PENDUGA *LEAST TRIMMED SQUARE* (LTS) PADA DATA YANG MENGANDUNG *OUTLIER*

SKRIPSI

Oleh
Hanief Anggara Atmagenta
NIM 121810101062

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2016

PENDUGA *LEAST TRIMMED SQUARE* (LTS) PADA DATA YANG MENGANDUNG *OUTLIER*

SKRIPSI

diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat untuk menyelesaikan Program Studi Matematika (S1) dan mencapai gelar sarjana sains

Oleh
Hanief Anggara Atmagenta
NIM 121810101062

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2016

PERSEMBAHAN

Dengan segala kerendahan hati dan puji syukur yang tak terhingga pada Allah SWT, skripsi ini saya persembahkan untuk:

1. Allah SWT yang telah memberikan kehidupan, pertolongan, rahmat berserta hidayah-Nya, dan kasih sayang yang tak ternilai harganya;
2. Ibunda Ratna Kurniasari dan Ayahanda Nyono Soemardi, yang telah mendidik, memberi dukungan, doa dan pengorbanan selama ini, serta memberi kasih sayang dan motivasi yang luar biasa;
3. Guru-guruku sejak taman kanak-kanak hingga para dosen Perguruan Tinggi yang terhormat, yang telah memberikan ilmu dan membimbing dengan penuh keasabaran;
4. Teman-teman seperjuangan angkatan 2012 (Bathics'12) yang telah menemani selama kurang lebih 4 tahun ini;
5. Almamater jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember;

MOTTO

"Barang siapa bertakwa kepada Allah maka Dia akan menjadikan jalan keluar baginya, dan memberinya rizki dari jalan yang tidak ia sangka, dan barang siapa yang bertawakkal kepada Allah melaksanakan kehendak-Nya, Dia telah menjadikan untuk sesuatu kadarnya"

(*Q.S. Ath-Thalaq: 2-3*)[*)]

"Kesuksesanmu tak bisa dibandingkan dengan orang lain, melainkan dibandingkan dengan dirimu sebelumnya"

(Jaya Setiabudi)

[*)] Departemen Agama Republik Indonesia 1998. *Al Qur'an dan Terjemahannya, Semarang: PT Kumudasmoro Grafindo*.

PERNYATAAN

Saya yang beranta tangan di bawah ini:

Nama : Hanief Anggara Atmagenta

NIM : 121810101062

Menyatakan dengan sesungguhnya bahwa karya ilmiah yang berjudul: "Penduga *Least Trimmed Square* (LTS) pada Data yang Mengandung *Outlier*" adalah benar-benar hasil karya sendiri, kecuali kutipan yang sudah saya sebutkan sumbernya, belum pernah diajukan pada institusi mana pun dan bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenarnya, tanpa ada tekanan dan paksaan dari pihak mana pun serta bersedia mendapat sanksi akademik jika ternyata di kemudian hari pernyataan ini tidak benar.

Jember, Desember 2016

Yang menyatakan,

Hanief Anggara Atmagenta

NIM 121810101062

SKRIPSI

PENDUGA *LEAST TRIMMED SQUARE* (LTS)
PADA DATA YANG MENGANDUNG *OUTLIER*

Oleh
Hanief Anggara Atmagenta
NIM 121810101062

Pembimbing

Dosen Pembimbing Utama : Dr. Alfian Futuhul Hadi S.Si, M.Si

Dosen pembimbing Anggota : Dian Anggraeni, S.Si, M.Si

PENGESAHAN

Skripsi berjudul "Penduga *Least Trimmed Square* (LTS) pada Data yang Mengandung *Outlier*" telah diuji dan disahkan pada:

Hari :

Tanggal :

Tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember

Tim Penguji

Ketua,

Dr. Alfian Futuhul Hadi, S.Si., M.Si
NIP. 197407192000121001

Sekretaris,

Dian Anggraeni, S.Si., M.Si
NIP. 198202162006042002

Penguji I,

Prof. Drs. I Made Tirta, M.Sc., Ph.D.
NIP. 195912201985031002

Penguji II,

Prof. Drs. Kusno, DEA., Ph.D.
NIP. 196101081986021001

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Jember,

Drs. Sujito, Ph.D.
NIP. 196102041987111001

RINGKASAN

Penduga *Least Trimmed Square* (LTS) pada Data yang Mengandung *Outlier*; Hanief Anggara Atmagenta, 121810101062; 2016; Halaman; Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Dalam analisis regresi, menduga parameter regresi secara otomatis juga mengestimasi model regresi. Metode yang umum digunakan dalam menduga parameter regresi adalah *Ordinar Least Square* (OLS). Namun jika terjadi penyimpangan asumsi pada data seperti adanya pencilan (*Outlier*) maka metode OLS ini tidak lagi efisien untuk digunakan. Data pencilan yaitu suatu data yang terletak jauh dari garis regresi. Adanya pencilan dapat mengakibatkan estimasi parameter regresi menjadi bias dan tidak efisien sehingga model regresi yanhg diperoleh tidak cocok (*fit*) terhadap data yang dimodelkan. Ada dua jenis data pencilan yang dapat mempengaruhi model regresi, diantaranya yaitu pencilan pada variabel bebas (x) dan pencilan pada variabel respon (y). Pencilan yang disebabkan karena variabel bebas (x) disebut *Laverage Point* dan pencilan yang disebabkan variabel respon (y) disebut *Vertical Outlier*. Model linier yang mengandung *Laverage Point* dan *Vertical Outlier* dapat diatasi dengan metode robust. Salah satu metode *robust* yaitu *Least Trimmed Square* (LTS).

Penelitian dilakukan terhadap dua jenis data yaitu data tanpa pencilan dan data yang mengandung pencilan (*Laverage Point* dan *Vertical Outlier*) dengan dua presentase pencilan yaitu 5% dan 10% pada 4 skenario simulasi, dengan masing-masing skenario simulasi 1 pencilan diberikan pada data variabel X_1, skenario simulasi 2 pencilan diberikan pada data variabel X_2, skenario simulasi 3 pencilan diberikan pada data variabel X_1 dan X_2, dan skenario simulasi 4 pencilan diberikan

pada data variabel X_1, X_2, dan Y. Jumlah data yang ditetapkan adalah 100. Data tersebut merupakan data simulasi dengan cara membangkitkan data dengan parameter yang telah ditetapkan. Simulasi dilakukan menggunakan program R versi 3.1.2 dengan bantuan fungsi lm, *ltsreg*, *sortlist* dan *Boxplot*. Data pada setiap skenario simulasi dianalisis menggunakan metode *Ordinary Least Square* dan metode *Least Trimmed Square* untuk mendapatkan estimasi parameter dan model. Selain untuk mengetahui pengaruh estimasi parameter dari kedua metode, penelitian ini juga bertujuan untuk mengetahui signifikan β dengan uji Hipotesis β melalui Uji T dan juga untuk mengetahui metode mana yang lebih baik untuk memodelkan data dengan membandingkan nilai Mean Square Error (MSE) model pada data pencilan (*Laverage Point* dan *Vertical Outlier*).

