PENGGUNAAN REGRESI *ROBUST* PADA DATA YANG MENGANDUNG PENCILAN DENGAN METODE MOMEN

Skripsi

Oleh:

NURMIATI NURDIN

H 121 09 279

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2013

“Penggunaan Regresi *Robust* pada Data yang Mengandung Pencilan dengan Metode Momen”

S K R I P S I

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin Makassar

Oleh:

NURMIATI NURDIN

H 121 09 279

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2013

P E R N Y A T A A N

Saya yang bertanda tangan di bawah ini menyatakan dengan sesungguh-sungguhnya bahwa skripsi yang saya buat dengan judul:

"Penggunaan Regresi *Robust* pada Data yang Mengandung Pencilan dengan Metode Momen"

adalah benar hasil kerja saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 26 April 2013

NURMIATI NURDIN

NIM : H 121 09 279

"Penggunaan Regresi *Robust* pada Data yang Mengandung Pencilan dengan Metode Momen"

Disetujui Oleh :

Pembimbing Utama	Pembimbing Pertama
Anna Islamiyati, S.Si., M.Si.	**Drs. Raupong, M.Si.**
NIP. 19770808 200501 2 002	**NIP. 19621015 198810 1 001**

Pada tanggal : 26 April 2013

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

Pada hari ini, Jum'at tanggal 26 April 2013, Panitia Ujian Skripsi menerima dengan baik skripsi yang berjudul :

"Penggunaan Regresi *Robust* pada Data yang Mengandung Pencilan dengan Metode Momen"

yang diajukan untuk memenuhi salah satu syarat guna memperoleh gelar Sarjana Sains pada Program Studi Statistika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Makassar, 26 April 2013

PANITIA UJIAN SKRIPSI

			Tanda Tangan
1. Ketua	:	**Prof. Dr. Syamsuddin Toaha, M.Sc.**	(..................................)
2. Sekretaris	:	**Andi Kresna Jaya, S.Si., M.Si.**	(..................................)
3. Anggota	:	**Andi Galsan Mahie, S.Si., M.Si.**	(..................................)
4. Anggota	:	**Anna Islamiyati, S.Si., M.Si.**	(..................................)
5. Anggota	:	**Drs. Raupong, M.Si.**	(..................................)

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Alhamdulillahirabbil'alamin penulis panjatkan atas ke hadirat ***Allah SWT*** atas limpahan rahmat, hidayah, nikmat dan ridho-Nya sehingga penulis dapat menyelesaikan tugas akhir dengan judul **"Penggunaan Regresi *Robust* pada Data yang Mengandung Pencilan dengan Metode Momen"** dengan segala kekurangan dan kelebihan. Salam dan sholawat penulis juga hanturkan kepada Baginda Rasulullah ***Muhammad SAW*** sebagai satu-satunya suri tauladan dalam menjalankan kehidupan dunia dan akhirat.

Penyusunan skripsi ini tentunya tidak lepas dengan bantuan berbagai pihak baik moriil maupun materiil. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang tulus serta penghargaan yang tak terhingga kepada Ayahanda **H. Nurdin** dan Ibunda tercinta **Hj. Bungatia** yang telah mendidik penulis dengan penuh kesabaran dengan cinta, kasih sayang dan penuh ketulusan hati. Serta kepada Kakanda dan Adinda ku tercinta **Bripka Ibrahim, Brigpol Amri, Khaeruddin, S.T.** dan **Nurhijrah Nurdin** yang selalu menjadi sahabat terbaik dan memberi motivasi yang tiada hentinya.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada :

1. Bapak **Prof. Dr. dr. Ir. H. Idrus Andi Patturusi, Sp.B., Sp.BO.** selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.

2. Bapak **Prof. Dr. H. Abd. Wahid Wahab, M.Sc.** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin dan **para staf Fakultas MIPA Universitas Hasanuddin** yang telah membekali ilmu dan kemudahan-kemudahan kepada penulis.
3. Ibu **Dr. Hasmawati, M.Si.** selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin dan **staf Jurusan Matematika** (*Pak* ***Nasir*** dan *Pak* ***Sutamin, S.Sos.***) yang telah membekali ilmu dan kemudahan-kemudahan kepada penulis.
4. Bapak **Andi Galsan Mahie, S.Si., M.Si.** selaku Penasehat Akademik penulis sekaligus sebagai Anggota Tim Penguji dalam penulisan tugas akhir ini. Terima kasih atas atas segala masukan positif dan motivasi yang diberikan selama penulis menjalani pendidikan.
5. Ibu **Anna Islamiyati, S.Si., M.Si** dan Bapak **Drs. Raupong, M.Si.** selaku dosen pembimbing. Terima kasih telah meluangkan waktunya dengan penuh kesabaran memberikan bimbingan, arahan dan saran kepada penulis dalam penyusunan tugas akhir ini. Smoga Allah senantiasa membalas segala kebaikan beliau dengan limpahan nikmat-Nya.
6. Bapak **Prof. Dr. Syamsuddin Toaha, M.Sc.** selaku Ketua Tim Penguji. Terima kasih atas segala koreksi dan saran yang diberikan dalam penyusunan tugas akhir ini.
7. Bapak **Andi Kresna Jaya, S.Si., M.Si.** selaku Sekretaris Tim Penguji. Terima kasih telah memberikan kritikan membangun dalam penyempurnaan penyusunan tugas akhir ini serta segala bentuk *support* yang telah diberikan.

