 20% pencilan dengan 1000 kali pengulangan.....	30
11. Korelasi antar variabel bebas dengan $n = 60$ .....	31
12. VIF (Variance Inflation Factor) dengan $n=60$ .....	32
13. Nilai VIF pada $n=60$ dan 5% pencilan dengan metode PRM .....	33
14. Rata-rata $\beta$ , <i>Standard Error</i> (SE), MSE pada OLS dan PRM pada $n=60$ dan 5% pencilan dengan 1000 kali pengulangan.....	34
15. Nilai VIF pada $n=60$ dan 10% pencilan dengan metode PRM.....	36
16. Rata-rata $\beta$ , <i>Standard Error</i> (SE), MSE pada OLS dan PRM pada $n=60$ dan 10% pencilan dengan 1000 kali pengulangan.....	36

17.	Nilai VIF pada $n=60$ dan 20% pencilan dengan metode PRM.....	39
18.	Rata-rata $\beta$ , <i>Standard Error</i> (SE), MSE pada OLS dan PRM pada $n=60$ dan 20% pencilan dengan 1000 kali pengulangan .....	39
19.	Korelasi antar variabel bebas dengan $n = 100$ .....	41
20.	VIF (Variance Inflation Factor) dengan $n=100$ .....	41
21.	Nilai VIF pada $n=100$ dan 10% pencilan dengan metode PRM.....	43
22.	Rata-rata $\beta$ , <i>Standard Error</i> (SE), MSE pada OLS dan PRM pada $n=100$ dan 10% pencilan dengan 1000 kali pengulangan.....	43
23.	Nilai VIF pada $n=100$ dan 10% pencilan dengan metode PRM .....	45
24.	Rata-rata $\beta$ , <i>Standard Error</i> (SE), MSE pada OLS dan PRM pada $n=100$ dan 10% pencilan dengan 1000 kali pengulangan.....	46
25.	Nilai VIF pada $n=100$ dan 10% pencilan dengan metode PRM .....	48
26.	Rata-rata $\beta$ , <i>Standard Error</i> (SE), MSE pada OLS dan PRM pada $n=100$ dan 20% pencilan dengan 1000 kali pengulangan.....	48
27.	Perbandingan MSE OLS dan PRM.....	50
28.	Korelasi antarvariabel Bebas pada data Indeks Pembangunan Manusia (IPM) Provinsi Jawa Tengah tahun 2020.....	53
29.	Nilai VIF data IPM Provinsi Jawa Tengah tahun 2020.....	53
30.	Nilai VIF PRM data IPM Provinsi Jawa Tengah tahun 2020 dengan metode PRM .....	55
31.	Nilai $\hat{\beta}$ dan $SE(\hat{\beta})$ pada OLS dan PRM dari data Indeks Pembangunan Manusia (IPM) Provinsi Jawa Tengah tahun 2020.....	56
32.	Nilai uji- $t$ metode OLS dan metode PRM.....	57



## DAFTAR GAMBAR

Gambar	Halaman
1. <i>Boxplot</i> pada data dengan $n=40$ dan pencilan 5%.....	23
2. Perbandingan <i>Standard Error</i> (SE) penduga OLS dan PRM pada $n = 40$ dan 5% pencilan.....	25
3. <i>Boxplot</i> pada data dengan $n=40$ dan pencilan 10%.....	26
4. Perbandingan <i>Standard Error</i> (SE) penduga OLS dan PRM pada $n = 40$ dan 10% pencilan.....	28
5. <i>Boxplot</i> data simulasi $n= 30$ dan pencilan 20%.....	29
6. Perbandingan <i>Standard Error</i> (SE) penduga OLS dan PRM pada $n= 30$ dan 20% pencilan.....	30
7. <i>Boxplot</i> pada Data dengan $n=60$ dan pencilan 5%.....	33
8. Perbandingan <i>Standard Error</i> (SE) penduga OLS dan PRM pada $n = 60$ dan 5% pencilan.....	34
9. <i>Boxplot</i> data simulasi $n= 60$ dan pencilan 10%.....	35
10. Perbandingan <i>Standard Error</i> (SE) penduga OLS dan PRM pada $n=60$ dan 10% pencilan.....	37
11. <i>Boxplot</i> data simulasi $n= 60$ dan pencilan 20%.....	38
12. Perbandingan <i>Standard Error</i> (SE) penduga OLS dan PRM pada $n=60$ dan 20% pencilan.....	40
13. <i>Boxplot</i> pada Data dengan $n=100$ dan pencilan 5%.....	42
14. Perbandingan <i>Standard Error</i> (SE) penduga OLS dan PRM pada $n = 100$ dan 5% pencilan.....	44

15. <i>Boxplot</i> data simulasi $n= 100$ dan pencilan 10%.....	45
16. Perbandingan <i>Standard Error</i> (SE) penduga OLS dan PRM pada $n=100$ dan 10% pencilan.....	46
17. <i>Boxplot</i> data simulasi $n= 100$ dan pencilan 20%.....	47
18. Perbandingan <i>Standard Error</i> (SE) penduga OLS dan PRM pada $n = 100$ dan 20% pencilan.....	49
19. Perbandingan MSE OLS dan PRM.....	50
20. <i>Boxplot</i> pada data Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Tengah tahun 2020.....	54
21. Perbandingan $SE(\hat{\beta})$ OLS dan $SE(\hat{\beta})$ PRM dari data Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Tengah tahun 2020.....	56

## I. PENDAHULUAN

### 1.1 Latar Belakang dan Masalah

Analisis regresi merupakan studi yang membahas hubungan fungsional antara variabel bebas dan variabel terikat. Analisis regresi linier terdiri dari dua jenis, yaitu analisis regresi linier sederhana dan analisis regresi linear berganda (Neter, *et al.*, 1997). Analisis regresi linear sederhana yaitu analisis regresi linear yang hanya melibatkan satu variabel bebas, sedangkan analisis regresi linear berganda analisis regresi linear dengan variabel terikat dipengaruhi oleh lebih dari satu variabel bebas (Myers, 1990).