Hasil analisis menunjukkan bahwa pada data tanpa pencilan, metode OLS adalah model regresi yang lebih baik memodelkan data. Dapat dilihat hasil uji Hipotesis β melalui Uji T pada metode OLS mengalami signifikan β di semua skenario simulasi artinya variabel bebas (X) berpengaruh nyata terhadap variabel respon (Y). Sedangkan pada metode LTS terjadi ketidakstabilan terhadap pengujian Hipotesis β, karena dapat dilihat pada metode LTS tidak semua skenario simulasi mengalami signifikansi β. Dapat dilihat juga dari nilai MSE pada model regresi OLS lebih kecil daripada nilai MSE pada model LTS pada data tanpa pencilan. Namun pada data sudah terkontaminasi pencilan (*Laverage Point* dan *Vertical Outlier*) dengan masing-masing presentase yang telah ditetapkan sebelumnya, nilai MSE dari metode LTS lebih kecil dibanding dengan nilai MSE pada metode OLS. Maka dalam hal ini metode LTS yang paling baik dalam memodelkan data. Karena metode LTS lebih baik dalam mengatasi data pada pencilan (*Laverage Point* dan *Vertical Outlier*) maka metode LTS dapat dikatakan *robust* terhadap pencilan.

PRAKATA

Alhamdulillah, puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat beserta hidayah-Nya sehingga penyusunan skripsi yang berjudul "Penduga *LEAST TRIMMED SQUARE* (LTS) pada Data yang Mengandung *Outlier*" ini dapat terselesaikan. Sholawat serta salam semoga tetap tercurahkan kepada Nabi Muhammad SAW. Skripsi ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan strata 1 (S1) di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Penyusunan skripsi ini tidak terlepas dari bantuan berbagai pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Allah SWT yang telah memberikan kehidupan, pertolongan, rahmat berserta hidayah-Nya, dan kasih sayang yang tak ternilai harganya;
2. Ibunda Ratna Kurniasari dan Ayahanda Nyono Soemardi, yang telah mendidik, memberi dukungan, doa dan pengorbanan selama ini, serta memberi kasih sayang dan motivasi yang luar biasa;
3. Bapak Dr. Alfian Futuhul Hadi, S.Si, M.Si selaku Dosen Pembimbing Utama dan Ibu Dian Anggraeni, S.Si, M.Si selaku Dosen Pembimbing Anggota yang telah meluangkan waktu, pikiran, dan tenaga untuk membimbing saya dalam penulisan skripsi ini dengan sangat sabar, tulus dan perhatian;
4. Bapak Prof. Drs. I Made Tirta, M.Sc., Ph.D_selaku Dosen Penguji I dan Bapak Prof. Drs. Kusno, DEA., PhD selaku Dosen Penguji II yang telah memberikan kritik dan saran yang sangat bermanfaat dan membangun untuk penyempurnaan skripsi ini;

5. Bapak Kusbudiono, S.Si. M.Si selaku Dosen Pembinmbing Akademik yang memberikan banyak masukan dan dukungan selama menjalani perkuliahan;
6. Seluruh Dosen dan Karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember;
7. Kelurga besar PMII khusunya rayon FMIPA Universitas Jember yang selalu memberikan dukungan, bantuan, dan inspirasi-inspirasi selama 4 tahun ini;
8. Lailatul Nurfadila yang telah menjadi sahabat, teman maupun orang terkasih yang selalu memberikan dukungan, bantuan serta menjadi penyemangat saat suka dan duka;
9. Sahabat-sahabat BALSEM yaitu Shella Ariska Susianti, Ilham Agung Nugroho dan Lailatul Nurfadila yang selalu mengingatkan untuk saling memberi semangat, dukungan, bantuan dan inspirasi-nspirasi serta rasa syukur yang selalu saya rasakan setiap bersama kalian;
10. Teman-teman BALAKENEM yang telah menjadi partner dalam kuliah maupun diluar kuliah dan teman-teman se-angkatan BHATICS’12 serta kakak maupun adik angkatan matematika, terimakasih atas bantuan, kekompakan, dan persahabatan yang diberikan selama ini;
11. Teman-teman KKN 79 yang sudah memberikan rasa kebersamaan dan pengalaman-pengalaman baru kepada saya;
12. Serta semua pihak yang tidak dapat disebutkan satu per satu.

Semoga semua kebaikan yang telah diberikan akan dibalas oleh Allah SWT.

Penulis sangat mengharapkan kritik dan saran yang membangun dari semua pihak demi penyempurnaan skripsi ini. Akhir kata,semoga skripsi ini dapat memberi manfaat bagi semua pihak.

Jember, Desember 2016 Hanief Anggara Atmagenta

DAFTAR ISI

DAFTAR TABEL

Halaman

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Regresi linier merupakan metode Statistika yang digunakan untuk mengetahui hubungan antara variabel terikat (*dependent; Y*) dengan satu atau lebih variabel bebas (*independent; X*). Model regresi linier merupakan model yang sering digunakan dalam analisis statistika. Pendugaan parameter regresi secara otomatis juga mengestimasi model regresi. Pendugaan parameter dalam model regresi dapat dilakukan dengan Metode Kuadrat Terkecil atau *Ordinary Least Square* (OLS). Metode OLS (*Ordinary Least Square*) merupakan metode penduga terbaik untuk analisis regresi, namun metode ini sangat sensitif terhadap pencilan (*outlier*). Bahkan jika hanya satu saja pencilan ekstrim dalam data, maka akan mengakibatkan penyimpangan pada pendugaan OLS.

Penggunaan metode OLS dalam analisis regresi memerlukan beberapa asumsi klasik tentang galat dalam model yang dihasilkan. Asumsi yang harus dipenuhi yaitu kenormalan galat, tidak terjadi autokorelasi, dan kehomogenan. Penduga parameter yang diperoleh akan bersifat *best linier unbiased estimator* (BLUE) apabila asumsi galat terpenuhi. Data yang melandasi pembentukan model regresi kadang memiliki pencilan (*Outlier*) sehingga metode OLS tidak akan efisien lagi untuk digunakan. Sehingga apabila terjadi pelanggaran asumsi, penduga yang dihasilkan metode OLS tidak bersifat *best linier unbiased estimator* (BLUE) dan model regresi yang dihasilkan tidak cocok terhadap data yang dimodelkan (Srinadi, 2014).

