PERBANDINGAN KEEFEKTIFAN METODE REGRESI ROBUST ESTIMASI-M DAN ESTIMASI-MM KARENA PENGARUH *OUTLIER* DALAM ANALISIS REGRESI LINEAR

(CONTOH KASUS DATA PRODUKSI PADI DI JAWA TENGAH TAHUN 2007)

skripsi

disajikan sebagai salah satu syarat
untuk memperoleh gelar Sarjana Sains
Program Studi Matematika

oleh

Hanna Ardiyanti
4150406542

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI SEMARANG

2011

PERNYATAAN

Dengan ini saya menyatakan bahwa isi skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya tidak terdapat karya yang diterbitkan oleh orang lain, kecuali yang secara tertulis dirujuk dalam skripsi ini dan disebutkan dalam daftar pustaka.

Semarang, 23 Agustus 2011

Hanna Ardiyanti
NIM 4150406542

PENGESAHAN

Skripsi yang berjudul

Perbandingan Keefektifan Metode Regresi Robust Estimasi-M dan Estimasi-MM karena Pengaruh *Outlier* dalam Analisis Regresi Linear (Contoh Kasus Data Produksi Padi di Jawa Tengah tahun 2007)

disusun oleh

Nama : Hanna Ardiyanti

NIM : 4150406542

telah dipertahankan dihadapan sidang Panitia Ujian Skripsi FMIPA UNNES pada tanggal 23 Agustus 2011.

Panitia:

Ketua

Dr. Kasmadi Imam S., M.S.
NIP. 195111151979031001

Sekertaris

Drs. Edy Soedjoko, M.Pd
NIP. 195604191987031001

Ketua Penguji

Iqbal Kharisudin, S.Pd, M.Sc
NIP. 197908052005011003

Anggota Penguji/
Pembimbing Pendamping

Prof. Dr. YL. Sukestiyarno, M.S.
NIP. 195904201984031002

Anggota Penguji/
Pembimbing Utama

Dra. Sunarmi, M.Si
NIP. 195506241988032001

MOTO DAN PERSEMBAHAN

Moto:

- Allah akan meninggikan orang-orang yang beriman di antara kamu dan orang-orang yang diberi ilmu pengetahuan beberapa derajat...? (Sunan At-Tirmidzi)
- TIdak ada yang mampu menolak takdir kecuali do'a. (Sunan At-Tirmidzi)
- Hargailah waktu dengan sebaik-baiknya, karena ketika semua telah berlalu hanya ada penyesalan yang terjadi.
- Terimalah segala resiko perjalanan hidup dengan setulus hati, dengan kelapangan hati, dan dengan rasa syukur yang besar karena semua akan indah pada waktunya.

Persembahan:

Skripsi ini kupersembahkan kepada:

1. Ayah dan Ibu tercinta, atas semua doa, kasih sayang dan motivasi sepanjang perjalanan hidupku.
2. Adik-adikku (Intan dan Vian) yang selalu kusayang.
3. Hony, yang selalu kukangenin, yang jauh disana tetapi selalu mengiringi langkahku.
4. Teman-teman Kost memberi dukungan dan motivasi.
5. Teman-teman MatPar'06 yang tak hanya memberiku kebahagiaan dan kenyamanan ketika aku belajar, tetapi juga membuka mataku betapa indahnya kebersamaan.

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kehadirat Tuhan Yang Maha Kuasa yang telah melimpahkan karunia-Nya, sehingga penulis masih diberi kekuatan untuk menyelesaikan skripsi dengan judul "Perbandingan Keefektifan Metode Regresi Robust Estimasi-M dan Estimasi-MM karena Pengaruh *Outlier* dalam Analisis Regresi Linear". Penyusunan skripsi ini sebagai syarat akhir untuk memperoleh gelar Sarjana Sains.

Penyelesaian skripsi ini tidak terlepas dari bantuan dan dukungan berbagai pihak yang sangat berguna bagi penulis. Oleh karena itu, perkenankanlah penulis mengucapkan terima kasih kepada:

1. Prof. Dr. H. Sudijono Sastroadjmojo, M. Si, Rektor Universitas Negeri Semarang.
2. Dr. Kasmadi Imam S.,M.S, Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang.
3. Drs. Edy Soedjoko, M.Pd, Ketua Jurusan matematika yang telah memberikan izin dalam penyusunan skripsi ini.
4. Dra. Sunarmi, M.Si, Pembimbing utama yang telah memberikan bimbingan dan pengarahan dalam penyusunan skripsi ini.
5. Prof. Drs. YL. Sukestiyarno, M.S.,Ph.D, Pembimbing pendamping yang telah memberikan bimbingan dan pengarahan dalam penyusunan skripsi ini.
6. Bapak dan Ibu dosen yang telah memberikan bekal ilmu yang tak ternilai harganya selama belajar di Fakultas Matematika dan Ilmu pengetahuan Alam Universitas Negeri Semarang.
7. Keluargaku tersayang yang senantiasa mendukung langkahku dengan iringan doa dan belain kasih saying.
8. Seseorang yang secara tidak langsung telah memberikan perhatian, kasih sayang dan doanya.
9. Teman-teman Matematika angkatan 2006 atas doa, bantuan, dan dukungan yang telah diberikan.

Penulis menyadari bahwa masih banyak kekurangan dalam penulisan skripsi. Oleh karena itu, kritik dan saran sangat diharapkan guna sempurnanya skripsi ini. Akhirnya, semoga skripsi ini dapat bermanfaat bagi para pembaca.

Semarang, 23 Agustus 2011

Penulis

ABSTRAK

Ardiyanti, Hanna. 2011. *Perbandingan Keefektifan Metode Regresi Robust Estimasi-M dan Estimasi-MM karena Pengaruh Outlier dalam Analisis Regresi Linear*. Skripsi, Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang. Pembimbing Utama: Dra. Sunarmi, M.Si. dan Pembimbing Pendamping: Prof. Drs. YL. Sukestiyarno, M.S.,Ph.D.

Kata kunci: *Outlier*, OLS, Estimasi-M, Estimasi-MM.

Analisis regresi linier adalah analisis terhadap hubungan satu variabel tak bebas (Y) dengan satu atau lebih variabel bebas (X). Estimasi parameter biasanya diselesaikan dengan metode kuadrat terkecil. Akan tetapi, apabila terdapat *outlier*, maka estimasi koefisien garis regresi dengan metode kuadrat terkecil menjadi tidak tepat. Hal ini mendorong penelitian ke dalam pendekatan yang lebih robust. Estimasi-M dan Estimasi-MM adalah metode-metode dalam regresi robust. Permasalahan yang dikaji dalam penelitian ini adalah metode manakah yang lebih efektif dalam mengatasi permasalahan *outlier* pada metode kuadrat terkecil. Tujuan dari penelitian ini adalah untuk mengetahui metode manakah yang lebih efektif, antara metode Estimasi-M dan metode Estimasi-MM.

Dalam penelitian ini mengambil simulasi pada suatu kasus dengan mengggunakan data dari BPS (Badan Pusat Statistik) Provinsi Jawa Tengah yaitu data tentang produksi padi tiap kabupaten atau kota di Provinsi Jawa Tengah tahun 2007, dimana variabel–variabel tersebut meliputi jumlah produksi padi, luas panen, dan jumlah penduduk. Proses analisis dimulai dengan menggunakan metode kudrat terkecil, identifikasi *outlier*, dan analisis dengan dua metode robust. Dalam menilai hasil kedua metode dengan membandingkan standar error kedua metode dengan OLS yang terdapat *outlier*. Apabila standar error yang dihasilkan metode regresi robust lebih kecil dari OLS, maka regresi robust dapat menganalisis data tanpa membuang *outlier* dan menghasilkan estimasi yang resisten terhadap *outlie*r. Sehingga dapat dikatakan regresi robust dapat mengatasi kelemahan OLS terhadap pengaruh *outlier*.

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan bahwa baik Estimasi-M maupun Estimasi-MM mempunyai keefektifan yang sama dalam mengatasi *outlier* pada OLS, karena keduanya dapat mengecilkan standar error yang dihasilkan OLS. Dilihat dari efek *breakdown point*, Estimasi-M kurang efektif daripada Estimasi-MM dalam mengatasi pengaruh *outlier* pada variabel prediktor. Berdasarkan hasil penelitian disarankan bagi peneliti yang menjumpai *outlier* dalam data observasi, tidak perlu membuang *outlier* tersebut, karena regresi robust dapat menghasilkan model regresi yang resisten terhadap *outlier*.

DAFTAR ISI

DAFTAR TABEL

Tabel

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1 Latar Belakang

Regresi merupakan salah satu teknik analisis statistika yang paling banyak digunakan. Banyak sekali teknik analisis statistika yang diturunkan atau didasarkan pada prinsip-prinsip regresi. Adanya analisis regresi sangat menguntungkan bagi banyak pihak, baik bidang sains, sosial, industri maupun bisnis.

> Istilah regresi pertama kali diperkenalkan pada tahun 1886 oleh Sir Francis Galton dalam penelitian biogenetisnya. Galton menemukan adanya tendensi bahwa orang tua yang memiliki tubuh tinggi memiliki anak-anak yang tinggi dan orang tua yang memiliki tubuh pendek memiliki anak-anak yang pendek pula. Meskipun demikian, Galton mengamati bahwa ada kecenderungan tinggi anak cenderung bergerak menuju rata-rata tinggi populasi secara keseluruhan. Dengan kata lain, ketinggian anak yang amat tinggi atau orang tua yang amat pendek cenderung bergerak kearah rata-rata tinggi populasi (Supranto, 2005:35).

Interprestasi modern mengenai regresi agak berlainan dengan regresi versi Galton. “Secara umum analisis regresi pada dasarnya adalah studi mengenai ketergantungan satu variabel dependen (terikat) dengan satu atau lebih variabel independen (bebas) dengan tujuan untuk mengestimasi atau memperkirakan rata-rata populasi atau nilai rata-rata variabel dependen berdasarkan nilai variabel independen yang diketahui” (Gujarati, 1995:16).

Hasil dari analisis regresi berupa koefisien regresi untuk masing-masing variabel independen. Koefisien ini diperoleh dengan cara memprediksi nilai variabel dependen dengan suatu persamaan. koefisien regresi dihitung dengan dua tujuan sekaligus. Pertama, meminimumkan penyimpangan antara nilai aktual dan nilai estimasi variabel dependen. Kedua, mengoptimalkan korelasi antara nilai aktual dan nilai estimasi variabel dependen berdasarkan data yang ada.

Analisis regresi merupakan suatu analisis statistika yang mengukur kekuatan hubungan dan menunjukan arah hubungan antara sekelompok variabel. Dalam analisis regresi dibedakan dua jenis variabel yaitu variabel bebas (independen) dan variabel terikat (variabel dependen). Hubungan antara variabel–variabel tersebut dapat dinyatakan dalam model matematika.

Bentuk umum model regresi linear

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon \ .$$

Keterangan : $\beta_{0,} \beta_1, \ldots, \beta_k$ = koefisien regresi,

$X_1, X_{2,} \ldots, X_k$ = variabel bebas

Y = variabel terikat

. ε = error

Salah satu tujuan dalam analisis regresi adalah mengestimasi koefisien regresi dalam model. Pada umumnya digunakan metode estimasi kuadrat terkecil atau *Ordinary Least Square* (OLS) *Method* untuk mengestimasi koefisien regresi dalam model regresi. Metode kuadrat terkecil adalah suatu metode yang digunakan untuk mengestimasi koefisien garis regresi dengan cara meminimumkan jumlah kuadrat residual. Penggunaan metode kuadrat terkecil memerlukan beberapa asumsi klasik yang harus dipenuhi. Beberapa asumsi itu antara lain:

(1) ε_i merupakan variabel random dan mengikuti distribusi normal;

(2) varians dari ε_i adalah konstan dan homokedastisitas;

(3) tidak ada autokorelasi; dan

(4) tidak ada multikolienaritas di antara variabel independen.

Jika asumsi-asumsi klasik dalam metode kuadrat terkecil terpenuhi maka penduga parameter yang diperoleh bersifat *Best Linear Unbiased Estimasi* (BLUE). Pada kenyataannya, asumsi ini tidak selalu dipenuhi sehingga penggunaan metode kuadrat terkecil perlu dihindari. Salah satu penyebab tidak terpenuhinya asumsi klasik (asumsi normalitas) adalah adanya *outlier*. *Outlier* adalah satu atau beberapa data yang terlihat jauh dari pola kumpulan data keseluruhan.

Adanya *outlier* dalam Metode Kuadrat Terkecil mengakibatkan estimasi koefisien garis regresi yang diperoleh tidak tepat. Hal ini berarti nilai estimasi parameter-parameter dalam model regresi linear dapat dipengaruhi oleh satu titik data ekstrim yang merupakan *outlier*. Pendeteksian *outlier* merupakan tahapan diagnosis yang perlu dilakukan terutama jika estimasi modelnya dengan metode kuadrat terkecil, yang dikenal cukup peka terhadap *outlier*. Metode pendeteksian pencilan dilakukan dengan beberapa metode, antara lain metode *boxplot*, *Leverage value*, *Cook's Distance*, dan *Standardized residual*.

Terdapatnya *outlier* dalam data akan mengakibatkan bentuk sebaran data tidak lagi simetrik tetapi cenderung menjulur ke arah *outlier* sehingga melanggar asumsi normalitas. Terkadang untuk mengatasi hal ini, seorang peneliti melakukan transformasi pada data dengan maksud agar asumsi terpenuhi. Namun, seringkali transformasi yang dilakukan terhadap data tidak dapat memperkecil nilai *leverage outlier* yang akhirnya membiaskan pendugaan. Dalam kasus seperti ini, analisis regresi robust merupakan metode yang paling layak digunakan.

Regresi robust diperkenalkan oleh Andrews (1972). "Regresi robust merupakan metode regresi yang digunakan ketika distribusi dari error tidak normal dan atau adanya beberapa *outlier* yang berpengaruh pada model" (Olive, 2005:3). Metode ini merupakan alat penting untuk menganalisis data yang dipengaruhi *outlier* sehingga dihasilkan model yang robust atau resisten terhadap *outlier*. Suatu estimasi yang resisten adalah estimasi yang relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data. Prosedur robust ditujukan untuk mengakomodasi adanya keanehan data, sekaligus mengidentifikasi adanya data *outlier*.

Dalam regresi robust terdapat beberapa metode estimasi, antara lain adalah Estimasi-M, *Least Median of Squares* (LMS), *Least Trimmed Squares* (LTS), Estimasi-S, dan Estimasi-MM. Kelima metode regresi robust tersebut mempunyai kelemahan dan kelebihan masing-masing. Estimasi-M mempunyai efisiensi yang tinggi, tetapi nilai *breakdown point* = 0. LMS, LTS, dan Estimasi-S mempunyai *breakdown point* yang tinggi (BDP = 0,5), akan tetapi efisiensinya sangat rendah. Estimasi-MM mempunyai efisiensi tinggi dan *breakdown point* yang tinggi pula. Pada penelitian ini, penulis hanya menggunakan metode regresi robust dengan estimasi-M dan estimasi-MM.

Pemilihan kedua metode tersebut karena estimasi-M dan estimasi-MM merupakan suatu teknik robust yang popular dan paling umum serta mudah dalam pengaplikasiannya daripada metode regresi robust yang lain. Selain itu terdapat perbedaan nilai *breakdown point* dari kedua metode tersebut sehingga mendorong penulis untuk mendalami kedua metode tersebut.

Berdasarkan uraian di atas, maka penulis mencoba mengambil judul "Perbandingan Keefektifan Metode Regresi Robust Estimasi-M dan Estimasi-MM karena Pengaruh *Outlier* dalam Analisis Regresi Linear".

1.2 Permasalahan

Berdasarkan latar belakang yang telah diuraikan di atas, maka permasalahan yang timbul adalah sebagai berikut.

(1) Bagaimana kriteria metode-metode yang dapat dipergunakan dalam mendeteksi keberadaan *outlier*?

(2) Bagaimana hasil model regresi robust dengan metode estimasi-M?

(3) Bagaimana hasil model regresi robust dengan metode estimasi-MM?

(4) Metode manakah yang lebih efektif antara estimasi-M dan estimasi-MM jika ditinjau dari efek *breakdown point* dan standar eror?

1.3 Pembatasan Masalah

Dalam skripsi ini, penulis memberikan batasan masalah sebagai berikut.

(1) Data yang digunakan adalah data yang memuat *outlier*.

(2) Model regresi yang dipakai adalah model regresi linear.

(3) Metode yang digunakan adalah regresi robust dengan estimasi-M dan estimasi-MM.