Metode *Ordinary Least Square* (OLS) adalah metode regresi yang biasa digunakan untuk analisis regresi linear. Menurut Gujarati (2004), metode OLS memiliki asumsi-asumsi tertentu. Asumsi tersebut yaitu parameter bersifat linear, nilai rata-rata eror sebesar nol, homoskedasitas, tidak terjadi autokorelasi, tidak terjadi multikolinearitas pada variabel bebas dan *error* berdistribusi normal. Apabila salah satu asumsi tidak terpenuhi maka hasil analisis kurang akurat (Montgomery, *et al.*, 2012).

Analisis regresi linear berganda yang mempunyai banyak variabel sering terjadi masalah seperti multikolinearitas. Multikolinearitas disebabkan oleh analisis regresi linear berganda yang memiliki derajat korelasi yang tinggi antara dua atau lebih variabel bebas. Munculnya multikolinearitas mengakibatkan estimasi parameter tidak stabil (Myers, 1990). Untuk mengatasi masalah multikolinearitas dapat diatasi dengan salah satu metode yaitu *Partial Least Square* (PLS).

Selain masalah multikolinearitas sering terjadi juga masalah pencilan. Pencilan adalah data yang menyimpang dari kelompok mayoritas datanya, bila digambarkan secara grafik data tersebut terletak di luar data mayoritas. Adanya pencilan dapat menyebabkan residual yang besar dari model yang diperoleh, ragam menjadi lebih besar dan taksiran interval memiliki rentang yang lebar. Menurut Chen (2002), regresi *robust* adalah metode yang penting untuk menganalisis data yang mengandung pencilan. Dalam regresi *robust* terdapat beberapa metode penduga parameternya antara lain, Penduga-R, Penduga-L, Penduga-M, Penduga-S, Penduga-LMS.

Pada satu set data, masalah regresi bisa timbul sendiri-sendiri atau bahkan muncul secara bersamaan. Jika masalah regresi muncul secara bersamaan dimana pada data terjadi multikolinearitas dan terdapat pencilan yang dapat mengakibatkan hasil semakin tidak akurat. Untuk bisa menganalisis data yang memiliki masalah multikolinearitas dan pencilan tidak bisa hanya menggunakan metode PLS dan regresi *Robust* saja. Oleh karena itu, diperlukan suatu metode yang dapat menangani adanya masalah multikolinearitas dan pencilan. Salah satu metode yang bisa mengatasi masalah multikolinearitas dan pencilan adalah metode *Partial Robust-M* (PRM). Metode *Partial Robust-M* diperkenalkan oleh Serneels, *et al.* (2005) merupakan pendekatan PLS *Robust* berdasarkan penduga-M dengan pembobotan kembali secara iteratif dan memberikan bobot terkecil pada pengamatan berpengaruh. Metode ini menggunakan algoritma SIMPLS dan dapat digunakan jika terjadi respon univariat. Metode PRM sangat efisien dalam hal biaya komputasi dan sifat statistik (Liebmann, *et al.*, 2010).

Berdasarkan hal-hal tersebut metode *Partial Robust-M* (PRM) akan diteliti menggunakan data simulasi menggunakan model regresi linear berganda yang mengandung masalah multikolinearitas dan data pencilan yang akan dibandingkan dengan metode OLS berdasarkan nilai MSE dari masing-masing metode dan menerapkan metode terbaik pada data *real*.

## 1.2 Tujuan Penelitian

Adapun tujuan penelitian ini yaitu :

1. Mengetahui performa metode *Partial Robust-M* (PRM) dalam menangani masalah multikolinearitas dan pencilan pada analisis regresi linear berganda.
2. Menentukan metode terbaik dengan membandingkan nilai *standard error* penduga dan MSE pada metode *Partial Robust-M* (PRM) dan metode *Ordinary Least Square* (OLS).

## 1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah menambah pengetahuan serta diharapkan dapat menjadi masukan bagi peneliti, mahasiswa dan para pembaca tentang metode *Partial Robust-M* dalam mengatasi multikolinearitas dan pencilan.

## II. TINJAUAN PUSTAKA

### 2.1 Analisis Regresi

Analisis regresi merupakan suatu metode yang digunakan untuk menentukan hubungan fungsional suatu variabel yang disebut variabel terikat dengan satu atau lebih variabel bebas (Gujarati & Porter, 2011). Variabel terikat yaitu variabel yang keberadaannya dipengaruhi oleh variabel lain dan dinotasikan dengan Y. Variabel bebas yaitu variabel yang keberadaannya tidak dipengaruhi oleh variabel lain, dinotasikan dengan X. Dalam analisis regresi suatu persamaan regresi hendak ditentukan dan digunakan untuk menggambar pola atau bentuk fungsi hubungan yang terdapat antar variabel (Montgomery, *et al.*, 2012).

Konsep regresi linear sederhana dapat digambarkan dalam fungsi sebagai berikut :

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (2.1)$$

dengan:

$\beta_0$  = Intersep,

$\beta_1$  = koefisien garis regresi atau *slope*,

Y = variabel terikat,

X = variabel bebas,

$\varepsilon$  = galat yang bersifat acak dan berdistribusi normal  $N(0, \sigma^2)$

Menurut Gujarati (2004), ada tiga tujuan dari model regresi yaitu yang pertama menjelaskan pola hubungan sebab akibat antara variabel terikat dan variabel bebas, tujuan kedua untuk mengetahui kontribusi tiap variabel bebas dalam menjelaskan variabel terikat, dan tujuan ketiga adalah menduga nilai variabel terikat untuk beberapa nilai variabel bebas tertentu.

