METODE *LEAST TRIMMED SQUARE* (LTS) DAN *MM-ESTIMATION* UNTUK MENGESTIMASI PARAMETER REGRESI KETIKA TERDAPAT *OUTLIER*

Skripsi

disusun sebagai salah satu syarat

untuk memperoleh gelar Sarjana Sains

Program Studi Matematika

oleh

Elok Tri Kusuma Dewi

4111411016

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI SEMARANG

2015

PERNYATAAN

Saya menyatakan bahwa skripsi ini bebas plagiat, dan apabila di kemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya bersedia menerima sanksi sesuai ketentuan peraturan perundang-undangan.

Semarang, 11 Agustus 2015

Elok Tri Kusuma Dewi

(4111411016)

HALAMAN PENGESAHAN

Skripsi yang berjudul

Metode *Least Trimmed Square* (LTS) dan MM-*Estimation* Untuk Mengestimasi Parameter Regresi Ketika Terdapat *Outlier*

disusun oleh

Elok Tri Kusuma Dewi

4111411016

telah dipertahankan di hadapan sidang Panitia Ujian Skripsi FMIPA UNNES pada tanggal 6 Agustus 2015.

Panitia:

Ketua	Sekretaris
Prof. Dr. Wiyanto, M.Si	Drs. Arief Agoestanto, M.Si
196310121988031001	196807221993031005

Ketua Penguji

Drs. Supriyono, M.Si
195210291980031002

Anggota Penguji Pembimbing 1	Anggota Penguji Pembimbing 2
Drs. Arief Agoestanto, M.Si	Dra. Sunarmi, M.Si
196807221993031005	195506241988032001

MOTTO DAN PERSEMBAHAN

MOTTO

- Karena ketika keputusan yang kamu ambil akan ada resiko yang mengikutinya. Just do it and don’t afraid.
- Senyum adalah ibadah.

PERSEMBAHAN

- Dosen-dosen Jurusan Matematika dan dosen pembimbing yang sudah memberikan saya ilmu yang bermanfaat dan membantu dalam menyelesaikan skripsi.
- Seluruh staff TU FMIPA UNNES, staff perpustakaan Matematika UNNES yang telah membantu dalam berbagai kebutuhan akademis.
- Seluruh staff Puskesmas Sekaran Kota Semarang atas segala bantuan data yang diperlukan.
- Mama,Papa,kakak dan keluarga yang saya cintai dan selalu mendoakanku.
- Sahabat Matematika Murni 2011 yang selalu memberikan semangat.
- Terimakasih untuk Styfanda, Atmira, Ruliana, Nurul, Nilam, Mila, Enggar, Puji, Eko, Afi yang telah membantu penyusunan skripsi ini.
- Teman-teman Full Colour atas perhatian dan kebersamaan selama ini.

KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah memberikan nikmat dan karunia-Nya serta kemudahan sehingga penulis dapat menyelesaikan skripsi yang berjudul ”Metode *Least Trimmed Square* (LTS) dan MM-*Estimation* Untuk Mengestimasi Parameter Regresi Ketika Terdapat *Outlier*”.

Penyusunan skripsi ini dapat diselesaikan berkat kerjasama, bantuan, dan bimbingan dari berbagai pihak. Oleh karena itu penulis mengucapkan terima kasih kepada:

1. Prof. Dr. Fathur Rokhman M.Hum, Rektor Universitas Negeri Semarang.
2. Prof. Dr. Wiyanto, M.Si, Dekan FMIPA Universitas Negeri Semarang.
3. Drs. Arief Agoestanto, M.Si, Ketua Jurusan Matematika FMIPA Universitas Negeri Semarang.
4. Dra Kristina Wijayanti MS, Ketua Prodi Matematika Jurusan Matematika FMIPA Universitas Negeri Semarang.
5. Drs. Arief Agoestanto, M.Si, dan Dra. Sunarmi, M.Si sebagai Dosen Pembimbing yang telah banyak memberikan arahan dan bimbingan kepada penulis.
6. Drs. Supriyono, M.Si sebagai Dosen Penguji yang telah banyak memberikan masukan kepada penulis.
7. Mama, Papa dan kakak tercinta yang senantiasa mendoakan serta memberikan dorongan baik secara moral maupun spiritual.
8. Semua pihak yang telah membantu dalam penelitian ini.

Dengan segala keterbatasan, penulis menyadari bahwa penulis masih banyak kekurangan. Oleh karena itu penulis berharap perlu dikembangkan penelitian selanjutnya di masa mendatang.

Semarang, 6 Agustus 2015

Penulis

ABSTRAK

Elok Tri Kusuma Dewi. 2015. *Metode Least Trimmed Square (LTS) dan MM-Estimation untuk Mengestimasi Parameter Regresi ketika Terdapat Outlier*. Skripsi Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang. Pembimbing : Drs. Arief Agoestanto, M.Si. dan Dra. Sunarmi, M.Si.

Kata kunci : Analisis Regresi Berganda, *Outlier, Regresi Robust*, *Least Trimmed Square (LTS),* MM -*Estimation.*

Analisis regresi linear berganda merupakan salah satu metode statistik yang digunakan untuk memodelkan dan menyelidiki hubungan antara satu variabel dependen dengan dua atau lebih variabel independen. Salah satu metode yang digunakan untuk mengestimasi parameter regresi adalah metode Odinary Least Square (OLS). Namun metode OLS sangat peka terhadap adanya penyimpangan asumsi pada data misalnya disebabkan adanya outlier. Akibatnya penduga OLS yang diperoleh menjadi tidak efisien.

Outlier (*outlier*) adalah data yang tidak mengikuti pola umum pada model regresi yang dihasilkan, atau tidak mengikuti pola data secara keseluruhan. Keberadaan outlier dalam data dapat mengganggu proses analisis data, sehingga mengakibatkan residual dan varians pada data menjadi lebih besar. Oleh karena itu, diperlukan metode lain yang dapat digunakan untuk mengatasi outlier yaitu Regresi Robust.

Regresi *robust* merupakan alat yang penting untuk menganalisis data yang terdeteksi sebagai data *outlier*. Pada regresi *robust* ada beberapa metode yang dapat digunakan yaitu M-*Estimation, Least Median Square*, *Least Trimmed Square, S-Estimation* serta MM-*Estimation.*

Penelitian ini bertujuan untuk mengetahui keefektifan regresi *robust* dengan metode *Least Trimmed Square* (LTS) dan *MM-estimation* pada regresi liniear berganda dilihat dari nilai R^2 dan residual masing-masing metode.

Data yang digunakan dalam penelitian ini merupakan hasil penelitian dari 40 responden yakni pasien dari Puskesmas Sekaran Gunung Pati Kota Semarang. Data ini terdiri dari usia (X_1) dan indeks massa tubuh (X_2) sebagai variabel independent sedangkan tekanan darah sistolik (Y) sebagai variabel dependent. Karena data pengaruh usia dan indeks massa tubuh (IMT) terhadap tekanan darah sistolik terdeteksi adanya outlier maka dilakukan estimasi parameter dengan metode MM -*estimation* dan *Least Trimmed Square.* Model yang dihasilkan menggunakan metode LTS yaitu $\hat{y} = 67{,}141 + 0{,}649X_1 + 0{,}587X_2$. Sedangkan model yang dihasilkan menggunakan metode MM-*Estimation* yaitu $\hat{y} = 65{,}308 + 0{,}666X_1 + 0{,}618X_2$.

Karena pada metode *Least Trimmed Square* (LTS) memperoleh nilai R^2 lebih besar dan residual lebih kecil dibandingkan metode MM-*estimation* maka dapat disimpulkan bahwa metode *Least Trimmed Square* (LTS) lebih efisien dalam mengestimasi parameter regresi dibandingkan metode MM-*estimation.*

DAFTAR ISI

Halaman

DAFTAR TABEL

Halaman

DAFTAR GAMBAR

Halaman

DAFTAR LAMPIRAN

BAB 1
PENDAHULUAN

1.1 Latar Belakang

Matematika merupakan ilmu pengetahuan dasar dalam proses berkembangnya ilmu pengetahuan yang lain. Hampir setiap ilmu pengetahuan membutuhkan peranan matematika. Tidak hanya itu, perkembangan dari kajian matematikapun kian pesatnya dan terbagi menjadi dua arah yakni murni dan terapan. Matematika terapan memiliki pengertian bahwa matematika digunakan di luar bidang matematika. Lain halnya dengan matematika murni yang mengkaji tentang seluk beluk dalam matematika serta memecahkan kasus-kasus dalam matematika. Dengan kata lain matematika merupakan ratu dari ilmu pengetahuan. Salah satu cabang dari matematika terapan adalah statistika.

Statistika memegang peranan penting dalam memecahkan masalah yang terjadi pada bidang-bidang ilmu lainnya. Seperti bidang ekonomi, kependudukan, kesehatan, dan kemiliteran. Dengan adanya permasalahan-permasalahan yang terjadi tersebut, maka statistikawan berusaha memberikan solusi berupa suatu hasil analisis yang berkualitas yang pada akhirnya dapat digunakan untuk pengambilan keputusan.

Analisis regresi liniear berganda merupakan salah satu metode statistik yang digunakan untuk memodelkan dan menyelidiki hubungan antara satu variabel terikat dengan satu variabel bebas. Hubungan yang menggambarkan antara variabel-variabel dalam regresi sering disebut sebagai model regresi klasik. Variabel sendiri dibedakan menjadi 2 macam, yakni variabel bebas (*variabel independent*) dan variabel terikat (*variabel dependent*). Variabel bebas adalah variabel yang tidak dipengaruhi variabel lain, sedangkan variabel terikat adalah variabel yang keberadaannya dipengaruhi oleh variabel lainnya.

Regresi linear berganda mempunyai bentuk persamaan :

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \qquad (1.1)$$

dimana β_0, $\beta_1$dan β_2 merupakan parameter yang belum diketahui nilainya. Dengan menggunakan metode kuadrat terkecil, maka akan diperoleh nilai parameter tersebut.

Dalam menetukan estimator terbaik dari analisis regresi sangat dipengaruhi oleh penggunaan metode. Metode untuk menaksir koefisien regresi klasik salah satunya dengan cara meminimumkan jumlah kuadrat tekecil, yang disebut dengan metode kuadrat terkecil atau metode *Ordinary Least Square* (OLS). Namun metode ini tidak dapat bekerja dengan baik apabila terdapat data *outlier*.

Berbagai kaidah telah diajukan untuk menolak *outlier* (memutuskan untuk menghilangkan data yang ada *outlier*nya, setelah itu data dianalisis ulang tanpa *outlier*). Penolakan *outlier* yang begitu saja bukanlah langkah yang bijaksana. Adakalanya *outlier* dapat memberikan informasi yang tidak bisa diberikan oleh titik data lainnya, misalnya karena *outlier* timbul dari kombinasi keadaan yang tidak biasa yang mungkin saja sangat penting dan perlu diselidik lebih jauh. Sebagai kaidah umum, *outlier* baru ditolak jika setelah ditelusuri ternyata merupakan akibat dari kesalahan-kesalahan seperti memasukkan ukuran atau analisis yang salah, ketidaktepatan pencatatan data, dan terjadi kerusakan alat pengukuran. Bila ternyata bukan akibat dari kesalahan-kesalahan semacam itu, penyelidikan yang seksama harus dilakukan (Drapper and Smith,1992).

Outlier adalah data pengamatan yang berada jauh (*ekstrim*) dari pengamatan-pengamatan lainnya (Makkulau *et al*., 2010). Untuk mengidentifikasi apakah terdapat data *outlier* atau tidak, dapat menggunakan beberapa metode salah satunya yang akan dibahas adalah metode DfFITS (*Difference fitted value* FITS). *Difference fitted value FITS* merupakan metode yang menampilkan nilai perubahan dalam harga yang diprediksi bilamana kasus tertentu dikeluarkan dan yang sudah distandarkan. Selain itu juga bisa menggunakan metode grafis dan metode *Cook's Distance*.

Salah satu metode untuk menganalisa data yang dipengaruhi oleh *outlier* sehingga dihasilkan model yang kekar adalah regresi robust. Regresi robust merupakan metode regresi yang tepat digunakan ketika ada beberapa *outlier* pada model. Suatu estimator yang kekar adalah relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data.

Prosedur regresi *robust* cenderung mengabaikan sisaan-sisaan (*error*) yang berhubungan dengan *outlier* yang besar. Ada beberapa metode dalam regresi robust yang dapat digunakan untuk mengatasi data *outlier*, diantaranya *Least Median of Square* (LMS), *Least Trimmed Squares* (LTS), *M-estimation*, *S-Estimation* dan *MM-estimation.* Namun pada skripsi hanya akan dibahas metode *Least Trimmed Squares* (LTS) dan *MM-estimation.* Dengan melihat nilai R^2 dan residual yang didapat dengan menggunakan rumus ataupun dengan bantuan *software Microsoft Excel* dan SPSS 19 akan dapat terlihat metode mana yang lebih efektif.

Metode *Least Trimmed Squares* (LTS) sebagai salah satu metode penaksiran parameter model regresi robust terhadap kehadiran nilai *outlier*. Adapun tujuan yang ingin dicapai yakni mendapatkan nilai parameter model regresi yang *robust* terhadap kehadiran nilai *outlier*. Metode ini tidak membuang bagian data *outlier* tapi menemukan model *fit* dari mayoritas data.

MM-*Estimation* adalah metode yang pertama kali diperkenalkan oleh Yohai pada tahun 1987 yakni dengan yang menggabungkan estimasi *high breakdown point* dan efisiensi statistik. Langkah pertama dalam estimasi ini adalah mencari estimator S dengan menjamin nilai *breakdown point*, kemudian menetapkan parameter-parameter regresi menggunakan estimasi M. Pada umumnya digunakan fungsi Tukey Bisquare β baik pada estimasi S maupun estimasi M. Metode ini juga mempertahankan kekekaran dari metode estimasi S, serta efisiensi dari metode estimasi M. Diharapkan melalui kedua metode regresi robust tersebut dapat diperoleh estimator yang baik sehingga menghasilkan model yang lebih baik dari model hasil *Ordinary Least Square* (OLS).

Dari penelitian terdahulu, diperoleh bahwa metode LTS merupakan metode estimasi parameter yang baik dari pada metode OLS ketika terdapat *outlier* dalam data sebesar 5%, 10% dan 20% (Maharani *et al*,. 2014). Selain itu, pada penelitian lain yang ditulis sebelumnya oleh Elen Dwi Pradewi (2012) dengan memodelkan regresi liniear berganda “Ketahanan Pangan di Jawa

Tengah Tahun 2007" menggunakan regresi robust estimasi M-IRLS dengan fungsi pembobot Huber dan Tukey Bisquare, Hanna Mahiroh (2011) mendapatkan bahwa metode LTS lebih baik dibanding dengan metode M-*Estimation* pada regresi liniear sederhana, Andhika Tegar Permana (2013) membandingkan metode LTS dengan metode penduga-S pada regresi berganda dan mendapatkan hasil bahwa penduga-S lebih baik dibandingkan metode LTS dan Ory Ade Maulana (2012) menggunakan regresi robust LTS dengan algoritma Fast-LTS dan C-Step untuk mengatasi *outlier* pada regresi liniear. Dengan pertimbangan itulah penulis tertarik untuk melakukan penelitian tentang keefektifan regresi *robust* dengan metode *Least Trimmed Square* (LTS) dan *MM-estimation* pada regresi liniear.

Untuk menentukan metode yang lebih efektif ada berbagai kriteria yang bisa ditetapkan sebagai acuannya, namun pada penelitian ini akan dilihat dari kriteria nilai R^2 dan residualnya. Jika nilai R^2 besar atau mendekati satu berarti variabel-variabel independen memberikan hampir semua informasi yang dibutuhkan untuk memprediksi variansi variabel dependen dan jika semakin kecil nilai residualnya maka semakin baik kecocokan suatu persamaan dengan data nilai duga Y semakin mendekati sebenarnya.

1.2 Rumusan Masalah

Berdasarkan uraian diatas, maka permasalahan yang dikaji dalam penelitian ini adalah:

1. Bagaimana pengidentifikasian *outlier* dari data regresi berganda?
2. Bagaimana bentuk model regresi *robust* terbaik dengan adanya data *outlier* dengan menggunakan metode *Least Trimmed Squares (LTS)* dan metode *MM-estimation* pada data regresi tersebut?
3. Bagaimanakah perbandingan yang dihasilkan dengan menggunakan metode *Least Trimmed Squares (LTS)* dan *MM-estimation* pada proses regresi *robust* dilihat dari nilai R^2 dan residualnya?