8. Seluruh **Dosen Jurusan Matematika** Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin. Terima kasih atas ilmu yang telah diberikan kepada penulis selama menjalani pendidikan.
9. Teman-teman terbaik **EKSTRIM'09** yaitu **Statistika 2009** (Uni, kiki, Ayu, Yanti, Mimi, Fitri, Yuni, Anda, Evi, Isna, Ida, Tenri, Vinni, Iva, Risma, Niki, Ira, Hesty, Whay, Yuli, Jumi, Jejen, Try, Naser, Endy, Iman, Fahrun, Mirsam, Juned, Firman, Irzan, Ivin) dan **Matematika 2009** (Dedel, Iche, Rina, Cia, Nur, Icha, Fifik, Mery, Devita, Arni, Erika, Apri, Nida, Inggrid, Lesdi, Edi, Taufik, Sadno, Faisal, Jamal, Ali, Fairus, Iksan). Terima kasih atas kebersamaan dan persaudaraan terindah yang terjalin selama perkuliahan ini. Ada banyak kenangan yang berkesan yang akan sulit untuk dilupakan.
10. Seluruh warga **HIMATIKA** tanpa terkecuali, yang telah mengajarkan arti kebersamaan dalam berorganisasi. ***BRAVO HIMATIKA.***
11. Seluruh pemain **DRUM CORP PRAMUKA UNHAS** tanpa terkecuali terutama teman-teman seperjuangan dalam **GPMB 2012.** Terima kasih atas kebersamaan termanis dan pengalaman yang telah diberikan. ***One Band One Sound.***
12. *My Bestfriend* Amma, Fika, Dian, Ayat, Yatto, Uga dan Ajju selaku teman seperjuangan dari Parepare. Terima kasih atas persahabatan yang telah terjalin dan motivasi yang diberikan kepada penulis selama ini. Selamat berjuang meraih mimpi kawan.
13. Teman-teman **KKN Gel. 82 Kab. Wajo, Kec. Sajoanging, Desa Sakkoli** (Ami, Martin, Komang, kak Nining, kak Eko, kak Roy, kak Diaz, Mas

Roman, dan Apri). Terima kasih atas pengalaman termanis yang telah diberikan.dan segala bentuk *support* yang telah diberikan.

14. Semua pihak yang tak sempat disebutkan satu persatu atas segala bentuk bantuan dan perhatiannya hingga terselesaikannya tugas akhir ini.

Penulis menyadari bahwa tugas akhir ini masih jauh dari kesempurnaan sehingga kritik dan saran yang membangun akan penulis terima dengan tangan terbuka demi perbaikan lebih lanjut.

Akhir kata, Semoga tugas akhir ini dapat bermanfaat dan menambah ilmu pengetahuan bagi semua pihak yang membacanya.

Amin Yaa Rabbal'alamin.

Makassar, April 2013

Penulis

PENGGUNAAN REGRESI *ROBUST* PADA DATA YANG MENGANDUNG PENCILAN DENGAN METODE MOMEN

ABSTRAK

Analisis regresi merupakan sebuah alat statistika yang memberikan tentang pola hubungan antara dua variabel atau lebih. Salah satu metode yang umumnya digunakan dalam mengestimasi parameter pada analisis regresi linear adalah metode kuadrat terkecil (OLS). Namun metode ini mempunyai kelemahan apabila data terdeteksi mengandung *outlier*. Maka regresi *robust* disarankan dapat mengatasi masalah *outlier* dalam data untuk mengestimasi parameter, salah satunya adalah *Metode Momen* (MM) yang digunakan untuk data yang terdeteksi *outlier* pada variabel bebas dan variabel terikat serta memiliki nilai *breakdown point* yang tinggi.

Dalam skripsi ini dikaji tentang penggunaan Metode Momen dengan metode iterasi *Iteratively Reweighted Least Square* (IRLS). Metode Momen merupakan gabungan antara estimasi S dan estimasi M. Pada Metode Momen ini digunakan fungsi pembobot *Tukey Bisquare.*

Kata Kunci : Analisis Regresi, OLS, *Outlier*, *Breakdown Point*, Regresi *Robust*, Metode Momen, *Tukey Bisquare.*

USING ROBUST REGRESSION ON DATA CONTAINING OUTLIERS BY THE METHOD OF MOMENTS

ABSTRACT

Regression analysis is a statistical tool that provides about the relationship between two or more variables. One of the methods commonly used in estimating the parameters of the linear regression analysis is a Ordinary Least Squares (OLS). But this method has a weakness if the data contains outliers detected. Then the robust regression suggested to solve the problem of outliers in the data to estimate the parameters, one of which is the Method of Moments (MM) used for data detected outlier on the independent variable and the dependent variable and also has a high value of the breakdown point.

In this thesis examined the use of method of moments with the Iteratively Reweighted Least Square (IRLS). The method of moment is a combination of S-estimates and M-estimates. At the Method of Moment is used Tukey Bisquare weighting function.

Keywords : *Regression Analysis, OLS, outlier, Breakdown Point, Robust Regression, Method of Moments, Tukey Bisquare.*