1.4 Tujuan Penelitian

Berdasarkan latar belakang di atas, maka tujuan dari penelitian ini adalah agar pembaca dapat:

(1) mengetahui kriteria metode-metode yang dapat dipergunakan dalam mengidentifikasi keberadaan *outlier*;

(2) mengetahui penggunaan regresi robust dengan estimasi-M dan estimasi-MM dalam mengatasi permasalahan *outlier* pada OLS;

(3) membandingkan hasil model metode regresi robust estimasi-M dan estimasi-MM; dan

(4) mengetahui metode yang lebih efektif dalam dalam regresi robust jika ditinjau dari efek *breakdown point* dan standar eror.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini di antaranya adalah sebagai berikut.

(1) Menambah pengetahuan Matematika bidang Statistika khususnya tentang data *outlier*.

(2) Menambah perbendaharaan hasil penelitian murni, khususnya dapat digunakan sebagai alternatif pemilihan solusi persoalan statistika.

(3) Dapat memberikan pengetahuan untuk mengidentifikasi serta menanggulangi pengaruh *outlier* dalam data yang akan dianalisis.

(4) Dapat memberikan pengetahuan tentang perbandingan estimasi-M dan estimasi-MM.

1.6 Sistematika Penulisan

Secara garis besar skripsi ini dibagi menjadi tiga bagian yaitu bagian awal skripsi, bagian isi skripsi dan bagian akhir skripsi. Berikut ini dijelaskan masing-masing bagian skripsi.

(1) Bagian awal skripsi

Bagian awal skripsi meliputi halaman judul, abstrak, halaman pengesahan, halaman motto dan persembahan, kata pengantar, daftar isi, daftar gambar, daftar tabel, daftar lampiran dan abstrak.

(2) Bagian isi skripsi

Bagian isi terdiri dari lima bab. Adapun lima bab tersebut adalah sebagai berikut.

Bab 1 Pendahuluan

Pada bab Pendahuluan ini dikemukakan tentang alasan pemilihan judul, permasalahan, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan skripsi.

Bab 2 Landasan Teori

Dalam bab ini dikemukakan konsep-konsep yang dijadikan landasan teori seperti skala data, regresi linear, metode kuadrat terkecil (OLS), pencilan, identifikasi *outlier*, regresi robust, fungsi-fungsi ukuran robust, kerangka berpikir, dan hipotesis. Teori-teori tersebut mendasari pemecahan masalah yang diajukan

Bab 3 Metode Penelitian

Pada bab ini berisi penentuan masalah, perumusan masalah, studi pustaka, analisis dan pemecahan masalah, serta penarikan simpulan.

Bab 4 Pembahasan

Bab ini berisi tentang hasil penelitian dan pembahasan, sebagai jawaban dari permasalahan.

Bab 5 Penutup

Dalam bab ini dikemukakan simpulan dari pembahasan dan saran yang berkaitan dengan simpulan.

(3) Bagian akhir skripsi

Bagian akhir skripsi meliputi daftar pustaka dan lampiran-lampiran yang mendukung.

BAB II
LANDASAN TEORI

2.1 Skala Data

Data penelitian dapat diskala atau dikategorikan ke dalam dua tipe, yaitu:

2.1.1 Skala Non-metrik

Skala data non-metrik digunakan untuk penelitian kualitatif. Menurut Sukestiyarno (2008: 3-4), tipe data yang termasuk dalam jenis ini adalah :

(1) Data Nominal

Data ini berbentuk bilangan diskrit dan merupakan hasil konversi data kualitatif. Tiap bilangan dari jenis data ini tidak mempunyai arti menurut besarnya ataupun posisinya, melainkan hanya sebagai simbolisasi data saja. Datanya dapat secara bebas disusun tanpa memperhatikan urutan, dan dapat dipertukarkan sesuai kesepakatan.

Contoh:

- Data dari variabel jenis agama: Islam=1, Kristen=2, Katolik=3, Hindu=4, Budha=5.
- Data dari variabel status diri : Single=1, Kawin=2, Cerai=3.
- Data dari variabel jenis kelamin: Pria=1, Wanita=0.

(2) Data Ordinal

Seperti data nominal, data ini juga merupakan hasil konversi dari data kualitatif. Namun bilangan dari jenis data ini menunjukkan urutan yang berbeda menurut kualitas atributnya.

Contoh:

- Data dari variabel kinerja mahasiswa: 1=sangat jelek, 2=jelek, 3=cukup, 4=bagus, 5=sangat bagus. Disini, urutan data 1 sampai dengan 5 menyimbolkan kualitas. Bilangan pengganti kualitas tersebut mempunyai suatu tingkatan atribut.

2.1.2 Skala Metrik

Skala data metrik digunakan untuk penelitian kuantitatif. Menurut Sukestiyarno (2008: 4), tipe data yang termasuk dalam jenis ini adalah :

(1) Kardinal

Data ini berbentuk diskrit dan berasal dari hasil membilang atau menghitung dari suatu variabel. Data ini berupa bilangan numerik yang bulat. Contoh: Jumlah buku yang dimiliki mahasiswa, jumlah barang dagangan tiap koperasi, jumlah tendangan pemain sepak bola.

(2) Interval

Data ini merupakan hasil dari pengukuran suatu variabel. Data interval diasumsikan berbentuk bilangan kontinu yang mempunyai urutan. Pada data jenis ini tidak mempunyai nol mutlak. Artinya, jika responden mempunyai variabel bernilai nol (0) bukan berarti tidak memiliki substansi sama sekali. Misalkan pada variabel suhu/temperatur suatu ruangan. Terdapat ruangan yang mempunyai suhu 0^0 C, disini nilai nol bukan berarti ruangan tersebut tidak mempunyai suhu sama sekali tetapi suhu 0^0 C masih bermakna mempunyai subtansi suhu, terdapat juga suhu negatif.

(3) Rasio

Sama dengan jenis data interval, data ini juga merupakan hasil dari pengukuran suatu variabel dan merupakan data berbentuk kontinu. Perbedaan jenis data rasio dengan jenis data interval adalah jenis data ini mempunyai nol mutlak, artinya jika suatu responden mempunyai variabel bernilai nol (0) berarti tidak memiliki substansi sama sekali. Misalnya, variabel massa benda, jika suatu benda massa 0 kg berarti tidak ada substansi yang diukur massanya.

2.2 Matriks

Menurut Hadley (1992: 51), ”Matriks didefinisikan sebagai susunan persegi panjang dari bilangan-bilangan yang diatur dalam baris dan kolom”. Matriks dapat ditulis sebagai berikut:

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

Susunan di atas disebut matriks m kali n (ditulis mxn), karena memiliki m baris dan n kolom. Elemen-elemen matriks berupa bilangan real maupun fungsi bilangan real.

2.2.1 Perkalian Matriks dengan Skalar

Menurut Hadley (1992: 53), Jika diberikan sebuah matriks A dan sebuah skalar λ, maka hasil perkalian λ dan A ditulis λA didefinisikan sebagai

$$\lambda A = \begin{bmatrix} \lambda a_{11} & \lambda a_{21} & \cdots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \cdots & \lambda a_{2n} \\ \vdots & \vdots & & \vdots \\ \lambda a_{m1} & \lambda a_{m2} & \cdots & \lambda a_{mn} \end{bmatrix}$$

2.2.2 Perkalian Dua Matriks

Jika diberikan matriks A_{mxn} dan matriks B_{nxm} maka hasil kali AB didefinisikan sebagai C_{mxm} yang elemen-elemennya dihitung dari elemen-elemen dari A, B. Dapat ditulis sebagai berikut:

$$C_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}, \quad i = 1, \ldots, m; \quad j = 1, \ldots, r$$

(Hadley, 1992: 57)

Perkalian matriks A dan matriks B terdefinisi jika dan hanya jika jumlah kolom matriks A sama dengan jumlah baris matriks B.

Contoh:

Diketahui matriks $A=\begin{bmatrix}1 & 2\\0 & 1\end{bmatrix}$ dan $B=\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}$.

Tentukan perkalian antara matriks A dan matriks B

Penyelesaian:

$$AB=\begin{bmatrix}1 & 2\\0 & 1\end{bmatrix}\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}$$

$$=\begin{bmatrix}1.1+2.3 & 1.2+2.4\\0.1+1.3 & 0.2+1.4\end{bmatrix}$$

$$=\begin{bmatrix}7 & 10\\3 & 4\end{bmatrix}.$$

2.2.3 Transpose Matriks

Menurut Hadley (1992: 51) ”Transpose dari matriks A adalah matriks yang dibentuk dari A dengan mempertukarkan baris-baris dan kolom-kolom sehingga baris i dari A menjadi kolom i dari matriks transpose”. Transpose dinotasikan dengan A'. Pandang A adalah matriks mxn, maka A' adalah matriks nxm.

Contoh:

$$A=\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\end{bmatrix} \text{ maka } A'=\begin{bmatrix}1 & 4\\2 & 5\\3 & 6\end{bmatrix}.$$

2.2.4 Determinan Matriks

Determinan suatu matriks A biasanya dilambangkan dengan det(A) atau $|A|$. Pandang A matriks persegi berordo nxn

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

Jika elemen pada baris ke-i dan kolom ke-j dihapus, maka determinan matriks persegi sisanya (berordo n-1) disebut minor dari a_{ij}, dan dinyatakan oleh $|M_{ij}|$. Minor bertanda, $(-1)^{ixj}|M_{ij}|$ disebut kofaktor a_{ij} dan dinyatakan oleh a_{ij}. Nilai determinan $|A|$, dengan A matriks berordo nxn adalah jumlah hasil kali yang diperoleh dari perkalian tiap elemen suatu baris (kolom) $|A|$ dengan kofaktornya, yaitu:

$$|A| = \sum_{k=1}^{n} (\pm) a_{1i} a_{2j} \ldots a_{nr}$$

Sebuah unsur diberi tanda (+) jika (i, j, k,… ,r) adalah permutasi genap dari (1, 2, … , n), dan tanda (-) jika permutasi ganjil.

(Hadley, 1992: 72).

2.2.5 Matriks Identitas

Menurut Hadley (1992: 62), “Matriks identitas ordo n, yang ditulis dengan I atau I_n adalah matriks bujur sangkar yang mempunyai angka-angka satu sepanjang diagonal utama (diagonal kiri atas menuju kanan bawah) dan nol di mana-mana”. Secara umum dapat ditulis:

$$I = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 1 \end{bmatrix}$$

2.2.6 Matriks Adjoint

Matriks adjoint biasanya dilambangkan dengan adj(A). Pandang A matriks persegi berordo nxn

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

Jika setiap elemen matriks A diganti oleh kofaktornya, maka diperoleh matriks kofaktor K sebagai berikut.

$$K = \begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{22} & \cdots & k_{2n} \\ \vdots & \vdots & & \vdots \\ k_{n1} & k_{n2} & \cdots & k_{nn} \end{bmatrix}$$

dimana kofaktor $a_{ij} = k_{ij} = (-1)^{i+j} \det(A_{ij})$. Tanda kofaktor minus (-) kalau (i+j) ganjil, dan plus (+) kalau (i+j) genap. Transpose dari matriks kofaktor disebut adjoint. Jadi adj(A) = K'.

(Supranto, 2005: 314)

2.2.7 Invers Matriks

Diberikan matriks bujur sangkar A. Jika terdapat matriks bujur sangkar A^{-1} yang memenuhi hubungan

$$A^{-1}A = AA^{-1} = I$$

maka A^{-1} disebut invers kebalikan dari A.

Invers dari matriks A dapat dinyatakan dengan:

$$A^{-1} = \frac{AdjA}{|A|}, \qquad |A| \neq 0$$

(Hadley, 1992: 89)

Contoh:

Diketahui matriks $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$

Tentukan invers dari matriks A

Penyelesaian:

$$A^{-1} = \frac{1}{|A|} \text{ Adj A}$$

$$= \frac{1}{(1x1)-(2x0)} \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$$

$$= \frac{1}{1} \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$$

2.3 Regresi Linear

2.3.1 Model Regresi Linear Sederhana

Analisis regresi adalah suatu metode yang berguna untuk menentukan hubungan suatu variabel yang disebut variabel dependen dengan satu atau lebih variabel yang menerangkan atau yang sering disebut variabel independen. Salah satu tujuan analisis regresi adalah menentukan model regresi yang baik, sehingga model dapat digunakan untuk menerangkan dan memprediksi hal-hal yang berhubungan dengan variabel-variabel yang terlibat di dalam model regresi.

Menurut Sembiring (1995:32), “model regresi adalah model yang memberikan gambaran mengenai hubungan antara variabel bebas dengan variabel terikat”. Jika analisis dilakukan untuk satu variabel bebas dengan variabel terikat, maka regresi ini dinamakan regresi sederhana dengan model:

$$Y = \alpha + \beta X + \varepsilon \,. \qquad (2.1)$$

Keterangan : α, β = koefisien garis regresi,

Y = variabel terikat

X = variabel bebas

ε = error / sesatan.

2.3.2 Model Regresi Linear Berganda

Suatu masalah mungkin melibatkan beberapa variabel $X_1, X_2, \ldots, X_k$ dan satu variabel terikat Y yang diduga nilainya tergantung pada nilai-nilai $X_1, X_2, \ldots, X_k$. Regresi linear ganda menjelaskan hubungan fungsional linear antara kelompok variabel bebas $\left[X_1, X_2, \ldots, X_k\right]$ dan variabel terikat Y. Secara umum, model regresi linear ganda melibatkan satu variabel terikat Y dan variabel bebas $X_1, X_2, \ldots, X_k$ dinyatakan sebagai berikut

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon \qquad (2.2)$$

Keterangan : $X_1, X_2, \ldots, X_k$ = variabel bebas

β_j = koefisien regresi

Y = variabel terikat

ε = error

2.3.3 Asumsi Model Regresi Linear

Asumsi-asumsi yang harus dipenuhi agar OLS dapat menghasilkan estimasi yang baik pada model regresi yaitu sebagai berikut.

(1) Nilai rata-rata dari kasalahan pengganggu sama dengan nol

$E[\varepsilon_i] = 0$ untuk i=1, 2,… , n.

(2) Tidak ada autokorelasi antara kasalahan pengganggu yang satu dengan yang lainnya

$\mathrm{kov}[\varepsilon_i \varepsilon_j] = 0$ untuk $i \neq j$.

(3) Semua kesalahan penggangu mempunyai varian sama atau disebut dengan homoskedastisitas

$\mathrm{var}[\varepsilon_i] = \sigma^2$ untuk i=1, 2,… , n.

(4) Variabel bebas X adalah suatu himpunan bilangan yang tetap dan bebas terhadap kesalahan pengganggu ε_i.

(5) Tidak terdapat hubungan antara variabel bebas X atau tidak terdapat multikolienaritas antara variabel bebas X.

(6) Gangguan berdistribusi normal dengan rata-rata nol dan varians σ^2

$\varepsilon_i N \sim (0, \sigma^2)$.

2.4 Metode Kuadrat Terkecil (Ordinary Least Square Method)

Metode kuadrat terkecil pertama kali dikemukakan oleh Carl Freidrich Gauss, seorang ahli matematika Jerman. Metode kuadrat terkecil merupakan metode yang lebih banyak digunakan dalam pembentukan model regresi atau mengestimasi parameter-parameter regresi dibandingkan metode-metode lain.

Metode kuadrat terkecil adalah metode yang digunakan untuk mengestimasi nilai □ dengan cara meminimumkan jumlah kuadrat dari residu.

Menurut Sembiring (1995:40), estimasi koefisien garis regresi □ dan □ pada n data pengamatan dengan metode kuadrat terkecil diperoleh dengan meminimumkan fungsi:

$$J = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 . \qquad (2.3)$$

Pada persamaan (2.3), x_i dan y_i bilangan yang berasal dari pengamatan, sedangkan □ dan □ berubah bila garis regresinya berubah. Jika J diturunkan terhadap □ dan □, kemudian menyamakannya dengan nol, maka diperoleh

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

atau,

$$\sum_{i=1}^{n} y_i - n\alpha - \beta \sum_{i-1}^{n} x_i = 0 \qquad (2.4)$$

dan

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

atau,

$$\sum_{i=1}^{n} y_i x_i - \alpha \sum_{i=1}^{n} x_i - \beta \sum_{i=1}^{n} x_i^2 = 0 \qquad (2.5)$$

Jika nilai □ dan □ pada persamaan (2.4) dan (2.5) diganti dengan a dan b, maka persamaannya menjadi suatu sistem persamaan linear. Nilai a dan b merupakan estimasi (taksiran) dari □ dan □̃

$$\sum_{i=1}^{n} a + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{2.6}$$

$$a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i$$

Dari persamaan (2.6) yang pertama diperoleh :

$$\begin{aligned} a &= \frac{\sum_{i=1}^{n} y_i}{n} - b\,\frac{\sum_{i=1}^{n} x_i}{n} \\ &= \bar{y} - b\,\bar{x}, \end{aligned}$$

dengan $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ dan $\bar{y} = \dfrac{\sum_{i=1}^{n} y_i}{n}$.