1.3 Batasan Masalah

Batasan masalah yang dilakukan pada penelitian ini adalah sebagai berikut,

1. Pendeteksian *Outlier* hanya menggunakan metode grafis, metode *Cook's Distance* dan DfFITS.
2. Penelitian hanya menggunakan metode robust *Least Trimmed Squares (LTS)* dan metode robust *MM-Estimation.*
3. Paket program yang mendukung penelitian adalah *software Microsoft Excel* dan SPSS 19.

1.4 Tujuan Penelitian

Berdasarkan permasalahan yang timbul, penelitian ini mempunyai tujuan:

1. Mengetahui bagaimana cara pengidentifikasian *outlier* dari data regresi.

2. Mengetahui bentuk model regresi *robust* terbaik dengan adanya data *outlier* dengan menggunakan metode *Least Trimmed Squares (LTS)* dan metode *MM-estimation* pada data regresi tersebut.
3. Mangetahui bagaimanakah perbandingan yang dihasilkan dengan menggunakan metode *Least Trimmed Squares (LTS)* dan *MM-estimation* pada proses regresi *robust* dilihat dari nilai R^2 dan residualnya.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dalam penulisan proposal ini diantaranya :

1.5.1 Bagi Mahasiswa

1. Mahasiswa mendapatkan pengetahuan tentang data *outlier*.
2. Mahasiswa mendapat pengetahuan tentang metode *Least Trimmed Square* (LTS).
3. Mahasiswa dapat mengetahui prosedur penggunaan metode *Least Trimmed Square* (LTS).
4. Mahasiswa mendapat pengetahuan tentang metode *MM-Estimation.*
5. Mahasiswa dapat mengetahui prosedur penggunaan metode *MM-Estimation.*

1.5.2 Bagi Pembaca

1. Dapat menambah atau memperkaya khasanah kepustakaan Jurusan Matematika.
2. Menambah topik kajian tentang metode *Least Trimmed Square* (LTS) dan *MM-Estimation*

.

1.6 Sistematika Penulisan Skripsi

Secara garis besar skripsi ini dibagi menjadi tiga bagian (bab) 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, pernyataan keaslian tulisan, pengesahan, motto dan persembahan, kata pengantar, abstrak, daftar isi, daftar gambar, daftar tabel, dan daftar lampiran.

(2) Bagian isi skripsi

Bagian isi skripsi secara garis besar terdiri dari lima bab, yaitu:

BAB 1 PENDAHULUAN

Bab ini berisi mengenai latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan skripsi.

BAB 2 TINJAUAN PUSTAKA

Bab ini berisi kajian teori yang mendasari dan berhubungan dengan pemecahan masalah. Teori-teori tersebut digunakan untuk memecahkan masalah yang diangkat dalam skripsi ini. Teori yang digunakan adalah *cook's distance*, DfFITS, metode *Least Trimmed Square*, dan metode *MM-Estimator*.

BAB 3 METODE PENELITIAN

Bab ini mengulas metode yang digunakan dalam penelitian yang berisi langkah-langkah yang dilakukan untuk memecahkan masalah yaitu pengumpulan data, analisis data dan kesimpulan.

BAB 4 HASIL PENELITIAN DAN PEMBAHASAN

Bab ini berisi mengenai penyelesaian dari permasalahan yang diungkapkan.

BAB 5 PENUTUP

Bab ini berisi tentang simpulan dari pembahasan dan saran yang berkaitan dengan simpulan.

(3) Bagian akhir skripsi

Bagian akhir skripsi meliputi daftar pustaka yang memberikan informasi tentang buku sumber serta literatur yang digunakan dan lampiran-lampiran yang mendukung skripsi.

BAB 2
LANDASAN TEORI

Ada dua sub bab yang akan dibahas pada landasan teori ini, yaitu tinjauan pustaka serta kerangka berfikir. Tinjauan pustaka berisi tentang pengertian-pengertian yang bersangkutan dengan pendeteksian *outlier* dan pemodelan regresi *robust* dengan metode *Least Trimmed Square* (LTS) dan *MM-Estimation* secara teoritis dan dengan bantuan *software Microsoft Excel* dan SPSS 19. Sedangkan kerangka berfikir menggambarkan tentang arah penulisan untuk mencapai tujuan penelitian.

2.1 Tinjauan Pustaka

2.1.1 Regresi Liniear Berganda

Regresi liniear berganda adalah regresi dimana variable Y dihubungkan dengan dua variabel bebas (X_1, X_2), namun masih menunjukkan diagram hubungan yang liniear. Penambahan variabel bebas diharapkan dapat lebih menjelaskan karakteristik hubungan yang ada, walaupun masih saja ada variabel yang terabaikan. Bentuk umum persamaan regresi liniear berganda dapat dituliskan sebagai berikut :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, i = 1,2, \dots, n \quad (2.1)$$

dimana

Y_i = variabel terikat (nilai duga Y)

X_{1i}, X_{2i} = variabel bebas

β_0, β_1 dan β_2 = koefisien regresi liniear berganda

2.1.2 Residual

Residual atau sisaan dalam regresi linear sederhana merupakan selisih dari nilai prediksi dengan nilai yang sebenarnya atau $\varepsilon_i = y_i - \hat{y}_i$. Namun penggunaan jarak $\varepsilon_i = y_i - \hat{y}_i$ tidaklah memuaskan. Dengan meminimumkannya didapat hasil yang wajar seperti berikut:

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2.2}$$

Jika nilai pengamatan terletak dalam garis regresi maka nilai residualnya sama dengan nol. Jadi jika total jarak atau nilai mutlak dari residual sama dengan nol ($\sum_{i=1}^{n} |\varepsilon_i| = 0$) berarti semua nilai pengamatan berada pada garis regresi. Semakin besar nilai residualnya maka garis regresi semakin kurang tepat digunakan untuk memprediksi. Yang diharapkan adalah total residualnya kecil sehingga garis regresi cukup baik untuk digunakan.

2.1.3 Metode *Ordinary Least Square (*Metode Kuadrat Terkecil*)*

Salah satu penduga model untuk bentuk regresi linear berganda adalah dengan *Ordinary Least Square*. Konsep dari metode ini adalah meminimumkan jumlah kuadrat sisa (selisih antara data sebenarnya dengan data dugaan) dari model regresi yang terbentuk. Metode *Ordinary Least Square* pertama kali diperkenalkan oleh Carl Freidrich Gauss, seorang ahli matematika dari Jerman. Metode ini merupakan metode yang paling banyak digunakan dalam pembentukan model regresi mengestimasi parameter regresi dibandingkan dengan metode-metode yang lain.

Menurut Sembiring (1995: 40) dalam mengestimasi koefisien regresi β_0, $\beta_1$dan β_2 pada *n* data suatu penelitian adalah

$$J = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i})^2 \tag{2.3}$$

dan itu harus bernilai minimum. Pada persamaan (2.1) nilai *x* dan *y* berasal dari pengamatan, sedangkan β_0, $\beta_1$dan β_2 berubah bila garis regresinya berubah. Jika J berubah diturunkan terhadap β_0, $\beta_1$dan β_2, kemudian menyamakannya dengan nol, sehingga diperoleh

$$\frac{\partial J}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i}) = 0$$

atau

$$\sum_{i=1}^{n} y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} x_{1i} - \beta_2 \sum_{i=1}^{n} x_{2i} = 0 \tag{2.4}$$

dan

$$\frac{\partial J}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i})\, x_{1i} = 0$$

atau

$$\sum_{i=1}^{n} y_i x_{1i} - \beta_0 \sum_{i=1}^{n} x_{1i} - \beta_1 \sum_{i=1}^{n} x_{1i}^2 - \beta_2 \sum_{i=1}^{n} x_{1i} x_{2i} = 0 \tag{2.5}$$

dan

$$\frac{\partial J}{\partial \beta_2} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i}) x_{2i} = 0$$

atau

$$\sum_{i=1}^{n} y_i x_{2i} - \beta_0 \sum_{i=1}^{n} x_{2i} - \beta_1 \sum_{i=1}^{n} x_{1i} x_{2i} - \beta_2 \sum_{i=1}^{n} x_{2i}^2 = 0 \tag{2.6}$$

Persamaan (2.4), (2.5), dan (2.6) disebut persamaan normal. Untuk langkah selanjutnya persamaan (2.1) akan disusun dalam bentuk matriks seperti dibawah ini dengan

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \cdot \\ \cdot \\ \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ \vdots & & \vdots \\ \\ 1 & x_{n1} & x_{n2} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \cdot \\ \cdot \\ \\ \varepsilon_n \end{pmatrix}$$

$$X\beta = \begin{pmatrix} \beta_0 + x_{11}\beta_1 + x_{12}\beta_2 \\ \beta_0 + x_{21}\beta_1 + x_{22}\beta_2 \\ \beta_0 + x_{31}\beta_1 + x_{32}\beta_2 \\ \vdots \\ \\ \beta_0 + x_{n1}\beta_1 + x_{n2}\beta_2 \end{pmatrix}$$

$$X\beta + \varepsilon = \begin{pmatrix} \beta_0 + x_{11}\beta_1 + x_{12}\beta_2 \\ \beta_0 + x_{21}\beta_1 + x_{22}\beta_2 \\ \beta_0 + x_{31}\beta_1 + x_{32}\beta_2 \\ \vdots \\ \\ \beta_0 + x_{n1}\beta_1 + x_{n2}\beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \cdot \\ \cdot \\ \\ \varepsilon_n \end{pmatrix}$$

$$= \begin{pmatrix} \beta_0 + x_{11}\beta_1 + x_{12}\beta_2 + \varepsilon_1 \\ \beta_0 + x_{21}\beta_1 + x_{22}\beta_2 + \varepsilon_2 \\ \beta_0 + x_{31}\beta_1 + x_{32}\beta_2 + \varepsilon_3 \\ \vdots \\ \\ \beta_0 + x_{n1}\beta_1 + x_{n2}\beta_2 + \varepsilon_n \end{pmatrix}$$

Maka dapat dituliskan persamaan

$$Y = X\beta + \varepsilon$$

Secara umum, untuk model regresi garis lurusnya diperoleh

$$X'X = \begin{pmatrix} n & \sum x_{i1} & \sum x_{i2} \\ \sum x_{i1} & \sum {x_i}^2{}_1 & \sum x_{i1}x_{i2} \\ \sum x_{i2} & \sum x_{i1}x_{i2} & \sum {x_i}^2{}_2 \end{pmatrix},$$

$$X'Y = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots. & 1 \\ 1 & x_{21} & x_{22} & \dots. & x_{n1} \\ 1 & x_{31} & x_{32} & \dots. & x_{n2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \cdot \\ \cdot \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_{i1} y_i \\ \sum x_{i2} y_i \end{pmatrix}.$$

Jadi bentuk matriks dari persamaan normal (2.4), (2.5), dan (2.6) dapat dituliskan sebagai

$$X'X\beta = X'Y \tag{2.7}$$

dari persamaan (2.7) didapatkan nilai β yaitu

$$\beta = (X'X)^{-1}X'Y \tag{2.8}$$

2.1.4 *Outlier* (*outlier*)

Outlier adalah data yang tidak mengikuti pola umum pada model regresi yang dihasilkan, atau tidak mengikuti pola data secara keseluruhan. Dalam suatu himpunan data biasanya terdapat 10% amatan yang merupakan *outlier* (Hampel *et al*., 1986). Jumlah maksimum *outlier* dalam data yang diperbolehkan adalah 50% (Rousseeuw *et al*., 1987: 303).

Pada umumnya *outlier* mempunyai sisaan (*error*) berjarak tiga simpangan baku. *Outlier* merupakan suatu keganjilan yang menandakan suatu titik data yang sama sekali tidak tipikal dari data yang lainnya.

Apabila dalam pengamatan terdapat data *outlier*, maka alternatif langkah yang diambil adalah menghilangkan atau membuang data *outlier* tersebut secara langsung terlebih dahulu sebelum dilakukan analisis lebih lanjut. Data *outlier* tersebut dapat dibuang secara lansung jika data tersebut diperoleh dari kesalahan teknis peneliti, seperti kesalahan mencatat amatan atau ketika menyiapkan peralatan (Smith, 1992).

Keberadaan data *outlier* akan mengganggu dalam proses analisis data dan harus dihindari dari beberapa hal. Dalam kaitannya dalam analisis regresi, *outlier* dapat menyebabkan hal-hal berikut (Soemartini, 2007: 7):

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

Selain itu adanya *outlier* berpengaruh akan memberikan nilai penduga parameternya bersifat bias sehingga berakibat interpretasi hasil yang diperoleh menjadi tidak valid. Namun menghindari *outlier* berpengaruh (menghapus *outlier* berpengaruh) dalam melakukan analisis bukanlah hal yang tepat untuk dilakukan. Adakalanya *outlier* memberikan informasi yang tidak bisa diberikan oleh titik data lainnya, misalnya *outlier* timbul karena kombinasi keadaan yang tidak biasa yang mungkin saja sangat penting dan perlu diselidiki lebih jauh (Draper & Smith, 1992).

2.1.5 Deteksi Outlier

Ketika peneliti mendeteksi outlier, perlakuan pertamanya adalah melihat kemungkinan bahwa outlier merupakan data yang terkontaminasi. Data outlier dapat dikenali dengan pemeriksaan visual dari data mentahnya (*raw*) atau dari diagram pencar dari variabel dependen (Jacob, 2003: 394). Jika terdapat lebih dari dua variabel independen, beberapa outlier mungkin akan sangat sulit dideteksi dengan pemeriksaan visual. Oleh karena itu, dibutuhkan alat bantu pada pemeriksaan visual yang dapat membantu dalam pendeteksian outlier.

Dalam statistik ruang, data *outlier* harus dilihat terhadap posisi dan sebaran data yang lainnya sehingga akan dievaluasi apakah data *outlier* tersebut perlu dihilangkan atau tidak. Ada berbagai macam metode yang dapat digunakan untuk mendeteksi adanya data *outlier* yang berpengaruh dalam koefisien regresi diantaranya adalah metode grafis, *boxplot*, *laverage values*, DfFITS, *cook's distance*, DfBETA(s). Namun pada skripsi ini pendeteksian oulier yang akan dibahas menggunakan metode grafis, metode *Cook's Distance* dan metode DfFITS.

2.1.5.1 Metode Grafis

Untuk melihat apakah terdapat data *outlier* pada data, dapat dilakukan dengan memplot antara data dengan observasi $ke - i$ ($i = 1,2, \ldots, n$), jika sudah didapatkan model regresi maka dapat dilakukan dengan memplot antara residual (*error*) dengan nilai prediksi Y $(\hat{y})$. Jika terdapat data yang letaknya jauh dari pola yang terbentuk dari keseluruhan data, maka data tersebut merupakan data *outlier*.

Kelemahan dari metode ini adalah keputusan yang memperlihatkan data tersebut merupakan *outlier* atau tidak bergantung pada kebijakan peneliti, karena pengamatannya dilakukan hanya dengan visualisasi gambar.

2.1.5.2 Cook's Distance

Cook's Distance merupakan salah satu metode pendeteksian *oulier* dengan cara menampilkan nilai jarak *cook* atau dengan kata lain menunjukkan besarnya pengaruh adanya data *outlier* terhadap semua estimator koefisien regresi. Perhitungan *Cook's Distance* di rumuskan sebagai berikut :

$$(\text{cook's distance})_i = \left[\frac{R^2_{standardi}}{2}\right] \times \left[\frac{h_{ii}}{1 - h_{ii}}\right]$$

dimana h_{ii} adalah nilai pengaruh untuk kasus ke$-i$.

Suatu data yang mempunyai nilai jarak *cook* lebih besar dari $\mathrm{F}(0{,}5; \mathrm{k};\ \mathrm{n} - \mathrm{k})$ maka didefinisikan sebagai *outlier*, dengan k banyaknya variabel independen dan n banyaknya observasi (Soemartini: 2007).

2.1.5.3 Metode DfFITS (Difference fitted value FITS)

Difference fitted value FITS merupakan metode yang menampilkan nilai perubahan dalam harga yang diprediksi bilamana kasus tertentu dikeluarkan, yang sudah distandarkan. Perhitungan DfFITS di rumuskan sebagai berikut :

$$(DfFITS)_i = t_i \left(\frac{h_{ii}}{1 - h_{ii}}\right)^{\frac{1}{2}}$$

dimana t, adalah *studentized deleted residual* untuk kasus ke$-i$ dan h_{ii} adalah nilai pengaruh untuk kasus ke$-i$.

dengan,

$$t_i = e_i \sqrt{\frac{n - k - 1}{JKG(1 - h_{ii}) - e_i^2}},$$

e_i adalah residual ke$-i$ dan JKG adalah jumlah kuadrat galat.