DAFTAR ISI

HALAMAN JUDUL i

LEMBAR KEOTENTIKAN iii

LEMBAR PERSETUJUAN PEMBIMBING iv

LEMBAR PENGESAHAN PENGUJI v

KATA PENGANTAR vi

ABSTRAK x

ABSTRACT xi

DAFTAR ISI xii

DAFTAR GAMBAR xv

DAFTAR TABEL xvi

DAFTAR LAMPIRAN.. xvii

BAB I PENDAHULUAN 1

1.1 Latar Belakang 1

1.2 Rumusan Masalah 3

1.3 Batasan Masalah 4

1.4 Tujuan Penelitian 4

1.5 Manfaat Penelitian 4

BAB II TINJAUAN PUSTAKA 5

2.1 Regresi 5

2.2 Model Regresi Linier Sederhana 5

2.3 Model Regresi Linier Berganda 6

2.4 Pencilan (*Outlier*) 6

2.4.1 Defenisi Pencilan 6

2.4.2 Dampak Pencilan... 7

2.4.3 Tipe Pencilan 7

2.4.4 Identifikasi Pencilan.. 8

2.5 Metode Kuadrat Terkecil 11

2.6 Regresi *Robust* 13

2.6.1 *Breakdown Point* 15

2.6.2 Fungsi Obyektif... 15

2.6.3 Estimasi S 16

2.6.4 Estimasi M 17

2.6.5 Estimasi MM 18

2.7 Uji Signifikansi Parameter 19

BAB III METODOLOGI PENELITIAN 22

3.1 Sumber Data 22

3.2 Identifikasi Variabel 22

3.3 Metode Analisis 23

3.3 Diagram Alur Kerja 24

BAB IV HASIL DAN PEMBAHASAN 25

4.1 Estimasi Parameter Regresi *Robust* Menggunakan

Metode Momen 25

4.2 Pengolahan Data 27

4.2.1 Identifikasi Pencilan 27

4.2.2 Regresi *Robust* Estimasi MM 28

4.2.3 Koefesien Determinasi (R^2) 30

4.2.2 Uji Signifikansi Parameter 31

BAB V PENUTUP 35

5.1 Kesimpulan 35

5.2 Saran 36

DAFTAR PUSTAKA.. 37

LAMPIRAN.. 38

DAFTAR GAMBAR

Gambar 1 Diagram Alur Kerja 24

DAFTAR TABEL

Tabel 1 Hasil Iterasi Memperoleh Nilai Koefesien Parameter *Robust* 30

Tabel 2 Nilai T_{hitung} Model Regresi *Robust* 32

DAFTAR LAMPIRAN

Lampiran 1 Data IPK dan Nilai UAN 39

Lampiran 2 Nilai *DfFITS* dan *Leverage Value* (h_{ii}) 40

Lampiran 3 Output Program SAS 9.1 Estimasi-S 41

Lampiran 4 Hasil Perhitungan Nilai Estimasi, Nilai Residual dan Nilai Pembobot untuk Iterasi Pertama 42

Lampiran 5 Hasil Perhitungan Nilai Estimasi, Nilai Residual dan Nilai Pembobot untuk Iterasi Kedua 43

Lampiran 6 Hasil Perhitungan Nilai Estimasi, Nilai Residual dan Nilai Pembobot untuk Iterasi Ketiga 44

Lampiran 7 Hasil Perhitungan Nilai Estimasi, Nilai Residual dan Nilai Pembobot untuk Iterasi Keempat 45

Lampiran 8 Hasil Perhitungan Nilai Estimasi, Nilai Residual dan Nilai Pembobot untuk Iterasi Kelima 46

Lampiran 9 Hasil Perhitungan Nilai Estimasi, Nilai Residual dan Nilai Pembobot untuk Iterasi Keenam 47

Lampiran 10 Hasil Perhitungan Nilai Estimasi, Nilai Residual dan Nilai Pembobot untuk Iterasi Ketujuh 48

Lampiran 11 Hasil Perhitungan Nilai Estimasi, Nilai Residual dan Nilai Pembobot untuk Iterasi Kedelapan 49

Lampiran 12 Hasil Perhitungan Nilai Estimasi, Nilai Residual dan Nilai Pembobot untuk Iterasi Kesembilan 50

Lampiran 13 Output Regresi *Robust* Estimasi MM 51

Lampiran 14 Output Hasil Estimasi Menggunakan OLS 52

BAB I

PENDAHULUAN

1.1 Latar Belakang

Regresi linier adalah metode statistika yang digunakan untuk membentuk model hubungan antara variabel terikat (*dependent*; respon; *Y*) dengan satu atau lebih variabel bebas (*independent*; prediktor; *X*). Secara umum regresi linear terdiri dari dua yaitu regresi linear sederhana dimana terdapat satu variabel terikat Y dan satu variabel bebas X sedangkan regresi linear berganda dimana terdapat satu variabel terikat Y dan beberapa variabel bebas X. Regresi linear banyak digunakan dalam berbagai bidang dalam hal analisis untuk melihat pengaruh suatu kondisi atau kejadian. Dalam menaksir parameter model regresi ini maka penaksir yang umum digunakan adalah penaksir kuadrat terkecil (*ordinary least square)*. Hal ini disebabkan oleh mudahnya penghitungan penaksir ini dan sifatnya sebagai penaksir tak bias terbaik untuk parameter model regresi jika data yang digunakan memenuhi asumsi klasik (Draper & Smith,1998: 34-38).

Pelanggaran asumsi yang sering terjadi pada data, biasanya disebabkan oleh adanya data pencilan. Menurut Soemartini (2007) pencilan adalah pengamatan yang jauh dari kelompok data yang mungkin berpengaruh besar terhadap koefesien regresi. Beragam faktor yang dapat menyebabkan adanya pencilan, diantaranya kekeliruan pada sistem pengukuran (*measurement system error*), kesalahan input data (*human* *error*) atau karena terjadinya peristiwa yang luar biasa (misalnya krisis atau bencana).

Cara mengatasi pencilan diantaranya membuang data pencilan dari proses analisis dengan pertimbangan sudah bisa terwakili oleh sebagian besar data lainnya dan tidak mengurangi informasi serta beranggapan bahwa bisa saja pencilan tersebut disebabkan kekeliruan, bukan data sebenarnya. Dengan menghilangkan data pencilan diharapkan telah hilang pula penyebab pelanggaran asumsi, sehingga peneliti dapat menggunakan metode analisis standar.