Pencilan (*Outlier*) adalah data yang secara numerik berbeda dengan data lainnya (Barnett dan Lewis, 1994). Hair *et al* (1995) juga menyatakan bahwa pencilan (*Outlier*) adalah data yang muncul memiliki karakteristik unik yang terlihat sangat jauh berbeda dari observasi lainnya dan muncul dalam bentuk nilai ekstrim, baik untuk sebuah variabel tunggal ataupun variabel kombinasi. Pencilan lebih sederhananya dapat diartikan sebagai data yang menyimpang jauh dari data lainnya dalam suatu rangkaian data. Adanya pencilan pada data regresi mengakibatkan model regresi tidak memenuhi asumsinya dan model regresi tidak cocok (*fit*) terhadap data yang dimodelkan. Hal ini menyebabkan model yang dihasilkan tidak dapat digunakan untuk memprediksi.

Data pencilan yang mempengaruhi model regresi ada dua jenis yaitu pencilan pada variabel bebas (X) dan pencilan pada variabel respon (Y). Menurut Midi dan Mohammed (2015), pengamatan data pencilan di variabel bebas (X) disebut *Leverage Point*. Pencilan ini dapat diklasifikasikan menjadi 2 jenis yaitu *Good Leverage Point* dan *Bad Leverage Point*. *Good Leverage Point* adalah suatu titik pencilan pada variabel bebas tetapi titik pencilannya terletak dekat dengan garis linier. Sedangkan *Bad Leverage Point* adalah suatu titik pencilan pada variabel bebas tetapi titik pencilannya terletak jauh dari garis linier. Pencilan yang disebabkan varibel respon (Y) disebut *Vertical Outlier*. *Vertical Outlier* adalah suatu titik pencilan karena koordinat Y yang ekstrim.

Metode yang dapat digunakan untuk mengatasi pencilan pada regresi adalah metode analisis regresi *robust*. Regresi *robust* adalah metode yang penting untuk menganalisis data yang terkontaminasi oleh pencilan. Regresi *robust* terdiri dari beberapa metode antara metode *M-Estimator*, *S-Estimator*, *Least Absolute Value* (LAV), *Least Trimmed Square* (LTS) *dan Least Median Square* (LMS). Metode-metode penduga regresi *robust* tersebut memiliki kelebihan dan kelemahan masing-masing. Regresi *robust* dengan metode LTS lebih efisien dibanding LMS, karena LTS memiliki fungsi yang objektif yang lebih halus sehingga lebih sensitif terhadap efek lokal dan mempunyai nilai *breakdown* yang paling tinggi (Yaffee, 2002).

Metode LTS (*Least Trimmed Square*) adalah metode penaksir parameter regresi *robust* dengan meminimumkan jumlah sisaan kuadrat terhadap sebaran data yang sudah terpotong. Penelitian yang dilakukan oleh Suyanti (2014) menunjukkan bahwa metode LTS lebih efektif dari *M-Estimation* karena LTS menunjukkan nilai koefisien determinasi (R^2) yang lebih besar dari metode *M-Estimation*. Penelitian yang dilakukan oleh Putri (2014) juga menyimpulkan bahwa metode LTS lebih efektif daripada metode MM karena nilai salah baku penduga yang dihasilkan metode LTS lebih kecil dibanding metode MM.

Berdasarkan keterangan di atas penelitian ini bertujuan untuk menganalisis sifat penduga *Least Trimmed Square* (LTS) pada data yang mengandung pencilan di variabel respon (Y) dan variabel bebas (X). Menurut Rousseeuw (1984) metode LTS mampu mengatasi pencilan yang disebabkan baik oleh variabel bebas (X) dan variabel respon (Y) (Suyanti, 2014). Penelitian ini menggunakan data simulasi yang dibantu dengan program R.

1.2 Rumusan Masalah

Rumusan masalah yang akan dipecahkan dalam penelitian ini adalah

a. Bagaimana pengaruh *Laverage Point* dan *Vertical Outlier* terhadap pendugaan parameter model regresi linier berganda dilihat dari pencilan satu arah yaitu pencilan kanan/atas?

b. Bagaimana estimasi metode *Least Trimmed Square* untuk data yang mengandung *Leverage Point* dan *Vertical Outlier* pada regresi linier berganda?

1.3 Tujuan

Berdasarkan rumusan masalah di atas, maka tujuan dari penelitian ini adalah

a. Mengetahui informasi apa yang terjadi pada pengaruh *Laverage Point* dan *Vertical Outlier* terhadap pedugaan parameter regresi linier berganda dengan metode *Ordinary Least Square*.

b. Mengestimasi/menduga model regresi linier berganda pada data dengan *Leverage Point* dan *Vertical Outlier* menggunakan metode *Least Trimmed Square*.

c. Membandingkan pengaruh *Laverage Point* dan *Vertical Outlier* menggunakan metode *Ordinary Least Square* dan *Least Trimmed Square* pada model regresi linier berganda.

1.4 Manfaat

Manfaat yang diharapkan oleh penulis adalah dapat digunakan sebagai referensi tambahan tentang regresi *robust* khususnya metode *Least Trimmed Square* (LTS).

BAB 2. TINJAUAN PUSTAKA

2.1 Model Regresi Linier

Model regresi merupakan metode yang digunakan untuk menganalisis hubungan antara variabel. Hubungan tersebut dapat dituliskan kedalam bentuk persamaan yang menghubungkan variabel tak bebas (Y) dan peubah bebas (X). Terdapat dua jenis regresi yang terkenal menurut Gujarati (2003), yaitu regresi linier sederhana dan regresi linier berganda *(multiple linier regression model)* atau sering disebut juga dengan regresi klasik. Regresi linier klasik digunakan untuk menggambarkan hubungan antara peubah tak bebas (Y) dan peubah bebas (X) (Ummah, 2016).

Bentuk persamaan linier sederhan dapat dituliskan

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (2.1)$$

dengan $\varepsilon = error$

Sedangkan, bentuk umum dari regresi linier berganda dengan p variabel bebas adalah (Kurtner et.al, 2004)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip+1} + \varepsilon_i \qquad (2.2)$$

dengan:

Y_i adalah variabel tak bebas untuk pengamatan ke-i, untuk i = 1,2,3, ... , n.