Suatu data yang mempunyai nilai *absolute DfFITS* lebih besar dari $2\sqrt{\frac{k+1}{n}}$ maka didefinisikan sebagai *outlier*, dengan k banyaknya variabel independen dan n banyaknya observasi (Soemartini: 2007).

2.1.6 *Goodness of FIT*

Menurut Lungan (2006: 267) uji kesesuaian (*goodness of fit*) bertujuan untuk mengambil kesimpulan tentang sebaran populasi. Ketepatan fungsi regresi sampel dalam menaksir nilai aktual dapat diukur dari *Goodness of FIT*nya. Nilai *Goodness of FIT* dapat diukur dari nilai koefisien determinasi (R^2). Koefisien determinasi pada umumnya mengukur seberapa jauh kemampuan model dalam menerangkan variasi variabel dependen. Nilai koefisien determinasi adalah antara nol dan satu. Nilai R^2 yang kecil berarti kemampuan variabel-variabel independen dalam menjelaskan variasi variabel dependen amat terbatas, sedangkan jika nilai R^2 mendekati satu berarti variabel-variabel independen memberikan hampir semua informasi yang dibutuhkan untuk memprediksi variasi variabel-variabel dependen (Imam, 2006: 87).

2.1.7 Fungsi Obyektif

Fungsi obyektif adalah fungsi yang digunakan untuk mencari fungsi pembobot pada regresi robust. Fungsi pembobot yang digunakan antara lain adalah fungsi pembobot *Huber* dan fungsi pembobot *Tukey Bisquare* (Montgomery & Peck, 1992: 369).

2.1.7.1 Fungsi Pembobot Huber

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

dengan,

$$\psi(e_i) = \rho'^{(e_i)} = \frac{\partial(\rho(e_i))}{\partial e_i} = \begin{cases} e_i, & |e_i| \leq c \\ c, & e_i > c \\ -c, & e_i < -cc \end{cases}$$

dan fungsi pembobot,

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

2.1.7.2 Fungsi Pembobot Tukey Bisquare

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

dengan,

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

dan fungsi pembobot,

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

Konstanta yang menghasilkan efisiensi tinggi dengan residual berdistribusi normal dan dapat memberikan perlindungan terhadap *outlier* yaitu konstanta dengan nilai $c = 1.345$ untuk fungsi pembobot *Huber* dan $c = 4{,}685$ untuk pembobot *Tukey Bisquare*.

2.1.8 Regresi Robust

Regresi *robust* merupakan alat yang penting untuk menganalisis data yang terdeteksi sebagai data *outlier*. Regresi *robust* digunakan untuk mendeteksi *outlier* dan memberikan hasil yang resisten terhadap adanya data *outlier* (Chen: 2002). Sedangkan menurut Aunuddin (1989), regresi *robust* ini ditujukan untuk mengatatasi adanya data ekstrim serta meniadakan pengaruhnya terhadap hasil pengamatan tanpa terlebih dulu mengadakan identifikasi.

Metode ini merupakan metode yang mempunyai sifat :

1. Sama baiknya dengan *ordinary least square* ketika semua asumsi terpenuhi dan tidak terdapat titik data yang berpengaruh.
2. Dapat menghasilkan model regresi yang lebih baik daripada *ordinary least square* ketika asumsi tidak dipenuhi dan terdapat tititk data yang berpengaruh.
3. Perhitungannya cukup sederhana dan mudah dimengerti, tetapi dilakukan secara iteratif sampai diperoleh dugaan terbaik yang mempunyai standar *error* parameter yang paling kecil.

Dalam regresi *robust* terdapat beberpapa estimasi, yaitu :

2.1.8.1 M-Estimation

Salah satu regresi *robust* yang penting dan paling luas digunakan adalah *M-Estimation*. Menurut Montgomery (1992), pada prinsipnya *M-Estimation* merupakan estimasi yang meminimumkan suatu fungsi residual ρ dan residualnya.

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

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

2.1.8.2 Least Median Squares (LMS)

Metode LMS merupakan metode *High Breakdown Value* yang diperkenalkan oleh Rousseeuw pada tahun 1984. Metode LMS adalah suatu metode estimasi parameter regresi *robust* dengan meminimumkan median dari kuadrat sisaan.

2.1.8.3 Least Trimmed Squares (LTS)

LTS diusulkan oleh Rousseuw (1998) sebagai alternatif *robust* untuk mengatasi kelemahan *ordinary least squares (OLS)*, yaitu dengan menggunakan sebanyak $h(h \leq n)$ kuadrat residual yang diturunkan nilainya.

$$\min_{b} \sum_{i=1}^{h} \varepsilon_i^2 \qquad (2.8)$$

dengan

$$h_0 = \left[\frac{n}{2}\right] + \left[\frac{(k+1)}{2}\right] \qquad (2.9)$$

keterangan:

$\varepsilon_i^2 =$ kuadrat residual yang diurutkan dari terkecil ke terbesar $\varepsilon_1^2 < \varepsilon_2^2 < \cdots < \varepsilon_i^2 < \cdots < \varepsilon_h^2 < \cdots < \varepsilon_n^2$

$n =$ banyaknya sampel

$k =$ parameter regresi

Jumlah h menunjukkan sejumlah subset data dengan kuadrat fungsi objektif terkecil. Untuk mendapatkan nilai residual pada LTS, digunakan algoritma LTS menurut Rousseeauw dan Van Driessen (1999) dalam Willems dan Aels (2005) adalah gabungan FAST-LTS dan C-*step*, yaitu dengan mengestimasi parameter β_0, $\beta_1$dan β_2, kemudian menentukan *n* residual dengan menggunakan rumus $\varepsilon_i^2 = (y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i})^2$. Setelah itu menghitung $\sum_{i=1}^{h_0} \varepsilon_i^2$, dengan $h_0 = \left[\frac{n}{2}\right] + \left[\frac{(k+1)}{2}\right]$pengamatan dengan nilai ε_i^2 terkecil. Tahapan-tahapan di atas dilakukan sampai diperoleh nilai residual terkecil dan konvergen.

2.1.8.4 S-Estimation

Metode *robust S-Estimation* juga merupakan metode *High Breakdown Value* yang diperkenalkan pertama kali oleh Rousseeuw dan Yohai pada tahun 1984. S-*Estimation* adalah suatu metode estimasi parameter regresi robust dengan meminimumkan *scale* S.

Menurut Rousseeuw & Yohai (1987), *S-Estimation* didefinisikan sebagai

$$\hat{\theta}_n = \arg\min_{\theta} S_n(\theta) \qquad (2.10)$$

dimana $S_n(\theta)$ adalah skala M-*estimator* dari residual $r_i(\theta)$, yaitu:

$$S_n(\theta) = \inf\left\{ s > 0; n^{-1} \sum_{i=1}^{n} \mathcal{X}\left(\frac{r_i}{s}\right) < 0 \right\}$$

Untuk sebuah nilai fungsi $\mathcal{X}$ yang tepat, kemungkinan yang lain adalah

$$\hat{\theta}_n = \arg\min_{\theta} \sum_{l=1}^{[n(1-x)]} \{r_i(\theta)^2; i = 1, 2, \dots, n\}_{(l)} \qquad (2.11)$$

2.1.8.5 **MM-*Estimation***

MM-*Estimation* adalah metode yang pertama kali diperkenalkan oleh Yohai pada tahun 1987 yaitu dengan menggabungkan estimasi S dan estimasi M. Langkah pertama dalam estimasi ini adalah mencari estimator S dengan menghitung estimator awal, kemudian menetapkan parameter-parameter regresi menggunakan estimasi M. Pada umumnya digunakan fungsi *Tukey Bisquare* β baik pada estimasi S maupun estimasi M.

Bentuk dari metode MM-*Estimation* adalah

$$\tilde{\beta}_{MM} = \arg\min \sum_{i=1}^{n} \rho\left(\frac{\varepsilon_i}{\hat{\sigma}}\right) = \arg min \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}}\right) \quad (2.12)$$

MM-*Estimation* juga menggunakan *Iteratively Reweighted Least Squares (IRLS)* untuk mencari estimasi parameter regresi.

Adapun langkah-langkah dalam proses *MM-Estimation* adalah:

(a) Menghitung estimator awal koefisien $\hat{\beta}_j^{(1)}$dan residual $\varepsilon_i^{(1)}$ dari regresi robust (estimasi S) dan dengan bobot huber / bisquare.

(b) Residual $\varepsilon_i^{(1)}$ pada langkah pertama digunakan untuk menghitung skala estimasi $\hat{\sigma}_s^{(1)}$ dan dihitung pula pembobot awal $w_i^{(1)}$.

(c) Residual $\varepsilon_i^{(1)}$ dengan skala estimasi $\hat{\sigma}_s$ pada langkah kedua digunakan dalam iterasi awal sebagai penaksir WLS untuk menghitung koefisien regresi $\sum_{i=1}^{n} w_i^{(1)} \left(\frac{\varepsilon_i^{(1)}}{\hat{\sigma}_s^{(1)}}\right) x_i = 0,\ w_i^{(1)}$ yang merupakan pembobot Huber/bisquare.

(d) Menghitung bobot baru $w_i^{(2)}$ dengan skala estimasi dari iterasi awal WLS.

(e) Mengulang langkah (b),(c),(d) (dengan skala estimasi tetap konstan) sampai mendapatkan $\sum_{i=1}^{n}\left|\varepsilon_i{}^{(m)}\right|$ konvergen (selisih $\hat{\beta}_j^{(m+1)}$ dan $\hat{\beta}_j^{(m)}$ mendekati 0, dengan banyak m iterasi).

2.1.9 Estimasi Parameter $\boldsymbol{\beta}$

Untuk meminimumkan ρ (fungsi obyektif) dari residualnya, dicari turunan parsial pertama dari ρ terhadap β_j disamadengankan 0. Ini memberikan $\rho = k + 1$ sistem persamaan

$$\sum_{i=1}^{n} x_{ij}\,\psi\left[\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}_S}\right] = 0 \qquad (2.13)$$

dengan $\psi = \rho'$ dan ψ merupakan fungsi *influence* yang digunakan dalam memperoleh bobot, x_{ij} adalah observasi ke$-i$ pada regresi ke$-j$ dan $x_{i0} = 1$.

Didefinisikan fungsi pembobot:

$$w(e_i) = \frac{\psi\left[\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}_S}\right]}{\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}_S}} \qquad (2.14)$$

Dan $w_i = w(e_i)$, maka persamaaan (2.13) dapat ditulis:

$$\sum_{i=1}^{n} x_{ij}\, w_i\left[y_i - \sum_{j=0}^{k} x_{ij}\beta_j\right] = 0 \qquad (2.15)$$

Menurut Montgomery & Peck (1992), estimasi pada regresi *robust* yang dilakukan dengan estimasi *Iteratively Reweighted Least Squares (IRLS)* membutuhkan proses iterasi dimana nilai w_i akan berubah nilainya disetiap iterasi. Iterasi akan berhenti sampai didapatkan nilai $\hat{\beta}_j$ yang konvergen yaitu selisih nilai $\hat{\beta}_j^{(m+1)}\ dan\ \hat{\beta}_j^{(m)}$ mendekati 0.

2.1.10 SPSS

Menurut Sukestiyarno (2013: 8) program aplikasi statistik SPSS (*Statistical Package for Social Sciences*) merupakan salah satu program yang relatif popular saat ini. Pada perkembangannya sekarang SPSS sudah meluas penggunaannya tidak hanya di bidang sosial saja tetapi juga lebih banyak digunakan di bidang eksakta.

SPSS memuat perangkat-perangkat statistik dasar, sehingga cukup baik dipergunakan untuk memahami sifat-sifat suatu data dan pengolahan data secara sederhana. Variasi analisisnya sangat luas. SPSS merupakan *software* yang dapat digunakan untuk mengolah data dalam statistik. Ada beberapa pilihan menu yang ada pada SPSS, diantaranya menu *File, Edit, View, Data, Translate, Anlyze, Graphs, Utilities, Add-ons, Window* dan *Help*. Untuk menganalisis regresi dengan bantuan SPSS menu yang digunakan adalah *Analyze* lalu *Regresion* pilih *Linear*. Setelah itu input variabel dependent, variabel independent dan bobot yang terlibat didalamnya.

2.2 Kerangka Berfikir

Berdasarkan tinjauan pustaka dapat dibuat kerangka berfikir bahwa dalam analisis regresi hubungan yang sebenarnya tidak dapat diketahui secara pasti, tetapi model hubungan model tersebut dapat diestimasi berdasarkan data pengamatan. Menurut Sembiring (1995) adanya *outlier* dalam data dapat mengakibatkan estimator koefisien regresi yang diperoleh kurang tepat. Sehingga, diperlukan suatu metode regresi yang kekar terhadap *outlier*, yaitu estimasi regresi *robust* dengan metode *Least Trimmed Square* (LTS). Estimasi regresi *robust* metode *Least Trimmed Square* (LTS) merupakan metode yang diperkenalkan oleh Rousseeuw pada tahun 1984. Metode *Least Trimmed Square* (LTS) adalah suatu metode estimasi parameter regresi *robust* dengan menggunakan konsep pengepasan OLS untuk meminimumkan jumlah kuadrat sisaan. Selain metode *Least Trimmed Square* (LTS), metode *MM-Estimation* juga merupakan metode yang mampu mengatasi data *outlier*. Dengan mencari estimator S terlebih dahulu, kemudian menetapkan parameter-parameter regresi menggunakan estimasi M. Kedua metode ini akan dibandingkan dengan melihat nilai R^2 dan residualnya yang diperoleh dengan menggunakan rumus atau dengan bantuan *software Microsoft Excel* dan SPSS 19 untuk mencari metode mana yang paling efektif untuk mengatasi adanya data *outlier*.

BAB 3
METODE PENELITIAN

Metode penelitian merupakan salah satu langkah yang dillakukan penulis dalam penelitian sehingga data yang diperoleh semakin lengkap untuk memecahkan masalah yang dihadapi. Metode penelitian yang digunakan dalam penulisan skripsi ini adalah metode kajian pustaka dengan melalui beberapa tahapan, yaitu :

3.1 Menentukan Masalah

Menentukan masalah dimulai dari studi pustaka. Studi pustaka merupakan penelaahan dari beberapa sumber yang relevan untuk mengumpulkan informasi yang diperlukan dalam penulisan skripsi ini. Setelah beberapa sumber pustaka terkumpul, maka akan dilanjutkan penelaahan isi dari sumber-sumber pustaka tersebut. Dari penelaahan tersebut akan memunculkan ide-ide yang kemudian dijadikan landasan teori dari penulisan skripsi ini. Permasalahan yang muncul pada penulisan skripsi ini adalah tentang regresi *robust* menggunakan metode *MM-estimation* dan *Least Trimmed Square (LTS).*

3.2 Merumuskan Masalah

Perumusan masalah ini bertujuan untuk membatasi permasalahan sehingga diperoleh kajian yang jelas. Dari penelaahan yang dilakukan, muncul suatu ide yang kemudian dijadikan sebagai landasan untuk penulisan skripsi ini. Beberapa masalah yang telah ditentukan, selanjutnya akan dirumuskan ke dalam beberapa pertanyaan.

Rumusan masalah dalam penulisan skripsi ini adalah:

1. Bagaimana pengidentifikasian *outlier* dari data regresi berganda?
2. Bagaimana bentuk model regresi *robust* terbaik dengan adanya data *outlier* dengan menggunakan metode *Least Trimmed Squares (LTS) dan* metode *MM-Estimation* pada data regresi tersebut?
3. Bagaimanakah perbandingan yang dihasilkan dengan menggunakan metode *Least Trimmed Squares (LTS)* dan *MM-Estimation* pada proses regresi *robust* dilihat dari nilai R^2 dan residualnya?

3.3 Studi Pustaka

Pada tahapan ini dilakukan kajian sumber-sumber pustaka dengan cara mengumpulkan data serta informasi dari perpustakaan, serta internet yang berkaitan dengan permasalahan yang timbul yaitu persamaan regresi *robust* dengan metode *MM-Estimation* serta *Least Trimmed Squares* (LTS). Pada tahapan ini juga dilakukan pengumpulan konsep pendukung seperti definisi serta teorema-teorema yang mendukung untuk menyelesaikan permasalahan yang muncul, sehingga diperoleh suatu ide mengenai pemecahan masalah tentang metode *MM-Estimation* dan metode *Least Trimmed Squares* (LTS).