Jika data pencilan merupakan data yang sangat berpengaruh dan menyimpan informasi penting dari sebuah peristiwa, maka peneliti tidak diperkenankan membuang data pencilan begitu saja. Data pencilan tersebut tetap dipertahankan dalam analisis dengan melakukan transformasi terhadap data dengan maksud agar asumsi terpenuhi. Namun seringkali transformasi yang dilakukan terhadap data tidak dapat menghilangkan atau memperkecil nilai *leverage outlier* yang akhirnya membiaskan pendugaan. Dalam keadaan seperti ini, pendekatan yang biasa digunakan adalah regresi *robust*.

Regresi *robust* merupakan alat yang penting untuk menganalisis data yang terkontaminasi oleh pencilan. Regresi *robust* digunakan untuk mendeteksi pencilan dan memberikan hasil yang resisten terhadap adanya pencilan (Chen 2000). Regresi *robust* terdiri dari 5 metode yaitu estimasi M (*M Estimation*), estimasi LMS (L*east Median of Square*), estimasi LTS

(*Least Trimmed Square*), estimasi S (*Scale Estimation*) dan estimasi MM (*Method of Moment*).

Penelitian tentang regresi *robust* dengan metode LTS telah dilakukan oleh Nuraidah (2012), yang dimana metode LTS digunakan hanya ketika variabel bebasnya terdapat pencilan.

Dalam penelitian ini penulis membahas menggunakan metode estimasi-MM karena metode ini mempunyai kelebihan yaitu dapat digunakan untuk data yang terdeteksi pencilan pada variabel bebas dan variabel terikat .

Estimasi MM (*Method of Moment*), dikenalkan oleh Yohai (1987). Metode ini menggabungkan estimasi S (estimasi dengan *high breakdown point*) dan estimasi M.

Berdasarkan uraian dan penelitian sebelumnya tersebut maka penulis tertarik untuk mengangkat judul **"Penggunaan Regresi *Robust* Pada Data yang Mengandung Pencilan dengan Metode Momen"**

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, dapat dirumuskan masalah sebagai berikut :

1. Bagaimana mengestimasi parameter regresi *robust* menggunakan estimasi Metode Momen?
2. Bagaimana model regresi pada data indeks prestasi kumulatif pada mahasiswa yang mengandung pencilan dengan regresi *robust* melalui estimasi Metode Momen?

1.3 Batasan Masalah

Dalam penelitian ini dibatasi pada penggunaan estimasi Metode Momen dalam penerapannya pada data yang mengandung pencilan dengan menggunakan pembobot *Tukey Bisquare*.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini berdasarkan permasalahan yang telah dirumuskan sebelumnya adalah :

1. Untuk mengestimasi parameter regresi *robust* menggunakan estimasi Metode Momen.
2. Untuk mendapatkan model regresi pada data indeks prestasi kumulatif mahasiswa yang mengandung pencilan dengan regresi *robust* melalui estimasi Metode Momen.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat akademisi maupun praktisi bagi pengguna ilmu statistik sebagai gambaran dan alternatif pertimbangan dalam menganalisis model regresi yang di dalamnya terdapat data pencilan.

BAB II

TINJAUAN PUSTAKA

2.1 Regresi

Analisis regresi merupakan sebuah alat statistika yang memberikan penjelasan tentang pola hubungan antara dua variabel atau lebih. Dalam analisis regresi, dikenal dua jenis variabel yaitu :

- Variabel terikat disebut juga variabel *dependent* yaitu variabel yang keberadaannya diperngaruhi oleh variabel lainnya dan dinotasikan dengan Y.
- Variabel bebas disebut juga variabel *independent* yaitu variabel yang tidak dipengaruhi oleh variabel lainnya dan dinotasikan dengan X.

2.2 Model Regresi Linier Sederhana

Bentuk hubungan yang paling sederhana antara variabel X dengan variabel Y adalah bentuk garis lurus atau berbentuk hubungan linier yang disebut dengan regresi linier sederhana atau sering disebut regresi linier saja dengan persamaan matematikanya sebagaimana diungkapkan (Walpole, Ronald E, dkk. 1995) adalah sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad ; \quad i = 1,2,\dots,n \tag{2.1}$$

di mana :

Y_i : variabel terikat pada pengamatan ke-i

X_i : variabel bebas pada pengamatan ke-*i*

β_0 : *intercept*

β_1 : koefisien regresi

ε_i : galat (*error*)

2.3 Model Regresi Linier Berganda

Hubungan fungsional atau hubungan kausal antara dua atau lebih variable yang dinyatakan dalam suatu bentuk fungsi linier pada umumnya dapat dinyatakan dalam bentuk persamaan matematika yang dibahas dalam analisis regresi. Untuk hubungan fungsional yang linier dapat dirumuskan dalam bentuk persamaan regresi sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i \quad ; \quad i = 1,2,\dots,n \tag{2.2}$$

di mana :

Y_i : variabel terikat pada pengamatan ke-*i*

X_{ik} : variabel bebas pada pengamatan ke-*i*

β_0 : *intercept*

$\beta_1, \dots, \beta_k$: koefisien-koefisien regresi atau koefisien kemiringan

ε_i : galat (*error*)

2.4 Pencilan (*Outlier*)

2.4.1 Defenisi Pencilan (*Outlier*)

Pencilan adalah pengamatan yang jauh dari pusat data yang mungkin berpengaruh besar terhadap koefesien regresi. Pencilan dapat muncul karena kesalahan dalam memasukkan data, kesalahan pengukuran, analisis, atau kesalahan-kesalahan lain. Pengaruh pencilan dalam analisis data dapat dibedakan berdasarkan asal pencilan tersebut yaitu yang berasal dari peubah respon (*y-outliers*; titik *influence*) atau berasal dari peubah bebasnya (*x-outliers*; titik *leverage*).