β_i , β_1, β_2, … , β_p adalah parameter yang menentukan koefisien dari variabel bebas.

ε_i adalah galat ($error$) untuk pengamatn ke-i yang diasumsikan berdistribusi normal yang salig bebas dan identik dengan rata-rata 0 (nol) dan variansi σ^2.

Persamaan (2.2) dapat dituliskan dalam bentuk matriks menjadi

$$\mathrm{Y} = \mathrm{X}\beta + \varepsilon$$

dengan:

$$\underline{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},\ \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{2n} & \dots & x_{np} \end{bmatrix},\ \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix},\ \text{dan}\ \boldsymbol{\varepsilon} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

dengan :

Y adalah vektor variabel tidak bebas (variabel terikat) berukuran $n \times 1$

X adalah matriks variabel bebas (penduga berukuran $n \times p$

β adalah vektor parameter (koefisien regresi) berkuran $p \times 1$

ε adalah vektor *error* berukuran $n \times 1$

Menurut gujarati (2003) asumsi-asumsi pada model regresi linier berganda adalah sebagai berikut :

1. Model regresinya adalah linier dalam parameter
2. Nilai rata-rata dari *error* adalah nol. Asumsi $E(\varepsilon) = 0$ yang berarti bahwa nilai harapan atau rata-rata vektor yang setiap komponennya bernilai nol. ε adalah vektor kolom n × 1 dan 0 adalah vektor nol, sehingga E(ε) = 0.
3. Variansi dari *error* adalah konstan, yaitu bahwa setiap kesalahan penggunaan mempunyai varian yang sma $E({\varepsilon_i}^2) = \sigma^2$ untuk semua i.
4. Tidak terjadi autorelasi pada *error*. Artinya kesalahan antara pengganggu yang satu dengan yang lainnya bebas, $kov(\varepsilon_i , \varepsilon_j) = 0$.
5. Tidak terjadi multikolinieritas pada variabel bebas.
6. *Error* berdistribusi normal dengan varians konstan $\varepsilon \sim N(0, \sigma^2 I)$.

2.2 Analisis Residual

Umumnya pengamatan yang dicurigai sebagai pencilan, *influential observation* dan *high laverage* dikategorikan ke dalam pelanggaran asumsi, maka lebih tepat jika digunakan analisis residual. Metode yang digunakan dalam hubungan dengan pencilan (*Outlier*), *influential observation* (pengamatan berpengaruh), dan *high laverage* (pengaruh tinggi) adalah analisis residual. Residual banyak memegang peranan penting dalam pengujian untuk model regresi karena residual merupakan sisa pada suatu pengamatan. Sisa didefinisikan sebagai

$$\varepsilon_i = Y_i - \hat{Y}, \quad i = 1,2, \dots , n \tag{2.3}$$

Sisa memiliki banyak informasi, karena merupakan bagian yang amat penting dalam setiap analisis data. Informasi dari data semula yang tidak terserap oleh model akan menjadi sisa. Jika semua yang ada pada data telah masuk ke dalam model maka sisa akan berbentuk secara acak, tetapi jika model yang digunakan tidak mampu mengambil semua pola yag ada pada data maka sisa akan mempunyai kecenderungan tertentu. Dalam hal itu, model belum dikatakan benar-benar baik, dalam arti masih dapat disempurnakan. Jika sisa sudah berbentuk acak maka anggapan tentang kenormalan dan kesamaan variansi dapat diuji dari sisa.

Baiknya model dapat dilihat dari R^2 dan pengujian hipotesis mengenai koefisien regresi. Ketidakcocokan model dengan data dilihat dengan mengamati sisa. Adanya pencilan dalam data dapat dilihat dengan mengamati sisa. Secara umum, sisa memberi keterangan tentang data yang tidak mengikuti pola umum model yang digunakan, ditandai sisanya yang relatif besar. Sisa yang relatif besar dapat merupakan petunjuk bahwa modelnya belum cocok ataupun pengamatannya merupakan pencilan. Belum ada patokan yang disepakati para statistikawan kapan suatu pengamatan dapat dikategorikan sebagai pencilan. Secara umum, pencilan adalah data yang tidak mengikuti pola umum model, sehingga dapat diambil patokan yaitu yang sisnya berjarak 3 simpangan baku atau lebih rata-ratanya (yaitu nol)(Sembiring, 1995).

2.3 Metode Kuadrat Terkecil (*Ordinary Least Square*)

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. Langkah-langkah yang dilakukan untuk mengestimasi dengan metode kuadrat tekecil adalah sebagai berikut:

a. Meminimumkan kuadrat galat dengan mengubah model linier menjadi eksplisit terhadap galat

$$Yi = \beta 0 + \beta 1Xi + \varepsilon i$$

$$\boldsymbol{\varepsilon_i} = Y_{\boldsymbol{i}} - (\beta_0 + \beta_1 X_i) \quad (2.4)$$

b. Mengkuadratkan kesalahan yang diperoleh dan menjumlahkannya untuk seluruh pasangan data. Diperoleh bentuk jumlah kuadrat kesalahan sebagai berikut

$$\sum_{i=1}^{n} {\varepsilon_i}^2 = \sum_{i=1}^{n} [Y_i - (\beta_0 + \beta_1 X_i]^2 = \sum_{i=1}^{n} \left[Y_i - \sum_{j=0}^{1} \beta_j X_{ij}\right]^2 \quad (2.5)$$

c. Menurunkan persamaan (2.4) terhadap parameter yang menjadi kepentingan estimasi dengan metode kuadrat terkecil diproses dengan mencari minimum $\sum_{i=1}^{n} \varepsilon_i^2$ terhadap β_0 dan β_1

$$\frac{\partial(\sum_{i=1}^{n} {\varepsilon_i}^2)}{\partial \beta_0} = -2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i) \quad (2.6)$$

$$\frac{\partial(\sum_{i=1}^{n} {\varepsilon_i}^2)}{\partial \beta_1} = -2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i) X_i \quad (2.7)$$

d. Persamaan (2.5 dan (2.6) disamakan dengan nol sehingga diperoleh persamaan normal

$$\left.\begin{array}{l} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_1) = 0 \\ \sum_{i=1}^{n} Y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} X_i = 0 \end{array}\right\} \quad (2.8)$$