Persamaan (2.6) yang kedua menjadi

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

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

Jadi,

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

Estimasi persamaan regresi $\hat{y}_i$ adalah $\hat{y}_i = a + bx_i$ dan nilai sisaan $e_i = y_i - \hat{y}_i$. Jadi taksiran persamaan regresi dapat ditulis sebagai

$$\hat{y}_i = a + bx_i$$

$$= \overline{y} - b\overline{x} + bx_i$$

$$= \overline{y} + b(x_i - \overline{x}).$$

Menurut Sembiring (1995:93), estimasi parameter dengan metode kuadrat terkecil untuk regresi berganda sebagai berikut.

Dari persamaan regresi sederhana dapat ditulis

$$J = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 X_{i1} - \ldots - \beta_k X_{ik}\right)^2. \tag{2.7}$$

Untuk meminimumkan (2.7), dicari turunan J secara parsial terhadap β_j, j = 0, 1, 2,… ,k dan disamakan dengan nol sehingga diperoleh

$$\frac{\partial J}{\partial \beta_0} = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_i X_{i1} - \ldots - \beta_k X_{ik}) = 0$$

$$\frac{\partial J}{\partial \beta_1} = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_i X_{i1} - \ldots - \beta_k X_{ik}) x_{i1} = 0,$$

$$\frac{\partial J}{\partial \beta_2} = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_i X_{i1} - \ldots . - \beta_k X_{ik}) x_{i2} = 0,$$

$$\vdots$$

$$\frac{\partial J}{\partial \beta_k} = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_i X_{i1} - \ldots - \beta_k X_{ik}) x_{ik} = 0. \tag{2.8}$$

Persamaan (2.8) menghasilkan p persaman normal berikut ini

$$n\hat{\beta}_0 + \hat{\beta}_1\sum_{i=1}^{n} x_{i1} + \hat{\beta}_2\sum_{i=1}^{n} x_{i2} + \ldots + \hat{\beta}_k\sum_{i=1}^{n} x_{ik} = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0\sum_{i=1}^{n} x_{i1} + \hat{\beta}_1\sum_{i=1}^{n} x_{i1}{}^2 + \hat{\beta}_2\sum_{i=1}^{n} x_{i1}x_{i2} + \ldots + \hat{\beta}_k\sum_{i=1}^{n} x_{i1}x_{ik} = \sum_{i=1}^{n} x_{i1}y_i$$

$$\hat{\beta}_0\sum_{i=1}^{n} x_{i2} + \hat{\beta}_1\sum_{i=1}^{n} x_{i1}x_{i2} + \hat{\beta}_2\sum_{i=1}^{n} x_{i1}x_{i2}{}^2 + \ldots + \hat{\beta}_k\sum_{i=1}^{n} x_{i2}x_{ik} = \sum_{i=1}^{n} x_{i2}y_i$$

$$\vdots$$

$$\hat{\beta}_0\sum_{i=1}^{n} x_{ik} + \hat{\beta}_1\sum_{i=1}^{n} x_{i1}x_{ik} + \hat{\beta}_2\sum_{i=1}^{n} x_{i2}x_{ik} + \ldots + \hat{\beta}_k\sum_{i=1}^{n} x_{ik}{}^2 = \sum_{i=1}^{n} x_{ik}y_i. \quad (2.9)$$

Jika disusun dalam bentuk matrik maka persamaan (2.9) menjadi

$$X'X\hat{\beta} = X'Y \qquad (2.10)$$

dengan

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \; X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ 1 & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}, \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}$$

$$X'X = \begin{pmatrix} n & \sum_{i=1}^{n} x_{i1} & \cdots & \sum_{i=1}^{n} x_{ik} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 & \cdots & \sum_{i=1}^{n} x_{i1}x_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_{ik} & \sum_{i=1}^{n} x_{i1}x_{ik} & \cdots & \sum_{i=1}^{n} x_{ik}^2 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_1 \\ \sum_{i=1}^{n} x_{i1} y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ik} y_i \end{pmatrix}$$

Untuk menyelesaikan persamaan (2.10) maka harus dikalikan dengan invers dari $(X'X)$. Sehingga estimasi kuadrat terkecil dari β adalah

$$(X'X)^{-1} X'X\hat{\beta} = (X'X)^{-1} X'Y$$

$$\hat{\beta} = (X'X)^{-1} X'Y.$$

Model persamaan regresi berganda dapat ditulis

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon$$

Penyelesaian estimasi parameter dengan metode kuadrat terkecil untuk regresi berganda dengan menggunakan matriks sebagai berikut. Dengan menggunakan notasi matriks, model persamaan regresi berganda dapat ditulis

$$Y = X\beta + \varepsilon \tag{2.11}$$

Prinsip dari OLS adalah mengestimasi nilai □ dengan cara meminimumkan jumlah kuadrat error. Jumlah kuadrat error dalam aljabar matrik dinotasikan dengan $\varepsilon'\varepsilon$.

Dari persamaan (2.11), $\varepsilon = Y - X\beta$. Kemudian untuk meminimumkan jumlah kuadrat error dapat ditunjukkan sebagai berikut.

$$\min \varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta)$$

$$= Y'Y - Y'X\beta - Y'X\beta + (X\beta)'X\beta$$

$$= Y'Y - 2Y'X\beta + \beta'X'X\beta \qquad (2.12)$$

Untuk meminimumkan (2.12), dicari turunan $\varepsilon'\varepsilon$ terhadap β dan disamakan dengan nol sehingga diperoleh

$$\frac{\partial \varepsilon'\varepsilon}{\partial \beta} = -2X'Y + 2X'X\beta = 0$$

$$= X'Y + X'X\beta = 0$$

$$\Leftrightarrow X'X\beta = X'Y$$

$$\Leftrightarrow (X'X)^{-1}\, X'X\beta = (X'X)^{-1}\, X'Y$$

$$\Leftrightarrow I\beta = (X'X)^{-1}\, X'Y$$

$$\Leftrightarrow \beta = (X'X)^{-1}\, X'Y.$$

2.5 Pencilan (Outlier)

"*Outlier* adalah kasus atau data yang memiliki karakteristik unik yang terlihat sangat berbeda jauh dari observasi-observasi lainnya dan muncul dalam bentuk nilai ekstrim, baik untuk sebuah variabel tunggal maupun variabel kombinasi" (Ghozali, 2009: 40).

Menurut Hampel, Rousseeaw dan Stahel, sebagaimana dikutip oleh Olive (2006:4), mendefinisikan *outlier* adalah observasi yang menyimpang dari pola yang terbentuk oleh sebagian besar data.

Menurut Ghozali (2009: 40), Terdapat empat penyebab timbulnya data *outlier* antara lain:

Kesalahan dalam memasukan data, gagal dalam menspesifikasi adanya missing value dalam program komputer, *outlier* bukan merupakan anggota populasi yang di ambil sebagai sampel, dan *outlier* berasal dari populasi yang di ambil sebagai sampel, tetapi ditribusi dari variabel dalam populasi tersebut memiliki nilai ekstrim serta tidak terdistribusi secara normal.

Pada analisis regresi terdapat 3 tipe *outlier* yang berpengaruh terhadap estimasi OLS. Roesseuw dan Leroy (1987) sebagaimana dikutip oleh Croux (2008:2), mengenalkan 3 jenis *outlier* tersebut sebagai *vertical outliers*, *good leverage points* dan *bad leverage points*.

Vertical outliers adalah semua pengamatan yang terpencil pada variabel respon, tetapi tidak terpencil dalam variabel prediktor. Keberadaan *vertical outliers* berpengaruh terhadap estimasi *Least Squares*, khususnya pada estimasi intersep. *Good leverage points* adalah pengamatan yang terpencil pada variabel prediktor tetapi terletak dekat dengan garis regresi. Hal ini berarti x_i menjauh tetapi y_i cocok dengan garis linear. Keberadaan *good leverage points* tidak berpengaruh terhadap estimasi *Least Squares*, tetapi berpengaruh terhadap inferensi statistik karena good leverage points meningkatkan estimasi standar error. *Bad leverage points* adalah pengamatan yang terpencil pada variabel prediktor dan terletak jauh dari garis regresi. Keberadaan *bad leverage points* berpengaruh signifikan pada estimasi *Least Squares*, baik terhadap intersep maupun slope dari persamaan regresi.

Menurut Soemartini (2007: 14), kombinasi residu robust dan jarak robust mencirikan 4 model titik, yaitu:

(1) observasi biasa yaitu suatu titik yang memiliki residu robust kecil dan nilai jarak robust kecil;

(2) *vertical outlier* yaitu suatu titik yang memiliki nilai residu robust besar dan nilai jarak robust kecil;

(3) *good leverage points* yaitu suatu titik yang memiliki niliai residu robust kecil dan nilai jarak robust besar; dan

(4) *bad leverage points* yaitu suatu titik yang memiliki nilai residu robust dan nilai jarak robust besar.

Outlier berpengaruh terhadap proses analisis data, misalnya terhadap nilai mean dan standar deviasi. Oleh karena itu, keberadaan *outlier* dalam suatu pola data harus dihindari. *Outlier* dapat menyebabkan varians pada data menjadi lebih besar, interval dan range menjadi lebar, mean tidak dapat menunjukan nilai yang sebenarnya (bias) dan pada beberapa analisis inferensi, *outlier* dapat menyebakan kesalahan dalam pengambilan keputusan dan kesimpulan.

Berbagai kaidah telah diajukan untuk menolak *outlier* (dengan kata lain untuk memutuskan menyisihkan *outlier* tersebut dari data, kemudian menganalisis kembali tanpa *outlier* tersebut). Penolakan begitu saja suatu *outlier* bukanlah prosedur yang bijaksana. Adakalanya *outlier* memberikan informasi yang tidak bisa diberikan oleh data lainnya, misalnya karena *outlier* timbul dari kombinasi keadaan yang tidak biasa yang mungkin saja sangat penting dan perlu diselidiki lebih jauh. Secara filosofi *outlier* seharusnya tetap dipertahankan jika data *outlier* tersebut memang representasi dari populasi. Sebagai kaidah umum *outlier* baru kita tolak jika setelah ditelusuri ternyata merupakan akibat dari kesalahan-kesalahan seperti kesalahan mencatat amatan bersangkutan atau kesalahan ketika menyiapkan peralatan.

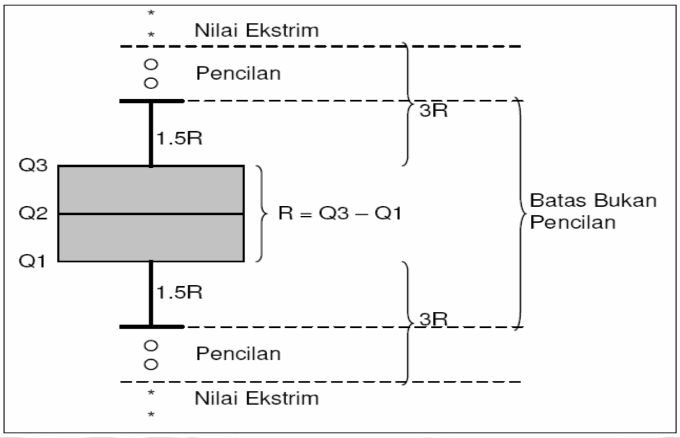
2.6 Identifikasi *Outlier*

Dalam statistik, tahapan diagnosis terhadap data *outlier* perlu dilakukan karena data outlier tersebut berpengaruh terhadap koefisien regresi. Terdapat beberapa metode untuk menentukan batasan *outlier* dalam sebuah analisis, yaitu sebagai berikut.

2.6.1 Metode *Boxplot*

Boxplot merupakan metode grafis yang dikembangkan oleh Tukey dan sering digunakan untuk analisis data dan diintepretasikan untuk memperoleh informasi dari sebuah sampel. *Boxplot* bisa dibuat relatif mudah secara manual atau dengan bantuan program komputer statistika.

Metode ini merupakan yang paling umum yakni dengan mempergunakan nilai kuartil dan jangkauan. Kuartil 1, 2, dan 3 akan membagi sebuah urutan data menjadi empat bagian. Jangkauan (IQR, *Interquartile Range*) didefinisikan sebagai selisih kuartil 1 terhadap kuartil 3, atau IQR = Q3 – Q1. ”Data-data *outlier* dapat ditentukan yaitu nilai yang kurang dari 1.5*IQR terhadap kuartil 1 dan nilai yang lebih dari 1.5*IQR terhadap kuartil atas” (Soemartini, 2007:9).



Gambar 2.1 Skema Identifikasi *Outlier* menggunakan *Boxplot*

2.6.2 Metode *Leverage Value*

Menurut Soemartini (2007: 14) ”*Leverage* adalah pengamatan dengan nilai ekstrim pada variabel tak bebas atau ukuran jauhnya variabel tak bebas menyimpang dari rata-ratanya”. Nilai *leverage* merupakan nilai pengaruh yang terpusat. Pada observasi variabel independen dapat juga menggunakan nilai *leverage* (h_{ii}) untuk mendeteksi adanya *outlier*. Nilai *leverage* (h_{ii}) merupakan elemem-elemen diagonal dari matriks H. H disebut matriks Hat karena mentransformasi vektor respon observasi y ke dalam vektor respon pencocokan $\hat{y}$

$$\hat{y} = X\hat{\beta}$$

$$= \mathrm{X}((X'X)^{-1}X'y)$$

$$= X\ (X'X)^{-1}X'y$$

$$= \mathrm{H\ y}$$

H adalah matriks n x n dan H = $X\ (X'X)^{-1}X'$.

"Observasi yang mempunyai nilai *leverage* (h_{ii}) yang melebihi (2p-1)/n, dengan p adalah banyaknya variabel independen ditambah konstan dan n jumlah observasi maka akan mengindikasikan terdapat *outlier*" (Soemartini, 2007: 8).

2.6.3 Metode *Cook's Distance*

Metode lain untuk mendeteksi *outlier* adalah dengan suatu ukuran pengaruh yang diperkenalkan oleh Cook (1977) yang dinamakan *Cook's Distance*. *Cook's distance* merupakan suatu ukuran untuk mendeteksi besarnya pengaruh adanya outlier terhadap semua estimasi koefisien regresi, yaitu

$$\mathrm{D}_i = \frac{ei^2 h_{ii}}{pMSE(1-h_{ii})^2}$$

"Dengan h_{ii} adalah nilai *leverage* untuk kasus ke-*i* dan n jumlah data pengamatan, suatu data disebut *outlier* apabila nilai D_i >4/n" (Yaffe, 2002:44).

2.6.4 *Standardized Residual*

Suatu metode yang sederhana dan efektif untuk mendeteksi *outlier* adalah dengan memeriksa residual. Residual ke-i didefinisikan sebagai berikut:

$$e_i = y_i - \hat{y}_i$$

Sesuai dengan residual ke-i di atas, dapat didefinisikan standardized residual ke-i sebagai.

$$e_{is} = \frac{e_i}{\sqrt{MSE}}$$

dengan MSE = $\dfrac{\sum_{i=1}^{n} e_i^2}{n-2}$

MSE adalah rata-rata residual kuadrat dan akar dari MSE disebut standar eror. Standar eror merupakan ukuran kebaikan model regresi. Standar eror mengukur besarnya variansi model regresi, semakin kecil nilainya semakin baik model regresinya.

Untuk melakukan identifikasi *outlier*, diperhatikan nilai-nilai dari *standardized residual*. "Jika nilai dari *standardized residual* lebih dari 3,5 atau kurang dari -3,5 maka data tersebut dikatakan sebagai data *outlier*" (Yaffe, 2002:35).

2.7 Breakdown Point

Menurut Huber (1981: 13), "*Breakdown point* adalah fraksi terkecil atau persentase dari *outlier* yang dapat menyebabkan nilai estimator menjadi besar". *Breakdown point* untuk sebuah estimator T di F didefinisikan sebagai:

$$\varepsilon^* = \varepsilon^*(F_0, T) = sup\{\varepsilon;\ b(\varepsilon) < b(1)\},$$

dengan $b(1) = \infty$. *Breakdown point* digunakan untuk menjelaskan ukuran kerobustsan dari tehnik robust. Kemungkinan tertinggi *breakdown point* untuk sebuah estimator adalah 0,5. Jika *breakdown point* lebih dari 0,5 berarti estimasi model regresi tidak dapat menggambarkan informasi dari mayoritas data.

2.8 Regresi Robust

Regresi robust diperkenalkan oleh Andrews (1972). "Regresi robust merupakan metode regresi yang digunakan ketika distribusi dari error tidak normal dan atau adanya beberapa *outlier* yang berpengaruh pada model" (Olive, 2005:3)."Regresi robust digunakan untuk mendeteksi *outlier* dan memberikan hasil yang resisten terhadap adanya *outlier*" (Chen, 2002:1). Efisiensi dan *breakdown point* digunakan untuk menjelaskan ukuran kerobust-an dari tehnik robust. Efisiensi menjelaskan seberapa baiknya suatu tehnik robust sebanding dengan *Least Square* tanpa *outlier*. Semakin tinggi effisiensi dan *brekdown point* dari suatu estimator maka semakin robust (resisten) terhadap *outlier*.

ur statistik yang bersifat robust ini ditujukan untuk mengakomodasi keberadaan data ekstrim dan sekaligus meniadakan pengaruhnya terhadap hasil analisis tanpa terlebih dulu mengadakan identifikasi terhadapnya.