3.4 Analisis Pemecahan Masalah

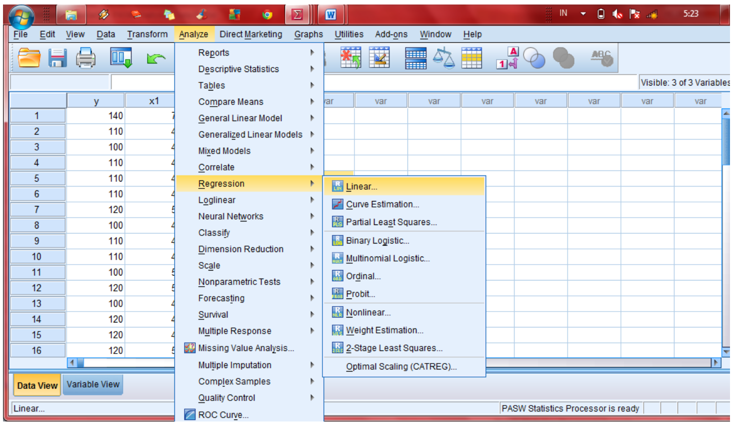
Dari beberapa sumber pustaka yang menjadi kajian dalam penulisan skripsi ini, diperoleh suatu pemecahan dari masalah yang muncul. Analisis dan pemecahan masalah dari permasalahan yang muncul adalah sebagai berikut:

Pendeteksian adanya data *outlier* menggunakan metode *Cook's Distance* dengan kriteria nilai *cook's* > $\mathrm{F}(0.5; \mathrm{k}, \mathrm{n}-\mathrm{k})$ dan *DfFITS* dengan kriteria nilai

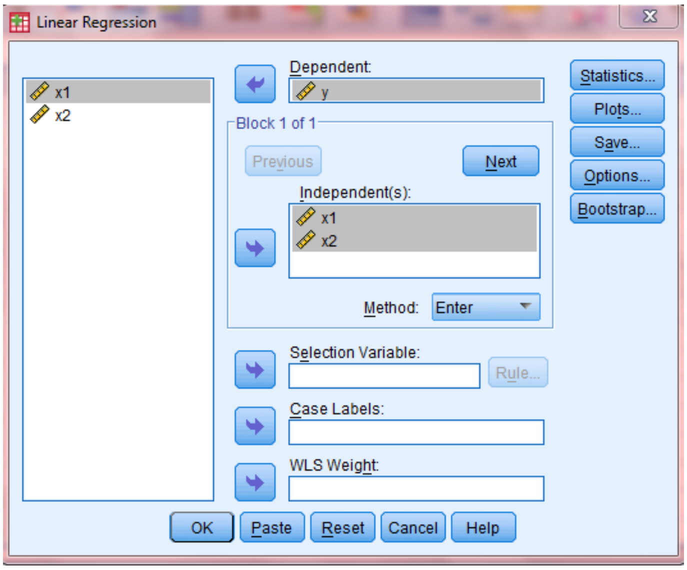
DfFITS > $2\sqrt{\frac{k+1}{n}}$. Untuk mendapatkan nilai *Cook's Distance* dan *DfFITS* dapat menggunakan bantuan SPSS 19. Adapun langkah menggunakan SPSS sebagai berikut:

1. Klik Analyze → Regression → Liniear

   Seperti tampilan dibawah ini:

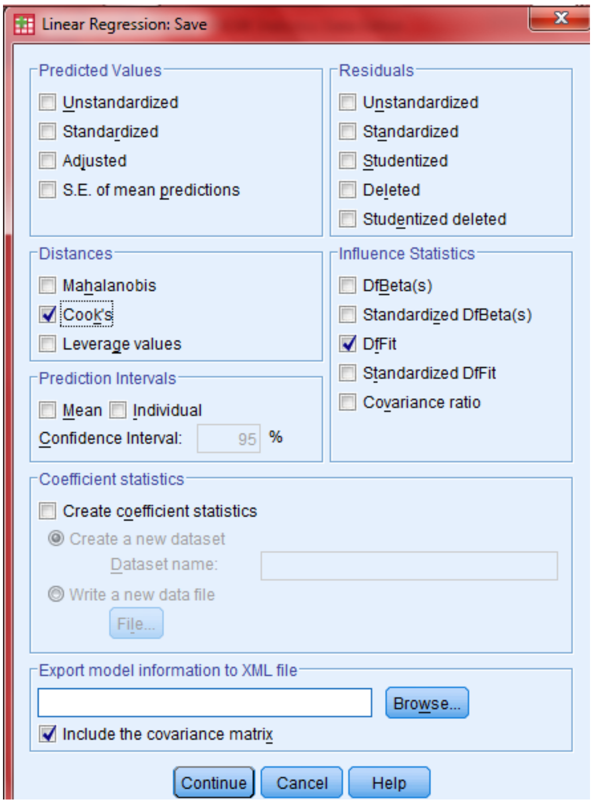


2. Klik variabel terikat → pindahkan ke kotak Dependent.
3. Klik variabel bebas → pindahkan ke kotak Independent(s).



4. Klik tombol *save*, ceklis *cook's* pada kolom *distance* dan DfFITS pada kolom *Influence Statistic*.

   Berikut adalah tampilannya:



5. Klik *Continue*
6. OK

Langkah selanjutnya menaksir nilai parameter dengan metode *Least Trimmed Squares* (LTS).

Adapun langkah-langkah pada metode metode *Least Trimmed Squares* (LTS) adalah menggunakan gabungan FAST LTS, C-*Step* dan FWLS yaitu :

a. menghitung estimasi parameter β_0, $\beta_1$dan β_2
b. menentukan n residual $\varepsilon_i^2 = (y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i})^2$ yang bersesuaian dengan β_0, $\beta_1$dan β_2

c. kemudian menghitung sejumlah $h_0 = \left[\frac{n}{2}\right] + \left[\frac{(k+1)}{2}\right]$ pengamatan dengan nilai ε_i^2 terkecil.

d. menghitung $\sum_{i=1}^{h_0} \varepsilon_i^2$

e. melakukan estimasi parameter β_0, $\beta_1$dan β_2 dari h_0 pengamatan.

f. menentukan n kuadrat residual $\varepsilon_i^2 = (y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i})^2$ kemudian menghitung sejumlah h_{baru} pengamatan dengan nilai ε_i^2 terkecil.

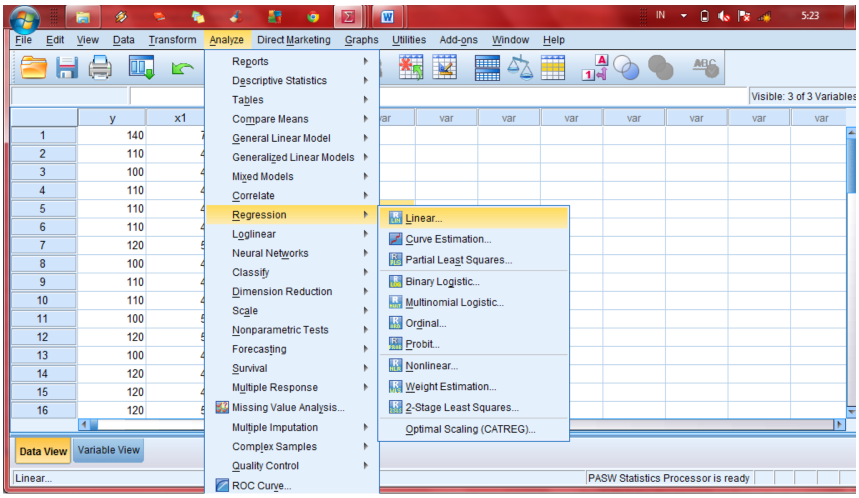
g. menghitung $\sum_{i=1}^{h_{baru}} \varepsilon_{(i)}^2$

h. melakukan C-*step* yaitu tahap *d* sampai *f* untuk mendapatkan fungsi objektif yang kecil dan konvergen.

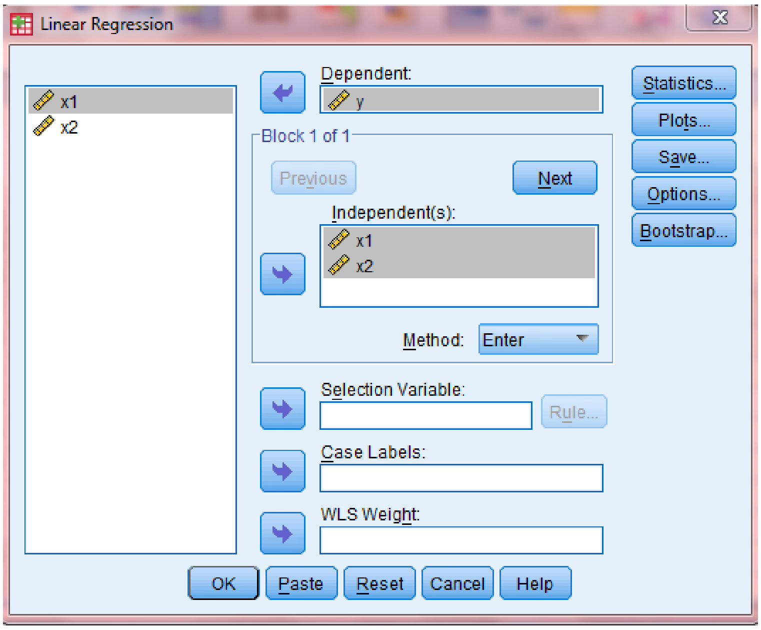
Tahapan yang digunakan FAST LTS, C-*Step* dan FWLS diatas digunakan untuk menghitung iterasi-iterasi dari metode *Least Trimmed Squares* (LTS), kemudian untuk estimasi parameternya digunakan bantuan *software* SPSS 19. Adapun langkah-langkah sebagai berikut:

1. Klik Analyze → Regression → Liniear

Seperti tampilan dibawah ini:



2. Klik variabel terikat → pindahkan ke kotak Dependent.
3. Klik variabel bebas → pindahkan ke kotak Independent(s).



4. Klik OK

Langkah selanjutnya menaksir nilai parameter dengan menggunakan metode *MM-estimation* dilakukan dengan menggunakan estimasi kuadrat terkecil dengan pembobot iteratif. Prosedur ini dinamakan *Iteratively Reweighted Least Squares (IRLS)*.

Adapun langkah-langkah dalam proses *MM-estimation* adalah:

a. Menghitung estimator awal koefisien $\hat{\beta}_j^{(1)}$dan residual $\varepsilon_i^{(1)}$ dari regresi robust (estimasi S) dengan bobot huber / bisquare.

b. Residual $\varepsilon_i^{(1)}$ pada langkah pertama digunakan untuk menghitung skala estimasi $\hat{\sigma}_s^{(1)}$ dan dihitung pula pembobot awal $w_i^{(1)}$.

c. Residual $\varepsilon_i^{(1)}$ dengan skala estimasi $\hat{\sigma}_s$ pada langkah kedua digunakan dalam iterasi awal sebagai penaksir WLS untuk menghitung koefisien regresi $\sum_{i=1}^{n} w_i^{(1)} \left(\frac{\varepsilon_i^{(1)}}{\hat{\sigma}_s^{(1)}}\right) x_i = 0,\ w_i^{(1)}$ yang merupakan pembobot Huber/bisquare.

d. Menghitung bobot baru $w_i^{(2)}$ dengan skala estimasi dari iterasi awal WLS.

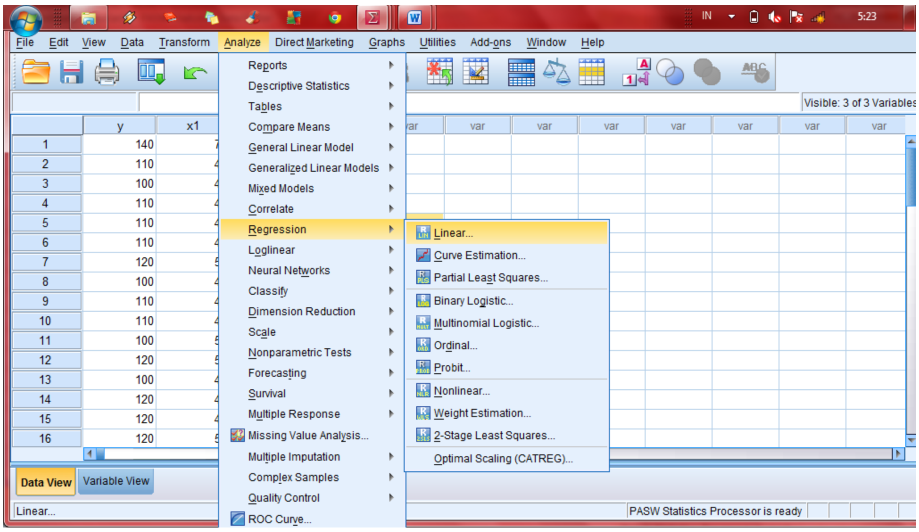
e. Mengulang langkah b, c, d (dengan skala estimasi tetap konstan) sampai mendapatkan $\sum_{i=1}^{n} \left|\varepsilon_i{}^{(m)}\right|$ konvergen (selisih $\hat{\beta}_j^{(m+1)}$ dan $\hat{\beta}_j^{(m)}$ mendekati 0, dengan banyak m iterasi).

Dalam mengestimasi parameter MM-*estimation* metode iterasi sangat diperlukan, karena residual tidak dapat dihitung sampai diperoleh model terbaik dan parameter regresi juga tidak dapat dihitung tanpa mengetahui nilai bobotnya. Kemudian untuk estimasi parameternya digunakan bantuan *software* SPSS 19.

Adapun langkah-langkah sebagai berikut:

1. Klik Analyze → Regression → Liniear

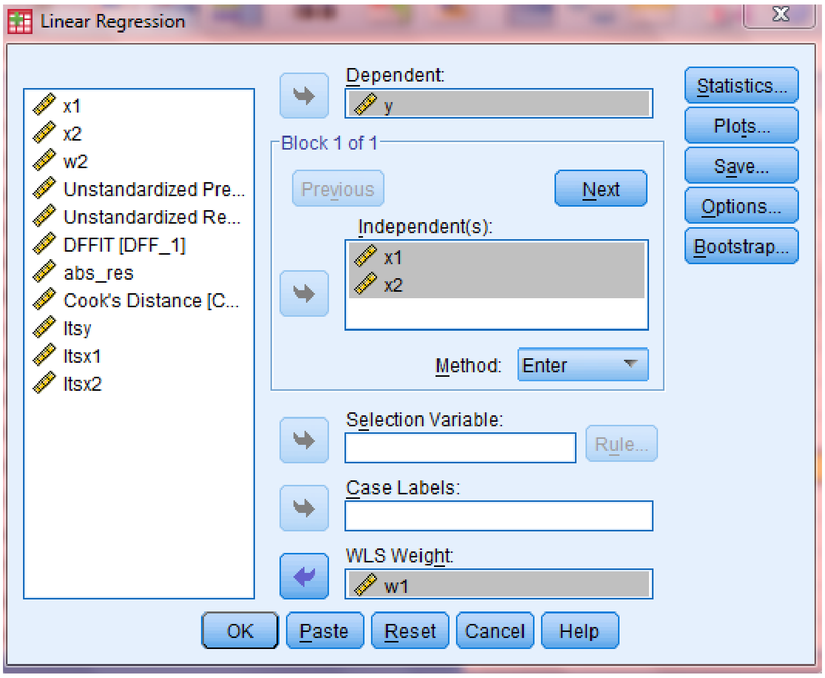
   Seperti tampilan dibawah ini:



2. Klik variabel terikat → pindahkan ke kotak Dependent.

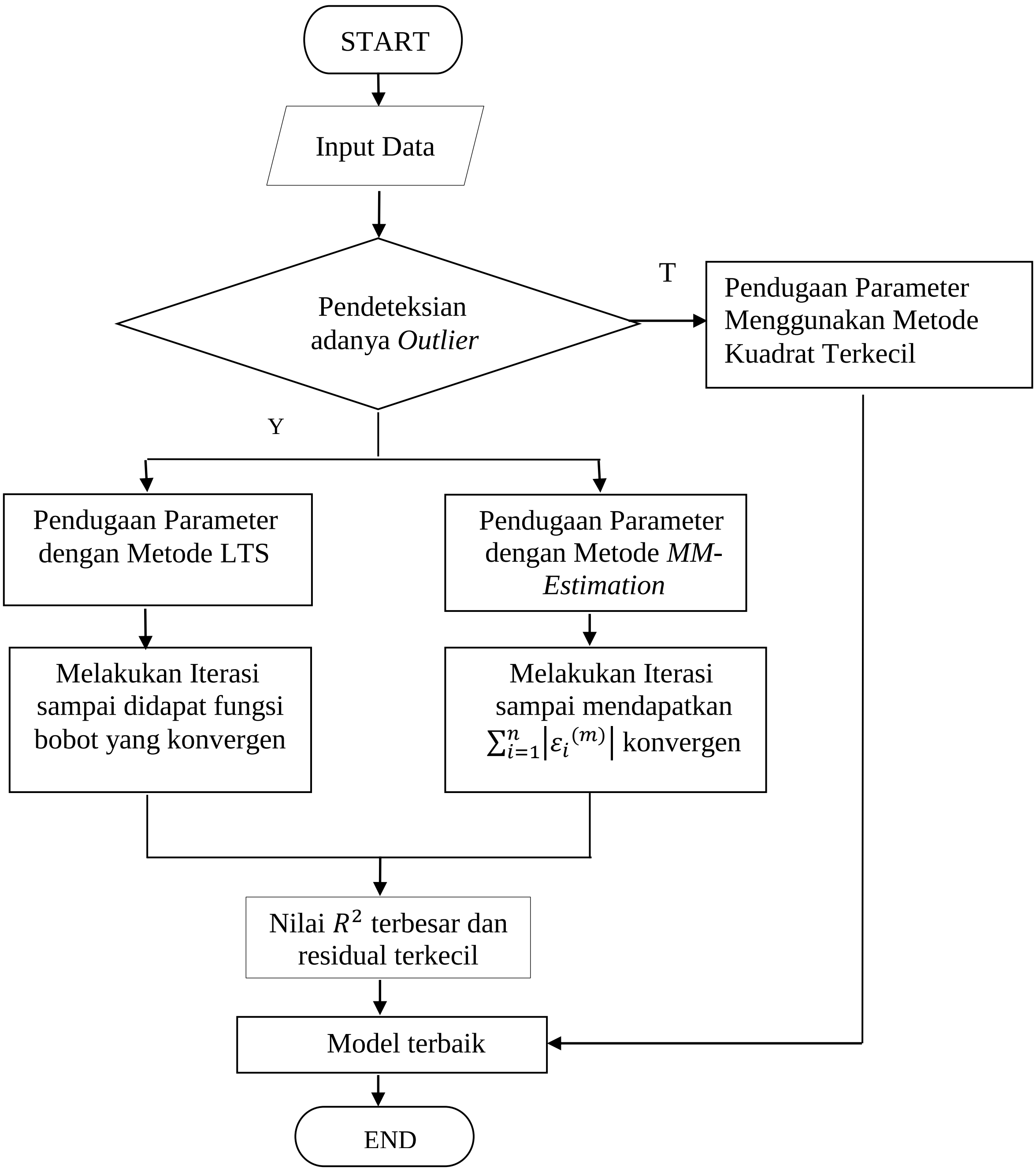
3. Klik variabel bebas → pindahkan ke kotak Independent(s).
4. Klik variabel pembobot → pindahkan kekotak WLS weight

   Berikut adalah tampilannya:



5. Klik OK

Setelah itu, membandingkan nilai R^2 dan residualnya yang diperoleh dari metode *Least Trimmed Squares (LTS)* dan metode *MM-estimation.* Perbandingan diperoleh dari hasil output hasil pengujian menggunakan *software* SPSS 19.



Gambar 3.1 Diagram Alir Langkah-langkah Estimasi Model

3.5 Penarikan Simpulan

Penarikan kesimpulan merupakan tahapan terakhir dalam metode penelitian. Penarikan kesimpulan diperoleh dari bahan-bahan pustaka dan pembahasan.

BAB 4
HASIL DAN PEMBAHASAN

Setelah dijabarkan mengenai beberapa hal tentang data *outlier*, selanjutnya pada bab IV ini berisi hasil penelitian dan pembahasan. Pembahasan difokuskan pada data yang diperkirakan terdapat data *outlier*. Disini akan dibahas mengenai cara mendeteksi apakah data tersebut terdapat *outlier* atau tidak, serta metode yang digunakan untuk mengatasi data *outlier* tersebut.

4.1 Hasil

Berikut adalah hasil analisis dalam penelitian yang telah dilakukan

4.1.1 Analisis Deskriptif

Contoh kasus dalam penelitian ini menggunakan data yang merupakan hasil penelitian dari 40 responden yaitu pengaruh usia dan indeks massa tubuh (IMT) terhadap tekanan darah sistolik. Responden yang dituju adalah pasien dari Puskesmas Sekaran Kota Semarang. Data ini terdiri dari data yaitu : usia dan indeks massa tubuh (IMT) sebagai variabel independent dan tekanan darah sistolik sebagai variabel dependent. Data terlampir pada lampiran 1.

Sebelum menganalisis menggunakan regresi akan diperiksa terlebih dahulu apakah data tersebut terdapat *outlier*. Jika dideteksi terdapat data *outlier*, untuk mengatasinya dapat digunakan regresi *robust* metode *Least Trimmed Square* (LTS) dan MM-*estimation.* Pada pengolahan data ini menggunakan bantuan *software Microsoft Excel* dan SPSS 19.

4.1.2 Uji Distribusi Normal Variabel Residual dengan *Kolmogorov Smirnov*

Metode *Kolmogorov Smirnov* menggunakan data dasar yang belum diolah dalam tabel distribusi frekuensi.

Hipotesis:

H_0 = Sampel berasal dari populasi yang berdistribusi Normal

H_1 = Sampel tidak berasal dari populasi yang berdistribusi Normal

Setelah dilakukan pengecekan menggunakan software SPSS didapatkan output yang telah disajikan dalam tabel 4.1 seperti berikut:

Tabel 4.1 Uji Test Distribusi Normal Variabel Residual

One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual
N		40
Normal Parameters[a,b]	Mean	.0000000
	Std. Deviation	5.84902572
Most Extreme Differences	Absolute	.080
	Positive	.045
	Negative	-.080
Kolmogorov-Smirnov Z		.504
Asymp. Sig. (2-tailed)		.961

a. Test distribution is Normal.

b. Calculated from data.

Intepretasi data

Dari uji *Kolmogorov Smirnov* diperoleh hasil nilai signifikan untuk uji distribusi normal yaitu 0,961 > 0,05.

Kesimpulan

Karena nilai signifikansi untuk uji test distribusi normal lebih dari 0,05 maka dapat disimpulkan bahwa data mempunyai distribusi normal.

4.1.3 Uji Asumsi Multikolinearitas

Pengujian selanjutnya adalah uji multikoliniearitas sebagai uji asumsi yang perlu dipenuhi dalam regresi berganda.

Hipotesis

H_0: model regresi memiliki masalah multikoliniearitas

H_1: model regresi tidak memiliki masalah multikoliniearitas

Setelah dilakukan pengecekan menggunakan *software* SPSS 19 didapatkan *output* yang telah disajikan dalam tabel 4.2 seperti berikut:

Tabel 4. 2 Nilai *Tolerance* dan VIF

Model	Collinearity Statistics	
	Tolerance	VIF
(Constant)		
Usia	0,602	1,660
IMT	0,602	1,660

Intepretasi Data

Dari tabel 4.2 terlihat bahwa variabel usia nilai *tolerance* nya adalah 0,602 dan VIFnya adalah 1,660. Variabel indeks massa tubuh (IMT) nilai *tolerance* nya adalah 0,602 dan VIFnya adalah 1,660.

Kesimpulan

Dari tabel 4.2 terlihat bahwa semua variabel prediktor mempunyai nilai *tolerance* lebih dari 0,1 dan memiliki nilai VIF kurang dari 10 sehingga bisa diduga bahwa antar variabel prediktor tidak terjadi persoalan multikolinearitas. Untuk itu pada penelitian ini digunakan semua variabel prediktor untuk pemodelan.

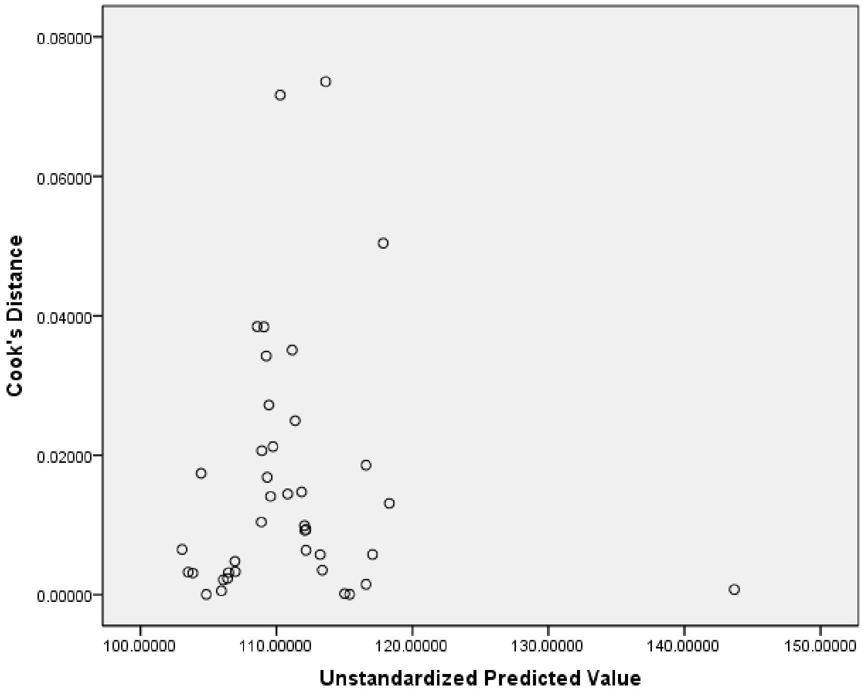
Karena antar variabel tidak mengandung multikolinearitas maka langkah selanjutnya yaitu pendeteksian *outlier*.

4.1.4 Pendeteksian *Outlier*

Pada penelitian ini, untuk mendeteksinya digunakan metode *cook's distance* dan metode *DfFITS.*

4.1.4.1 Metode Cook's Distance

Deteksi *outlier* yang pertama adalah melihat nilai *Cook's Distance*. Sebelumnya akan disajikan gambar *scatter plot* yang menyajikan *Cook's Distance* dan variabel *Unstandaized Predicted Value.*



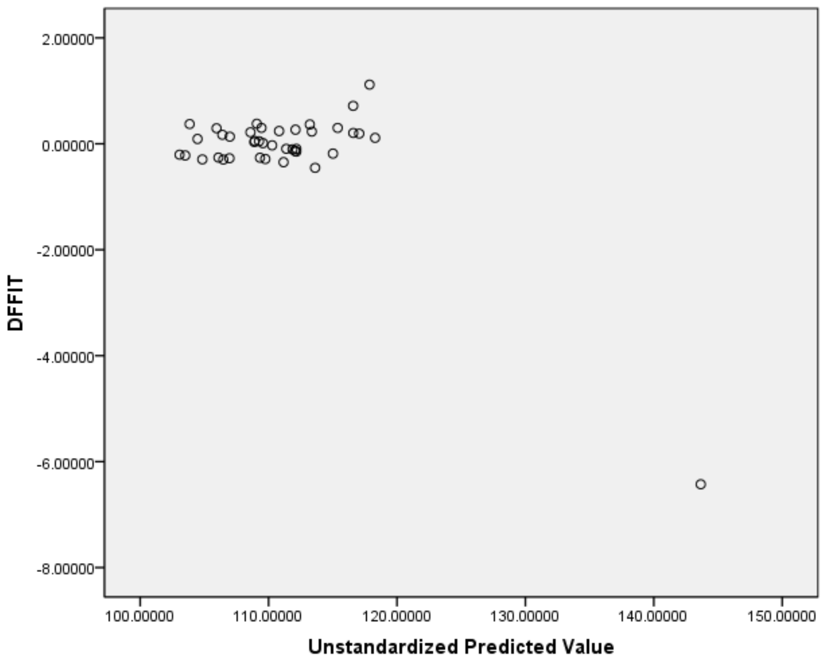
Gambar 4.1. *Scatter Plot* antara *Cook's* vs *Unstandaized Predicted Value*

Dari plot diatas menunjukan bahwa ada titik yang menjauh dari titik lainnya, artinya ada data yang teridentifikasi sebagai *outlier*. Maka perlu dilakukan pengecekan data.

Pada metode *Cook's Distance* berlaku ketentuan jika nilai *Cook's Distance* dari masing-masing data lebih dari $F(0{,}5; p, n-p)$, maka data tersebut merupakan *outlier*. Diperoleh bahwa $F(0{,}5; p, n-p)$ sebesar $0{,}803395$, jadi dari lampiran 2 terlihat bahwa tidak ada data yang mempunyai nilai *Cook's Distance* lebih dari 0,803395. Data dilampirkan pada lampiran. Untuk lebih jelasnya akan disajikan gambar yang menyajikan *Cook's Distance* dan *Unstandaized Predicted Value*.

***4.1.4.2 Metode* DfFITS**

Deteksi *outlier* selanjutnya adalah melihat nilai DfFIT (*Difference in fit Standardized*). Sebelumnya akan disajikan gambar *scatter plot* yang menyajikan DfFITS dan variabel *Unstandaized Predicted Value.*



Gambar 4.2. *Scatter Plot* antara DfFITS vs *Unstandaized Predicted Value*

Dari plot diatas menunjukan bahwa ada titik yang menjauh dari titik lainnya, artinya ada data yang teridentifikasi sebagai *outlier*. Maka perlu pengecekan data.

Selanjutnya pada hasil pengolahan data menggunakan metode DfFIT untuk masing-masing data seperti pada lampiran 3. Dengan ketentuan jika nilai DfFIT masing-masing data yang lebih dari $2\sqrt{\frac{k+1}{n}}$ maka dikategorikan sebagai *outlier*. Batas nilai penentuan berdasarkan DfFITS > $0{,}547723$ merupakan data *outlier*. Dari data pada lampiran 3 terlihat data yang mempunyai nilai DfFIT > 0,447221 dan menjadi *outlier* adalah data ke-1, ke-23 dan ke-28. Dengan nilai DfFITS data ke-1 = 6,42822, data ke-23 = 0,71722 dan data ke-28 = 1,11782. Oleh karena itu, perlu dilakukan analisis regresi menggunakan metode yang *robust* untuk data yang menggandung *outlier*, agar hasil regresi yang dihasilkan lebih tepat dan efisien.

Langkah selanjutnya adalah melakukan analisis regresi untuk mendapatkan nilai estimasi parameter dari data tersebut menggunakan metode *robust Least Trimmed Square* (LTS) dan metode *MM-Estimation.* Selanjutnya dapat dibandingkan metode mana yang lebih efektif digunakan untuk menyelesaikan masalah regresi, perbandingan metode dengan melihat nilai R^2 dan residualnya.

4.1.5 Metode *Least Trimmed Square*

Penerapan metode *Least Trimmed Square* (LTS) memerlukan beberapa iterasi untuk mendapatkan model terbaik. Pada iterasi 1 diperleh persamaan model $\hat{y} = 65{,}323 + 0{,}665X_2 + 0{,}618X_1$ dan $\sum_{i=1}^{ho} \varepsilon_i^2 = 1334{,}233$. Karena

$h_o = \left[\frac{n}{2}\right] + \left[\frac{(k+1)}{2}\right] = 21$, maka pada iterasi selanjutnya digunakan data sebanyak 21 dengan ε_i^2 dari yang terkecil. Estimasi dilakukan dengan menggunakan bantuan *software Microsoft Excel* dan SPSS 19, dapat dilihat pada lampiran 4.

Selanjutnya iterasi 2 yang terdiri dari 12 data, diperoleh persamaan model $\hat{y} = 67{,}141 + 0{,}649X_1 + 0{,}587X_2$ dan $\sum_{i=1}^{ho} \varepsilon_i^2 = 199{,}2524$. Estimasi dilakukan dengan menggunakan bantuan *software Microsoft Excel* dan SPSS 19, dapat dilihat pada lampiran 5. Iterasi 2 ini merupakan iterasi terakhir untuk data tersebut. Hal ini dikarenakan data *outlier*nya tidak termasuk dalam iterasi selanjutnya.

Jika ditulis dalam satu tabel, penyelesaian menggunakan metode *Least Trimmed Square* (LTS) diperoleh penaksir robust sebagai berikut :

Tabel 4.3 Hasil iterasi *Least Trimmed Square*

Tahap	*N*	*H*	$\hat{\beta}_0^{(n)}$	$\hat{\beta}_1^{(n)}$	$\hat{\beta}_2^{(n)}$	$\sum_{i=1}^{ho} \varepsilon_i^2$
1	40	21	65,323	0,665	0,618	1334,233
2	21	12	67,141	0,649	0,587	199,2524

Berdasarkan tabel diatas, jelas terlihat $\sum_{i=1}^{ho} \varepsilon_i^2$ yang terkecil adalah pada iterasi kedua. Itu artinya persamaan yang paling baik yang diperoleh menggunakan metode *Least Trimmed Square* (LTS) adalah model $\hat{y} = 67{,}141 + 0{,}649X_1 + 0{,}587X_2$.