2.4.2 Dampak Pencilan (*Outlier*)

Dampak keberadaan pencilan ini dapat mengganggu dalam proses analisa data dan harus dihindari dalam banyak hal. Dalam kaitannya dengan analisis regresi, pencilan dapat menyebabkan hal – hal berikut :

1. Galat yang besar dari model yang terbentuk atau $E[e] \neq 0$.
2. Varians pada data tersebut menjadi lebih besar.
3. Taksiran interval memiliki rentang yang lebar.

2.4.3 Tipe Pencilan (*Outlier*)

Tipe-tipe dari pencilan diantaranya adalah:

1. Pencilan regresi adalah sebuah pengamatan yang menyimpang dari hubungan kelinearan ditentukan dari $(n-1)$ pengamatan yang lainnya, atau paling tidak dari mayoritas pengamatan tersebut.
2. Pencilan galat adalah sebuah pengamatan yang memiliki standarisasi galat yang besar ketika digunakan dalam perhitungan.
3. Pencilan x adalah sebuah pengamatan yang menyimpang hanya pada koordinat x atau disebut titik *leverage* baik (*good leverage points*).

4. Pencilan y adalah sebuah pengamatan yang menyimpang hanya pada koordinat y atau disebut pencilan vertical (*vertical outlier*s).
5. Pencilan x dan y adalah sebuah pengamatan yang menyimpang pada kedua koordinat atau disebut titik *leverage* jelek (*bad leverage points*).

(Puput, 2011 : 13).

2.4.4 Identifikasi Pencilan (*Outlier*)

Terdapat beberapa metode untuk mengidentifikasi adanya pencilan yang berpengaruh dalam koefisien regresi antara lain :

1. **Metode Grafis**

   Keuntungan dari metode ini yaitu mudah dipahami karena menampilkan data secara grafis (gambar) dan tanpa melibatkan perhitungan yang rumit. Sedangkan menurut Soemartini, kelemahan dari metode ini adalah keputusan bahwa suatu data merupakan pencilan sangat bergantung pada *judgement* peneliti, karena hanya mengandalkan visualisasi grafis, untuk itu dibutuhkan seseorang yang ahli dan berpengalaman dalam menginterpretasikan plot tersebut.

   a. **Scatter Plot**

      Metode ini dilakukan dengan cara memplot data dengan pengamatan ke-i $(i = 1,2,\ldots,n)$. Selain itu, jika sudah didapatkan model regresi maka dapat dilakukan dengan cara memplot antara galat (e) dengan nilai penaksir $Y(\hat{Y})$. Jika terdapat satu atau beberapa data yang terletak jauh dari pola kumpulan data keseluruhan maka hal ini mengindikasikan adanya pencilan.

b. Box Plot

Metode ini mempergunakan nilai kuartil dan jangkauan untuk mendeteksi pencilan. Dengan menggunakan nilai kuartil 1,2 dan 3 yang akan membagi sebuah urutan data menjadi beberapa bagian

$$IQR = Q3 - Q1 \tag{2.3}$$

di mana :

Q1 : Kuartil ke 1

Q2 : Kuartil ke 2

Q3 : Kuartil ke 3

IQR : Jangkauan (*Interquartile Range*)

Data-data yang merupakan pencilan yaitu nilai yang kurang dari 1,5xIQR terhadap kuartil 1 dan nilai yang lebih dari 1,5xIQR terhadap kuartil 3.

2. Metode *DfFITS* (*Difference fitted value FITS*) atau *Standardized DfFITS*

DfFITS merupakan suatu ukuran berpengaruh yang ditimbulkan oleh pengamatan ke-i terhadap nilai taksiran $\hat{y}_i$. Nilai $DfFITS_i$ diperoleh dari persamaan berikut:

$$(DfFITS)_i = \frac{\hat{y}_i - \hat{y}_{i-1}}{S_{i-1}^2 - \sqrt{h_{ii}}} \tag{2.4}$$

di mana :

$\hat{y}_i$: nilai taksiran y_i

$\hat{y}_{i-1}$: nilai taksiran y_i tanpa pengamatan ke-*i*

S_{i-1}^2 : jumlah kuadrat galat tanpa pengamatan ke-*i*

h_{ii} : elemen diagonal ke-*i* dari matriks $H = X_i^T (X^T X)^{-1} X_i$

Suatu pengamatan ke-*i* data diidentifikasikan sebagai pencilan apabila nilai :

$$|DfFITS_i| > 1 \qquad untuk\ n \le 30$$
$$|DfFITS_i| > 2\left(\frac{p}{n}\right)^{1/2} \qquad untuk\ n > 30$$

dengan *p* banyaknya parameter dan *n* banyaknya pengamatan .

3. Nilai Pengaruh (*Leverage Point*)

Metode yang digunakan dalam mengidentifikasi pencilan terhadap variabel X adalah nlai pengaruh (*Leverage Point*). Nilai pengaruh (h_{ii}) dari penamatan (X_i, Y_i) menunjukan besarnya peranan Y_i terhadap $\hat{Y}_i$ dan didefinisikan sebagai:

$$h_{ii} = X_i^T (X^T X)^{-1} X_i \ \ ; \ i: 1,2,\ldots,n \tag{2.5}$$

dengan $X_i = \begin{bmatrix} X_{i1} & X_{i2} & \cdots & X_{ik} \end{bmatrix}$ adalah vektor baris yang berisi nilai-nilai dari peubah variabel bebas dalam pengamatan ke-*i*. Nilai h_{ii} berada diantara 0 dan 1 dengan rumus

$$\sum_{i=1}^{n} h_{ii} = k \tag{2.6}$$

dengan *k=p-1*. Sehingga dapat dituliskan menjadi

$$2\bar{h}_{ii} = \frac{2\sum_{i=1}^{n} h_{ii}}{n} = \frac{2k}{n} = \frac{2(p-1)}{n} \tag{2.7}$$

Suatu pengamatan ke-i data diidentifikasikan sebagai pencilan apabila nilai $h_{ii} > 2\bar{h}_{ii}$ .Sehingga pengamatan ke-i dikatakan pencilan terhadap X.