Persamaan normal diatas selanjutnya dapat disederhanakan menjadi

$$\sum_{i=1}^{n} (Y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} X_i = 0 \quad (2.9a)$$

$$\sum_{i=1}^{n} X_i Y_i - \beta_0 \sum_{i=1}^{n} X_i - \beta_1 \sum_{i=1}^{n} {X_i}^2 = 0 \quad (2.9b)$$

e. Dari persamaan normal (2.8a) diatas diperoleh

$$n\beta_0 = \sum_{i=1}^{n} Y_i - \beta_i \sum_{i=1}^{n} X_i$$

$$\beta_0 = \frac{1}{n}\sum_{i=1}^{n} Y_i - \beta_i \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{Y} - \beta_1 \bar{X} \qquad (2.10)$$

Selanjutnya, hasil persamaan (2.9) ini di subtitusikan pada persamaan normal (2.8b) sehingga diperoleh:

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} {X_i}^2 - n(\bar{X})^2}$$

$$= \frac{\sum Y_1(X_i - \bar{X})}{\sum {X_i}^2 - n(\bar{X})^2}$$

Dengan $\sum(X_i - \bar{X})^2 = \sum {X_i}^2 - \sum \bar{X}^2 = \sum {X_i}^2 - n(\bar{X})^2$, maka

$$\hat{\beta}_1 = \frac{\sum Y_1(X_i - \bar{X})}{\sum(X_i - \bar{X})^2} \qquad (2.11)$$

(Tirta, 2009).

Persamaan (2.9) dan (2.10) merupakan penaksir kuadrat terkecil.

Penaksir kuadrat terkecil akan memenuhi sifat BLUE (*best, linier, unbiased, estimator*) jika memenuhi asumsi-asumsi regresi linier berganda. Dimana sifat BLUE meliputi,

1. *Best* : Varian ε nya paling minimum
2. *Linear* : $\beta_0, \beta_1, \beta_2$ adalah kombinasi linear dari y
3. *Unbiased* : Rata-rata, nila aktual $\beta_0, \beta_1,$ dan β_2 sama dengan $\beta_0, \beta_1, \beta_2$ $(E(\beta_0) = \beta_0, E(\beta_1) = \beta_1,$ dan $E(\beta_2) = \beta_2)$
4. *Estimator* : $\beta_0, \beta_1, \beta_2$ adalah estimator dari nilai sebenarnya

2.4 Pencilan (*Outlier*)

Pencilan adalah data yang tidak mengikuti pola umum pada model regresi yang dihasilkan, atau tidak mengikuti pola data secara keseluruhan. Pencilan juga dapat diartikan sebagai suatu keganjilan pada data amatan yang menunjukkan ketidaksesuaian dengan mayoritas sisaan data (Sembiring, 1995). Jumlah maksimum pencilan dalam data yang diperbolehkan adalah 50% (Rousseeuw *et al*., 1987: 303).

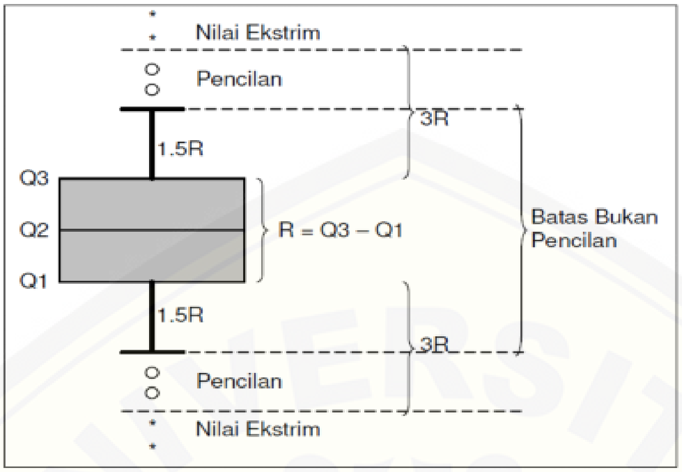
Pada umumnya pencilan mempunyai sisaan (*error*) berjarak tiga simpangan baku. Pencilan merupakan suatu keganjilan yang menandakan suatu titik data yang sama sekali tidak tipikal dari data yang lainnya. Keberadaan data pencilan akan mengganggu dalam proses analisis data dan harus dihindari dari beberapa hal. Dalam kaitannya dalam analisis regresi, pencilan dapat menyebabkan hal-hal berikut (Soemartini, 2007: 7):

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

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

2.5 Identifikasi dan Deteksi Pencilan dengan *Boxplot* (*outlier*)

Menurut Paludi (2009), salah satu metode yang digunakan untuk mengidentifikasi adanya pencilan yang berpengaruh dalam koefisien regresi adalah *boxplot*. Metode ini merupakan yang paling umum yaitu dengan mempergunakan nilai kuartil dan jangkauan. Kuartil 1, 2 dan 3 akan membagi sebuah urutan data menjadi empat bagian. Jangkauan (IQR, *Interquartile Range*) didefinisikan sebagai selisih kuartil 1 terhadap kuartil 3, atau $\mathrm{IQR} = \mathrm{Q3} - \mathrm{Q1}$. Data-data pencilan dapat ditentukan yaitu nilai dengan kuartil yang kurang dari 1,5 × IQR terhadap kuartil 1 dan nilai kuartil lebih dari 1,5 × IQR terhadap kuartil 3 (Ummah, 2016).



Gambar 2.1 Skema Identifikasi Pencilan Menggunakan IQR atau *boxplot*

2.6 Metode *Least Trimmed Square* (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.

$$\min \sum_{i=1}^{h} {\varepsilon_i}^2 \tag{2.12}$$

dengan

$$h = \left[\frac{n}{2}\right] + \left[\frac{k+1}{2}\right] \tag{2.13}$$

Keterangan:

ε_i^2 = kuadrat residual yang diturunkan dari terkecil ke terbesar $\varepsilon_1^2 < \varepsilon_2^2 < \cdots < \varepsilon_n^2$

$\varepsilon_i^2 < \cdots < \varepsilon_h^2 < \cdots < \varepsilon_n^2$

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,β_1 dan β_2,kemudian menentukan n residual dengan mengunakan rumus $(y_1 - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i})^2$. Setelah itu menghitung $\sum_{i=1}^{h} \varepsilon_i^2$, dengan $h = \left[\frac{n}{2}\right] + \left[\frac{k+1}{2}\right]$ pengamatan dengan ε_i^2 terkecil. Tahapan-tahapan di atas dilakukan sampai diperoleh nilai residual terkecil dan konvergen.

2.7 *Goodness of Fit*

Melalui uji kecocokan (*goodness of fit*) dapat diperoleh gambaran apakah model yang dipilih sesuai atau tidak. Dalam *goodness of fit* dapat ditentukan model terbaik dan bagaimana signifikannyasi parameternya.