4.1.5.1 Uji Parameter LTS serentak

Uji parameter serentak digunakan untuk mengetahui ada tidaknya pengaruh variabel prediktor terhadap variabel respon.

Hipotesis

$H_0 : \beta_0 = \beta_1 = \beta_2 = 0$

$H_1 : \exists\ \beta_j \neq 0$; untuk suatu $j = 1,2$

Tabel 4.4 Analisis Variansi LTS

Source	Jumlah kuadrat	Df	Rataan	F	Sig
Regresi	1354,202	2	677,101	107,88	0,00
Error	119,252	19	6,276		
Total	1473,455	21			

Taraf signifikansi

$$\alpha = 0{,}05$$

Statistik Uji

$$F_{hitung} = \frac{RK_{regresi}}{RK_{residual}}$$

Kriteria Uji

Tolak H_0 jika $F_{hitung} > F(0{,}05; p; n - p - 1)$ atau tolak H_0 jika nilai signifikansi < α

Keputusan

$F(0{,}05; p; n - p - 1)$ = 3,554557

Karena $F_{hitung} > F(0{,}05; p; n - p - 1)$ yaitu 107,88 > 3,554557 maka tolak H_0

Kesimpulan

Karena H_0 ditolak maka $\exists\ \beta_j \neq 0, j = 1{,}2$ yang berpengaruh terhadap model. Dari tabel 4.4 terlihat bahwa nilai signifikan adalah 0,00 < $0{,}05$. Ini berarti variabel prediktor memberikan pengaruh secara serentak pada model.

4.1.5.2 Uji Parsial Parameter LTS

Uji parsial parameter digunakan untuk mengetahui pengaruh masing-masing variabel prediktor terhadap variabel respon.

Hipotesis

H_0: $\beta_j = 0$; untuk suatu j = 1,2

H_1: $\exists\ \beta_j \neq 0$; untuk suatu $j = 1{,}2$

Tabel 4.5 *Coefficient* LTS

Model	Unstandardize Coefficient		Standardize Coefficient	t	Sig
	B	Std. Error			
Constant	67,141	3,449		19,469	0,00
x1	0,649	0,092	0,594	7,068	0,00
x2	0,587	0,106	0,466	5,544	0,00

Taraf signifikansi

$$\alpha = 0{,}05$$

Statistik Uji

$$t_{hitung} = \frac{\widehat{\beta_j}}{\sqrt{stdev(\widehat{\beta_j})}}$$

dengan

$\widehat{\beta_j}$ adalah taksiran parameter β_j

$stdev(\widehat{\beta_j})$ adalah taksiran standar deviasi dari β_i.

Kriteria pengujian

H_0 ditolak jika $|t_{hit}| > t_{\alpha/2,v}$ atau tolak H_0 jika nilai signifikansi kurang dari α

Keputusan

Dari tabel 4.5 diperoleh nilai signifikan β_0 adalah 0,00. Nilai signifikan β_1 adalah 0,00. Nilai signifikan β_2 adalah 0,00.

Kesimpulan

Parameter β_0 nilai signifikannya 0,00 < 0,05 maka tolak H_0. Parameter β_1 nilai signifikannya 0,00 < 0,05 maka tolak H_0 atau H_1 diterima, $\beta_1 \neq 0$, artinya parameter β_1 mempunyai pengaruh terhadap variabel respon. Parameter β_2 nilai signifikannya 0,00 < 0,05 maka tolak H_0 atau H_1 diterima, $\beta_2 \neq 0$, artinya parameter β_2 mempunyai pengaruh terhadap variabel respon. Jadi variabel prediktor x_1, x_2 berpengaruh secara parsial terhadap model.

4.1.6 Metode MM-*Estimator*

Sama halnya dengan metode *Least Trimmed Square* (LTS), metode MM-*Estimator* juga memerlukan beberapa iterasi dalam pengerjaannya. Pada metode MM-Estimator, peneliti menggunakan pembobot *tukey*, maka $c = 4{,}685$. Karena metode MM-*estimation* merupakan gabungan dari metode M-*estimation* dan S-*estimation* maka untuk menyelesaikannya, langkah pertama yaitu mencari estimator S, kemudian menetapkan parameter-parameter regresi menggunakan metode M-*estimation*.

Pada iterasi 1, diperoleh parameter dari S-*estimator* seperti tabel berikut ini

Tabel 4.6 Parameter S-*estimator*

Parameter	Nilai
$\hat{\beta}_0^{(1)}$	65,323
$\hat{\beta}_1^{(1)}$	0,665
$\hat{\beta}_2^{(1)}$	0,618
Scale	33,356

Parameter dari S-*estimator* tersebut kemudian digunakan untuk mencari nilai residual $\varepsilon_i^{(1)}$ yang selanjutnya digunakan untuk memperoleh nilai pembobot awal $w_i^{(1)}$. Nilai $w_i^{(1)}$ pada iterasi 2 digunakan sebagai nilai WLS pada iterasi 3 untuk mendapatkan persamaan regresinya. Dengan bantuan *sofware* SPSS diperoleh output seperti lampiran 6.

Dari tabel tersebut diperoleh persamaan $\hat{y} = 65{,}308 + 0{,}666X_1 + 0{,}618X_2$. Dengan persamaan tersebut diperoleh nilai $\sum_{i=1}^{n}\left|\varepsilon_i{}^{(2)}\right| = 186{,}6879$. Selanjutnya lakukan iterasi lagi dengan $w_i^{(2)}$ sebagai WLS dan diperoleh output seperti pada lampiran 7.

Ternyata diperoleh persamaan yang sama dengan iterasi sebelumnya, itu berarti iterasinya cukup sebanyak dua kali. Dan persamaan yang dipakai adalah persamaan regresi yang terakhir, yaitu $\hat{y} = 65{,}308 + 0{,}666X_1 + 0{,}618X_2$ dan $\sum_{i=1}^{n}\left|\varepsilon_i{}^{(2)}\right| = 186{,}6879$.

Dengan kata lain, untuk data tersebut diperoleh penaksir *robust* sebagai berikut:

Tabel 4.7 Hasil Iterasi MM-*Estimation*

Tahap	$\hat{\beta}_0^{(n)}$	$\hat{\beta}_1^{(n)}$	$\hat{\beta}_2^{(n)}$	$\sum_{i=1}^{n}\left\|\varepsilon_i{}^{(m)}\right\|$
Estimator S	65,323	0,665	0,618	186,687
1	65,308	0,666	0,618	186,688
2	65,308	0,666	0,618	186,688

Berdasarkan tabel diatas, jelas terlihat $\sum_{i=1}^{n}\left|\varepsilon_i{}^{(m)}\right|$ sudah konvergen. Itu artinya persamaan yang paling baik yang diperoleh menggunakan metode MM-*Estimation* adalah $\hat{y} = 65{,}308 + 0{,}666X_1 + 0{,}618X_2$.

4.1.6.1 Uji Parameter MM-Estimation serentak

Uji parameter serentak digunakan untuk mengetahui ada tidaknya pengaruh variabel prediktor terhadap variabel respon.

Hipotesis

$H_0: \beta_0 = \beta_1 = \beta_2 = 0$

$H_1: \exists\ \beta_j \neq 0$; untuk suatu $j = 1{,}2$

Tabel 4.8 Analisis Variansi *MM-Estimation*

Source	Jumlah kuadrat	Df	Rataan	F	Sig
Regresi	1727,439	2	863,72	24,129	0,00
Error	1324,433	37	35,795		
Total	3051,872	39			

Taraf signifikansi

$$\alpha = 0{,}05$$

Statistik Uji

$$F_{hitung} = \frac{RK_{regresi}}{RK_{residual}}$$

Kriteria Uji

Tolak H_0 jika $F_{hitung} > F(0{,}05; p; n-p-1)$ atau tolak H_0 jika nilai signifikansi < α

Keputusan

$F(0{,}05; p; n-p-1)$ = 3,554557

Karena $F_{hitung} > F(0{,}05; p; n-p-1)$ yaitu 24,129 > 3,251924 maka tolak H_0

Kesimpulan

Karena H_0 ditolak maka $\exists\ \beta_j \neq 0, j = 1{,}2$ yang berpengaruh terhadap model. Dari tabel 4.8 terlihat bahwa nilai signifikan kurang dari 0,05. Ini berarti variabel prediktor memberikan pengaruh secara serentak pada model.

4.1.6.2 Uji Parsial Parameter MM-Estimation

Uji parsial parameter digunakan untuk mengetahui pengaruh masing-masing variabel prediktor terhadap variabel respon.

Hipotesis

H_0: $\beta_j = 0$; untuk suatu j = 1,2

H_1: $\exists\ \beta_j \neq 0$; untuk suatu $j = 1{,}2$

Tabel 4.9 Coefficient *MM-Estimation*

Model	Unstandardize Coefficient		Standardize Coefficient	T	Sig
	B	Std. Error			
Constant	65,308	7,208		9,06	0,00
x1	0,666	0,196	0,474	3,396	0,002
x2	0,616	0,241	0,358	2,563	0,015

Taraf signifikansi

$$\alpha = 0{,}05$$

Statistik Uji

$$t_{hitung} = \frac{\widehat{\beta_J}}{\sqrt{stdev(\widehat{\beta_J}\,)}}$$

dengan

$\widehat{\beta_J}$ adalah taksiran parameter β_j

$stdev(\widehat{\beta_J}\,)$ adalah taksiran standar deviasi dari β_i.

Kriteria pengujian

H_0 ditolak jika $|t_{hit}| > t_{\alpha/2,v}$ atau tolak H_0 jika nilai signifikansi kurang dari α

Keputusan

Dari tabel 4.9 diperoleh nilai signifikan β_0 adalah 0,00. Nilai signifikan β_1 adalah 0,002. Nilai signifikan β_2 adalah 0,015.

Kesimpulan

Parameter β_0 nilai signifikannya 0,00 < 0,05 maka tolak H_0. Parameter β_1 nilai signifikannya 0,002 < 0,05 maka tolak H_0 atau H_1

diterima, $\beta_1 \neq 0$, artinya parameter β_1 mempunyai pengaruh terhadap variabel respon. Parameter β_2 nilai signifikannya 0,015 < 0,05 maka tolak H_0 atau H_1 diterima, $\beta_2 \neq 0$, artinya parameter β_2 mempunyai pengaruh terhadap variabel respon. Jadi variabel prediktor x_1, x_2 berpengaruh secara parsial terhadap model.

4.1.7 Pemilihan Model Regresi Terbaik

Tahapan pemilihan model regresi terbaik dimulai dengan melihat nilai R^2 dan nilai residual dari model regresinya.

Tabel 4.10 Perbandingan Nilai R^2 dari metode LTS dan metode MM-*Estimation*

No	Metode Regresi	$\hat{\beta}_1^{(n)}$	$\hat{\beta}_1^{(n)}$	$\hat{\beta}_2^{(n)}$	R^2
1	LTS	67,141	0,649	0,587	0,919
2	MM-Estimation	65,308	0,666	0,618	0,566

Dari tabel 4.10 dapat dilihat bahwa pada model regresi pada metode *Least Trimmed Square* nilai R^2 nya adalah 0,919. Sedangkan pada metode MM-*Estimation* nilai R^2 nya adalah 0,566. Ini berarti model regresi pada metode *Least Trimmed Square* memberikan pengaruh yang lebih besar yaitu sebanyak 91,9% dibandingan dengan metode MM-*Estimation* yang hanya memberikan pengaruh sebanyak 56,6%. Dengan kata lain, metode *Least Trimmed Square* (LTS) merupakan metode terbaik untuk mengestimasi parameter pada saat data terdeteksi mengandung *outlier* karena memiliki R^2 yang lebih banyak.

4.2 Pembahasan

Berdasarkan hasil penelitian, diketahui bahwa data yang diperoleh dari Puskesmas Sekaran Kota Semarang merupakan data diskrit yang meliputi 3 variabel yaitu usia, indeks masa tubuh (IMT) dan tekanan darah sistolik. Variabel residual dari data tersebut berdistribusi normal.

Hasil uji asumsi multikolinearitas menunjukkan bahwa data tidak mengalami multikolinearitas. Hal ini dilihat dari output VIF dan *tolerance* pada tabel 4.2. Dari tabel 4.2 terlihat bahwa semua nilai VIF kurang dari 10 dan nilai *tolerance* lebih dari 0,1 karena data tidak mengalami multikolinearitas maka semua variabel prediktor masuk ke dalam model.

Setelah memenuhi semua asumsi dalam regresi berganda, dilakukan pengecekan adanya *outlier* pada data tersebut. Pengecekan *outlier* pada penelitian ini menggunakan metode *Cook's Distance* dan metode DfFIT (*Difference in fit Standardized*). Hasil pengecekan metode *Cook's Distance* adalah data tidak terdeteksi adanya *outlier*. Karena tidak ada data yang mempunyai nilai *Cook's Distance* > 0,803395. Sedangkan pada metode DfFIT (*Difference in fit Standardized*) ada 3 data yang terlihat terdeteksi *outlier*. Batas nilai penentuan berdasarkan DfFITS adalah data yang nilainya lebih dari 0,547723 akan dikategorikan data *outlier*. Dari data pada lampiran 3 terlihat data yang mempunyai nilai DfFIT > 0,547723 dan menjadi *outlier* adalah data ke-1, ke-23 dan ke-28. Dengan nilai DfFITS data ke-1 = 6,42822, data ke-23 = 0,71722 dan data ke-28 = 1,11782. Oleh karena itu, perlu dilakukan analisis regresi menggunakan metode yang *robust* untuk data yang menggandung *outlier*.

Pada metode yang pertama dalam proses regresi robust *Least Trimmed Square* (LTS) dihasilkan model regresi $\hat{y} = 67{,}141 + 0{,}649X_1 + 0{,}587X_2$ dan $\sum_{i=1}^{ho} \varepsilon_i^2 = 199{,}2524$. Persamaan itu diperoleh dari beberapa iterasi. Iterasi yang terjadi pada data tersebut sebanyak 2 iterasi. Hal ini karena pada iterasi ke-3, data *outlier* tidak termasuk didalamnya, ini tidak sesuai dengan konsep regresi *robust* yaitu tetap mengikut sertakan data *outlier* dalam menemukan model persamaan regresi. Pada metode *Least Trimmed Square* (LTS) juga terjadi pemangkasan sejumlah data sebesar h, dimana nilai h didapat dari rumus $h_o = \left[\frac{n}{2}\right] + \left[\frac{(k+1)}{2}\right]$. Inilah yang menyebabkan nilai jumlah kuadrat residual pada metode ini semakin kecil dari iterasi 1 sampai iterasi akhir. Nilai R^2 yang didapatkan dalam metode ini adalah 0,919. Ini menunjukan bahwa variabel independent memberikan pengaruh yang cukup besar terhadap variabel dependen.

Proses regresi selanjutnya menggunakan metode MM-*Estimation* dan menghasilkan model $\hat{y} = 65{,}308 + 0{,}666X_1 + 0{,}618X_2$. Pada metode ini juga mengalami 2 iterasi untuk sampai pada model regresi terbaik metode MM-*Estimation.* Nilai residual yang di dapat dari metode ini termasuk residual yang cukup besar yaitu $\sum_{i=1}^{n} \left|\varepsilon_i^{(2)}\right| = 186{,}6879$. Sedangkan nilai R^2 yang didapatkan dalam metode MM-*Estimation* adalah 0,566. Ini menunjukan bahwa model regresi yang diperoleh tadi lemah atau kurang baik digunakan untuk memprediksi.

Pada penelitian ini, peneliti membandingkan nilai R^2 dari masing-masing model regresi pada metode *Least Trimmed Square* dan metode MM-*Estimation.* Karena nilai R^2 dari metode *Least Trimmed Square* lebih besar dibandingkan metode MM-*Estimation*, maka metode *Least Trimmed Square* lebih efektif jika diabndingkan metode MM-*Estimation*. Untuk nilai residual, jika nilai residualnya semakin besar atau dengan kata lain menjauhi nol (0), maka persamaan yang dihasilkan kurang baik.