Beberapa peneliti menyarankan penggunaan metode regresi robust sebagai pengontrol hasil pendugaan menggunakan metode kuadrat terkecil, bila kedua hasil tersebut tidak berbeda jauh maka hasil metode kuadrat terkecil dapat digunakan dengan lebih yakin, sedangkan kalau terdapat perbedaan yang mencolok maka sisaan dari hasil metode regresi robust lebih menjelaskan dalam menggambarkan pengamatan mana yang perlu mendapat perhatian lebih lanjut tanpa memerlukan tehnik diagnostik yang khusus.

Menurut Chen (2002:1), terdapat 3 kelas masalah yang dapat menggunakan tekhnik regresi robust yaitu:

(1) masalah dengan *outlier* yang terdapat pada peubah y (respon);
(2) masalah dengan *outlier* yang terdapat pada peubah x (*leverage points*); dan
(3) masalah dengan *outlier* yang terdapat pada keduanya yaitu pada peubah y (respon) dan peubah x (penjelas).

Banyak metode yang dikembangkan dalam regresi robust untuk mengatasi masalah *outlier*. Dalam regresi robust terdapat beberapa metode estimasi yaitu:

2.8.1 Estimasi-M

Wilcox (2005: 51) menjelaskan "estimasi-M pertama kali diperkenalkan oleh Huber pada tahun 1973 dan merupakan penggambaran dari suatu percobaan yang menggabungkan sifat efisiensi OLS dan ketahanan dari estimasi LAV (LAD)". LAV merupakan estimasi yang meminimumkan jumlah nilai mutlak dari residual.

$$\sum_{i=1}^{n}|e_i| = \min \sum_{i=1}^{n}\left|y_i - \sum_{j=0}^{k} x_{ij}\beta_j\right|.$$

Metode LAV lebih resisten terhadap *outlier* daripada OLS karena pengaruh dari *outlier* dibatasi. Hal ini dapat dilihat dari bentuk fungsi *influence* sebagai berikut.

$$\psi(u_i) = \begin{cases} 1, & u_i \le 0 \\ 0, & u_i = 0 \\ -1, & u_i < -0 \end{cases}$$

dimana u_i nilai skala residual.

Penggabungan LAV dan OLS dalam Estimasi-M dapat dilihat dari fungsi *influence* dari Estimasi-M sebagai berikut.

$$\psi(u_i) = \begin{cases} c, & u_i > c \\ u_i, & |u_i| \le c \\ -c, & u_i < -c \end{cases}$$

Esimasi-M mempunyai sifat seperti OLS pada fungsi tengah, tetapi pada nilai ekstrim, Estimasi-M seperti LAV. Estimasi-M dikembangkan untuk mrmperbaiki kelemahan yang tidak robust terhadap outlier pada variabel prediktor maupun pada variabel prediktor. Sehingga Estimasi-M resisten terhadap outlier pada variabel respon sama seperti LAV, dan tidak resisten terhadap outlier pada variabel prediktor. Estimasi-M merupakan suatu metode robust yang luas dan terkenal serta dapat di análisis dengan mudah secara teoritis maupun komputer. Estimasi-M mempunyai *breakdown point* sebesar nol (0).

Estimasi-M merupakan estimasi yang meminimumkan suatu fungsi residual ρ .

$$\hat{\beta}_m = \min \sum_{i=1}^{n} \rho(e_i) = \min \sum_{i=1}^{n} \rho(y_i - \sum_{j=0}^{k} x_{ij}\beta_j)$$

Fungsi ρ dipilih sebagai representasi pembobot dari residual. Solusi di atas bukan merupakan skala *equivariant*. Oleh karena itu untuk memperoleh skala residual harus distandarkan dengan sebuah skala estimasi robust $\hat{\sigma}$. Sehingga persamaannya menjadi:

$$\hat{\beta}_m = \min \sum \rho\left(\frac{e_i}{\hat{\sigma}}\right) = \min \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{i=1}^{n} x_{ij}\beta_j}{\hat{\sigma}}\right) \quad (2.13)$$

dimana $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ merupakan nilai estimasi-M dari $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ yang meminimumkan

$$\sum \rho(u_i) = \sum \rho(\frac{e_i}{\hat{\sigma}}) \quad (2.14)$$

Dipilih estimasi yang popular untuk $\hat{\sigma}$ adalah

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

dimana MAD adalah *Median Absolute Deviatian*. ”Pemilihan konstan 0,6745 membuat $\hat{\sigma}$ merupakan suatu estimasi yang mendekati tak bias dari σ jika *n* besar dan residu berdistribusi normal” (Fox, 2002:2).

Prosedur estimasi-M sebagai berikut.

(1) Dihitung penaksir β, dinotasikan b menggunakan metode kuadrat terkecil, sehingga didapatkan $\hat{y}_{i,0}$ dan $\varepsilon i_{,0}$ = yi $- \hat{y}_{i,0}$, (i = 1, 2, ... n) yang diperlakukan sebagai nilai awal .

(2) Menghitung nilai $\hat{\sigma}$

$$\hat{\sigma} = \frac{MAD}{0{,}6745} = \frac{median\{|e_i - median(e_i)|\}}{0{,}6745}.$$

(3) Mencari nilai skala residual (u_i)

$$u_i = \frac{(y_i - \hat{y}_i)}{\hat{\sigma}} = \frac{e_i}{\hat{\sigma}}.$$

(4) Mendefinisikan pembobot berdasarkan fungsi pembobot $w_i = w(u_i)$, dengan konstanta untuk pembobot Huber sebesar 1,345 dan tukey bisquare sebesar 4,685.

(5) Memperbaiki estimasi $\hat{\beta}$ berdasarkan metode Weighted Least Squares (WLS) dengan pembobot w_i sehingga diperoleh $\hat{\beta}$ yang baru pada iterasi ke-1.

(6) Selanjutnya ulangi langkah 2 sampai langkah 5 sehingga nilai w_i akan berubah pada tiap iterasinya sehingga diperoleh $\hat{\beta}_m$ yang konvergen.

Estimasi kuadrat terkecil dapat digunakan sebagai nilai permulaan $\hat{\beta}_0$.

2.8.2 *Least Median Squares* (LMS)

Metode LMS merupakan metode *High Breakdown Value* yang diperkenalkan oleh Rousseeuw pada tahun 1984. Wilcox (2005: 51) menjelaskan "Metode LMS adalah suatu metode estimasi parameter regresi robust dengan meminimumkan median dari kuadrat residual". LMS sangat robust terhadap *outlier* pada variable X maupun Y. Metode LMS mengganti jumlah kuadrat residual yang merupakan karakteristik OLS dengan median kuadrat residual.

$$\min MED(e_i^{\ 2}) = \min MED(y_i - \sum_{j=0}^{k} x_{ij}\beta_j)^2.$$

Ide untuk dengan menggantikan penjumlahan dengan median, menghasilkan estimasi yang resisten terhadap *outliers*. Walau hasil ini dicapai (LMS mempunyai *breakdown point* = 0.5), akan tetapi LMS mempunyai kelemahan ketika pembatas itu digunakan. LMS mempunyai efisiensi sebesar 37%.

2.8.3 Least Trimmed Squares (LTS)

Sama halnya dengan metode LMS, metode robust LTS juga merupakan metode *High Breakdown Value* yang diperkenalkan oleh Rousseeuw pada tahun 1984. Metode LTS adalah suatu metode estimasi parameter regresi robust dengan untuk meminimumkan jumlah kuadrat h residual.

$$\min \sum_{i=1}^{h} e_{(i)}^2$$

dengan h = $[n/2]+[(k+2)/2]$

Keterangan : $e_{(i)}^2$ = Kuadrat residual yang diurutkan dari terkecil ke terbesar.

$$e_{(1)}^2 < e_{(2)}^2 < e_{(3)}^2 < \ldots. < e_{(i)}^2 < \ldots < e_{(h)}^2 < \ldots < e_{(n)}^2$$

n = Banyaknya pengamatan

k = Parameter regresi

Jumlah h menunjukkan sejumlah subset data dengan kuadrat fungsi objektif terkecil. Nilai h pada persamaan akan membangun *breakdown point* sebesar 0,5.

LTS merupakan mempunyai resisten yang paling tinggi terhadap *outlier*, akan tetapi LTS sangat tidak efisien (efisiensi relatif 8%) dan dapat mengakibatkan kesalahan dalam penggambaran model data jika dinilai dari pengelompokan *outlier*, atau jika jumlah data relatif kecil. Meskipun demikian, LTS masih mempunyai hubungan dalam perhitungan dengan estimasi lain. Antara lain, GM Estimasi Yang diajukan oleh Coakley dan Hettmansperger (1993) mempergunakan LTS untuk memperoleh taksiran nilai dari residual. Residual LTS juga dapat dipergunakan secara efektif pada plot diagnostik *outlier*.

Prosedur estimasi LTS dapat diuraikan sebagai berikut.

(1) Menghitung estimasi parameter β.

(2) Menentukan n residual $e_i^2 = (y_i - \sum_{j=0}^{k} x_{ij}\beta_j)^2$ yang bersesuaian dengan $\hat{\beta}$, kemudian menghitung h $= \dfrac{n+k+2}{2}$ pengamatan dengan nilai e_i^2 terkecil.

(3) Menghitung $\sum_{i=1}^{h} e_i^2$.

(4) Melakukan estimasi parameter β_{new} dari h pengamatan.

(5) menentukan n kuadrat residual yang baru yang bersesuaian dengan $\hat{\beta}_{new}$ kemudian menghitung sejumlah h_{new} pengamatan dengan nilai e_i^2 terkecil.

(6) Menghitung $\sum_{i=1}^{h_{new}} e_i^2$ dan mengulang langkah 4 sampai 6 untuk mendapatkan fungsi obyektif yang kecil dan konvergen.

2.8.4 Estimasi-S

Metode robust S merupakan metode *High Breakdown Value* yang diperkenalkan pertama kali oleh Rousseeuw dan Yohai pada tahun 1984. Menurut Wilcox (2005: 55), “Estimasi-S merupakan solusi dengan kemungkinan terkecil dari penyebaran residual”.

$$\min \hat{\sigma}(e_1(\hat{\beta}),\ldots,e_n(\hat{\beta}))$$

Selain meminimumkan varians dari residual, Estimasi-S juga meminimumkan skala residual dari estimasi-M.

Estimasi-S mempunyai *breakdown point* sebesar 0,5. Breakdown point sebesar 0,5 diperoleh dengan menggunakan fungsi Tukey Bisquare dan tuning constan sebesar 1,547. Meskipun Estimasi-S mempunyai *breakdown point* yang tinggi = 0,5, Estimasi-S tidak menarik untuk digunakan karena mempunyai efisiensi yang sangat rendah (kurang lebih sekitar 30% relatif terhadap OLS ketika distribusi error normal). Prosedur estimasi-S dapat diuraikan sebagai berikut:

(1) Dihitung penaksir β, menggunakan metode kuadrat terkecil, sehingga didapatkan $\hat{y}_{i,0}$ dan $\varepsilon i_{,0}$ = yi − $\hat{y}_{i,0}$, (i = 1, 2, ... n) yang diperlakukan sebagai nilai awal.

(2) Menghitung nilai $\hat{\sigma}$

$$\hat{\sigma} = \frac{MAD}{0{,}6745} = \frac{\frac{1}{n}\sum_{i=1}^{n}|e_i|}{0{,}6745}.$$

(3) Mencari nilai skala residual (u_i)

$$u_i = \frac{(y_i - \hat{y}_i)}{\hat{\sigma}} = \frac{e_i}{\hat{\sigma}}.$$

(4) Mendefinisikan pembobot berdasarkan fungsi pembobot

$w_i = w(u_i)$, dengan menggunakan fungsi bobot bisquare c = 1,547 untuk mendapatkan nilai breakdown sebesar 0,5.

(5) Memperbaiki estimasi $\hat{\beta}$ berdasarkan metode Weighted Least Squares (WLS) dengan pembobot w_i sehingga diperoleh $\hat{\beta}$ yang baru pada iterasi ke-1.

(6) Selanjutnya ulangi langkah 2 sampai langkah 5 sehingga nilai w_i akan berubah pada tiap iterasinya sehingga diperoleh $\hat{\beta}_m$ yang konvergen.

2.8.5 Estimasi-MM

Wilcox (2005: 56) menjelaskan "metode estimasi-MM dikenalkan oleh Yohai (1987) yang menggabungkan suatu *high breakdown point* (50%) dengan efisiensi tinggi (mencapai 95%)". Estimasi MM dimulai dengan mencari estimasi S yang sangat robust dan resisten yang meminimumkan suatu skala residual. Selanjutnya skala residual tetap konstan dan di akhiri dengan menetapkan parameter-parameter regresi menggunakan estimasi-M. Estimasi-MM mempunyai *breakdown point* yang sama dengan Estimasi-S yaitu sebesar 0,5. Estimasi-MM mempunyai *breakdown point* sebesar 0,5 menjelaskan bahwa banyaknya *outlier* hingga separuh data pengamatan tidak berpengaruh terhadap estimasi-MM.

Estimasi-MM didefinisikan sebagai berikut.

$$\hat{\beta}_{mm} = \min \sum \rho\left(\frac{e_i}{\hat{\sigma}}\right) = \min \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{i=1}^{n} x_{ij}\beta_j}{\hat{\sigma}}\right).$$

Estimasi-S sebagai permulaan dengan nilai *breakdown* yang tinggi dan di akhiri dengan estimasi-M yang membuat estimator mempunyai efisiensi yang tinggi. Pada umumnya digunakan fungsi Tukey Bisquare baik pada estimasi-S maupun estimasi-M.

Sebagaimana dalam kasus estimasi-M, estimasi MM menggunakan Iteratively Reweighted Least Square (IRLS) untuk mencari estimasi parameter regresi.

Prosedur estimasi-MM dapat diuraikan sebagai berikut.

(1) Mengestimasi koefisien $\hat{\beta}_j^{(1)}$, sehingga diperoleh residual $e_i^{(1)}$ yang diambil dari regresi robust dengan *high breakdown point*.

(2) Residual $e_i^{(1)}$ pada langkah pertama digunakan untuk menghitung skala residual Estimasi-M, $\hat{\sigma}$ dan dihitung pula bobot awal $w_i^{(1)}$.

(3) Residual $e_i^{(1)}$ dan skala residual $\hat{\sigma}$ dari langkah (2) digunakan dalam iterasi awal dengan metode WLS untuk menghitung koefisien regresi.

$$\sum_{i=1}^{n} w_i^{(1)}\left(\frac{e_i^{(1)}}{\hat{\sigma}}\right)x_i = 0$$

dimana w_i menggunakan pembobot Huber atau Tukey bisquare.

(4) Menghitung bobot baru $w_i^{(2)}$ menggunakan residual dari iterasi awal WLS (langkah 3).

(5) Langkah 2, 3, 4 diulang (reiterasi dengan skala residual tetap konstan) sampai $\sum_{i=1}^{n} \left| e_i^{\ m} \right|$ konvergen, yaitu selisih $\beta_j^{\ m+1}$ dengan $\beta_j^{\ m}$ kurang dari 10^{-4}, dengan m adalah banyaknya iterasi.

Dari kelima metode di atas, peneliti memilih dua metode robust, yaitu metode robust estimasi-M dan metode robust estimasi-MM karena kedua metode di atas yang popular digunakan, dan peneliti ingin membandingkan kedua estimasi robust tersebut. Perbandingan dari kelima metode tersebut dapat dilihat pada tabel 2.1 (Wilcox, 2005: 58).

Tabel 2.1 Perbandingan Beberapa Estimasi Regresi Robust

Estimasi	*Breakdown Point*	Efisiensi
M (Huber,biweight)	0	95%
LMS	0,5	37%
LTS	0,5	8%
S	0,5	33%
MM	0,5	95%

2.9 Fungsi-fungsi Ukuran Robust

Fungsi pembobot yang digunakan estimasi-M antara lain:

2.9.1 Fungsi Pembobot Huber

Menurut Cranmer (2005: 12) ”Fungsi Huber dikembangkan oleh Huber pada tahun 1964, fungsi objektif Huber adalah gabungan dari OLS dan *Least Absolute Value* (LAV)”. Fungsi objektif meminimumkan gabungan dari jumlah kuadrat residual dan jumlah mutlak residual. Fungsi huber lebih resisten terhadap *outlier* daripada OLS.