2.7.1 Uji Kebaikan Model dengan Hipotesis β Menggunakan Uji T

Kebaikan model dapat dilihat dari nilai alternatif uji t yang berfokus pada uji hipotesis $H_0: \beta^* = 0$ dimana β^* adalah koefisien X^* dalam persamaan regresi $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \beta^* X^* + E$. Kesetaraan pada statistika untuk menguji hipotesis ini adalah

$$T = \frac{\hat{\beta}^*}{S_{\hat{\beta}^*}} \tag{2.14}$$

dimana $\hat{\beta}^*$ adalah estimasi koefisien yang sesuai dan $S_{\hat{\beta}^*}$ adalah perkiraan standar eror $\hat{\beta}^*$, yang keduanya didapat dari program regresi standar.

Dalam melakukan test ini, menolak H_0: $\beta^* = 0$, jika

$$\begin{cases} |T| > t_{n-p-2,1-\propto/2} & \text{(Uji dua sisi; } H_A\text{: } \beta^* \neq 0 \\ T > t_{n-p-2,1-\propto} & \text{(Uji satu sisi; } H_A\text{: } \beta^* > 0 \\ T < -t_{n-p-2,1-\propto} & \text{(Uji satu sisi lebih rendah; } H_A\text{: } \beta^* < 0 \end{cases} \tag{2.15}$$

(Kleinbaum *et al*., 1998).

2.7.2 Uji Kebaikan Model dengan MSE (*Mean Square Error*)

Kebaikan Model dapat dilihat dari nilai MSE (*Mean Square Error*). Nilai MSE dapat diketahui dari persamaan

$$MSE = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n} \tag{2.16}$$

dengan

n = banyaknya sampel

$\hat{y}_i$ = nilai y dugaan ke-i

y_i = nilai y sebenarnya ke-i

Semakin kecil nilai MSE, maka model yang diperoleh semakin baik (Muzathik *et al*., 2011).