2.5 Metode Kuadrat Terkecil

Metode kuadrat terkecil (*Ordinary Least Square* = OLS) merupakan suatu metode untuk mendapatkan garis regresi yang baik yaitu sedekat mungkin dengan datanya sehingga nanti menghasilkan prediksi yang baik (Widarjono, 2005).

Dasar dari penaksiran koefisien regresi linier dalam regresi linier berganda adalah metode kuadrat terkecil yaitu meminimumkan jumlah kuadrat galat sedemikian sehingga didapat koefisien-koefisien regresi yang tak bias. Akan ditaksir koefisien-koefisien regresi dengan menggunakan metode kuadrat terkecil yaitu dengan meminimumkan jumlah kuadrat galat $\sum_{i=1}^{n} \varepsilon_i^2$.

Estimator OLS memberikan hasil cukup baik saat semua asumsi klasik dalam regresi dipenuhi, akan tetapi metode ini memiliki kelemahan yang sangat sensitif terhadap pencilan. Pengaruh pencilan dapat menyebabkan estimasi OLS mempunyai nilai variansi yang sangat besar dan mengakibatkan distribusi galat e_i tidak lagi berdistribusi normal. Dengan demikian maka pengujian statistik untuk melihat signifikansi hasil estimasi parameter regresi dan untuk pembuatan selang kepercayaan yang didasarkan pada distribusi normal tidak dapat dilakukan karena menjadi tidak dapat diandalkan lagi (Rosseeuw, 1987).

Model regresi berganda dalam bentuk matriks dapat dituliskan sebagai berikut :

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.8}$$

dengan,

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, dan \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

di mana $p = k + 1$

Dalam hal ini *y* adalah variabel bebas *X* adalah matriks konstanta, *β* adalah vektor parameter dan *ε* adalah vektor galat yang bersifat acak normal bebas dengan nilai ekspektasi, $E(\boldsymbol{\varepsilon}) = 0$ dan matriks varians kovarians,$\sigma^2(\boldsymbol{\varepsilon}) = \sigma^2 I_k$.

Untuk mendapatkan penaksir y yaitu $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ diperlukan nilai penaksir untuk parameter $\boldsymbol{\beta}$ yaitu $\hat{\boldsymbol{\beta}}$. Salah satu prosedur penaksir yang sering digunakan adalah metode kuadrat terkecil. Pada dasarnya, metode ini meminimumkan jumlah kuadrat simpangan y dari nilai ekspektasinya yaitu meminimumkan :

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{\varepsilon}_i^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\boldsymbol{y} - \boldsymbol{X\beta})^T (\boldsymbol{y} - \boldsymbol{X\beta}) \tag{2.9}$$

dan $S(\boldsymbol{\beta})$ dapat dinyatakan sebagai berikut :

$$S(\boldsymbol{\beta}) = \boldsymbol{y}^T \boldsymbol{y} - \boldsymbol{\beta}^T \boldsymbol{X}^T \boldsymbol{y} - \boldsymbol{y}^T \boldsymbol{X\beta} \quad + \boldsymbol{\beta}^T \boldsymbol{X}^T \boldsymbol{X\beta} \tag{2.10}$$

$$= \boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{y} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta}$$

Karena $\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{y}$ adalah sebuah matriks $(1\ x\ 1)$ atau sebuah skalar dan transposenya adalah $(\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{y})^T = \boldsymbol{y}^T\boldsymbol{X}\,\boldsymbol{\beta}$ merupakan skalar juga penaksir kuadrat terkecil harus memenuhi,

$$\left.\frac{\partial S}{\partial \boldsymbol{\beta}}\right|_{\widehat{\boldsymbol{\beta}}} = -2\boldsymbol{X}^T\boldsymbol{y} + 2\boldsymbol{X}^T\boldsymbol{X}\widehat{\boldsymbol{\beta}} = 0 \qquad (2.11)$$

maka penaksir kuadrat terkecil dari $\boldsymbol{\beta}$ adalah

$$\widehat{\boldsymbol{\beta}} = [\boldsymbol{X}^T\boldsymbol{X}]^{-1}\,\boldsymbol{X}^T\boldsymbol{y} \qquad (2.12)$$

2.6 Regresi *Robust*

Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi dari galat tidak normal dan atau adanya beberapa pencilan yang berpengaruh pada model (Ryan, 1997). Metode ini merupakan alat penting untuk menganalisa data yang dipengaruhi oleh pencilan sehingga dihasilkan model yang *robus*t atau *resistant* terhadap pencilan. Suatu estimasi yang *resistant* adalah relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data. Metode ini dikembangkan oleh Rousseuw dan Leroy (1987).

Metode *robust* lebih didekatkan pada parameter rata-rata dan variansikovariansi dari suatu penaksir tertentu, yaitu dengan menstandarisasikan penaksir untuk parameter rata-rata dan variansi-kovariansi sedemikian sehingga menghasilkan penaksir yang konsisten terhadap parameter-parameter tersebut. Dalam hal ini, dilakukan dengan bentuk pembatasan nilai pada penaksiran untuk parameter-parameternya. Dengan ke-*robust*-an, penaksirannya tidak akan menyimpang terlalu jauh.