Fungsi pembobot yang disarankan oleh Huber memakai fungsi obyektif

$$\rho(u_i) = \begin{cases} \frac{1}{2}u_i^{\,2}, & |u_i| \le c \\ c|u_i| - \frac{1}{2}c^2, & |u_i| > c \end{cases}$$

dengan

$$\psi(u_i) = \rho'(u_i) = \frac{\partial(\rho(u_i))}{\partial u_i} = \begin{cases} c, & u_i > c \\ u_i, & |u_i| \le c \\ -c, & u_i < -c \end{cases}$$

dan fungsi pembobot

$$w_i = w(u_i) = \frac{\psi(u_i)}{u_i} = \begin{cases} 1, & |u_i| \le c \\ \frac{c}{|u_i|}, & |u_i| > c \end{cases}$$

2.9.2 Fungsi Pembobot Tukey Bisquare

Menurut Cranmer (2005: 12) ”Fungsi tukey memiliki perbedaan daripada fungsi Huber. Khususnya pada tingkat residual yang besar”.

Fungsi pembobot yang disarankan oleh Tukey memakai fungsi obyektif

$$\rho(u_i) = \begin{cases} \dfrac{c^2}{6}\left\{1-\left[1-\left(\dfrac{u_i}{c}\right)^2\right]^3\right\}, & |u_i| \le c \\ \dfrac{c^2}{6}, & |u_i| > c \end{cases}$$

Sehingga untuk nilai mutlak skala residual yang lebih besar daripada c, tidak meningkat. Hal ini berarti pengaruh dari residual dibatasi.

$$\psi(u_i) = \rho'(u_i) = \frac{\partial(\rho(u_i))}{\partial u_i} = \begin{cases} u_i\left[1-\left(\dfrac{u_i}{c}\right)^2\right]^2, & |u_i| \le c \\ 0, & |u_i| > c \end{cases}$$

dan fungsi pembobot

$$w_i = w(u_i) = \frac{\psi(u_i)}{u_i} = \begin{cases} \left[1-\left(\dfrac{u_i}{c}\right)^2\right]^2, & |u_i| \le c \\ 0, & |u_i| > c \end{cases}$$

Secara ringkas, fungsi obyektif ρ dan fungsi pembobot dari estimasi Huber, dan Tukey bisquares dapat dilihat pada Tabel 2.2 (Fox, 2002: 3). Fungsi Huber memberikan pembobot sebesar 1 untuk $|u_i| \le c$ dan mengecil pada $|u_i| > c$ .Pada fungsi Tukey bisquares, diberi pembobot nol ketika $|u_i| > c$ pembobotnya mengecil dengan segera setelah u_i beranjak dari nol.

Tabel 2.2 Perbandingan Fungsi Huber dan Fungsi Tukey bisquare

Metode	Huber	Tukey Bisquare	Interval

Fungsi objektif	$\rho(u_i)=\begin{cases}\frac{1}{2}u_i^{2}\\ c\lvert u_i\rvert-\frac{1}{2}c^2\end{cases}$	$\rho(u_i)=\begin{cases}\frac{c^2}{6}\left\{1-\left[1-\left(\frac{u_i}{c}\right)^2\right]^3\right\}\\ \frac{c^2}{6}\end{cases}$	$\lvert u_i\rvert\le c$ $\lvert u_i\rvert> c$
Fungsi Pembobot	$w(u_i)=\begin{cases}1\\ \frac{c}{\lvert u_i\rvert}\end{cases}$	$w(u_i)=\begin{cases}\left[1-\left(\frac{u_i}{c}\right)^2\right]^2\\ 0\end{cases}$	$\lvert u_i\rvert\le c$ $\lvert u_i\rvert> c$

Nilai c untuk estimator Huber dan Tukey Bisquare disebut *tuning constan*. Semakin kecil nilai c menghasilkan lebih resisten terhadap *outlier*. Estimasi-M mempunyai effisiensi sekitar 95% ketika residual berdistribusi normal. ”Untuk bobot huber nilai c = 1,345 dan untuk bobot bisquare nilai c= 4,685 ”(Fox, 2002).

2.10 Kerangka Berpikir

Tujuan dalam analisis regresi linear adalah mengestimasi koefisien regresi dalam model. Pada umumnya digunakan metode estimasi kuadrat terkecil atau *Ordinary Least Square* (OLS) *Method* untuk mengestimasi koefisien regresi dalam model regresi. Namun metode ini sangat sensitif terhadap kehadiran *outlier*. Hasil estimasi koefisien garis regresi dengan estimasi OLS menjadi tidak tepat jika terdapat dalam data atau observasi terdapat *outlier*.

Pendeteksian *outlier* merupakan tahapan diagnosis yang perlu dilakukan terutama jika estimasi modelnya dengan metode kuadrat terkecil, yang dikenal cukup peka terhadap *outlier*. Metode pendeteksian *outlier* dilakukan dengan beberapa metode, antara lain metode *boxplot*, *Leverage Value*, *Cook's Distance* dan *Standardized Residual*. Jika dalam tahapan pendeteksian *outlier* tidak terdapat outlier maka estimasi model dengan metode kuadrat terkecil diterima, tetapi apabila terdapat *outlier* maka diperlukan suatu metode yang bersifat robust terhadap keberadaan *outlier* .

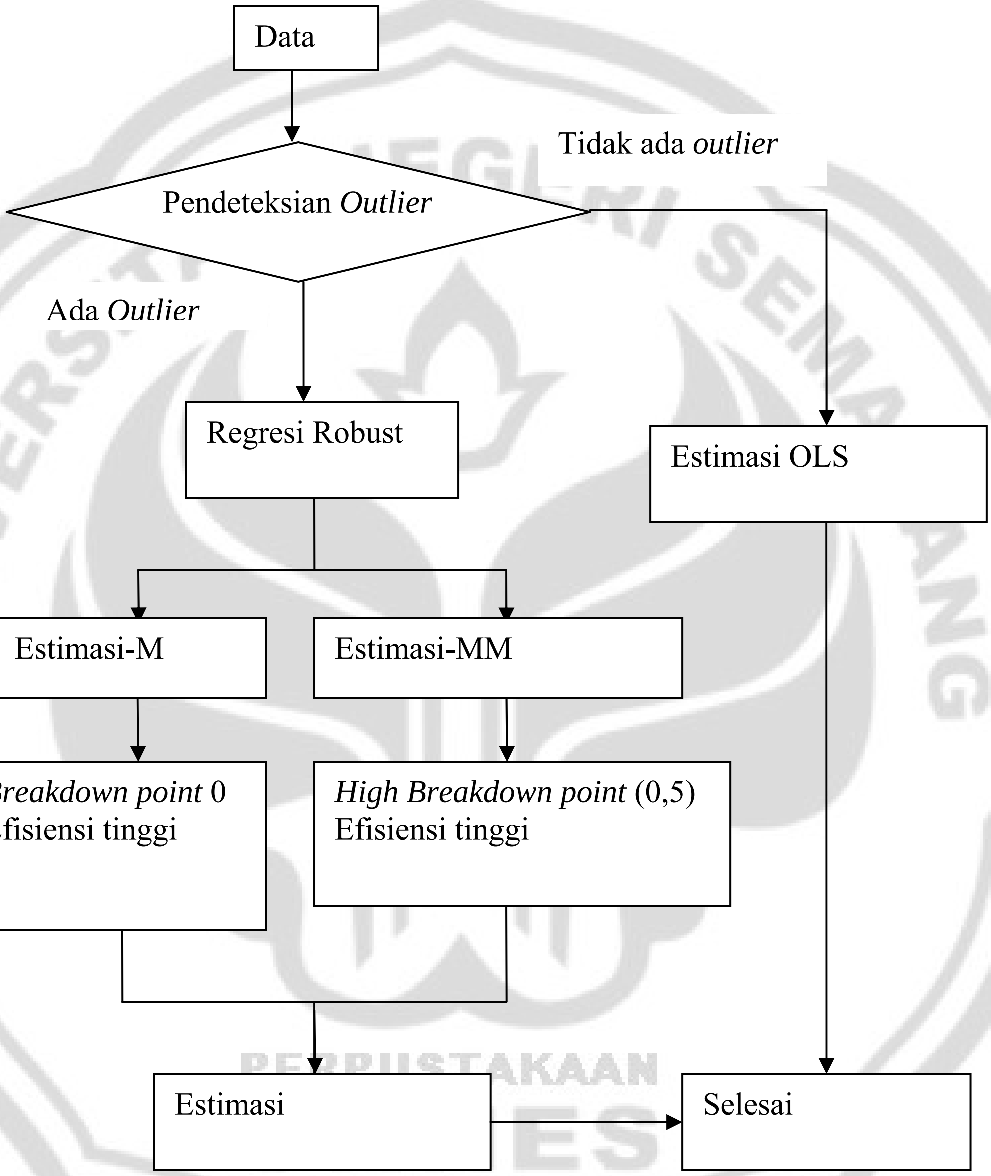
Metode regresi robust merupakan salah satu cara untuk mengatasi kelemahan OLS terhadap *outlier* pada sekumpulan data. Regresi robust menghasilkan estimasi model yang resisten terhadap pengaruh-pengaruh *outlier* dari observasi-observasi yang memuat *outlier*. Dalam regresi *robust* terdapat beberapa metode estimasi, antara lain adalah estimasi-M, *Least Median of Squares* (LMS), *Least Trimmed Squares* (LTS), Estimasi-S, dan Estimasi-MM. Kelima metode regresi robust tersebut mempunyai kelemahan dan kelebihan masing-masing. Disini peneliti memilih metode regresi robust M dan MM dan membandingkan kedua metode tersebut.

Estimasi-M merupakan suatu teknik *robust* yang luas dan terkenal daripada metode regresi *robust* yang lain. Estimasi-M mempunyai *breakdown point* sebesar 0. Karena estimasi-M mempunyai *breakdown point* sebesar 0 maka estimasi-M tidak bekerja dengan baik untuk mengestimasi parameter pada data yang terdapat *outlier* pada variabel prediktor. Disamping mempunyai *breakdown point* sebesar 0, estimasi-M juga mempunyai efisiensi yang tinggi sebesar (95%).

Berbeda dengan estimasi-M, metode estimasi-MM menggabungkan estimasi nilai high breakdown (50%) dengan efisiensi tinggi (mencapai 95%). Jadi estimasi-MM mempunyai nilai breakdown point yang tinggi sebesar 0,5 dan efisiensi yang tinggi. Nilai *breakdown point* yang tinggi pada estimasi-MM menyebabkan estimasi-MM bekerja dengan baik untuk mengestimasi parameter pada data yang terdapat *outlier* pada variabel prediktor maupun respon. Disamping melihat perbedaan pada nilai efek nilai *breakdown*nya, perbandingan keefektifan kedua fungsi tersebut pada regresi linear dapat dilihat dari standar error.

Apabila standar error yang dihasilkan dengan metode robust dapat memperkecil standar error yang dihasilkan dengan metode OLS, maka metode regresi robust dapat mengatasi permasalahan *outlier* pada OLS.

Tabel 2.3 Kerangka Berfikir



2.11 Hipotesis

Dari kerangka berpikir di atas dapat dibuat hipotesis penelitian yaitu regresi robust dapat mengatasi permasalahan OLS terhadap data atau observasi yang terdapat *outlier* dan regresi robust Estimasi-MM lebih efektif daripada Estimasi-M.

BAB III

METODE PENELITIAN

Pada penelitian ini, metode penelitian yang penulis gunakan adalah metode studi pustaka. Langkah-langkah yang dilakukan adalah sebagai berikut:

3.1 Penentuan Masalah

Dalam tahap ini dilakukan pencarian sumber-sumber pustaka yang relevan pustaka untuk mengumpulkan informasi yang diperlukan dan memilih bagian dalam sumber tersebut yang dapat dijadikan sebagai permasalahan. Permasalahan yang muncul di sini adalah tentang *outlier*.

3.2 Perumusan Masalah

Perumusan masalah dimaksudkan untuk membatasi permasalahan sehingga diperoleh bahan kajian yang jelas. Dan selanjutnya dirumuskan permasalahan sebagai berikut.

(5) Bagaimana kriteria metode-metode yang dapat dipergunakan dalam mendeteksi keberadaan *outlier*?

(6) Bagaimana hasil model regresi robust dengan metode estimasi-M?

(7) Bagaimana hasil model regresi robust dengan metode estimasi-MM?

(8) Metode manakah yang lebih baik antara estimasi-M dan estimasi-MM jika ditinjau dari nilai efek *breakdown point* dan standar error?

3.3 Studi Pustaka

Dalam tahap ini dilakukan kajian-kajian sumber pustaka dengan mengumpulkan data atau informasi yang berkaitan dengan permasalahan, mengumpulkan konsep pendukung seperti definisi dan teorema serta membuktikan teorema-teorema untuk menyelesaikan permasalahan, sehingga didapat ide mengenai bahan dasar pengembangan upaya pemecahan masalah.

3.4 Analisis dan Pemecahan Masalah

Tahap ini dimaksudkan untuk memberikan solusi-solusi dari permasalahan yang telah ditentukan seperti yang telah dikemukakan di atas.

Analisis dan pemecahan masalah dilakukan dengan langkah-langkah sebagai berikut.

b. Mengidentifikasi dan mengumpulkan materi-materi prasyarat yang nantinya digunakan sebagai pedoman dalam menganalisis data dengan metode robust-M dan metode robust-MM.

c. Mengambil data sekunder, untuk selanjutnya dilakukan estimasi model regresi dengan metode kuadrat terkecil.

d. Melakukan pendeteksian *outlier* pada data tersebut.

e. Mengatasi permasalahan *outlier* pada OLS dengan metode regresi robust estimasi-M dan estimasi-MM.

f. Membandingkan hasil model regresi robust yang terbentuk dari kedua metode robust tersebut.

g. Membandingkan efek nilai *breakdown point* dan standar eror regresi robust dari metode estimasi-M dan metode estimasi-MM.

3.5 Penarikan Simpulan

Tahap ini merupakan tahap akhir dalam penelitian. Penarikan simpulan dari permasalahan yang dirumuskan berdasarkan studi pustaka dan pembahasannya.

BAB IV
HASIL PENELITIAN DAN PEMBAHASAN

4.1 Regresi Robust Estimasi-M dan Estimasi-MM untuk Permasalahan *Outlier* pada OLS

Estimasi parameter regresi linear bertujuan untuk menjelaskan pengaruh satu atau lebih variabel x_i terhadap variabel respon y_i. Metode estimasi yang sering digunakan adalah *Ordinary Least Squares* (OLS). Akan tetapi OLS sangat sensitif terhadap *outlier*. Terdapatnya *outlier* dalam suatu data pengamatan mengakibatkan koefisien garis regresi yang dihasilkan dengan OLS tidak tepat. Sehingga kita mungkin berfikir secara gampang untuk membuang *outlier*, kemudian menganalisis kembali tanpa *outlier*. Akan tetapi, pengikutsertaan atau penyisihan *outlier* bukan masalah sederhana, tetapi butuh pertimbangan yang sangat hati-hati. *Outlier* dapat dibuang apabila setelah ditelusuri data *outlier* tersebut bukan bagian representatif dari data pengamatan (data *outlier* diperoleh dari kesalahan teknis peneliti dalam mencatat data). Namun secara statistik, membuang *outlier* bukanlah tindakan yang bijaksana, karena suatu *outlier* dapat memberikan informasi yang cukup berarti. Oleh karena itu, diperlukan suatu alternatif terhadap keberadaan *outlier*, yaitu dengan regresi robust.

Sebelum dilakukan analisis dengan regresi robust, sebaiknya dilakukan pendeteksian *outlier* untuk mengidentifikasi adanya *outlier* atau tidak. Metode pendeteksian *outlier* dilakukan dengan beberapa metode, antara lain metode *boxplot*, *Leverage value*, *Cook's Distance*, dan *Standardized residual*. Jika dideteksi terdapat data *outlier*, maka dapat digunakan regresi *robust*.

Regresi robust merupakan metode yang dapat menganalisis data yang mengandung *outlier* dan menghasilkan estimasi model yang resisten terhadap *outlier*. Dalam regresi robust terdapat beberapa metode estimasi, antara lain adalah Estimasi-M, *Least Median of Squares* (LMS), *Least Trimmed Squares* (LTS), Estimasi-S, dan Estimasi-MM. Kelima metode regresi robust tersebut mempunyai kelemahan dan kelebihan masing-masing. Estimasi-M mempunyai efisiensi yang tinggi, tetapi nilai *breakdown point* = 0. LMS, LTS, dan Estimasi-S mempunyai *breakdown point* yang tinggi (BDP = 0,5), akan tetapi efisiensinya sangat rendah. Sedangkan estimasi-MM merupakan gabungan efisiensi tinggi dari estimasi-M dengan *breakdown point* tinggi dari Estimasi-S.

Efisiensi dan *breakdown point* digunakan untuk menjelaskan ukuran kerobust-an dari tehnik robust. Efisiensi menjelaskan seberapa baiknya suatu tehnik robust sebanding dengan *Least Square* tanpa *outlier*. *Breakdown point* adalah suatu ukuran kestabilan dari estimator ketika data observasi mengandung *outlier* dalam jumlah besar. Semakin tinggi effisiensi dan *brekdown point* dari suatu estimator maka semakin robust (resisten) terhadap *outlier*. Pada penelitian ini, penulis hanya menggunakan metode regresi robust dengan estimasi-M dan estimasi-MM. Pemilihan kedua metode tersebut karena estimasi-M merupakan suatu teknik robust yang luas dan terkenal.