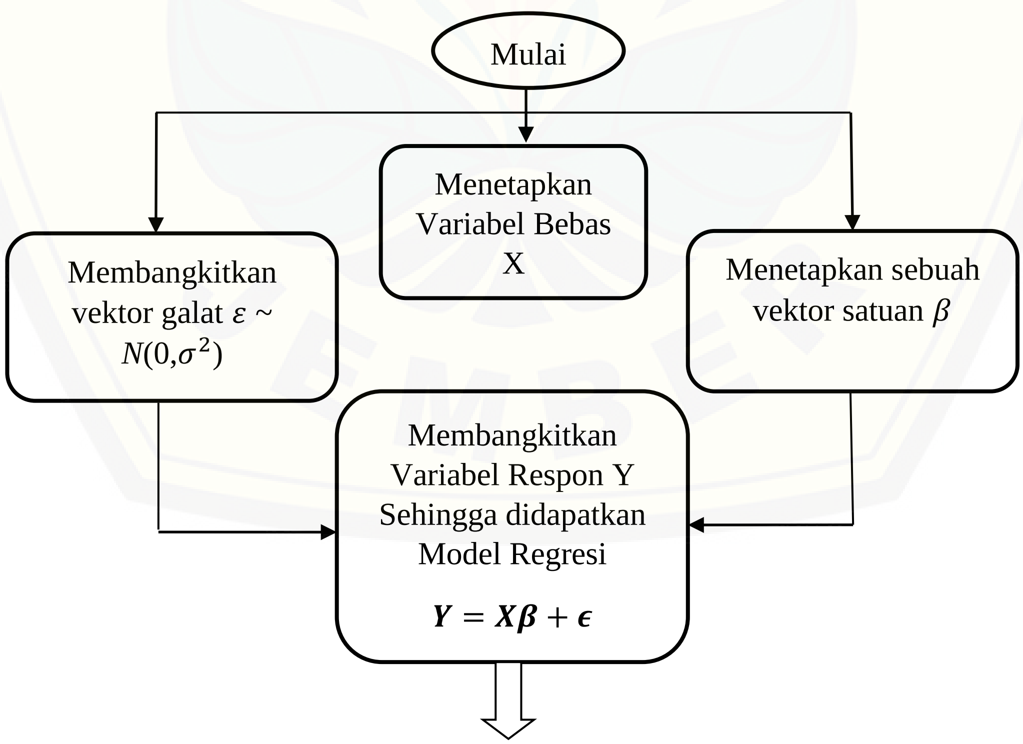
BAB 3. METODOLOGI PENELITIAN

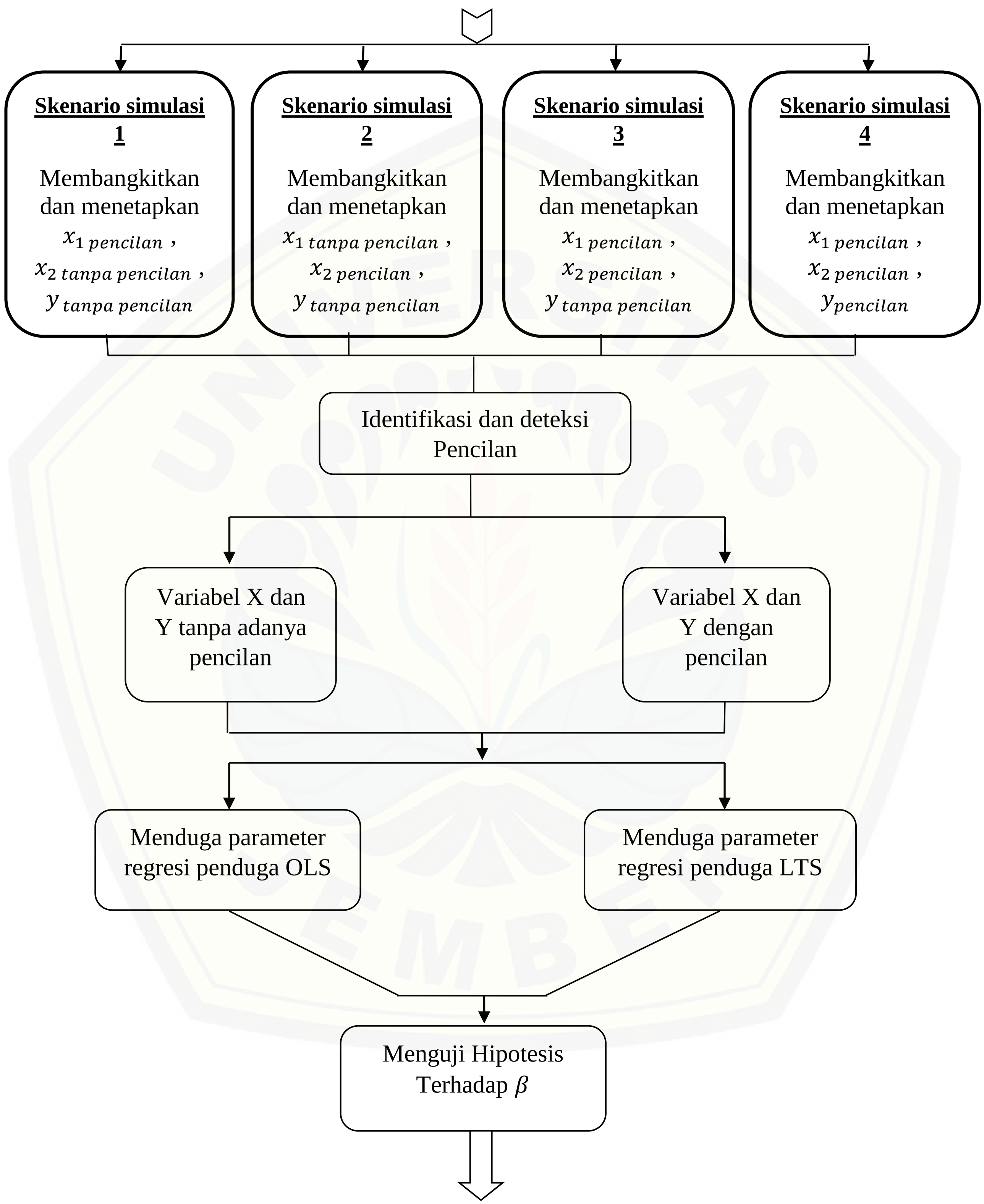
3.1 Data Penelitian

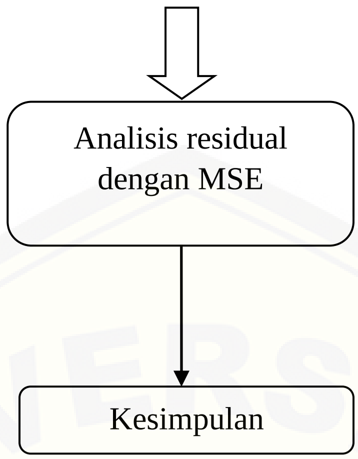
Data yang digunakan dalam penelitian ini merupakan data simulasi dengan membangkitkan data melalui program R. Ukuran data yang dipilih yaitu 100 dengan menetapkan 2 variabel X (x_1, x_2). Data tersebut adalah data awal yang akan digunakan untuk menghasilkan variabel **Y**. Kemudian dilakukan simulasi terhadap data X (x_1, x_2) dan **Y.** Hasil dari simulasi yang didapat akan diberikan 2 presentase pencilan pada X (x_1, x_2) dan **Y** sebesar 5% dan 10%.

3.2 Diagram Alir Penelitian

Langkah –langkah penelitian ini disajikan dalam diagram alir pada Gambar 3.2 berikut.




Skenario simulasi 1
Membangkitkan dan menetapkan
x_1 pencilan, x_2 tanpa pencilan, y tanpa pencilan
Skenario simulasi 2
Membangkitkan dan menetapkan
x_1 tanpa pencilan, x_2 pencilan, y tanpa pencilan
Skenario simulasi 3
Membangkitkan dan menetapkan
x_1 pencilan, x_2 pencilan, y tanpa pencilan
Skenario simulasi 4
Membangkitkan dan menetapkan
x_1 pencilan, x_2 pencilan, y pencilan
Identifikasi dan deteksi Pencilan
Variabel X dan Y tanpa adanya pencilan
Variabel X dan Y dengan pencilan
Menduga parameter regresi penduga OLS
Menduga parameter regresi penduga LTS
Menguji Hipotesis Terhadap β



Gambar 3.1 Skema Desain Penelitian

Deskripsi dan penjelasan dari masing-masing langkah penelitian pada skema desain penelitian di atas adalah sebagai berikut.

a. Tahap Simulasi

1) Menetapkan Variabel Bebas

Menetapkan x_1 dan x_2 dengan korelasi 0. Data yang diperoleh dimisalkan X.

2) Membangkitkan Vektor galat

Membangktikan vektor galat ($\boldsymbol{\varepsilon}$) dari sebaran normal $N(0,\sigma^2)$ dengan varian besar.

3) Menetapkan Vektor satuan

Menetapkan β_0, β_1, β_2 pada vektor satuan.

4) Membangkitkan Variabel respon **Y**

Membangkitkan Variabel respon **Y** dengan berdasarkan kombinasi linier x_1, x_2 , vektor satuan β_0, β_1, β_2 dan galat.

5) Membangkitkan Variabel bebas **X** dan variabel respon **Y** dengan pencilan ($\mathbf{X}_{\text{pencilan}}$),($\mathbf{Y}_{\text{pencilan}}$)

- Mencari nilai rata-rata dari Variabel bebas **X** dan Variabel respon **Y** yang telah dibangkitkan sebelumnya.

- Mencari nilai maksimum dari Variabel bebas **X** dan Variabel respon **Y** yang telah dibangkitkan sebelumnya.
- Menghitung standart deviasi dari Variabel bebas **X** dan Variabel respon **Y** yang telah dibangkitkan sebelumnya.
- Menambahkan nilai maksimum dari Variabel bebas **X** dan Variabel respon **Y** dengan lebih dari tiga kali standart deviasi yang bersesuaian.

$$\text{Pencilan } (x) \geq \max(x) + 3\times \text{ stdev}(x)$$

$$\text{Pencilan } (y) \geq \max(y) + 3\times \text{ stdev}(y)$$

- Pencilan ditambahkan secara proporsional pada setiap data dalam Variabel bebas **X** dan Variabel respon **Y** yang awal.

6) Menetapkan 4 skenario simulasi
   - Skenario simulasi 1

     Menetapkan dan membangkitkan $x_{1\ p}$, $x_{2\ tp}$, $y_{\ tp}$.
   - Skenario simulasi 2

     Menetapkan dan membangkitkan $x_{1\ tp}$, $x_{2\ p}$, $y_{\ tp}$.
   - Skenario simulasi 3

     Menetapkan dan membangkitkan $x_{1\ p}$, $x_{2\ p}$, $y_{\ tp}$.
   - Skenario simulasi 4

     Menetapkan dan membangkitkan $x_{1\ p}$, $x_{2\ p}$, $y_{\ p}$.