## 2.2 Analisis Regresi Linear Berganda

Menurut Montgomery & Runger (2011), analisis regresi linear berganda merupakan analisis hubungan secara linear antara dua atau lebih variabel bebas dengan satu variabel terikat. Persamaan model regresi linear berganda dapat ditunjukkan sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i \quad (2.2)$$

dengan:

$$i = 1, 2, 3, \dots, n$$

$Y_i$  = variabel terikat pengamatan ke-i,

$X_{i1}, X_{i2}, \dots, X_{ip}$  = variabel bebas pengamatan ke-i,

$\beta_0$  = konstanta atau intersep

$\beta_0, \beta_1, \dots, \beta_p$  = koefisien regresi

$\varepsilon_i$  = sisaan (galat) pengamatan ke-i

Persamaan regresi linear berganda dapat dinyatakan dalam notasi matriks

$$Y = X\beta + \varepsilon \text{ dengan } \varepsilon \sim N(0, \sigma^2) \quad (2.3)$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & \dots & x_{p1} \\ 1 & x_{21} & \dots & x_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{2n} & \dots & x_{pn} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

dengan:

$Y$  = vektor  $n \times 1$  variabel terikat

$X$  = matriks  $n \times (p + 1)$  variabel bebas

$\beta$  = vektor  $(p + 1) \times 1$  koefisien variabel bebas

$\varepsilon$  = vektor  $n \times 1$  variabel acak galat

## 2.3 Asumsi Analisis Regresi

Menurut Gujarati (2004), asumsi analisis regresi terdiri dari:

1. Model regresi linear dalam parameter.
2. Galat berdistribusi normal dengan rata-rata nol dan variansi  $\sigma^2$ ,  $\varepsilon_i \sim N(0, \sigma^2)$
3. Variansi galat adalah konstan atau homoskedastik yang berarti memiliki variansi yang sama.

$$\text{Var}(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2, i = 1, 2, 3, \dots, n$$

4. Tidak terdapat autokorelasi antar galat

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$$

5. Tidak terjadi multikolinearitas, artinya tidak ada hubungan linear yang kuat antara dua variabel bebas.

## 2.4 Ordinary Least Square (OLS)

Menurut Montgomery, *et al.* (2012), OLS merupakan suatu metode yang digunakan untuk mengestimasi  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$  dengan cara meminimalkan jumlah kuadrat galat. OLS adalah metode yang paling banyak digunakan dan dibandingkan dengan metode-metode lainnya dalam pembentukan model regresi atau mengestimasi parameter regresi. Untuk mengestimasi koefisien regresi  $\beta_0, \beta_1, \dots, \beta_p$  pada  $n$  data suatu penelitian adalah:

$$j = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} - \dots - \beta_p X_{ip})^2 \quad (2.4)$$

Dan  $j$  harus bernilai minimum. Jika  $j$  berubah untuk diturunkan terhadap  $b_0, b_1, \dots, b_p$ . Kemudian menyamakan dengan nol akan diperoleh

$$\frac{\partial j}{\partial \beta_0} = -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \dots - \beta_p X_{ip}) = 0$$

$$\frac{\partial j}{\partial \beta_1} = -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \dots - \beta_p X_{ip}) X_{i1} = 0$$



$$\frac{\partial j}{\partial \beta_2} = -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \dots - \beta_p X_{ip}) X_{i2} = 0$$

$$\vdots$$

$$\frac{\partial j}{\partial \beta_p} = -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \dots - \beta_p X_{ip}) X_{ip} = 0$$

Selanjutnya didapatkan turunan dari j secara parsial sebagai berikut:

$$n\beta_0 + \beta_1 \sum_{i=1}^n X_{i1} + \beta_2 \sum_{i=1}^n X_{i2} + \dots + \beta_p \sum_{i=1}^n X_{ip} = \sum_{i=1}^n Y_i$$

$$\beta_0 \sum_{i=1}^n X_{i1} + \beta_1 \sum_{i=1}^n X_{i1}^2 + \beta_2 \sum_{i=1}^n X_{i1} X_{i2} + \dots + \beta_p \sum_{i=1}^n X_{i1} X_{ip} = \sum_{i=1}^n Y_i X_{i1}$$

$$\beta_0 \sum_{i=1}^n X_{i2} + \beta_1 \sum_{i=1}^n X_{i1} X_{i2} + \beta_2 \sum_{i=1}^n X_{i2}^2 + \dots + \beta_p \sum_{i=1}^n X_{i2} X_{ip} = \sum_{i=1}^n Y_i X_{i2}$$

$$\vdots$$

$$\beta_0 \sum_{i=1}^n X_{ip} + \beta_1 \sum_{i=1}^n X_{i1} X_{ip} + \beta_2 \sum_{i=1}^n X_{i2} X_{ip} + \dots + \beta_p \sum_{i=1}^n X_{ip}^2 = \sum_{i=1}^n Y_i X_{ip}$$

Apabila dinyatakan dalam bentuk matriks, persamaan di atas menjadi

$$X^T X \beta = X^T Y \quad (2.5)$$

Dengan demikian  $\beta$  sebagai penduga dapat diperoleh melalui rumus:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (2.6)$$

Sifat-sifat penduga metode kuadrat terkecil adalah sebagai berikut:

1.  $\hat{\beta}$  linear

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

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

2.  $\hat{\beta}$  tak bias

$\hat{\beta}$  adalah penduga tak bias jika  $E(\hat{\beta}) = \beta$

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

karena  $E(\varepsilon) = 0$  dan  $(X^T X)^{-1} X^T X = I$ , sehingga didapat  $E(\hat{\beta}) = \beta$ . Maka dapat disimpulkan  $\hat{\beta}$  merupakan penduga tak bias dari  $\beta$ .

3.  $\hat{\beta}$  memiliki variansi minimum.

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

dimana  $(X^T X)^{-1} \sigma^2$  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 variansi minimum bersifat *Best Linear Unbiased Estimator* (BLUE).

## 2.5 Multikolinearitas

Multikolinearitas adalah terdapatnya hubungan linear antara beberapa atau semua variabel bebas dalam suatu model regresi linear berganda (Gujarati, 2004).

Masalah multikolinearitas hanya ditemukan pada regresi linear berganda.