Dalam membandingkan keefektifan kedua metode regresi robust tersebut, penulis membandingkan kedua metode robust tersebut dengan OLS. Secara statistik, apabila regresi robust dapat mengecilkan standar error yang dihasilkan dengan OLS dengan adanya *outlier*, maka dapat disimpulkan regresi robust dapat menghasilkan model yang resisten terhadap pengaruh *outlier*. Sehingga metode regresi robust dapat menjadi solusi permasalahan OLS terhadap data observasi yang terdapat *outlier*. Alternatif lain juga bisa dilakukan dengan membandingkan kedua metode regresi robust tersebut dengan OLS tanpa *outlier*.

Apabila hasil standar error yang dihasilkan metode regresi robust hampir sama dengan standar error yang dihasilkan OLS tanpa *outlier*, maka dapat disimpulkan bahwa regresi robust sama baiknya dengan OLS yang tidak ada *outlier*.

Selain melihat standar error dari kedua metode tersebut, akan dilihat pula nilai *breakdown point*nya. *Breakdown point* adalah fraksi terkecil dari *outlier* yang dapat mengakibatkan nilai suatu estimator menjadi besar.

Perbandingan Estimasi-M dan Estimasi-MM dilihat dari nilai *breakdown point* nya sebagai berikut.

4.1.1 *Breakdown Point* Estimasi-M

Fungsi ρ memberikan setiap residual terhadap fungsi objektif. Oleh karena itu, Fungsi ρ harus mempunyai sifat sebagai berikut.

(1) Selalu non negatif, $\rho(e) \geq 0$;

(2) $\rho(0) = 0$;

(3) *Symmetric*, $\rho(e) = \rho(-e)$; dan

(4) Monoton di $|e_i|, \rho(e_i) \geq \rho(e_i')$ untuk $|e_i| \geq |e_i'|$.

Menurut Huber (1981:52), Estimasi-M mempunyai fungsi lokasi $\psi(x;t) = \psi(x-t)$ dengan fungsi ψ monoton meningkat.

Breakdown point adalah fraksi terkecil atau persentase dari *outlier* yang dapat menyebabkan nilai estimator menjadi besar. *Breakdown point* untuk sebuah estimator T di F didefinisikan sebagai:

$$\varepsilon^* = \varepsilon^*(F_0, T) = sup\{\varepsilon;\ b(\varepsilon) < b(1)\},$$

dengan $b(1) = \infty$.

Untuk menghitung *breakdown point* berhubungan dengan maximum bias b_1 dengan $b_1(\varepsilon) = \sup|\mathrm{T(F)} - \mathrm{T(F_0)}|$.

Ambil $b_+(\varepsilon) = sup\{T(F)|d_i(F_0, F \le \varepsilon\}$; dan

$$b_-(\varepsilon) = inf\{T(F)|d_i(F_0, F \le \varepsilon\};$$

sehingga $b_1(\varepsilon) = max\{b_+(\varepsilon), (b_-(\varepsilon)\}$.

Fungsi lokasi Estimasi-M didefinisikan T(F) = $\int \psi(x - T(F))F(dx) = 0$

Maka $\lambda(t;F) = \int \psi(x - T(F))F(dx)$

karena ψ monoton, maka $T^*(F) \le T^{**}(F)$ dengan

$T^*(F) = sup\{t|\lambda(t;F) > 0\}$,

dan $T^{**}(F) = inf\{t|\lambda(t;F) < 0\}$,

Anggota terbesar dari himpunan $d_1(F_0, F) \le \varepsilon$ adalah kebalikan distribusi F_1:

$F_1(x) = (F_0(x-\varepsilon) - \varepsilon)^+$;

dimana $F_1(x) = 0$, untuk $x \le x_0 + \varepsilon$,

$F_1(x) = F_0(x-\varepsilon) - \varepsilon$, untuk $x > x_0 + \varepsilon$,

dengan nilai x_0 tetap diperoleh

$F_0(x_0) = \varepsilon$

Jadi

$$\lambda(t;F) \leq \lambda(t;F_1) = \int_{x_0}^{\infty} \psi(x-t+\varepsilon)F_0(dx) + \varepsilon\psi(\infty),$$

didefinisikan $b_+(\varepsilon) = inf\{t|\lambda(t;F_1) < 0\}$ dan $b_-(\varepsilon) = sup\{t|\lambda(t;F_1) > 0\}$

Menurut Huber (1981:53) F_0 *symmetric* sehingga

$$b_1(\varepsilon) = b_+(\varepsilon) = -b_-(\varepsilon).$$

Perlu diperhatikan $b_+(\varepsilon) < \infty$ jika $\lim_{t\to\infty} \lambda(t;F_1) < 0$,

dan $b_+(1) = \infty$

karena $\lim_{t\to\infty} \lambda(t;F_1) = (1-\varepsilon)\psi(-\infty) + \varepsilon\psi(+\infty) < 0$

sehingga diperoleh $b_+(\varepsilon) < \infty$.

Maka dapat disimpulkan $b_+(\varepsilon) < b_+(1) = \infty$,

Persamaan $(1-\varepsilon)\psi(-\infty) + \varepsilon\psi(+\infty) < 0$

$$\Leftrightarrow \frac{\varepsilon}{1-\varepsilon} < \frac{-\psi(-\infty)}{\psi(\infty)}$$

Jadi untuk meghindari *breakdown* di sisi kanan, seharusnya mempunyai $\frac{\varepsilon}{1-\varepsilon} < \frac{-\psi(-\infty)}{\psi(\infty)}$. Jika kita juga mengambil sisi kiri ke dalam perhitungan, maka dihasilkan *breakdown point* sebesar

$$\varepsilon^* = \frac{\eta}{1+\eta}$$

dengan $\eta = min\left\{-\frac{\psi(-\infty)}{\psi(+\infty)}, -\frac{\psi(+\infty)}{\psi(-\infty)}\right\}$;

nilai breakdown terbaik sebesar $\varepsilon^* = \frac{1}{2}$ jika $\psi(-\infty) = -\psi(+\infty)$, dan jika ψ tidak dibatasi (*unbounded*) maka $\varepsilon^* = 0$.

Karena Estimasi-M fungsi ψ (fungsi pengaruh) tidak dibatasi (*unbounded)* maka *breakdown point* Estimasi-M sebesar 0.

Dari uraian penjelasan di atas dapat diperoleh teorema sebagai berikut.

Teorema 4.1

Diketahui ψ suatu fungsi monoton naik, tetapi tidak kontinu, fungsi mempunyai dua tanda. Kemudian Estimasi-M didefinisikan oleh $\int \psi(x - T(F))F(dx) = 0$, adalah fungsi kontinu tidak kuat pada F_0 jika ψ tidak dibatasi. *Breakdown point* $\varepsilon^* = \frac{\varepsilon}{1+\varepsilon}$ dan nilai *breakdown point* yang dihasilkan sebesar 0.

Karena *breakdown point*nya sebesar 0 maka Estimasi-M resistan terhadap *outlier* pada variabel respon, tetapi tidak resisten terhadap variabel prediktor.

4.1.2 *Breakdown Point* Estimasi-MM

Estimasi-MM mempunyai sifat Fungsi ρ antara lain:.

(1) $\rho(0) = 0$ dan ρ fungsi kontinu.

(2) *Symmetric*, $\rho(e) = \rho(-e)$;

(3) $0 \leq e \leq v$ berakibat $\rho(e) \leq \rho(v)$;

(4) Ambil $a = \sup \rho(e),\ maka\ 0 < a < \infty$; dan

(5) Jika $\rho(e) < a\ dan\ 0 \leq e < v, maka\ \rho(e) < \rho(v)$.

Breakdown point adalah fraksi terkecil atau persentase dari *outlier* yang dapat menyebabkan nilai estimator menjadi besar. *Breakdown point* untuk sebuah estimator T di F didefinisikan sebagai:

$$\varepsilon^* = \varepsilon^*(F_0, T) = sup\{\varepsilon;\ b(\varepsilon) < b(1)\},$$

dengan $b(1) = \infty$.

Untuk menghitung *breakdown point* berhubungan dengan maximum bias b_1 dengan $b_1(\varepsilon) = \sup|\mathrm{T(F)} - \mathrm{T(F_n)}|$.

Ambil $b_+(s) = sup\{T(F) | d_1(F_0, F \leq s\}$; dan

$$b_-(s) = inf\{T(F) | d_1(F_0, F \leq s\};$$

sehingga $b_1(s) = max\{b_+(s), (b_-(s)\}$.

Akibatnya $b_1(s) = b(s)$ merupakan titik kontinu dari b_1.

Fungsi lokasi Estimasi-MM didefinisikan $T(F) = \int \psi\left(\frac{x-T(F)}{s}\right) F(dx) = 0$

Karena Estimasi-MM merupakan gabungan dari Estimasi-M dan suatu skala estimator maka persamaan untuk estimasi-MM adalah pasangan dari statistik (T, S) yang didefinisikan ke dalam dua bentuk persamaan:

$$T(F) = \int \psi\left(\frac{x-T(F)}{s}\right) F(dx) = 0,$$

$$S(F) = \int \chi\left(\frac{x-T(F)}{S}\right) F(dx) = 0$$

Sehingga $\lambda(t; F) = \int \psi\left(\frac{x-T(F)}{s}\right) F(dx)$,

Dan $\lambda(s; F) = \int \chi\left(\frac{x-T(F)}{s}\right) F(dx)$

karena ψ monoton, maka $T^*(F) \leq T^{**}(F)$ dengan

$T^*(F) = sup\{t | \lambda(t; F) > 0\}$,

dan $T^{**}(F) = inf\{t | \lambda(t; F) < 0\}$,

Anggota terbesar dari himpunan $d_1(F_0, F) \leq s$ adalah kebalikan distribusi F_1:

$F_1(x) = (F_0(x - s) - s)^+$;

dimana $F_1(x) = 0$, untuk $x \leq x_0 + s$.

$F_1(x) = F_0(x - s) - s$, untuk $x > x_0 + s$,

dengan nilai x_0 tetap diperoleh

$F_0(x_0) = s$

Ambil $\{F\}$ adalah sebuah barisan ε *outlier* dan $F = (1-\varepsilon)F_0 + \varepsilon H$, sehingga $T(F) \to \infty$, $S(F) \to \infty$.

Asumsikan bahwa $0 \leq lim\frac{T(F)}{S(F)} = y \leq \infty,$

dan $\varepsilon_s^* \leq \varepsilon_t^*$

Persaman T(F) dan S(F) menjadi

$$(1-\varepsilon)\int \psi\left(\frac{x-t}{s}\right)F_0(dx) + \varepsilon\int \psi\left(\frac{x-t}{s}\right)H(dx) = 0$$

$$(1-\varepsilon)\int \chi\left(\frac{x-t}{s}\right)F_0(dx) + \varepsilon\int \chi\left(\frac{x-t}{s}\right)H(dx) = 0$$

Jika koefisien dari ε diganti dengan $\psi(\infty)$ dan $\chi(\infty)$ maka

$$\lambda(t;F_1) = (1-\varepsilon)\int \psi\left(\frac{x-t}{s}\right)F_0(dx) + \varepsilon\psi(\infty) < 0,$$

$$\lambda(s;F_1) = (1-\varepsilon)\int \chi\left(\frac{x-t}{s}\right)F_0(dx) + \varepsilon\psi(\infty) < 0.$$

maka

$$\lambda(t;F) \leq \lambda(t;F_1) \text{ dan } \lambda(s;F) \leq \lambda(s;F_1)$$

Didefinisikan $b_+(\varepsilon) = inf\{t|\lambda(t;F_1) < 0\}$ dan $b_+(\varepsilon) = inf\{t|\lambda(s;F_1) < 0\}$.

Sedangkan $b_-(\varepsilon) = sup\{t|\lambda(t;F_1) > 0\}$ dan $b_-(\varepsilon) = sup\{t|\lambda(s;F_1) > 0\}$

Menurut Huber (1981:53) Γ_0 *symmetric*

Sehingga $b_1(\varepsilon) = b_+(\varepsilon) = \quad b_-(\varepsilon)$.

Dalam persaman limit menjadi

$$\lim_{r\to\infty} \lambda(t;F_1) = (1-\varepsilon)\psi(-y) + \varepsilon\psi(\infty) < 0,$$

$$\lim_{r\to\infty} \lambda(s;F_1) = (1-\varepsilon)\chi(-y) + \varepsilon\chi(\infty) < 0,$$

Penggunanaan sifat ψ *dan* χ yang monoton dan simetri maka persaman di atas menjadi

$$\chi^{-1}\left(-\frac{s}{1-s}\chi(\infty)\right) \leq y \leq \psi^{-1}\left(\frac{s}{1-s}\psi(\infty)\right)$$

Sehingga

$$\chi^{-1}\left(-\frac{s}{1-s}\chi(\infty)\right) = \psi^{-1}\left(\frac{s}{1-s}\psi(\infty)\right)$$

Akibatnya

$b_+(s_s) = b_+(s_t) = b_+(s)$

Dan *breakdown point* $s_s^* = s_t^* = s^*$

Karena $\lim_{t\to\infty} \lambda(t;F_1) < 0$ dan $\lim_{t\to\infty} \lambda(s;F_1) < 0$,

Perlu diperhatikan $b_+(s) < \infty$ jika $\lim_{t\to\infty} \lambda(t;F_1) < 0$,

dan $b_+(1) = \infty$

Maka dapat disimpulkan $b_+(s) < b_+(1) = \infty$ dan $b_+(s_s) < b_+(1) = \infty$,

Persamaan $\lim_{t\to\infty} \lambda(t;F_1)$ merupakan solusi breakdown point Estimasi-MM dihasilkan *breakdown point* sebesar

$$s^* = \frac{\eta}{1+\eta}$$

dengan $\eta = min\left\{-\frac{\psi(-\infty)}{\psi(+\infty)}, -\frac{\psi(+\infty)}{\psi(-\infty)}\right\}$;

Berdasarkan teorema 4.1, nilai breakdown terbaik sebesar $s^* = \frac{1}{2}$ jika $\psi(-\infty) = -\psi(+\infty)$, dan jika ψ tidak dibatasi (*unbounded*) maka $s^* = 0$.

Karena Estimasi-MM fungsi ψ (fungsi pengaruh) dibatasi (*bounded)* maka $\psi(-\infty) = -\psi(+\infty)$.

$$\eta = min\left\{-\frac{\psi(-\infty)}{\psi(+\infty)}, -\frac{\psi(+\infty)}{\psi(-\infty)}\right\}$$

$$\eta = min\{1,1\} = 1$$

$$\varepsilon^* = \frac{\eta}{1+\eta} = \frac{1}{1+1} = \frac{1}{2}$$

Jadi *breakdown point* Estimasi-M sebesar 0,5.

Karena *breakdown point*nya sebesar 0 maka Estimasi-M resistan terhadap *outlier* pada variabel respon, tetapi tidak resisten terhadap variabel prediktor.

4.2 Contoh Kasus

Dalam penelitian ini mengambil simulasi pada suatu kasus dengan mengggunakan data dari BPS (Badan Pusat Statistik) Provinsi Jawa Tengah. yaitu data tentang produksi padi yang ada pada tiap kabupaten/kota di Provinsi Jawa Tengah tahun 2007. Data yang digunakan dalam permasalahan ini ada 3 variabel, dimana variabel-variabel tersebut meliputi produksi padi di Jawa Tengah sebagai variabel dependen sedangkan luas panen dan jumlah penduduk sebagai variabel independen. Sedangkan obyek yang digunakan terdiri dari 29 kabupaten dan 6 kota di Jawa Tengah yaitu Kabupaten Cilacap, Banyumas, Purbalingga, Banjarnegara, Kebumen, Purworejo, Wonosobo, Magelang, Boyolali, Klaten, Sukoharjo, Wonogiri, Karanganyar, Sragen, Grobogan, Blora, Rembang, Pati, Kudus, Jepara, Demak, Semarang, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Brebes, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan, dan KotaTegal. Data dapat dilihat pada lampiran 1.

Proses analisis regresi robust dimulai dengan regresi kuadrat terkecil terlebih dahulu, kemudian pengidentifikasian *outlier* dan selanjutnya dengan metode regresi robust estimasi-M dan estimasi-MM. Pengolahan data komputasi yang digunakan sebagai alat bantu adalah program SPSS 16 dan SAS 9.

4.2.1 Metode Kuadrat Terkecil

Analisis dimulai dengan menganalisis regresi biasa menggunakan metode kuadrat terkecil. Berdasarkan hasil output tabel 4.1, diperoleh model regresi antara variabel independen dan variabel dependen data produksi padi di Jawa Tengah tahun 2007 sebagai berikut

$$\hat{Y}_i = -41{,}61685 + 0{,}10703X_1 + 298{,}18646X_2$$

dengan $\hat{Y}_i$ = produksi padi (ton)

X_1 = luas panen (hektar)

X_2 = jumlah penduduk (juta jiwa)

Model regresi tersebut mempunyai nilai R^2 sebesar 0,5332 = 53,32%. Tahapan selanjutnya adalah melakukan pendeteksian *outlier* untuk mengetahui ada atau tidaknya *outlier* dalam data observasi.