Tabel 4.11 Nilai residual metode LTS dan metode MM-E*stimation*

No	Metode Regresi	$\sum_{i=1}^{n} \lvert\varepsilon_i\rvert$
1	LTS	44,89375
2	MM-Estimation	186,6879

Dari tabel 4.11 terlihat bahwa metode *Least Trimmed Square* (LTS) mempunyai nilai residual yang lebih kecil, hal ini disebabkan adanya pemangkasan (*trimmed*) data. Jadi, sama halnya dengan nilai R^2, nilai residual metode *Least Trimmed Square* (LTS) juga lebih baik jika dibandingkan dengan metode MM-E*stimation.*

BAB 5
PENUTUP

5.1 Simpulan

Berdasarkan rumusan masalah dan dari hasil pembahasan pada BAB IV maka dapat diambil beberapa simpulan berikut:

(1) Pengecekan outlier pada penelitian ini menggunakan metode *Cook's Distance* dan metode DfFIT (*Difference in fit Standardized*). DfFIT menampilkan nilai perubahan dari variabel yang diprediksikan bila kasus tertentu dikeluarkan yang sudah distandartkan. Sedangkan *Cook's Distance* menampilkan nilai jarak *cook* atau dengan kata lain menunjukkan besarnya pengaruh adanya data *outlier* terhadap semua estimator koefisien regresi. Kriteria data dikategorikan *outlier* jika nilai *Cook's Distance* $> F(0,5; p; n-p)$ dan nilai DfFIT $> 2\sqrt{\frac{p}{n}}$.

(2) Model regresi *robust* terbaik dengan adanya data *outlier* dengan menggunakan metode *Least Trimmed Squares* (LTS) yaitu

$$\hat{y} = 67{,}141 + 0{,}649X_1 + 0{,}587X_2$$

Sedangkan model regresi robust terbaik dengan adanya data *outlier* dengan menggunakan metode *MM-estimation* yaitu

$$\hat{y} = 65{,}308 + 0{,}666X_1 + 0{,}618X_2$$

Pada regresi robust, metode *Least Trimmed Square* (LTS) lebih efektif jika dibandingkan dengan metode MM-E*stimation* dilihat dari nilai R^2

dan residualnya. Hal ini disebabkan adanya pemangkasan (*trimmed*) terhadap data yang mempunyai residual besar.

5.2 Saran

Adapun saran yang dapat diberikan peneliti pada penelitian ini adalah:

(1) Pada pembahasan ini hanya mengkaji pada metode *Least Trimmed Square* (LTS) dan metode MM-*Estimation* dalam regresi *robust*,sehingga ada baiknya dilakukan pengujian untuk mengatasi adanya data *outlier* dengan menggunakan metode lain yang ada pada regresi *robust*.

(2) Perhitungan estimasi parameter dalam penelitian ini hanya menggunakan *software Microsoft Excel* dan SPSS 19, diharapkan pada penelitian selanjutnya dapat menggunakan program lain seperti Matlab.

DAFTAR PUSTAKA

Aunuddin. 1989. *Pendeteksian Outlier.* Statisticsanalyst. Wordpress.com

Bungawati, D., K. A. Pratama., & S. D. Richard. 2011. Kajian Indeks Massa Tubuh (IMT) Terhadap Tekanan Darah Pada Perawat Di Rumah Sakit Baptis Kediri. *Jurnal STIKES RS. Baptis Kediri*, 4(2):94-103.

Candraningtyas, S., D. Safitri, & D. Ispriyanti. 2013. Regresi Robust MM-Estimator Untuk Menangani *Outlier* Pada Regresi Linear Berganda. *Jurnal Gaussian*, 2(4): 395-404.

Chen, C. 2002. *Robust Regression and Outlier Detection with the ROBUSTREG Procedure*. SAS Institute Inc*:* Cary NC.

Cizek, P. 2013. Reweighted least trimmed squares: an alternative to one-step estimators. *Springer*. 22: 514-533.

Coskuntuncel, C. 2013. The Use of Alternative Regression Methods in Social Sciences and the Comparison of Least Squares and M Estimation Methods in Terms of the Determination of Coefficient. *Educational Sciences: Theory & Practice.* 13(4). 2151-2158.

Draper, N.R., & H. Smith. 1992. *Analisis Regresi Terapan*, Diterjemahkan oleh Bambang Sumantri. Gramedia: Jakarta.

Gujarati, N. D & Porter, C. D. 2010. *Dasar – Dasar Ekonometrika*. Translated by Mardanugraha, E. , Wardhani, S., Mangunsong, C.2010. Jakarta : Penerbit Salemba Empat.

Ghozali, Imam. 2006. *Aplikasi Analisis Multivariat Dengan Program SPSS.* Semarang: Badan Penerbit Universitas Diponegoro.

Huber, P. J. 1973. *Robust Regression: Asymptotics, conjecture and Monte Carlo, Ann*. Stat. 1(5):799-821.

Hampel. 1986. *Multivariate aregression S-estimation for Robust Estimation and Inference.* Ghent University.

Herawati, N., N. Khoirin, & S. Eri. 2011. Analisis Ketegaran Regresi Robust Terhadap Letak *Outlier*: Studi Perbandingan. *Bulletin of Mathematics*, 3(1): 49-60.

Larson, J. 2011. Our statistical intuitions may be misleading us: Why we need robust statistics. *Cambridge University Press*. 45(4). 460-474.

Lungan, R. 2006. *Aplikasi Statistika dan Hitung Peluang*. Yogyakarta : Graha Ilmu.

Maharani, I. F., N. Satyahadewi, & D. Kusnandar. 2014. Metode Ordinary Least Squares dan Least Trimmed Squares Dalam Mengestimasi Parameter Regresi Ketika Terdapat Outlier. *Buletin Ilmiah Mat. Stat. Dan Terapannya*. 3(3): 163-168.

Makkulau., L. Susanti., Purhadi, & M. Muhammad. 2010. Pendeteksian Outlier Dan Penentuan Faktor-Faktor Yang Mempengaruhi Produksi Gula Dan Tetes Tebu Dengan Metode Likelihood Displacement Statistic-Lagrange. *Jurnal TI*, 12(2): 25-100.

Montgomery, D. C., & E. A. Peck. 1992. *Introduction to Linear Regression Analysis* (2^{th} ed). John Wiley & Sons Inc: New York.

Permana, A. T. 2012. Perbandingan Metode Least Trimmed Square (LTS) dan Penduga-S Sebagai Metode Pendugaan Parameter Regresi Robust. *Jurnal Matematika Universitas Brawijaya*. Diakses tanggal 8 Maret 2015.

Pradewi, E. D. & Sudarno. 2012. Kajian Estimasi-M IRLS Menggunakan Fungsi Pembobot Huber Dan Bisquare Tukey Pada Data Ketahanan Pangan di Jawa Tengah. *Journal Media Statistik*. 5(1): 1-10.

Prasetyo, T., N. W. S. Wardhani, & W. H. Nugroho. 2012. Perbandingan Metode Robust Generalized-M Schweppe One-Step Estimator (GM-S1S) dan Metode Robust M-Estimator Untuk Menangani *Outlier* Pada Regresi Liniear Berganda. *Jurnal Matematika Universitas Brawijaya*. Diakses tanggal 8 Maret 2015.

Priyatno, D. 2013. *Analisis Korelasi, Regresi dan Multivariate dengan SPSS.* Yogyakarta: Penerbit Gava Media.

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

Sembiring, R. K. 1995.*Analisis Regresi. Bandung*: ITB.

Sembiring, R. K. 2003.*Analisis Regresi* (2^{th} ed). *Bandung*: ITB.

Smith. 1992. *Analisis Regresi Terapan.* Jakarta: PT. Gramedia Pustaka.

Soemarti. 2007. *Outlier (Outlier)*. Bandung: Universitas Padjadjaran wordpress.

Sukestiyarno. 2013. *Olah Data Penelitian Berbantuan SPSS* (4^{th} ed.). Semarang: Universitas Negeri Semarang.

Suyanti & Sukestiyarno. 2014. Deteksi Outlier Menggunakan Diagnosa Regresi Berbasis Estimator Parameter Robust. *Unnes Journal of Mathematics*. 3(2). 12-29.

Yohai, Victor J. 1987. *Hight Breakdown Point and Hight Efficiency Robust Estimates for Regression.* The Annals of Statistics.

Lampiran 1

Data Usia, Indeks Masa Tubuh (IMT) dan Tekanan Darah Sistolik (Puskesmas Sekaran Kota Semarang)

No	y	x_1	x_2
1	140	75	46,02
2	110	45	23,15
3	100	44	19,23
4	110	45	22,06
5	110	46	26,28
6	110	48	19,43
7	120	51	22,6
8	100	48	22,51
9	110	44	19,11
10	110	40	22,68
11	100	50	24,34
12	120	53	23,95
13	100	44	20
14	120	47	23,03
15	120	49	22,94
16	120	50	23,92
17	100	40	18,02
18	120	50	29,14
19	110	42	22,19
20	110	40	26,95

No	y	x_1	x_2
21	120	52	29,74
22	110	40	19,29
23	120	57	21,6
24	110	45	27,34
25	110	50	18,92
26	100	45	23,44
27	100	40	18,73
28	120	45	36,57
29	120	54	25,59
30	100	45	22,77
31	110	51	25,51
32	110	45	27,18
33	120	44	23,44
34	100	40	20,89
35	110	50	20,7
36	120	45	22,96
37	110	50	21,48
38	100	42	20,76
39	106	40	20,28
40	110	47	19,96

Keterangan :

Usia (X_1)
Indeks Masa Tubuh (X_2)
Tekanan Darah Sistolik (Y)

Lampiran 2

Nilai *Cook's Distance*

No	y	x_1	x_2	***Cook's***	**Jenis Data**
1	140	75	46,02	0,00075	Bukan
2	110	45	23,15	0,0141	Bukan
3	100	44	19,23	0,00316	Bukan
4	110	45	22,06	0,01042	Bukan
5	110	46	26,28	0,00637	Bukan
6	110	48	19,43	0,03425	Bukan
7	120	51	22,6	0,00576	Bukan
8	100	48	22,51	0,03508	Bukan
9	110	44	19,11	0,00233	Bukan
10	110	40	22,68	0,00054	Bukan
11	100	50	24,34	0,07358	Bukan
12	120	53	23,95	0,00006	Bukan
13	100	44	20	0,00481	Bukan
14	120	47	23,03	0,01446	Bukan
15	120	49	22,94	0,0092	Bukan
16	120	50	23,92	0,00354	Bukan
17	100	40	18,02	0,00648	Bukan
18	120	50	29,14	0,00148	Bukan
19	110	42	22,19	0,00331	Bukan
20	110	40	26,95	0,03846	Bukan
21	120	52	29,74	0,01312	Bukan
22	110	40	19,29	0,00309	Bukan
23	120	57	21,6	0,01858	Bukan
24	110	45	27,34	0,00938	Bukan
25	110	50	18,92	0,07166	Bukan
26	100	45	23,44	0,02122	Bukan
27	100	40	18,73	0,00324	Bukan
28	120	45	36,57	0,05043	Bukan
29	120	54	25,59	0,00578	Bukan
30	100	45	22,77	0,01686	Bukan
31	110	51	25,51	0,00015	Bukan
32	110	45	27,18	0,0099	Bukan
33	120	44	23,44	0,03841	Bukan
34	100	40	20,89	0,00004	Bukan
35	110	50	20,7	0,02494	Bukan
36	120	45	22,96	0,0272	Bukan
37	110	50	21,48	0,01474	Bukan
38	100	42	20,76	0,00214	Bukan
39	106	40	20,28	0,0174	Bukan
40	110	47	19,96	0,02066	Bukan

Lampiran 3

Nilai *DfFITS*

No	y	x_1	x_2	$DfFITS$	$\|DfFITS\|$	Jenis Data
1	140	75	46,02	-6,42822	6,42822	Outlier
2	110	45	23,15	0,01252	0,01252	Bukan
3	100	44	19,23	-0,29819	0,29819	Bukan
4	110	45	22,06	0,03192	0,03192	Bukan
5	110	46	26,28	-0,0941	0,09410	Bukan
6	110	48	19,43	0,04887	0,04887	Bukan
7	120	51	22,6	0,3656	0,36560	Bukan
8	100	48	22,51	-0,34652	0,34652	Bukan
9	110	44	19,11	0,17021	0,17021	Bukan
10	110	40	22,68	0,29283	0,29283	Bukan
11	100	50	24,34	-0,45273	0,45273	Bukan
12	120	53	23,95	0,30065	0,30065	Bukan
13	100	44	20	-0,27312	0,27312	Bukan
14	120	47	23,03	0,24129	0,24129	Bukan
15	120	49	22,94	0,26608	0,26608	Bukan
16	120	50	23,92	0,23061	0,23061	Bukan
17	100	40	18,02	-0,2059	0,20590	Bukan
18	120	50	29,14	0,20373	0,20373	Bukan
19	110	42	22,19	0,13535	0,13535	Bukan
20	110	40	26,95	0,21731	0,21731	Bukan
21	120	52	29,74	0,11222	0,11222	Bukan
22	110	40	19,29	0,37263	0,37263	Bukan
23	120	57	21,6	0,71722	0,71722	Outlier
24	110	45	27,34	-0,14637	0,14637	Bukan
25	110	50	18,92	-0,02869	0,02869	Bukan
26	100	45	23,44	-0,28639	0,28639	Bukan
27	100	40	18,73	-0,22	0,22000	Bukan
28	120	45	36,57	1,11782	1,11782	Outlier
29	120	54	25,59	0,19273	0,19273	Bukan
30	100	45	22,77	-0,26375	0,26375	Bukan
31	110	51	25,51	-0,18535	0,18535	Bukan
32	110	45	27,18	-0,13402	0,13402	Bukan
33	120	44	23,44	0,37873	0,37873	Bukan
34	100	40	20,89	-0,2945	0,29450	Bukan
35	110	50	20,7	-0,09336	0,09336	Bukan
36	120	45	22,96	0,30007	0,30007	Bukan
37	110	50	21,48	-0,10442	0,10442	Bukan
38	100	42	20,76	-0,26038	0,26038	Bukan
39	106	40	20,28	0,0916	0,09160	Bukan
40	110	47	19,96	0,05315	0,05315	Bukan

Lampiran 4 (Metode LTS)

Iterasi 1

NO	y	x_1	x_2	$\hat{y}$	e_i	e_i^2
1	140	75	46,02	143,655	-3,655	13,361
2	110	45	23,15	109,564	0,436	0,190
3	100	44	19,23	106,476	-6,476	41,934
4	110	45	22,06	108,890	1,110	1,232
5	110	46	26,28	112,164	-2,164	4,683
6	110	48	19,43	109,260	0,740	0,548
7	120	51	22,6	113,215	6,785	46,040
8	100	48	22,51	111,164	-11,164	124,629
9	110	44	19,11	106,401	3,599	12,950
10	110	40	22,68	105,948	4,052	16,419
11	100	50	24,34	113,625	-13,625	185,647
12	120	53	23,95	115,379	4,621	21,349
13	100	44	20	106,952	-6,952	48,325
14	120	47	23,03	110,820	9,180	84,271
15	120	49	22,94	112,095	7,905	62,494
16	120	50	23,92	113,366	6,634	44,015
17	100	40	18,02	103,067	-3,067	9,408
18	120	50	29,14	116,593	3,407	11,611
19	110	42	22,19	106,975	3,025	9,149
20	110	40	26,95	108,588	1,412	1,995
21	120	52	29,74	118,294	1,706	2,912
22	110	40	19,29	103,852	6,148	37,794
23	120	57	21,6	116,587	3,413	11,647
24	110	45	27,34	112,154	-2,154	4,641
25	110	50	18,92	110,275	-0,275	0,075
26	100	45	23,44	109,743	-9,743	94,932
27	100	40	18,73	103,506	-3,506	12,293
28	120	45	36,57	117,860	2,140	4,579
29	120	54	25,59	117,058	2,942	8,653
30	100	45	22,77	109,329	-9,329	87,033
31	110	51	25,51	115,014	-5,014	25,136
32	110	45	27,18	112,055	-2,055	4,224
33	120	44	23,44	109,078	10,922	119,286
34	100	40	20,89	104,841	-4,841	23,439
35	110	50	20,7	111,375	-1,375	1,891
36	120	45	22,96	109,447	10,553	111,375
37	110	50	21,48	111,857	-1,857	3,449
38	100	42	20,76	106,091	-6,091	37,103
39	106	40	20,28	104,464	1,536	2,358
40	110	47	19,96	108,922	1,078	1,162
∑	4446	1878	944,7	4446	-3,8E-13	1334,233

$$h_0 = \frac{(3n + p + 1}{4} = 21$$

$$\sum_{i=1}^{ho} \varepsilon_i^2 = 1334,233$$

Iterasi 2

No	y	x_1	x_2	$\hat{y}$	e_i	e_i^2
1	110	50	18,92	110,7056	-0,7056	0,4978
2	110	45	23,15	109,9440	0,0560	0,0031
3	110	48	19,43	109,7069	0,2931	0,0859
4	110	47	19,96	109,3690	0,6310	0,3982
5	110	45	22,06	109,3040	0,6960	0,4845
6	110	50	20,7	111,7508	-1,7508	3,0653
7	110	40	26,95	108,9300	1,0700	1,1449
8	106	40	20,28	105,0133	0,9867	0,9736
9	120	52	29,74	118,3574	1,6426	2,6981
10	110	50	21,48	112,2088	-2,2088	4,8789
11	110	45	27,18	112,3105	-2,3105	5,3385
12	120	45	36,57	117,8245	2,1755	4,7329
13	110	45	27,34	112,4045	-2,4045	5,7815
14	110	46	26,28	112,4311	-2,4311	5,9103
15	120	54	25,59	117,2186	2,7814	7,7360
16	110	42	22,19	107,4330	2,5670	6,5893
17	100	40	18,02	103,6862	-3,6862	13,5879
18	120	50	29,14	116,7069	3,2931	10,8445
19	120	57	21,6	116,8229	3,1771	10,0939
20	100	40	18,73	104,1031	-4,1031	16,8354
21	110	44	19,11	106,9226	3,0774	9,4704
22	140	75	46,02	142,8463	-2,8463	8,1016
$\sum$	2476	1050	540,44	2476,0000	-2E-13	119,2524

$$\sum_{i=1}^{h1} \varepsilon_i^2 = 119,2524$$

Lampiran 5 (Metode MM-*Estimation*)