Dampak adanya multikolinieritas dalam model regresi linier berganda adalah (Gujarati, 2004):

1. Penduga OLS masih bersifat BLUE, tetapi mempunyai varians dan kovarians yang besar sehingga sulit untuk mendapatkan taksiran yang tepat.
2. Akibat penduga OLS mempunyai varians dan kovarians yang besar, menyebabkan interval pendugaan akan cenderung lebih lebar dan nilai hitung statistik uji t akan kecil, sehingga membuat variabel bebas statistik tidak signifikan mempengaruhi variabel tidak bebas.
3. Walaupun variabel individu bebas tidak berpengaruh terhadap variabel terikat melalui uji t, tetapi nilai koefisien determinasi ( $R^2$ ) masih bisa relatif tinggi.

Salah satu cara menguji gejala multikolinearitas dalam model regresi adalah dengan melihat VIF (*Variance Inflation Factor*) dan TOL (*Tolerance*) dari masing-masing variabel bebas terhadap variabel terikatnya. Uji untuk mengetahui gejala multikolinearitas dengan melihat nilai VIF dan TOL dilakukan dengan langkah-langkah sebagai berikut:

1. Menghitung korelasi antar variabel bebas
2. Meregresikan variabel bebas selain  $X_1$  terhadap  $X_1$ .
3. Menghitung koefisien determinasi dari regresi variabel bebas selain  $X_1$  terhadap  $X_1$  dan diperoleh  $R_j^2$ .
4. Menghitung nilai TOL dengan rumus  $TOL = (1 - R_j^2)$
5. Menghitung nilai VIF dengan rumus  $VIF = \frac{1}{TOL}$

Apabila nilai *Variance Inflation Factor* (VIF)  $> 10$  maka mengindikasikan terdapat multikolinearitas (Myers, 1990).

## 2.6 Pencilan

Menurut Rousseeuw & Leroy (1987), pencilan merupakan data yang tidak mengikuti pola umum pada model regresi yang dihasilkan atau tidak mengikuti pola data secara keseluruhan. Pencilan juga dapat diartikan sebagai keganjilan pada data amatan yang menunjukkan ketidaksesuaian dengan sisa data (Sembiring, 2003). Keberadaan data pencilan akan mengganggu dalam proses analisis data. Menurut Draper & Smith (1998), adanya pencilan akan memberikan nilai penduga parameter yang bersifat bias sehingga berakibat pada interpretasi hasil yang diperoleh menjadi tidak valid. Dalam analisis regresi, pencilan dapat menyebabkan hal-hal berikut (Soemartini, 2007):

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

Data pencilan harus dilihat terhadap posisi dan sebaran data yang lainnya sehingga akan dievaluasi apakah data pencilan tersebut perlu dihilangkan atau tidak. Oleh karena itu digunakan pendeteksian pencilan, salah satunya dengan menggunakan *boxplot*. Metode *boxplot* sangat terkenal dalam mendeteksi pencilan. Metode ini menggunakan nilai kuartil. Kuartil 1, 2, dan 3 akan membagi sebuah urutan data menjadi empat bagian. Jangkauan (*Interquartile (IQR)*) didefinisikan sebagai selisih antara kuartil 1 dan kuartil 3 atau  $IQR = Q3 - Q1$ . Data yang mengalami masalah pencilan apabila nilai yang kurang dari  $1.5Q \cdot IQR$  terhadap kuartil 3 sehingga memenuhi kriteria sebagai data yang mengandung pencilan.

## 2.7 *Partial Least Square (PLS)*

Metode *Partial Least Square (PLS)* pertama kali diperkenalkan oleh Herman O. A Wold pada tahun 1960-an sebagai metode alternatif untuk mengatasi keterbatasan metode kuadrat terkecil ketika data mengalami masalah multikolinearitas. Metode

PLS adalah teknik statistik *multivariate* yang memungkinkan perbandingan antara beberapa variabel bebas dan beberapa variabel terikat. PLS merupakan *soft* model yang dapat menjelaskan struktur keragaman data. Model yang dihasilkan oleh metode PLS dapat mengoptimalkan hubungan antar dua kelompok variabel. Pendugaan model hubungan  $Y$  dengan  $X$  dan penduga nilai  $Y$  tertentu menggunakan suatu algoritma. Proses penentuan model dilakukan secara iterasi dengan melibatkan keragaman pada variabel  $X$  dan  $Y$ . Struktur ragam dalam  $Y$  mempengaruhi perhitungan komponen kombinasi linear dalam  $X$  dan sebaliknya, struktur ragam dalam  $X$  berpengaruh terhadap kombinasi linear dalam  $Y$  (Bilfarsah, 2005).

Matriks dan vektor yang digunakan dalam PLS dinotasikan dengan  $X$  dan  $y$  yang diasumsikan sebagai matriks data terpusat (*mean centered*) yaitu  $\tilde{X}$  dan  $\tilde{y}$ . Struktur model untuk metode PLS dalam dua persamaan, yaitu (Bastien, *et al.*, 2004)

$$\tilde{X} = \mathbf{TP}' + \mathbf{E} \quad (2.7)$$

dengan:

$\tilde{X}$  = matriks *mean centered* dari  $k$  variabel bebas yang berukuran  $n \times p$

$\mathbf{T}$  = matriks komponen berukuran  $n \times p$ ,

$\mathbf{P}'$  = matriks loading berukuran  $p \times p$ ,

$\mathbf{E}$  = matriks galat berukuran  $n \times p$ ,

Komponen-komponen yang terbentuk sebagai prediktor baru sebagai untuk menduga parameter yang kemudian digunakan untuk memprediksi peubah respon  $Y$ .

$$\tilde{y} = \mathbf{TA} + \mathbf{F} \quad (2.8)$$

dengan:

$\tilde{y}$  = vektor *mean centered* dari  $n$  observasi variabel terikat  $Y$  berukuran  $n \times 1$

$\mathbf{A}$  = vektor parameter berukuran  $p \times 1$ ,

$\mathbf{F}$  = vektor galat berukuran  $n \times 1$ ,

Salah satu algoritma metode PLS adalah menggunakan SIMPLS. Algoritma SIMPLS ini dikemukakan oleh De Jong pada tahun 1993. Metode SIMPLS mengasumsikan peubah X dan Y dihubungkan dalam model bilinear seperti berikut ini:

$$x_i = \bar{x} + Pt_i + g_i \quad (2.9)$$

$$y_i = \bar{y} + At_i + f_i \quad (2.10)$$

dengan:

$\bar{x}$  dan  $\bar{y}$  = rata-rata peubah X dan Y

$t_i$  = skor berdimensi k, dengan  $k < p$  dan  $i=1, \dots, n$

P = matriks loading X berdimensi dengan  $g_i$  dan  $f_i$

A = matriks berkoefisien regresi y berdimensi  $q \times k$

$g_i$  dan  $f_i$  = sisaan atau galat dari model

## 2.8 Regresi *Robust*

Regresi *robust* pertama kali dikenalkan oleh Andrews (1978) sebagai model regresi yang digunakan apabila distribusi dari galat tidak normal atau mengandung pencilan yang berpengaruh terhadap model regresi. Menurut Chen (2002), regresi *robust* merupakan alat yang penting untuk menganalisis data yang mengandung pencilan. Tujuan utama regresi *robust* adalah untuk memberikan hasil yang stabil karena adanya pencilan. Menurut Chen (2002), metode penduga dalam regresi *robust* sebagai berikut: Penduga-M, Penduga-S, Penduga LTS, Penduga-MM.

## 2.9 Regresi *Robust-M*

Model regresi berganda yang melibatkan k variabel bebas adalah

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.11)$$

dengan:

$\mathbf{y}$  = vektor pada variabel terikat yang berukuran  $(n \times 1)$ ,

$\mathbf{X}$  = matriks dari  $k$  variabel bebas yang berukuran  $(n \times p)$ ,

$\boldsymbol{\beta}$  = adalah vektor parameter regresi yang berukuran  $(p \times 1)$ ,

$\boldsymbol{\varepsilon}$  = adalah vektor *error* yang berukuran  $(n \times 1)$ .

Salah satu metode yang digunakan untuk mengestimasi parameter regresi dalam analisis regresi berganda adalah metode OLS. Konsep dari OLS adalah mengestimasi parameter regresi dengan meminimumkan kuadrat galat

$$\hat{\beta}_{OLS} = \min \sum_{i=1}^n (y_i - x_i \beta)^2 \quad (2.12)$$

Ketika model regresi memiliki galat yang tidak berdistribusi normal maka estimasi dengan OLS menjadi tidak akurat sehingga jumlah kuadrat galat dari OLS diganti dengan sebuah fungsi objektif  $\rho$  dari metode estimasi yang bersifat *robust* yaitu Penduga-M, hal ini karena metode *robust* Penduga-M merupakan metode yang paling sederhana dan metode yang meminimumkan fungsi objektif dari galat  $\rho(\varepsilon_i)$  (Montgomery, *et al.*, 2012). Adapun persamaan untuk metode *robust*-M adalah :

$$\hat{\beta}_M = \min \sum_{i=1}^n \rho(y_i - x_i \beta) \quad (2.13)$$

Dimana  $\rho$  yang memberikan kontribusi galat pada fungsi objektif. Dengan mengambil  $\rho(u) = u^2$  untuk  $u$  sembarang fungsi sehingga kriteria meminimumkan akan sama dengan persamaan (2.12). Untuk mengurangi pengaruh galat yang besar dipilih fungsi  $\rho$  tertentu sehingga menghasilkan estimasi *robust* dari kuadrat terkecil. Misalkan  $r_i = y_i - x_i \beta$  merupakan galat dari model dan  $w_y$  adalah bobot dari galat yang didefinisikan sebagai berikut:

$$w_y = \frac{\rho(y_i - x_i \beta)}{(y_i - x_i \beta)^2} = \frac{\rho(r_i)}{(r_i)^2} \quad (2.14)$$

Persamaan (2.12) dapat ditulis kembali

$$\hat{\beta}_M = \min \sum_{i=1}^n w_y (y_i - x_i \beta)^2 \quad (2.15)$$

Nilai  $w_y$  akan berubah pada tiap iterasinya sehingga diperoleh  $\hat{\beta}_M$ . Estimasi-M merupakan metode untuk menganalisa data yang dipengaruhi oleh pencilan sehingga dihasilkan model *robust* atau resisten terhadap pencilan. Suatu estimasi

yang resisten adalah *relative* tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan besar pada bagian besar data. Oleh karena itu agar resisten terhadap tipe pencilan lain yaitu pengamatan berpengaruh (*leverage point*) maka bobot pada persamaan (2.14) akan dikalikan dengan bobot pengamatan berpengaruh  $w_x$  (Serneels, *et al.*, 2005) yaitu:

$$\hat{\beta}_M = \min \sum_{i=1}^n w_y w_x (y_i - x_i \beta)^2 \quad (2.16)$$

Pengamatan berpengaruh dapat didiagnosa berdasarkan nilai *leverage*, yaitu pencilan ditinjau dari nilai-nilai variabel bebas. Semakin besar nilai *leverage*-nya, pengamatan tersebut semakin berpotensi berpengaruh dalam estimasi parameter regresi.

### 2.10 Partial Robust-M (PRM)

Penerapan *Partial Robust-M* bermula pada penulisan algoritma kuadrat terkecil parsial pembobotan berulang. Jika terdapat masalah multikolinearitas pada analisis regresi berganda, maka model yang sesuai adalah model Regresi PLS, (Serneels, *et al.*, 2005). Idenya adalah meregresikan variabel bebas pada jumlah peubah laten  $k$  terbatas. Nilai variabel laten  $T$  yang mempunyai vektor  $t_i$  sebagai kolom dengan  $1 \leq i \leq n$ . Model regresi laten diberikan sebagai berikut:

$$y_i = \mathbf{A}t_i + \varepsilon_i \quad (2.17)$$

Karena dimensi  $A$  rendah, yakni  $k$ , vektor  $A$  dapat diestimasi seperti semula dengan meregresikan variabel bebas terhadap variabel laten menggunakan regresi *robust-M*. Perbedaan utamanya terletak pada bobot  $w_r$  dihitung dari sisaan, yaitu  $r_i = \tilde{y}_i - t_i A$  dan bobot  $w_x$  untuk pengamatan berpengaruh akan dihitung dari skor  $t_i$  sebagai ganti dari variabel bebas asli. Pembobot yang dibutuhkan agar resisten terhadap pencilan dan pengamatan berpengaruh adalah:

$$w_i = w_r \cdot w_x \quad (2.18)$$



dan menghasilkan penduga yang disebut dengan *Partial Robust* penduga-M (PRM). Selanjutnya adalah untuk memperoleh matriks  $t_{n,k}$  yang tidak dapat diamati secara langsung. Matriks  $T_{n,k}$  diperoleh dengan menggunakan algoritma SIMPLS. Setelah A dihasilkan, maka akhir pendugaan bagi  $\beta$  adalah  $\hat{\beta} = R\hat{A}$ . PLS tampak sebagai kasus khusus jika semua  $w_i$  yang diambil sama, sehingga menghasilkan penduga tak kekar. Dengan menganggap bahwa bobot telah diterapkan, maka PLS tidak akan sulit untuk mendapatkan A yang merupakan penduga PLS yang dihitung dari amatan berbobot  $(\sqrt{w_i}x_i, \sqrt{w_i}y_i)$ .

Algoritma yang digunakan oleh Serneels, *et al.* (2005) untuk menyelesaikan regresi PRM menggunakan nilai awal yang kuat dan membuat bobot tidak hanya bergantung pada residual tetapi juga pada skor.

Disediakan fungsi bobot yang disebut fungsi “Fair”

$$f(z, c) = \frac{1}{1 - \frac{|z|}{c}} \quad (2.19)$$

Dengan  $c = 4$  adalah konstanta tuning, pembobot  $w_r$  telah dihitung sebagai

$$w_r = f\left(\frac{r_i}{\hat{\sigma}}, c\right) \quad (2.20)$$

Dimana  $r_i = y_i - \text{median}_j y_j$  adalah residual dan  $\hat{\sigma} = \text{MAD}(r) = \text{median}|r_i - \text{median}_j r_j|$  menjadi MAD dari residual vektor r. Bobot  $w_x$  setelah dilakukan penskalaan setiap vektor skor  $t_i$  dihitung sebagai berikut :

$$w_x = f\left(\frac{\|t_i - \text{med}_{L_1}(T)\|}{\text{median}_i \|t_i - \text{med}_{L_1}(T)\|}, c\right) \quad (2.21)$$

Dimana  $\|\bullet\|$  merupakan norm vektor dan pembobot  $f$ ,  $\text{med}_{L_1}(T)$  merupakan median- $L_1$  yang dihitung dari vektor skor  $\{t_1, \dots, t_n\}$ , yaitu penduga *robust* dari pusat data awal dari vektor skor berdimensi h. Median- $L_1$  merupakan median contoh berganda, juga disebut spasial median.

Algoritma PRM terdiri dari langkah-langkah sebagai berikut:

1. Hitung nilai awal *robust* yang untuk bobot  $w_i = w_r \cdot w_x$
2. Lakukan PLS Klasik (SIMPLS) pada data berbobot  $w_i x_i$  dan  $w_i y_i$

3. Perbarui bobot,  $w_i$  dengan menghitung ulang  $w_r$  dari residu PLS dan  $w_x$  dari skor PLS.
4. Iterasi langkah 2 dan 3 sampai konvergensi koefisien regresi yang diperkirakan. Konvergensi dicapai jika perbedaan norma relatif antara koefisien regresi yang diestimasi lebih kecil dari nilai ambang tertentu.
5. Estimasi regresi diperoleh dari PLS berbobot terakhir.

### 2.11 Mean Square Error (MSE)

Menurut Hanke & Wichers (2005), *Mean Square Error* (MSE) merupakan salah satu pengukuran kesalahan yang populer dan mudah digunakan. Nilai MSE dihitung dengan mengkuadratkan selisih antara ramalan dengan nilai aktual. Umumnya, semakin kecil MSE maka semakin akurat nilai suatu peramalan atau suatu pemodelan. Misalkan  $\hat{\beta}$  adalah penaksiran parameter  $\beta$ . MSE dari  $\hat{\beta}$  yaitu didefinisikan oleh:

$$MSE(\hat{\beta}) = \frac{1}{m} \sum_{j=1}^m (\hat{\beta}_j - \beta)^2 \quad (2.23)$$

dengan:

$\hat{\beta}$  = penduga parameter

$\beta$  = parameter regresi

$m$  = banyaknya ulangan

### III. METODOLOGI PENELITIAN

#### 3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil 2020/2021 di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

#### 3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data simulasi dan data *real*. Untuk data simulasi, diperoleh dengan membangkitkan data variabel bebas  $p = 6$  dan banyak data ( $n = 40, 60, \text{ dan } 100$ ) yang diulangi sebanyak 1000 kali. Data tersebut telah didesain sedemikian rupa sehingga memenuhi, asumsi-asumsi klasik yang dibutuhkan dalam analisis regresi *Partial Robust-M* yaitu memenuhi asumsi multikolinearitas menggunakan simulasi Monte Carlo dan menambahkan galat yang mengandung pencilan sebanyak 5%, 10%, dan 20% dari banyaknya data pada variabel terikat.

Dalam hal ini peneliti menggunakan *software* Rstudio 1.3.1093. untuk melakukan simulasi dari data tersebut. Untuk mendapatkan data multikolinearitas pada setiap himpunan data dibangkitkan menggunakan simulasi Monte Carlo dengan persamaan sebagai berikut:

$$x_{ij} = (1 - \rho^2)^{\frac{1}{2}}Z_{ij} + \rho z_{ij}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p$$

dimana  $z_{i,j}$  adalah pembangkit bilangan acak *independent* normal standar dan  $\rho$  adalah nilai tetapan sehingga korelasi antara dua variabel penjelas diberikan oleh  $\rho^2$ , sedangkan  $n$  pengamatan untuk variabel terikat telah ditetapkan melalui persamaan berikut:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + e_i, \quad i = 1, 2, 3, \dots, n$$

Dengan menetapkan nilai  $\beta_0 = 0$  dan  $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 1$  Untuk mendapatkan data yang mengandung pencilan yaitu dengan membangkitkan galat yang berdistribusi  $N(0,1)$  yang telah dikontaminasi dengan pencilan  $N(30,5)$ .