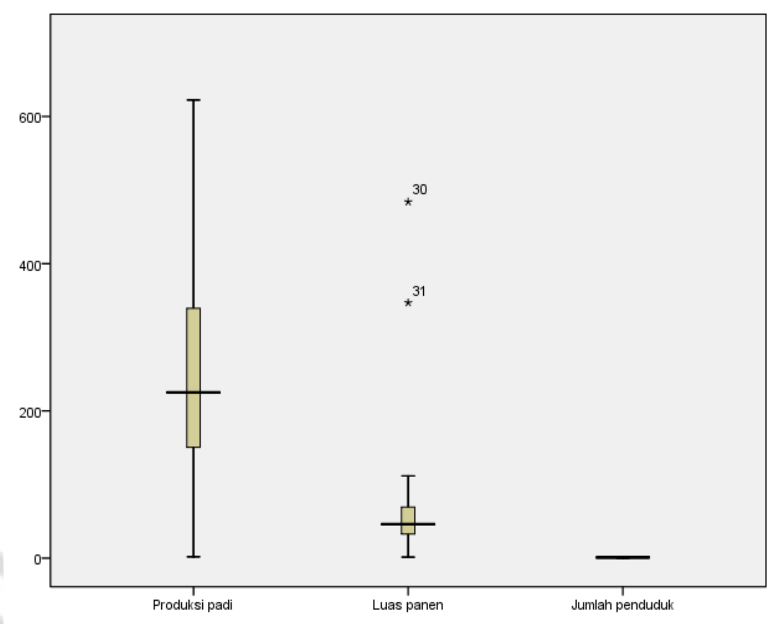
4.2.2 Pendeteksian *Outlier*

Suatu data diduga dan dinyatakan sebagai suatu *outlier* dapat dilakukan dengan berbagai macam metode. Beberapa metode diantaranya sebagai berikut.

4.2.2.1 Metode Boxplot

Metode *boxplot* dapat disajikan dalam bentuk perhitungan manual maupun dengan bantuan komputer. Hasil *Boxplot* dengan program SPSS dapat dilihat pada gambar 4.1. *Boxplot* dengan perhitungan manual dapat dilihat pada tabel 4.2.

Dengan metode *boxplot*, suatu data dikatakan *outlier* jika nilai data pengamatan lebih kecil dari Q1-(1,5*IQR) atau lebih besar dari Q3+(1,5*IQR). Berdasarkan tabel 4.2, terlihat bahwa pada variabel X_1 (luas panen) terdapat data *outlier*, yaitu pada data ke-30 (484>127,799) dan data ke-31(347>127,799). *Boxplot* yang disajikan dengan bantuan program SPSS 16 terlihat sebagai berikut



Gambar 4.1 *Boxplot* untuk ketiga variabel.

Dari ketiga tampilan *boxplot* di atas dapat terlihat bahwa data ke-30 dan 31 merupakan data *outlier*.

4.2.2.2 Metode Leverage Value

Untuk menentukan ada atau tidaknya *outlier*, dapat dilihat dari nilai leverage berikut ini. Nilai-nilai dari leverage yang melebihi nilai dari titik potong (cut off point) yaitu $\frac{2p-1}{n}$ dapat di indikasikan sebagai *outlier*.

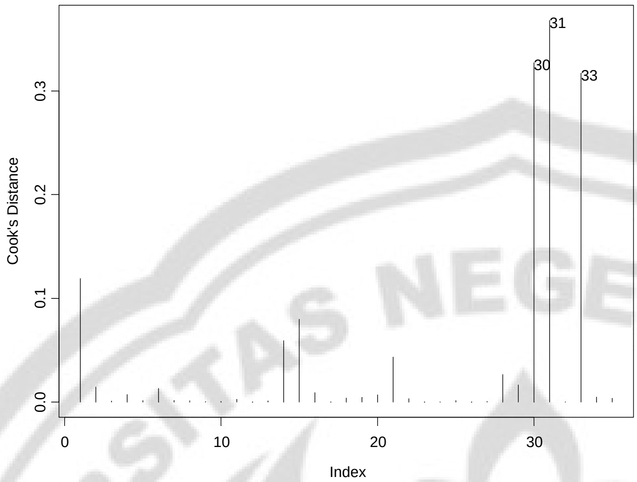
Hasil diagnosis metode *leverage* terhadap data itu sebagaimana dapat ditampilkan pada tabel 4.3, dengan p adalah banyaknya parameter dan n adalah banyaknya data diperoleh nilai (2p-1)/n= 5/35= 0,1429.

Berdasarkan hasil perhitungan, diperoleh bahwa data ke-30 dan 31 diduga sebagai *outlier* karena memiliki nilai *Leverage*>0,1429. Nilai leverage_{30} = 0,63695, nilai leverage_{31} = 0,27254.

4.2.2.3 Metode Cook's distance

Hasil diagnosis metode *cook's distance* pada tiap observasi terhadap data juga dapat dilihat pada tabel 4.4 maupun gambar 4.2. Suatu observasi diduga sebagai *outlier* apabila nilai C*ook's* distance>(4/*n*), dengan n adalah banyaknya data.

Berdasarkan hasil yang diperoleh pada tabel 4.4 , nilai *Cook's distance* pada data ke- 1, 30, 31 dan 33 diduga sebagai *outlier* karena memiliki nilai C*ook's distance* yang lebih besar dari nilai (4/*n*) = 0,11429.



Gambar 4.2 Hasil Plot Nilai *Cook's distance*

4.2.2.4 Standardized residual

Hasil diagnosis dengan standarisasi dalam nilai Z atau yang biasa disebut *Z-skore* pada tiap observasi terhadap data juga dapat dilihat pada tabel 4.5 Untuk melakukan identifikasi *outlier*, diperhatikan nilai-nilai dari *standardized residual*. Jika nilai dari *standardized residual* memiliki nilai yang lebih dari 3,5 atau kurang dari -3,5 maka data tersebut dikatakan sebagai data *outlier*.

Berdasarkan hasil pada tabel 4.5, diperoleh bahwa data ke-30 pada variabel X_1(luas panen) diduga sebagai *outlier* karena memiliki nilai z_{30} = 4,49740 > 3,5.

Dari hasil pengidentifikasian *outlier* dengan keempat metode di atas, diperoleh hasil yang tidak sama. Berbagai cara perhitungan yang ditawarkan diatas guna memberikan gambaran peneliti dalam mendeteksi *outlier*. Oleh karena itu, tergantung dari peneliti dalam memilih metode yang sesuai dengan kajian penelitian atau hal yang mendukung penelitian. Jadi dalam data luas produksi padi terdapat beberapa *outlier*.

Adanya *outlier* dalam data observasi mengakibatkan hasil estimasi koefisien garis regresi dengan metode kuadrat terkecil tidak tepat. Apabila tidak terdapat *outlier* maka hasil estimasi yang diperoleh dengan metode kuadrat terkecil tepat. Oleh karena itu, penulis menghilangkan *outlier* dari data observasi, kemudian menganalisis data tanpa *outlier* tersebut dengan metode kuadrat terkecil. Hal tersebut penulis lakukan untuk menilai berpengaruh atau tidaknya *outlier* dalam OLS. Dalam contoh kasus ini, penulis membuang data *outlier* observasi ke- ke- 1,30,31 dan 33, kemudian menganalisis data tanpa *outlier* dengan metode kuadrat terkecil.

Berdasarkan hasil output tabel 4.6, diperoleh model regresi sebagai berikut

$$\hat{Y}_i = -2{,}53173 + 5{,}50022X_1 + -6{,}42425X_2 .$$

Hasil estimasi model regresi tersebut diperoleh dengan OLS tanpa pengikutsertaan *outlier*. Pembuangan *outlier* dalam data pengamatan, ternyata dapat mengecilkan standar error yang diperoleh dengan OLS dengan data yang terdapat *outlier*. Hal ini berarti *outlier* sangat berpengaruh terhadap OLS karena dapat memperbesar standar error yang dihasilkan. Langkah untuk membuang *outlier* tersebut dari data pengamatan peneliti lakukan untuk menilai berpengaruh tidaknya *outlier* terhadap OLS. Di satu sisi membuang *outlier* dapat mengecilkan standar error. Akan tetapi di sisi lain pengikutsertaan *outlier* dapat memperbesar standar error. Pengikutsertaan atau penyisihan *outlier* bukan masalah sederhana, tetapi butuh pertimbangan yang sangat hati-hati. *Outlier* dapat dibuang apabila setelah ditelusuri data *outlier* tersebut bukan bagian representatif dari data pengamatan (data *outlier* diperoleh dari kesalahan teknis peneliti dalam mencatat data). Namun secara statistik, membuang *outlier* bukanlah tindakan yang bijaksana, karena suatu *outlier* dapat memberikan informasi yang cukup berarti. Oleh karena itu, diperlukan suatu alternatif terhadap keberadaan *outlier*, yaitu dengan regresi robust.

Analisis regresi robust memberikan solusi terhadap keberadaan *outlier*, dengan tetap memasukan *outlier* pada data dan menghasilkan model estimasi yang resisten terhadap pengaruh *outlier*. Secara statistik, apabila standar error yang dihasilkan regresi robust dapat mengecilkan standar error yang dihasilkan dengan OLS yang terdapat *outlier*, maka dapat disimpulkan regresi robust dapat menghasilkan model yang resisten terhadap pengaruh *outlier*. Alternatif lain juga dapat dilihat dari membandingkan standar error OLS tanpa *outlier* dengan regresi robust. Jika hasil estimasi atau standar error yang dihasilkan metode kuadrat terkecil tanpa *outlier* dengan regresi robust dengan adanya *outlier* hampir sama, maka dapat dikatakan hasil estimasi regresi robust sama baiknya dengan OLS ketika tanpa *outlier*. Sehingga regresi robust dapat mengatasi kelemahan OLS yang sangat peka terhadap kehadiran *outlier*. Disini peneliti menggunakan dua metode dalam regresi robust, yaitu metode estimasi-M dan metode estimasi-MM.

4.2.3 Regresi Robust Estimasi-M

Estimasi-M pertama kali diperkenalkan oleh Huber pada tahun 1973. Estimasi-M merupakan gabungan sifat efisiensi OLS dan ketahanan dari estimasi LAV (LAD). Sehingga Estimasi-M mempunyai efisiensi yang tinggi dan *breakdown point* yang sama dengan OLS, dan robust pada *outlier* pada variabel respon sama seperti estimasi LAD. Estimasi-M tersebut mempunyai *breakdown point* sebesar 0, sehingga mengakibatkan estimasi-M kurang resisten terhadap *outlier* pada variabel prediktor. Hasilnya estimasi M kurang robust terhadap pengaruh *bad leverage points*.

Berdasarkan hasil output estimasi-M pada tabel 4.7, diperoleh model regresi antara variabel independen dan variabel dependen data produksi padi di Jawa Tengah tahun 2007 sebagai berikut

$$\hat{Y}_i = -47{,}4168 - 0{,}0167X_1 + 301{,}3212X_2$$

Selain mengatasi *outlier*, regresi robust juga dapat mengidentifikasi *outlier*. Penilaian terdapatnya *outlier* pada variabel respon (*vertical outlier*), *outlier* pada variabel prediktor yang terdiri dari *good leverage points* dan *bad leverage points* dapat dilihat dari *hasil robust mcd distance* dan *robust residual*.

Berdasarkan hasil output tabel4.8, tidak terdapat *vertical outlier* dalam data. Estimasi-M mengidentifikasi observasi ke-14, 15, 21, dan 33 sebagai *bad leverage points* karena ke empat observasi tersebut mempunyai nilai residu robust besar dan jarak robust mcd besar. Sedangkan observasi ke-1, 30, dan 31 sebagai *good leverage points* karena mempunyai nilai residu robust kecil dan nilai jarak robust mcd besar. *Breakdown point* sebesar 0 mengakibatkan estimasi-M tidak resisten terhadap *outlier* pada variabel prediktor. Oleh karena dalam data tersebut terdapat beberapa *outlier* pada variabel prediktor khususnya *bad leverage points*, maka estimasi-M tidak bekerja dengan baik. Estimasi-M kurang resisten terhadap pengaruh *bad leverage points* sehingga standar error yang dihasilkan besar.

Model regresi tersebut mempunyai nilai R^2 sebesar 0,4971 = 49,71%. AICR dan BICR yang diperoleh dengan estimasi-M sebesar 60,801 dan 66,416. *Akaike Information Criterion Robust* (AICR) dan *Bayesian Information Criterion Robust* (BICR) digunakan untuk mengukur *goodness of fit* dalam sebuah model statistik. AICR dan BICR yang terendah menunjukan suatu model terbaik.

4.2.4 Regresi Robust Estimasi-MM

Metode estimasi-MM merupakan kombinasi antara kelas *High Breakdown Value* dengan estimasi-M, sehingga estimasi-MM mempunyai *breakdown point* sebesar 0,5 dan efisiensi yang tinggi pula. *Breakdown point 0,5* mengakibatkan estimasi-MM robust terhadap *outlier* pada variabel prediktor maupun respon.

Berdasarkan hasil output estimasi-MM di atas, diperoleh model regresi antara variabel independen dan variabel dependen data produksi padi di Jawa Tengah tahun 2007 sebagai berikut

$$\hat{Y}_i = -5{,}5555 + 5{,}5184 X_1 - 2{,}3668 X_2$$

Berdasarkan hasil output tabel 4.9, dapat dilihat bahwa estimasi-MM robust terhadap keberadaan *outlier*. Estimasi-MM dapat mengecilkan standar error yang dihasilkan metode OLS yang terdapat *outlier*. Penilaian terhadap estimasi-MM dapat juga dilihat dari standar error yang dihasilkan estimasi MM dengan pengikutsertaan *outlier* hasilnya hampir sama dengan metode OLS dengan membuang *outlier*. Jadi hasil estimasi-MM sama baiknya dengan OLS ketika tidak ada *outlier*. Jadi estimasi-MM dapat mengatasi kelemahan metode OLS yang sangat peka terhadap *outlier*.

Berdasarkan output tabel 4.10 terlihat tidak terdapat *vertical outlier*. Estimasi-MM mengidentifikasi observasi ke-30 dan 31 *bad leverage points* karena mempunyai nilai residu robust besar dan jarak robust mcd besar. Sedangkan observasi ke- 1, 14, 15, 21, dan 33 merupakan *good leverage points* karena mempunyai nilai residu robust kecil dan nilai jarak robust mcd besar.

Breakdown point 0,5 mengakibatkan estimasi-MM robust terhadap *outlier* pada variabel prediktor maupun respon. Hal ini dapat dilihat dari standar error yang dihasilkan dengan estimasi-MM kecil.

Model regresi tersebut mempunyai nilai R^2 sebesar 0,7848 = 78,48%. AICR dan BICR yang diperoleh dengan estimasi-MM sebesar 28,534 dan 36,60.

Dari perbandingan antara hasil estimasi OLS tanpa *outlier* dengan kedua metode robust yang dapat dilihat pada tabel 4.11, terlihat bahwa hasil estimasi robust M sangat berbeda jauh dengan OLS tanpa *outlier*. Hal ini dikarenakan adanya *outlier* pada variabel prediktor (*leverage point*), sehingga hasil estimasi dengan robust M tidak tepat. Sedangkan hasil estimasi robust MM hampir sama dengan hasil estimasi OLS tanpa *outlier*. Sehingga adanya *outlier* tidak mempengaruhi hasil estimasi robust MM.

Berdasarkan tabel 4.12 terlihat standar error untuk OLS tanpa *outlier* kemudian dengan adanya *outlier* meningkat untuk semua parameter regresi. Hal ini menunjukan bahwa *outlier* sangat berpengaruh terhadap OLS. Sedangkan hasil standar error untuk dengan estimasi-M juga besar, namun dapat mengecilkan standat error dari OLS dengan adanya *outlier*. Dalam kasus ini berarti estimasi-M kurang resisten untuk menangani *outlier* pada variabel prediktor. Sedangkan standar error dengan estimasi-MM lebih kecil jika dibandingkan dengan standar error darri OLS tanpa *outlier* maupun OLS dengan pengikutsertaan *outlier*. Hal ini berarti estimasi-MM robust terhadap pengaruh *outlier* pada variabel prediktor maupun respon.

4.3 Pembahasan

Berdasarkan hasil penelitian, diperoleh pembahasan sebagai berikut

(1) Metode pendeteksian *outlier* dilakukan dengan beberapa metode, antara lain metode boxplot, *Leverage value*, *Cook's Distance* dan *Standardized residual*.