Iterasi 1

NO	y	x_1	x_2	$\hat{y}$	e_i	e_i^2
1	140	75	46,02	143,655	-3,655	13,361
2	110	45	23,15	109,564	0,436	0,190
3	100	44	19,23	106,476	-6,476	41,934
4	110	45	22,06	108,890	1,110	1,232
5	110	46	26,28	112,164	-2,164	4,683
6	110	48	19,43	109,260	0,740	0,548
7	120	51	22,6	113,215	6,785	46,040
8	100	48	22,51	111,164	-11,164	124,629
9	110	44	19,11	106,401	3,599	12,950
10	110	40	22,68	105,948	4,052	16,419
11	100	50	24,34	113,625	-13,625	185,647
12	120	53	23,95	115,379	4,621	21,349
13	100	44	20	106,952	-6,952	48,325
14	120	47	23,03	110,820	9,180	84,271
15	120	49	22,94	112,095	7,905	62,494
16	120	50	23,92	113,366	6,634	44,015
17	100	40	18,02	103,067	-3,067	9,408
18	120	50	29,14	116,593	3,407	11,611
19	110	42	22,19	106,975	3,025	9,149
20	110	40	26,95	108,588	1,412	1,995
21	120	52	29,74	118,294	1,706	2,912
22	110	40	19,29	103,852	6,148	37,794
23	120	57	21,6	116,587	3,413	11,647
24	110	45	27,34	112,154	-2,154	4,641
25	110	50	18,92	110,275	-0,275	0,075
26	100	45	23,44	109,743	-9,743	94,932
27	100	40	18,73	103,506	-3,506	12,293
28	120	45	36,57	117,860	2,140	4,579
29	120	54	25,59	117,058	2,942	8,653
30	100	45	22,77	109,329	-9,329	87,033
31	110	51	25,51	115,014	-5,014	25,136
32	110	45	27,18	112,055	-2,055	4,224
33	120	44	23,44	109,078	10,922	119,286
34	100	40	20,89	104,841	-4,841	23,439
35	110	50	20,7	111,375	-1,375	1,891
36	120	45	22,96	109,447	10,553	111,375
37	110	50	21,48	111,857	-1,857	3,449
38	100	42	20,76	106,091	-6,091	37,103
39	106	40	20,28	104,464	1,536	2,358
40	110	47	19,96	108,922	1,078	1,162
$\sum$	4446	1878	944,7	4446	-3,8E-13	1334,233

Iterasi 2

$\hat{\sigma}_s = 33{,}356$

No	y	$\hat{y}$	$e_i^{(1)}$	$e_i^{(1)}/\hat{\sigma}_s$	$\psi(e_i^{(1)}/\hat{\sigma}_s)$	$w_i^{(1)}$
1	140	143,655	-3,655	-0,110	-0,109	0,999
2	110	109,564	0,436	0,013	0,013	1,000
3	100	106,476	-6,476	-0,194	-0,193	0,997
4	110	108,890	1,110	0,033	0,033	1,000
5	110	112,164	-2,164	-0,065	-0,065	1,000
6	110	109,260	0,740	0,022	0,022	1,000
7	120	113,215	6,785	0,203	0,203	0,996
8	100	111,164	-11,164	-0,335	-0,331	0,990
9	110	106,401	3,599	0,108	0,108	0,999
10	110	105,948	4,052	0,121	0,121	0,999
11	100	113,625	-13,625	-0,408	-0,402	0,985
12	120	115,379	4,621	0,139	0,138	0,998
13	100	106,952	-6,952	-0,208	-0,208	0,996
14	120	110,820	9,180	0,275	0,273	0,993
15	120	112,095	7,905	0,237	0,236	0,995
16	120	113,366	6,634	0,199	0,198	0,996
17	100	103,067	-3,067	-0,092	-0,092	0,999
18	120	116,593	3,407	0,102	0,102	0,999
19	110	106,975	3,025	0,091	0,091	0,999
20	110	108,588	1,412	0,042	0,042	1,000
21	120	118,294	1,706	0,051	0,051	1,000
22	110	103,852	6,148	0,184	0,184	0,997
23	120	116,587	3,413	0,102	0,102	0,999
24	110	112,154	-2,154	-0,065	-0,065	1,000
25	110	110,275	-0,275	-0,008	-0,008	1,000
26	100	109,743	-9,743	-0,292	-0,290	0,992
27	100	103,506	-3,506	-0,105	-0,105	0,999
28	120	117,860	2,140	0,064	0,064	1,000
29	120	117,058	2,942	0,088	0,088	0,999
30	100	109,329	-9,329	-0,280	-0,278	0,993
31	110	115,014	-5,014	-0,150	-0,150	0,998
32	110	112,055	-2,055	-0,062	-0,062	1,000
33	120	109,078	10,922	0,327	0,324	0,990
34	100	104,841	-4,841	-0,145	-0,145	0,998
35	110	111,375	-1,375	-0,041	-0,041	1,000
36	120	109,447	10,553	0,316	0,314	0,991
37	110	111,857	-1,857	-0,056	-0,056	1,000
38	100	106,091	-6,091	-0,183	-0,182	0,997
39	106	104,464	1,536	0,046	0,046	1,000
40	110	108,922	1,078	0,032	0,032	1,000

Iterasi 3

$\hat{\sigma}_s = 33{,}114$

No	y	$\hat{y}$	$e_i^{(2)}$	$e_i^{(2)}/\hat{\sigma}_s$	$\psi(e_i^{(2)}/\hat{\sigma}_s)$	$w_i^{(2)}$
1	140	143,6553	-3,6553	-0,1104	-0,1103	0,9989
2	110	109,5640	0,4360	0,0132	0,0132	1,0000
3	100	106,4756	-6,4756	-0,1956	-0,1949	0,9965
4	110	108,8902	1,1098	0,0335	0,0335	0,9999
5	110	112,1641	-2,1641	-0,0654	-0,0653	0,9996
6	110	109,2597	0,7403	0,0224	0,0224	1,0000
7	120	113,2147	6,7853	0,2049	0,2041	0,9962
8	100	111,1637	-11,1637	-0,3371	-0,3336	0,9897
9	110	106,4014	3,5986	0,1087	0,1086	0,9989
10	110	105,9479	4,0521	0,1224	0,1222	0,9986
11	100	113,6252	-13,6252	-0,4115	-0,4051	0,9846
12	120	115,3795	4,6205	0,1395	0,1393	0,9982
13	100	106,9516	-6,9516	-0,2099	-0,2091	0,9960
14	120	110,8201	9,1799	0,2772	0,2753	0,9930
15	120	112,0947	7,9053	0,2387	0,2375	0,9948
16	120	113,3656	6,6344	0,2004	0,1996	0,9963
17	100	103,0672	-3,0672	-0,0926	-0,0926	0,9992
18	120	116,5925	3,4075	0,1029	0,1028	0,9990
19	110	106,9752	3,0248	0,0913	0,0913	0,9992
20	110	108,5876	1,4124	0,0427	0,0426	0,9998
21	120	118,2937	1,7063	0,0515	0,0515	0,9998
22	110	103,8523	6,1477	0,1857	0,1851	0,9969
23	120	116,5872	3,4128	0,1031	0,1030	0,9990
24	110	112,1542	-2,1542	-0,0651	-0,0650	0,9996
25	110	110,2747	-0,2747	-0,0083	-0,0083	1,0000
26	100	109,7433	-9,7433	-0,2942	-0,2919	0,9921
27	100	103,5061	-3,5061	-0,1059	-0,1058	0,9990
28	120	117,8601	2,1399	0,0646	0,0646	0,9996
29	120	117,0584	2,9416	0,0888	0,0888	0,9993
30	100	109,3291	-9,3291	-0,2817	-0,2797	0,9928
31	110	115,0136	-5,0136	-0,1514	-0,1511	0,9979
32	110	112,0553	-2,0553	-0,0621	-0,0620	0,9996
33	120	109,0782	10,9218	0,3298	0,3266	0,9901
34	100	104,8414	-4,8414	-0,1462	-0,1459	0,9981
35	110	111,3750	-1,3750	-0,0415	-0,0415	0,9998
36	120	109,4466	10,5534	0,3187	0,3158	0,9908
37	110	111,8572	-1,8572	-0,0561	-0,0561	0,9997
38	100	106,0912	-6,0912	-0,1839	-0,1834	0,9969
39	106	104,4643	1,5357	0,0464	0,0464	0,9998
40	110	108,9222	1,0778	0,0325	0,0325	0,9999

Lampiran 6

1. Output Uji Distribusi Normal Variabel Residual dengan *Kolgomorov Smirnov*

One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual
N		40
Normal Parameters[a,b]	Mean	.0000000
	Std. Deviation	5.84902572
Most Extreme Differences	Absolute	.080
	Positive	.045
	Negative	-.080
Kolmogorov-Smirnov Z		.504
Asymp. Sig. (2-tailed)		.961

a. Test distribution is Normal.

b. Calculated from data.

2. Uji Multikoliniearitas

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	65.323	7.230		9.036	.000		
	x1	.665	.197	.473	3.382	.002	.602	1.660
	x2	.618	.242	.357	2.556	.015	.602	1.660

a. Dependent Variable: y

Lampiran 7

Iterasi 1 (Metode *Least Trimmed Square*)

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	x2, x1[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.751[a]	.564	.541	6.005

a. Predictors: (Constant), x2, x1

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1728.867	2	864.434	23.972	.000[a]
	Residual	1334.233	37	36.060		
	Total	3063.100	39			

a. Predictors: (Constant), x2, x1

b. Dependent Variable: y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	65.323	7.230		9.036	.000
	x1	.665	.197	.473	3.382	.002
	x2	.618	.242	.357	2.556	.015

a. Dependent Variable: y

Lampiran 8

Iterasi 2 (Metode *Least Trimmed Square*)

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	ltsx2, ltsx1[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: ltsy

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.959[a]	.919	.911	2.505

a. Predictors: (Constant), ltsx2, ltsx1

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1354.202	2	677.101	107.880	.000[a]
	Residual	119.252	19	6.276		
	Total	1473.455	21			

a. Predictors: (Constant), ltsx2, ltsx1

b. Dependent Variable: ltsy

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	67.141	3.449		19.469	.000
	ltsx1	.649	.092	.594	7.068	.000
	ltsx2	.587	.106	.466	5.544	.000

a. Dependent Variable: ltsy

Lampiran 9

Estimasi S (Metode MM-*Estimator*)

Model Information

Dependent Variable	y
Probability Distribution	Normal
Link Function	Identity

Case Processing Summary

	N	Percent
Included	40	100.0%
Excluded	0	.0%
Total	40	100.0%

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	y	40	100	140	111.15	8.862
Covariate	x1	40	40	75	46.95	6.300
	x2	40	18.02	46.02	23.6175	5.12379

Goodness of Fit[b]

	Value	df	Value/df
Deviance	1334.233	37	36.060
Scaled Deviance	40.000	37	
Pearson Chi-Square	1334.233	37	36.060
Scaled Pearson Chi-Square	40.000	37	
Log Likelihood[a]	-126.902		
Akaike's Information Criterion (AIC)	261.804		
Finite Sample Corrected AIC (AICC)	262.947		
Bayesian Information Criterion (BIC)	268.560		
Consistent AIC (CAIC)	272.560		

Dependent Variable: y
Model: (Intercept), x1, x2

a. The full log likelihood function is displayed and used in computing information criteria.

b. Information criteria are in small-is-better form.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
33.243	2	.000

Dependent Variable: y
Model: (Intercept), x1, x2

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	88.261	1	.000
x1	12.365	1	.000
x2	7.065	1	.008

Dependent Variable: y
Model: (Intercept), x1, x2

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	65.323	6.9531	51.695	78.951	88.261	1	.000
x1	.665	.1891	.294	1.036	12.365	1	.000
x2	.618	.2326	.162	1.074	7.065	1	.008
(Scale)	33.356[a]	7.4586	21.520	51.702			

Dependent Variable: y
Model: (Intercept), x1, x2

a. Maximum likelihood estimate.

Lampiran 10

Iterasi 1 (Metode MM-*Estimator*)

Model Information

Dependent Variable	y
Probability Distribution	Normal
Link Function	Identity
Scale Weight Variable	w1

Case Processing Summary

	N	Percent
Included	40	100.0%
Excluded	0	.0%
Total	40	100.0%

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	y	40	100	140	111.15	8.862
Covariate	x1	40	40	75	46.95	6.300
	x2	40	18.02	46.02	23.6175	5.12379
Scale Weight	w1	40	1	1	.9972732	.00355027

Goodness of Fit[b]

	Value	df	Value/df
Deviance	1324.574	37	35.799
Scaled Deviance	40.000	37	
Pearson Chi-Square	1324.574	37	35.799
Scaled Pearson Chi-Square	40.000	37	
Log Likelihood[a]	-126.812		
Akaike's Information Criterion (AIC)	261.623		
Finite Sample Corrected AIC (AICC)	262.766		
Bayesian Information Criterion (BIC)	268.379		
Consistent AIC (CAIC)	272.379		

Dependent Variable: y
Model: (Intercept), x1, x2

a. The full log likelihood function is displayed and used in computing information criteria.

b. Information criteria are in small-is-better form.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
33.389	2	.000

Dependent Variable: y
Model: (Intercept), x1, x2

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	88.739	1	.000
x1	12.465	1	.000
x2	7.100	1	.008

Dependent Variable: y
Model: (Intercept), x1, x2

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	65.308	6.9328	51.720	78.896	88.739	1	.000
x1	.666	.1886	.296	1.035	12.465	1	.000
x2	.618	.2318	.163	1.072	7.100	1	.008
(Scale)	33.114[a]	7.4046	21.364	51.328			

Dependent Variable: y
Model: (Intercept), x1, x2

a. Maximum likelihood estimate.

Lampiran 11

Iterasi 2 (Metode MM-*Estimator*)

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
33.391	2	.000

Dependent Variable: y
Model: (Intercept), x1, x2

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	88.746	1	.000
x1	12.467	1	.000
x2	7.101	1	.008

Dependent Variable: y
Model: (Intercept), x1, x2

Model Information

Dependent Variable	y
Probability Distribution	Normal
Link Function	Identity
Scale Weight Variable	w2

Case Processing Summary

	N	Percent
Included	40	100.0%
Excluded	0	.0%
Total	40	100.0%

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	y	40	100	140	111.15	8.862
Covariate	x1	40	40	75	46.95	6.300
	x2	40	18.02	46.02	23.6175	5.12379
Scale Weight	w2	40	1	1	.9972333	.00360220

Goodness of Fit[b]

	Value	df	Value/df
Deviance	1324.433	37	35.795
Scaled Deviance	40.000	37	
Pearson Chi-Square	1324.433	37	35.795
Scaled Pearson Chi-Square	40.000	37	
Log Likelihood[a]	-126.810		
Akaike's Information Criterion (AIC)	261.621		
Finite Sample Corrected AIC (AICC)	262.763		
Bayesian Information Criterion (BIC)	268.376		
Consistent AIC (CAIC)	272.376		

Dependent Variable: y
Model: (Intercept), x1, x2

a. The full log likelihood function is displayed and used in computing information criteria.

b. Information criteria are in small-is-better form.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	65.308	6.9326	51.720	78.896	88.746	1	.000
x1	.666	.1886	.296	1.035	12.467	1	.000
x2	.618	.2318	.163	1.072	7.101	1	.008
(Scale)	33.111[a]	7.4038	21.362	51.322			

Dependent Variable: y
Model: (Intercept), x1, x2

a. Maximum likelihood estimate.