Besarnya persentase pencilan pada setiap  $n$  data yaitu sebesar 5%, 10% dan 20%

$$e_i = G_h + P_l$$

$$G_h = N(0,1)$$

$$P_l = N(30,5)$$

dengan:

$e_i$  = data bangkitan *error* yang telah dikontaminasikan pencilan

$G_h$  = data bangkitan  $N(0,1)$  berukuran  $1 \times h$

$P_l$  = data bangkitan  $N(30,5)$  berukuran  $1 \times l$

Tabel 1. Simulasi Monte Carlo

Jumlah		Simulasi	Keterangan
Variabel (p)	Pengamatan (n)		
6	40	$x_{i1} = (1 - \rho_1^2)^{\frac{1}{2}} Z_{i1} + \rho_1 Z_{i7}$	$i = 1, 2, \dots, n$ $j = 1, 2, \dots, p$ $\rho_{1,2} = 0.99$ $\rho_{3,4,5,6} = 0.70$
	60	$x_{i2} = (1 - \rho_2^2)^{\frac{1}{2}} Z_{i2} + \rho_2 Z_{i7}$	
	100	$x_{i6} = (1 - \rho_6^2)^{\frac{1}{2}} Z_{i6} + \rho_6 Z_{i7}$	

Tabel 2. Simulasi Pencilan

Jumlah		Simulasi	Keterangan
Variabel (p)	Pencilan (o)		
6	5%	$y_i = x_{i1} + x_{i2} + \dots + x_{i6} + e_i$ $e_i = G_h + P_l$ $G_h = N(0,1)$ $P_l = N(30,5)$	$O = \text{floor}(o \times n)$ $O = \text{jumlah pencilan}$ $i = 1, 2, \dots, n$ $h = 1, 2, \dots, n - O$ $l = 1, 2, \dots, O$
	10%		
	20%		

Untuk data *real*, yakni menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik Provinsi Jawa Tengah. Data sekunder yang digunakan adalah data Indeks Pembangunan Manusia Provinsi Jawa Tengah pada tahun 2020 yang dipengaruhi dengan enam variabel bebas yaitu harapan lama sekolah ( $X_1$ ), rata-rata lama sekolah ( $X_2$ ), indeks kemiskinan ( $X_3$ ), jumlah rumah sakit ( $X_4$ ), jumlah dokter tiap kabupaten/kota ( $X_5$ ), kepadatan penduduk ( $X_6$ ).

### 3.3 Metodologi Penelitian

Penelitian ini dilakukan secara studi pustaka, yaitu dengan mempelajari buku-buku penunjang dan karya ilmiah yang disajikan dalam bentuk jurnal. Untuk mempermudah perhitungan dan hasil akurat penelitian ini menggunakan *software* R.Studio 1.3.1093. Adapun langkah-langkah yang dilakukan *untuk Partial Robust-M* sebagai berikut:

## 1. Data Simulasi

### a. Membangkitkan data yang mengandung multikolinearitas dengan simulasi

monte carlo dengan persamaan  $x_{ij} = (1 - \rho^2)^{\frac{1}{2}}Z_{ij} + \rho z_{ij}$ , dimana  $i=1,2,\dots,n$ ,  $j=1,2,\dots,p$  dengan banyaknya variabel bebas (p) 6 dan banyaknya pengamatan (n=40, 60, 100) dengan  $\rho_{1,2} = 0.99$  dan  $\rho_{3,4,5,6} = 0.70$ .

### b. Membangkitkan data pencilan dengan mengontaminasi galat yang berdistribusi normal $N(0,1)$ dengan $N(30,5)$ dengan persentase pencilan 5%, 10%, dan 20%.

### c. Misalkan $\beta_0 = 0$ dan $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 1$ bangkitkan nilai Y, dengan regresi $Y = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + \varepsilon$

### d. Melakukan uji multikolinearitas dengan melihat VIF pada data simulasi

### e. Melakukan pendeteksian adanya pencilan pada data simulasi dengan *boxplot*.

### f. Menghitung $\hat{\beta}_{OLS}$ dengan menggunakan metode *Ordinary Least Square* (OLS).

### g. Menghitung $\hat{\beta}_{PRM}$ dengan menggunakan regresi *Partial Robust-M* (PRM) dengan langkah-langkah sebagai berikut:

1. Menentukan data terpusat peubah  $\tilde{X}_{n,p}$  dan  $\tilde{Y}_{n,q}$
2. Menentukan bobot *robust* awal yaitu  $w_i = w_r * w_x$  dengan menggunakan penduga-M *Fair*.
3. Lakukan analisis regresi dengan algoritma SIMPLS pada matriks data  $\tilde{X}_{n,p}$  dan  $\tilde{Y}_{n,q}$  terboboti yang diperoleh dengan mengalikan setiap kolom dan terboboti yang diperoleh dengan mengalikan setiap kolom dengan  $\tilde{X}_{n,p}$  dan  $\tilde{Y}_{n,q}$  dengan  $\sqrt{w_i}$
4. Periksa skor SIMPLS ( $t_a$ ) dengan membagi ( $t_a$ ) dengan  $\sqrt{w_i}$
5. Regresikan skor SIMPLS dengan peubah respon (y) :  $y_i = At_i + \varepsilon_i$
6. Menghitung  $\hat{A}_{k,q} = (S_t)^{-1}S_{ty}$
7. Uji kekonvergenan  $\hat{A}_{k,q}$
8. Jika kekonvergenan tidak tercapai, hitung kembali bobot  $w_i = w_r * w_x$  dengan menggunakan penduga-M *Fair*, dimana  $r_i = y_i - At_i$
9. Kembali ke langkah 3 sampai  $\hat{A}$  konvergen. Konvergensi tercapai apabila perbedaan relatif (*relative difference*) dalam norm antara dua berturut-turut perkiraan  $\hat{A}$  lebih kecil daripada ambang batas yang ditentukan yaitu  $10^{-5}$