Menurut Chen (2002:1) metode-metode estimasi dalam regresi robust diantaranya adalah:

1. Estimasi M (*Maximum likelihood type*) yang dikenalkan oleh Huber (1973) adalah metode yang sederhana baik dalam penghitungan maupun secara teoritis. Estimasi ini menganalisis data dengan mengasumsikan bahwa sebagian besar yang terdeteksi pencilan pada variabel independen.
2. Estimasi LTS (*Least Trimmed Squares*) adalah metode dengan *high breakdown point* yang dikenalkan oleh Rousseeuw (1984). *Breakdown point* adalah ukuran proporsi minimal dari banyaknya data yang terkontaminasi pencilan dibandingkan seluruh data pengamatan.
3. Estimasi S (*Scale*) juga merupakan metode dengan *high breakdown point* yang dikenalkan oleh Rousseeuw and Yohai (1984). Dengan nilai *breakdown* yang sama, metode ini mempunyai efisiensi yang lebih tinggi dibanding estimasi LTS.
4. Estimasi MM (*Method of Moment*), dikenalkan oleh Yohai (1987). Metode ini menggabungkan estimasi S (estimasi dengan *high breakdown point*) dan estimasi M.

2.6.1 *Breakdown Point*

Breakdown point adalah salah satu cara yang digunakan untuk mengukur ke-*robust*-an (kekekaran) suatu estimator. *Breakdown point* merupakan proporsi minimal dari banyaknya pencilan dibandingkan seluruh data pengamatan.

Dengan kata lain, *breakdown point* sebagai suatu ukuran ke-r*obuts*-an dari suatu penaksir. Semakin besar nilai persen dari *breakdown point* pada suatu penaksir, maka penaksir tersebut semakin *robust*.

Regresi *robust* yang mempunyai *breakdown point* adalah regresi *robust* dengan metode estimasi S, LTS, LMS, dan MM. Metode estimasi MM mempunyai *breakdown point* 50%. *Breakdown point* 50% adalah *breakdown point* yang tinggi.

2.6.2 Fungsi Obyektif

Fungsi obyektif adalah fungsi yang digunakan untuk mencari fungsi pembobot pada regresi *robust.* Fungsi pembobot yang digunakan yaitu fungsi pembobot *Tukey Bisquare*.

Diberikan suatu fungsi obyektif sebagai berikut:

$$\rho(e_i^*) = \begin{cases} \frac{r^2}{6}\left[1-\left[1-\left(\frac{e_i^*}{r}\right)^2\right]^3\right], & |e_i^*| < r \\ \frac{r^2}{6}, & |e_i^*| \geq r \end{cases} \tag{2.13}$$

dengan fungsi *influence* yaitu:

$$\psi(e_i^*) = \rho'(e_i^*) = \frac{\partial\left(\rho(e_i^*)\right)}{\partial e_i^*} = \begin{cases} e_i^*\left[1-\left(\frac{e_i^*}{r}\right)^2\right]^2, & |e_i^*| < r \\ 0, & |e_i^*| \geq r \end{cases} \quad (2.14)$$

Sehingga diperoleh fungsi pembobot

$$w_i = w(e_i^*) = \frac{\psi(e_i^*)}{e_i^*} = \begin{cases} \left[1-\left(\frac{e_i^*}{r}\right)^2\right]^2, & |e_i^*| < r \\ 0, & |e_i^*| \geq r \end{cases} \quad (2.15)$$

di mana:

e_i^* : galat yang distandarisasi

r : 4,685

Nilai r pada fungsi objektif, *influence* dan pembobot adalah *tunning constant.* Kelly (2006) menyatakan permasalahan dalam estimasi regresi robust adalah perlu dilakukan pemilihan *tunning constant* agar estimasi yang diperoleh lebih spesifik dan memimimumkan jumlah kuadrat galat. Menurunkan *tunning constant* akan menaikan pembobot terhadap galat yang besar. Menaikkan *tunning constant* akan menurunkan pembobot terhadap galat yang besar. Semakin besar r maka estimasi robust akan mendekati *least square.*

2.6.3 Estimasi S

Jika data terkontaminasi pencilan pada variabel *X*, estimasi M tidak dapat bekerja dengan baik. Estimasi M tidak dapat mengidentifikasi *bad observation* yang berarti tidak dapat membedakan *good leverage point* dan *bad leverage point. Good leverage* merupakan pengamatan yang berada di ruang distribusi tetapi sudah tidak berada di daerah mayoritas data sedangkan *bad leverage* merupakan pengamatan yang tidak berada baik dalam ruang distribusi pengamatan maupun daerah mayoritas. Untuk mengatasi hal tersebut, estimasi *high breakdown* sangat diperlukan (Chen, 2002:5). Salah satu estimasi yang mempunyai nilai *high breakdown* adalah estimasi S. Bentuk estimator S adalah:

$$\hat{\beta}_s = \min_{\beta} \hat{\sigma}\left(e_1, e_2, ..., e_n\right) \tag{2.16}$$

di mana :

ρ : fungsi obyektif

$e_1, e_2, ..., e_n$: nilai galat hingga pengamatan ke-*n*

$\hat{\sigma}$: estimator skala *robust* yaitu

$$\hat{\sigma} = \sqrt{\frac{n\sum_{i=1}^{n}(e_i{}^2) - (\sum_{i=1}^{n} e_i)^2}{n(n-1)}} \tag{2.17}$$

2.6.4 Estimasi M

M-*Estimation* merupakan metode regresi *robust* yang sering digunakan. M-*Estimation* dipandang dengan baik untuk mengestimasi parameter yang disebabkan oleh x-*outlier* dan memiliki *breakdown point* $1/n$. Estimator M yang meminimumkan fungsi ρ (fungsi obyektif) dari galatnya.Bentuk estimatornya yaitu:

$$\hat{\beta}_M = \min_{\beta} \sum_{i=1}^{n} \rho(e_i^*) \quad ; i = 1,2, ..., n \tag{2.18}$$

di mana :

ρ : fungsi obyektif

e_i^* : galat yang distandarisai yaitu $\frac{e_i}{\hat{\sigma}_s}$

$\hat{\sigma}$: estimator skala *robust* yang memenuhi:

$$\hat{\sigma} = \frac{MAR}{0.6745} = \frac{\frac{1}{n}\sum_i^n |y_i - \hat{y}_i|}{0.6745} \tag{2.19}$$

dengan *MAR* adalah *Median Absolute Residual.*

y_i : variabel terikat pada pengamatan ke-*i*

$\hat{y}_i$: penaksir y_i

Dalam mengestimasi parameter regresi *robust* M metode iterasi diperlukan, karena galat tidak dapat dihitung sampai diperoleh model yang cocok dan parameter regresi juga tidak dapat dihitung tanpa mengetahui nilai galat. *Iteratively Reweighted Least Squares* (IRLS) adalah metode iterasi yang banyak digunakan.