-Metode *boxplot* mempergunakan nilai kuartil dan jangkauan. Data *outlier* dapat ditentukan yaitu nilai yang kurang dari 1.5*IQR terhadap kuartil 1 dan nilai yang lebih dari 1.5*IQR terhadap kuartil atas. Metode *boxplot* mendeteksi *outlier* pada data observasi ke-30 dan 31.

-*Leverage Value* merupakan nilai pengaruh yang terpusat. Observasi yang mempunyai nilai *leverage* (h_{ii}) yang melebihi (2p-1)/n, dengan p adalah banyaknya variabel independen ditambah konstan dan n jumlah observasi maka akan mengindikasikan terdapat *outlier*. Metode *Leverage Value* mendeteksi *outlier* pada data ke-30 dan 31.

-*Cook's distance* merupakan suatu ukuran untuk mendeteksi besarnya pengaruh adanya pencilan terhadap semua estimasi koefisien regresi. Dengan D_i adalah nilai *Cook's distance*, suatu data disebut *outlier* apabila nilai $D_i > 4/n$. Metode *Cook's distance* mendeteksi *outlier* pada data ke-1, 30, 31, dan 33.

-*Standardized residual* mendeteksi *outlier* dengan memeriksa residualnya. nilai dari *standardized residual* yang lebih dari 3,5 atau kurang dari -3,5 maka data tersebut dikatakan sebagai data *outlier*. *Standardized residual* mendeteksi *outlier* pada data ke-30.

Berbagai cara perhitungan yang diberikan untuk memberikan gambaran peneliti dalam mendeteksi *outlier*. Pengualifikasian *outlier* dengan ke empat metode di atas hasilnya tidak sama. Oleh karena itu tergantung peneliti dalam memilih metode yang sesuai dengan kajian penelitian atau hal yang mendukung penelitian.

(2) Dalam menilai hasil model regresi robust-m, peneliti membandingkan hasil model regresi estimasi-M dengan OLS tanpa *outlier*. Apabila hasil model regresi robust tersebut hampir sama dengan hasil model regresi yang dihasilkan OLS tanpa *outlier* untuk semua parameter regresi, maka estimasi model yang dihasilkan regresi robust dengan estimasi-M tersebut sudah tepat.

Dari contoh kasus, hasil model regresi robust estimasi-M data produksi padi di Jawa Tengah tahun 2007 sebagai berikut

$$\hat{Y}_i = -47{,}4168 - 0{,}0167X_1 + 301{,}3212X_2$$

Hasil model regresi robust tersebut sangat berbeda jauh dengan hasil model regresi yang dihasilkan OLS tanpa *outlier* untuk semua parameter regresi. Sehingga estimasi model yang dihasilkan regresi robust dengan estimasi-M tersebut kurang tepat. Hal ini karena adanya *outlier* pada variabel prediktor (*leverage points*).

(3) Dalam menilai hasil model regresi robust estimasi-MM, peneliti membandingkan hasil model regresi estimasi-mm dengan OLS tanpa *outlier*. Apabila hasil model regresi robust tersebut hampir sama dengan hasil model regresi yang dihasilkan OLS tanpa *outlier* untuk semua parameter regresi, maka estimasi model yang dihasilkan regresi robust dengan estimasi-MM tersebut sudah tepat.

Dari contoh kasus, Hasil model regresi robust estimasi-MM data produksi padi di Jawa Tengah tahun 2007 sebagai berikut

$$\hat{Y}_i = -5,5555 + 5,5184 X_1 - 2,3668 X_2$$

Hasil model regresi robust tersebut hampir sama dengan hasil model regresi yang dihasilkan OLS tanpa *outlier* untuk semua parameter regresi. Sehingga estimasi model yang dihasilkan regresi robust dengan estimasi-M tersebut sudah tepat.

(4) -*Breakdown point* adalah suatu ukuran kestabilan dari estimator ketika data observasi mengandung *outlier* dalam jumlah besar. Semakin tinggi *brekdown point* dari suatu estimator maka semakin robust (resisten) terhadap *outlier*. OLS mempunyai *breakdown point* sebesar 0 sehingga OLS sangat peka terhadap *outlier* pada variabel prediktor maupun respon. Estimasi-M merupakan gabungan dari OLS dan estimasi LAD. Oleh karena itu, estimasi-M mempunyai *breakdown point* sama dengan OLS dan ketahanan pada *outlier* pada variabel respon sama seperti LAD. Jadi estimasi-M resisten untuk *outlier* pada variabel respon, akan tetapi kurang resisten terhadap *outlier* pada variabel prediktor. Hal ini dapat dilihat dari hasil nilai standar error estimasi-M yang besar karena adanya *outlier* pada variabel prediktor.

Sedangkan estimasi-MM merupakan gabungan dari kelas *high breakdown value* yang mempunyai *breakdown point* tinggi sebesar 0,5 dengan efisiensi yang tinggi dari estimasi-m. Jadi estimasi-MM mempunyai *breakdown point* tinggi dan effisiensi yang tinggi pula. *Breakdown point* sebesar 0,5 mengakibatkan estimasi-MM robust terhadap *outlier* pada *variabel* prediktor maupun respon. Hal ini dapat dilihat dari nilai standar error yang dihasilkan estimasi-MM yang kecil. Berdasarkan efek nilai *breakdown point*nya maka estimasi-MM lebih efektif daripada estimasi-M.

-Standar error adalah ukuran kebaikan model regresi, semakin kecil nilainya semakin baik model regresinya. Disini peneliti membandingkan standar error yang dihasilkan regresi robust dengan standar error yang dihasilkan OLS yang terdapat *outlier*. Secara statistik, apabila regresi robust dapat mengecilkan standar error yang dihasilkan OLS yang terdapat *outlier* maka regresi robust tersebut dapat mengatasi kelemahan OLS yang sangat peka terhadap *outlier*.

Dari contoh kasus, standar error semua parameter regresi baik intersep, X_1, maupun X_2 dari estimasi-M lebih kecil dibandingkan OLS yang terdapat *outlier*. Hal ini berarti estimasi-M dapat mengecilkan error yang dihasilkan OLS, meskipun perbedaan standar error nya kecil. Hal ini dikarenakan adanya *outlier* pada variabel prediktor sehingga estimasi-M tidak bekerja degan baik. Jadi estimasi-M merupakan suatu metode alternatif pengganti OLS yang dapat digunakan apabila terdapat *outlier* pada data.

Sedangkan standar error semua parameter regresi baik intersep, X_1, maupun X_2 dari estimasi-MM lebih kecil dibandingkan OLS yang terdapat *outlier*. Hal ini berarti estimasi-M dapat mengecilkan error yang dihasilkan OLS, dan perbedaan standar error nya cukup besar. Jadi estimasi-MM merupakan suatu metode alternatif pengganti OLS yang dapat digunakan apabila terdapat *outlier* pada data.

Berdasarkan kedua penjelasan di atas, maka dapat disimpulkan bahwa baik estimasi-M maupun estimasi-MM dapat mengatasi permasalahan *outlier* pada OLS. Keduanya dapat menganalisis data yang mengandung *outlier* dan menghasilkan model yang resisten terhadap pengaruh *outlier* dengan mengecilkan standar error yang dihasilkan OLS. Sehingga regresi robust estimasi-M dan estimasi-MM dapat digunakan sebagai jembatan alternatif antara mengabaikan *outlier* atau menghapus *outlier* dalam data pengamatan. Namun apabila ditinjau dari adanya *outlier* pada variabel prediktor, estimasi-M tidak bekerja sebaik estimasi-MM sehingga kurang efektif daripada estimasi-MM.

Hal ini diperkuat dengan contoh kasus pengaruh luas panen dan jumlah penduduk terhadap produksi padi di Jawa Tengah tahun 2007.

BAB 5
PENUTUP

5.1 Simpulan

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan sebagai berikut.

(5) Metode pendeteksian *outlier* dilakukan dengan beberapa metode, antara lain metode *boxplot*, *Leverage value*, *Cook's Distance* dan *Standardized residual*. --Untuk Metode *boxplot*, data *outlier* dapat ditentukan yaitu nilai yang kurang dari 1.5*IQR terhadap kuartil 1 dan nilai yang lebih dari 1.5*IQR terhadap kuartil atas.

-Untuk metode *Leverage Value*, Observasi yang mempunyai nilai *leverage* (h_{ii}) yang melebihi (2p-1)/n, dengan p adalah banyaknya variabel independen ditambah konstan dan n jumlah observasi maka akan mengindikasikan terdapat *outlier*.

-Untuk metode *Cook's distance,* dengan D_i adalah nilai *Cook's distance,* suatu data disebut *outlier* apabila nilai $D_i > 4/n$.

-Untuk metode *Standardized,* nilai dari *standardized residual* yang lebih dari 3,5 atau kurang dari -3,5 maka data tersebut dikatakan sebagai data *outlier*.

Berbagai cara perhitungan yang diberikan untuk memberikan gambaran peneliti dalam mendeteksi *outlier*. Pengualifikasian *outlier* dengan ke empat metode di atas hasilnya tidak sama. Oleh karena itu tergantung peneliti dalam memilih metode yang sesuai dengan kajian penelitian atau hal yang mendukung penelitian.

(6) Hasil model regresi robust estimasi-m data produksi padi di Jawa Tengah tahun 2007 sebagai berikut.

$$\hat{Y}_i = -47,4168 - 0,0167X_1 + 301,3212X_2.$$

Hasil model regresi dengan estimasi-M tersebut kurang robust, karena terdapat *outlier* pada variabel prediktor (*leverage points*).

(7) Hasil model regresi robust estimasi-MM data produksi padi di Jawa Tengah tahun 2007 sebagai berikut.

$$\hat{Y}_i = -5{,}5555 + 5{,}5184 X_1 - 2{,}3668 X_2 .$$

Hasil model regresi dengan estimasi-MM tersebut robust karena menghasilkan estimasi model yang hampir sama dengan OLS yang tidak ada *outlier*.

(8) - Estimasi-M mempunyai *breakdown point* sebesar 0 sehingga resisten untuk *outlier* pada variabel respon, akan tetapi kurang resisten terhadap *outlier* pada variabel prediktor. Estimasi-MM mempunyai *breakdown point* sebesar 0,5 sehingga estimasi-MM resisten terhadap *outlier* pada *variabel* prediktor maupun respon. Berdasarkan efek nilai *breakdown point*nya, estimasi-MM lebih efektif daripada estimasi-M.

-Dalam menilai hasil standar error regresi robust dengan membandingkan hasil standar error yang diperoleh dengan OLS. Apabila standar error regresi robust lebih kecil daripada OLS, maka regresi robust dapat sebagai metode alternatif tanpa harus membuang *outlier* dan dapat menghasilkan estimasi model yang resisten terhadap *outlier*. Dengan demikian regresi robust dapat mengatasi permasalahan OLS terhadap pengaruh *outlier*.

Dari contoh kasus, diperoleh hasil standar error semua parameter regresi baik intersep, X_1, maupun X_2 dari estimasi-M lebih kecil dibandingkan OLS dengan adanya *outlier*. Namun selisih standar error yang dihasilkan kecil. Hal ini dikarenakan terdapatnya *outlier* pada variabel prediktor. Standar error semua parameter regresi baik intersep, X_1, maupun X_2 dari estimasi-MM lebih kecil dibandingkan OLS dengan adanya *outlier* dan selisih standar errornya pun cukup besar.

Berdasarkan kedua penjelasan di atas, maka dapat disimpulkan bahwa baik estimasi-M maupun estimasi-MM dapat digunakan sebagai metode alternatif dalam pemecahan permasalahan *outlier* yang berpengaruh pada OLS. Namun apabila ditinjau dari adanya *outlier* pada variabel prediktor, estimasi-M kurang efektif daripada estimasi-MM.

5.2 Saran

a. Apabila menjumpai data *outlier* dalam data observasi, tidak perlu membuang *outlier* tersebut, karena regresi robust dapat menghasilkan model regresi yang resisten terhadap *outlier*.
b. Peneliti seyogyanya memilih metode yang digunakan untuk mendeteksi *outlier* sesuai dengan hal yang mendukung tujuan penelitian atau olahan data.
c. Sebaiknya mencoba lagi metode-metode estimasi regresi robust yang lain sebagai alternatif untuk mengatasi permasalahan *outlier* yang tidak dapat diselesaikan dengan OLS.
d. Untuk mempermudah dalam melakukan analisis regresi robust dapat digunakan beberapa paket program diantaranya adalah S-Plus dan SAS.

DAFTAR PUSTAKA

Afrianto, D. 2010. *Analisis Pengaruh Stok Beras, Luas Panen, Rata-rata Produksi, Harga Beras, dan Jumlah konsumsi Beras Terhadap Ketahanan Pangan di Jawa Tengah*. Skripsi. Semarang: Fakultas Ekonomi Universitas Diponegoro Semarang.

Chen, C. 2002, Robust Regression and Outlier Detection with the Robustreg Procedure. *Statistics and Data Analysis*. SAS Institute: Cary, NC.

Cranmer, J.S. 2005. *Methods Exam Review Outliers and Influence*. Berlin: Springer Verlag.

Fox, J. 2002. *Robust Regression*. Tersedia di: http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-robust-regression.pdf [15 Desember 2010]

Ghozali, I. 2009. *Aplikasi Analisis Multivariat dengan program SPSS edisi 4*. Semarang: Penerbit Universitas Diponegoro.

Gujarati, D.N. 1995. *Basic Econometrics*. New York: McGraw Hill.

Hadley, G. 1992. *Aljabar Linier*. Jakarta: Erlangga.

Huber, P.J. 1981. *Robust Statistics*. New York: John Wiley and Sons.

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

Roesseuw, R.J and A.M. Leroy. 1987. *Robust Regression and Outlier Detection*. New York: John Wiley and Sons.

Sembiring, R.K. 1995. *Analisis regresi Edisi 2*. Bandung: Penerbit ITB.

Soemartini. 2007. *OUTLIER(Pencilan)*. Jatinangor: Penerbit Universitas Padjajaran.

Sukestiyarno. 2008. *Olah Data Penelitian dengan SPSS*. Semarang: Lembaga Penelitian UNNES.

Supranto. J. 2005. *Ekonometri*. Bogor: Ghalia Indonesia.

Wilcox, R.R. 2005. Introduction to Robust Estimation and Hypothesis. San Diego: Academic Press.

Yohai, V.J. 1987. High Breakdown Point and High efficiency Robust estimates for Regression. *The Annals Of Statistics*, Vol. 15, No.20, 642-656.

Yaffe, R.A, 2002. *Robust Regression Modelling with STATA Lecture Notes*. Avenue: Social Science and Mapping group Academic Computing Services.

Lampiran 1

DATA JAWA TENGAH DALAM ANGKA Tahun 2007				
No	**Kabupaten/Kota**	**Produksi padi(ton)**	**Luas panen (ha)**	**Jumlah penduduk (juta jiwa)**
1.	Kab. Cilacap	622,442	111,725	1,674
2.	Kab. Banyumas	351,340	64,989	1,532
3.	Kab. Purbalingga	188,644	35,590	0,863
4.	Kab. Banjarnegara	145,025	27,132	0,904
5.	Kab. Kebumen	360,331	67,959	1,208
6.	Kab. Purworejo	284,618	52,729	0,712
7.	Kab. Wonosobo	156,034	29,793	0,780
8.	Kab. Magelang	280,093	53,481	1,169
9.	Kab. Boyolali	225,248	41,717	0,942
10.	Kab. Klaten	327,522	58,505	1,139
11.	Kab. Sukoharjo	267,230	46,176	0,838
12.	Kab. Wonogiri	269,556	54,622	1,010
13.	Kab. Karanganyar	243,685	42,826	0,834
14.	Kab. Sragen	493,681	90,833	0,868
15.	Kab. Grobogan	571,485	101,994	1,334
16.	Kab. Blora	320,851	63,513	0,841
17.	Kab. Rembang	132,025	26,895	0,588
18.	Kab. Pati	385,164	76,608	1,214
19.	Kab. Kudus	127,543	24,992	0,760
20.	Kab. Jepara	198,981	38,020	1,078
21.	Kab. Demak	502,407	91,516	1,071
22.	Kab. Semarang	170,787	32,862	0,894
23.	Kab. Temanggung	177,551	32,624	0,717
24.	Kab. Kendal	214,111	40,063	0,897
25.	Kab. Batang	207,477	40,265	0,712
26.	Kab. Pekalongan	223,888	44,457	0,859
27.	Kab. Pemalang	357,467	70,694	1,372
28.	Kab. Tegal	298,062	55,898	1,471
29.	Kab. Brebes	458,518	84,696	1,814
30.	Kota Magelang	2,513	484	0,130
31.	Kota Surakarta	1,782	347	0,534
32.	Kota Salatiga	7,134	1,385	0,176
33.	Kota Semarang	24,689	5,046	1,436
34.	Kota Pekalongan	11,835	2,315	0,284
35.	Kota Tegal	7,135	1,347	0,250

Sumber: (Afrianto, 2010)