10. Pendugaan akhir dari  $\hat{A}$  secara langsung diperoleh dari langkah PLS terboboti terakhir.
  11. Menghitung koefisien regresi SIMPLS terhadap peubah asli (penduga parameter untuk regresi linier  $y_i = \beta_0 + qB'_p x_i + e_i$ )
    - h. Membandingkan nilai  $\hat{\beta}_{PRM}$  dengan  $\hat{\beta}_{OLS}$
    - i. Membandingkan nilai rata-rata *standard error*  $\hat{\beta}_{PRM}$  dengan  $\hat{\beta}_{OLS}$
    - j. Menghitung nilai MSE  $\hat{\beta}_{PRM}$  dan  $\hat{\beta}_{OLS}$  setelah 1000 pengulangan (m)
    - k. Membandingkan nilai MSE  $\hat{\beta}_{PRM}$  dan MSE  $\hat{\beta}_{OLS}$ .
2. Data Real atau Data Indeks Pembangunan Manusia Provinsi Jawa Tengah 2020
- a. Input data Indeks Pembangunan Manusia (IPM) Provinsi Jawa Tengah 2020.
  - b. Melakukan estimasi parameter regresi dengan metode *Ordinary Least Square* (OLS).
  - c. Melakukan uji asumsi klasik pada model regresi yang diperoleh.
  - d. Melakukan pendeteksian adanya pencilan dengan *boxplot*.
  - e. Melakukan estimasi parameter regresi dengan menggunakan metode *Partial Robust-M* (PRM).
  - f. Membandingkan hasil estimasi OLS dan PRM dengan melihat nilai *standard error*.
  - g. Melakukan uji signifikansi.
  - h. Menentukan variabel yang mempengaruhi Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Tengah pada tahun 2020 dengan metode terbaik.

## V. KESIMPULAN

Berdasarkan hasil penelitian, maka diperoleh kesimpulan:

1. Berdasarkan nilai MSE yang dihasilkan, model regresi yang diperoleh dengan metode *Partial Robust-M* (PRM) memiliki MSE yang lebih baik dibandingkan dengan metode *Ordinary Least Square* (OLS).
2. Semakin banyak jumlah data yang digunakan maka semakin kecil nilai MSE yang dihasilkan dengan metode *Partial Robusti-M* (PRM).
3. Pada data *real* yaitu data Indeks Pembangunan Manusia (IPM) Provinsi Jawa Tengah pada tahun 2020 yang mengalami masalah multikolinieritas dan pencilan ketika diolah dengan metode OLS menghasilkan model regresi linier berganda dimana hanya variabel rata-rata sekolah yang signifikan berpengaruh terhadap variabel terikat atau variabel Indeks Pembangunan Manusia (IPM). Menggunakan metode *Partial Robust-M* (PRM) terdapat empat variabel yang secara signifikan mempengaruhi variabel Indeks Pertumbuhan Manusia (IPM) yaitu variabel harapan lama sekolah, rata-rata lama sekolah, jumlah rumah sakit, jumlah dokter tiap kabupaten/kota.



## DAFTAR PUSTAKA

- Andrews, D. F. 1974. A Robust Method for Multiple Linear Regression. *Technometrics*. **16** (4): 523-531.
- Bastien, P., Vinzi, V., and Tenenhaus, M. 2004, Partial Least Square Generalized Linier Regression. *Computational Statistics & Data Analysis*. **48**: 17-46.
- Bilfarsah, A. 2005. Efektifitas Metode Aditif Spline Kuadrat Terkecil Parsial Dalam Pendugaan Model Regresi. *MAKARA of Science Series*. **9** (1): 28–33.
- Chen, C. 2002. *Robust Regression and Outlier Detection with the ROBUSTREG Procedure*. SAS Institute Inc., Cary NC.
- De Jong, S. 1993. SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*. **18**: 251–263.
- Draper, N. R. and Smith, H. 1998. *Applied Regression Analysis*. 3<sup>rd</sup> Edition. John Wiley & Sons, Inc., New York.
- Gujarati, D. 2004. *Basic Econometrics*. McGraw-Hill Companies, Inc., New York.
- Gujarati, D. dan Porter, D. C. 2011. *Dasar-Dasar Ekonometrika*. Ed. ke-5. Salemba Empat, Jakarta.
- Hanke, J.E. and Wichers, D.W. 2005. *Business Forecasting*. 8<sup>th</sup> Edition. Pearson Prentice Hall, New Jersey.
- Liebmann, B., Filzmoser, P., and Varmuza, K. 2010. Robust and classical PLS regression compared. *Journal of Chemometrics*. **24**: 111–120.

Montgomery, D. C., Peck, E. A., and Vining, G. 2012. *Introduction to Linear Regression Analysis*. 5<sup>th</sup> Edition. John Wiley & Sons, Inc., Canada.

Montgomery, D. C. and Runger, G. C. 2011. *Applied Statistics and Probability for Engineers*. John Wiley & Sons, Inc., Canada.

Myers, R. H. 1990. *Classical and Modern Regression with Applications*. PWS-KENT Pub. Co, Boston.

Neter, J., Wasserman, W., dan Kutner, M. H. 1997. *Model Linier Terapan*. Diterjemahkan oleh Ir. Bambang Sumantri. IPB, Bandung.

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

Sembiring, R. K. 2003. *Analisis Regresi*. Ed. Kedua. Institut Teknologi Bandung, Bandung.

Serneels, S., Christoper, C., and Filzmoser, P. 2005. Partial Robust M-Regression. *Chemometrics and Intelligent Laboratory Systems*. **79**: 55–64.

Soemartini. 2007. *Pencilan (Outlier)*. Universitas Padjajaran, Bandung.