2.6.5 Estimasi MM

Estimasi MM menggabungkan estimasi *high breakdown point* dan efisiensi statistik yang dikenalkan oleh Yohai (1987). Langkah pertama dalam estimasi ini adalah mencari estimator S, kemudian menetapkan parameter-parameter regresi menggunakan estimasi M. Estimasi S menjamin nilai *breakdown point* tinggi dan estimasi M membuat estimator mempunyai efisiensi tinggi. Pada umumnya digunakan fungsi *Tukey Bisquare* baik pada estimasi S maupun estimasi M.

Bentuk dari metode estimasi MM:

$$\hat{\beta}_{MM} = \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}}\right) = \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\,\beta_j}{\hat{\sigma}}\right) \qquad (2.20)$$

di mana :

ρ : fungsi obyektif

e_i : galat untuk pengamatan ke-*i*

$\hat{\sigma}$: estimator skala *robust*

Metode MM juga menggunakan IRLS (*Iteratively Reweighted Least Square*) untuk mencari estimasi parameter regresi.

Prosedur estimasi parameter pada model regresi linier ganda dengan regresi *robust* estimasi MM:

1. Menghitung estimator awal koefisien $\hat{\beta}_j^{(1)}$ dan galat $e_i^{(1)}$ dari regresi *robust* dengan *high breakdown point* (estimasi S) dengan bobot *Tukey Bisquare* .
2. Menghitung skala estimasi $\hat{\sigma}$ dan dihitung pula pembobot awal $w_i^{(1)}$ menggunakan galat $e_i^{(1)}$ pada langkah pertama.
3. Menghitung koefisien regresi menggunakan galat $e_i^{(1)}$ dengan skala estimasi $w_i^{(1)}$ pada langkah kedua.
4. Menghitung bobot baru $w_i^{(2)}$ dengan skala estimasi dari iterasi awal.
5. Mengulang langkah 2, 3, 4 (dengan skala estimasi tetap konstan) sampai mendapatkan $\sum_{i=1}^{n}\left|e_i^{(m)}\right|$ konvergen (selisih $\hat{\beta}_j^{(m+1)}$ dan $\hat{\beta}_j^{(m)}$ mendekati 0, dengan m banyaknya iterasi).

2.7 Uji Signifikansi Parameter

Pengujian signifikansi parameter dalam model regresi bertujuan untuk mengetahui hubungan antara variabel bebas dan variabel terikat. Disamping itu juga untuk mengetahui kelayakan parameter dalam menerangkan model. Terdapat dua tahap pengujian yaitu uji simultan dan uji parsial (individu).

a. Uji Simultan (Uji F)

Uji simultan merupakan pengujian secara bersama semua parameter dalam model regresi. Hipotesis yang digunakan adalah sebagai berikut :

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$

$H_1 : Ada\, \beta_j \neq 0 \quad ; j = 1,2, \dots , k$

Statistik uji yang digunakan untuk *Weighted Least Square* (WLS) adalah :

$$F_{hitung} = \frac{\left[\sum_{i=1}^{n} w_i (\hat{y}_i - \bar{y})^2\right] / (k)}{\left[\sum_{i=1}^{n} w_i (y_i - \hat{y}_i)^2\right] / (n-k-1)} \tag{2.21}$$

di mana:

w_i : Nilai pembobot untuk pengamatan ke-*i*

$\hat{y}_i$: Penaksir y untuk pengamatan ke-*i*

$\bar{y}_i$: Rataan dari y untuk pengamatan ke-*i*

Kriteria pengambilan keputusannya adalah:

Tolak H_0 jika $F_{hitung} \geq F_{tabel}$

Terima H_0 jika $F_{hitung} < F_{tabel}$

Nilai F_{tabel} dapat dilihat menggunakan Microsoft Office Excel dengan fungsi $= FINV(\alpha; n-k-1)$.

b. Uji Parsial (Uji T)

Uji parsial merupakan pengujian secara individu parameter dalam model regresi yang bertujuan untuk mengetahui parameter model regresi telah signifikan atau tidak. Hipotesis yang digunakan adalah sebagai berikut :

$H_0: \beta_j = 0 \quad ; j = 1,2, \dots, k$

$H_1: \beta_j \neq 0$

Statistik uji yang digunakan untuk *Weighted Least Square* (WLS) adalah :

$$t_{hitung} = \frac{\hat{\beta}_j}{S_e(\hat{\beta}_j)} \tag{2.22}$$

di mana:

$\hat{\beta}_j$: Penaksir parameter model regresi

$S_e(\hat{\beta}_j)$: Standard *error* dari β_j

Kriteria pengambilan keputusannya adalah:

Tolak H_0 jika $t_{hitung} \geq t_{tabel}$

Terima H_0 jika $t_{hitung} \geq t_{tabel}$

Nilai T_{tabel} dapat dilihat menggunakan Microsoft Office Excel dengan fungsi $= TINV(\alpha; n-k-1)$.