Dengan p = pencilan dan tp = tanpa pencilan. Skenario simulasi 1, 2 dan 3 hanya membangkitkan pencilan pada data variabel x_1 dan x_2 dan y tanpa pencilan untuk mengetahui pengaruh estimasi data yang mengandung *Laverage Point* (pencilan pada variable X). Sedangkan pada skenario simulasi 4 membangktikan pencilan data pada keseluruhan variabel, yaitu variabel x_1 dan x_1 dan y. Hal ini bertujuan untuk mengetahui pengaruh estimasi data yang mengandung *Laverage Point* (pencilan pada variabel X) dan *Vertical Outlier* (pencilan pada variabel Y).

b. Tahap Estimasi

1) Mengidentifikasi dan mendeteksi pencilan dengan memplot data menggunakan *Boxplot.*
2) Menduga parameter regresi penduga OLS pada **X** dan **Y** dengan data pada semua skenario simulasi.
3) Menduga Parameter regresi dengan penduga LTS pada **X** dan **Y** dengan data pada semua skenario simulasi.
4) Menghitung Hipotesis terhadap beta (β)

   ukuran kebaikan model untuk menguji $\hat{\beta}$ yang dihasilkan oleh variabel **X** adalah menggunakan uji hipotesis terhadap β, yaitu Uji T.
5) Menghitung Nilai MSE (*Mean Square Error*)

   Ukuran kebaikan model yang digunakan untuk menentukan metode pendugaan yang terbaik adalah Analisis Residual dengan MSE.

c. Tahap Analisis

1) Membandingkan Uji Hipotesis terhadap (β)

   Membandingkan Uji Hepotesis terhadap (β) dari model regresi OLS dan LTS pada variabel **X** dan **Y** pada data semua skenario simulasi.
2) Membandingkan MSE (*Mean Square Error*)

   Membandingkan MSE dari model regresi OLS dan LTS pada variabel **X** dan **Y** pada data semua skenario simulasi.
3) Kesimpulan

   Setelah semua tahapan penelitian dilakukan, dapat ditarik kesimpulan yang merupakan jawaban dari rumusan masalah pada bab 1.

BAB 5. PENUTUP

5.1 Kesimpulan

Berdasarkan hasil analisis data dan pembahsan pada bab sebelumnya, didapatlkan kesimpulan sebagai berikut.

a. Pengaruh *Laverage Point* dan *Vertical Outlier* terhadap pendugaan parameter model regresi dapat dilihat dari nilai parameter $(\beta_0, \beta_1, \beta_2)$. Pada metode OLS, data yang mengandung *Laverage Point* dan *Vertical Outlier* berpengaruh terhadap estimasi parameternya. Semakin banyak pencilan yang diberikan, hasil estimasi parameternya cenderung semakin besar pula. Overestimate yang dialami OLS tersebut kemungkinan disebabkan adanya tekanan pencilan atas/kanan. Beda halnya dengan metode LTS, semakin banyaknya pencilan tidak selalu memberikan hasil estimasi yang semakin besar pula. Hal ini berarti bahwa tekanan dari pencilan atas/kanan tidak terlalu berpengaruh pada hasil estimasi metode LTS. Selain itu, pada simulasi data tanpa pencilan, nilai MSE metode *Ordinary Least Square* lebih kecil dibanding *Least Trimmed Square*. Hal ini menunjukkan bahwa metode *Ordinary Least Squae* lebih baik memodelkan data tanpa pencilan.

b. Pada data yang mengandung *Laverage Point* dan *Vertical Outlier* nilai MSE metode *Least Trimmed Square* lebih kecil dibanding *Ordinary Least Square*. Dengan demikian, dapat disimpulkan bahwa metode *Least Trimmed Square* lebih baik memodelkan data pencilan pada variabel prediktor (x) dan variabel respon (y) (*Laverage Point* dan *Vertical Outlier*).

5.2 Saran

Pada penelitian ini penulis mencoba mengkaji hasil pendugaan parameter menggunkan metode *Least Trimmed Square* (LTS) pada yang mengandung pencilan

Laverage Point dan *Vertical Outlier* (Pencilan pada variabel X dan variabel Y). Penelitian selanjutnya dapat mencoba mengkaji hasil pendugaan parameter dengan menggunakan perbandingan dua atau lebih metode *robust* yang lainnya seperti *MM-Estimation*, *Least Median Square* (LMS) dan *Least Absolute Value* (LAV) atau metode selain metode *robust* seperti GLM (*Generealized Linier Model*) yaitu metode *Gamma* dan *Log* pada data yang mengandung pencilan (*Laverage Point* dan *Vertical Outlier*).

DAFTAR PUSTAKA

Drapper, N.R, & Smith, H. 1992. *Analisis Regresi Terapan.* Jakarta: PT Gramedia.

Barnett, V, & Lewis, T. (1994). *Outlier in statistical data* (3^{rd} edition). New York: Wiley.

Gujarati, D.N. 2003. *Basic Econometrics Fourth Edition.* New York: The McGraw-Hill Companies,Inc.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E., & Nizam A. 1998. *Applied Regression Analysis and other Multivariable Methods* (3^{rd} edition). New York: Brooks Publishing Company A divison of International Thomson Publishing, Inc.

Kurtner, M., Nachtsheim, C., dan Neter, J. 2004. *Applied Linear Regression Models Fifth Edition* [Only Chapters 1, 2, 13]. New York: The McGraw-Hill Companies,Inc.

Midi, H. dan Mohammed, A.M. 2015. The Identification of Good and Bad High Laverage Points in Multiple Linear Regression Model. *Mathematical Methods and System in Science and Engineering.* Malaysia: University Putra Malaysia.

Musafirah, Raupong, dan Sirajang. N. 2015. *Perbandingan Metode Robust Least Trimmed Square Dengan Metode Scale Dalam Mengestimasi Parameter Regresi Linear Berganda Untuk Data Yang Mengandung Pencilan*. Makassar:Universitas Hasanuddin.

Muzathik, dkk. 2011. Daily Global Solar Raditional Estimate Based on Sunshine Hours. *International Journal of Mechanical and Engineering* (IJMME), **6**(1): 75-80.

Myers, R.H., & Militon, J.S. 1991 *A First Course in the Theory of Linear Statistical Models.* Boston: PWS-KENT Publishing Company is a divison of Wadsworth, Inc.

Putri, D.E. 2014. *Perbandingan Regresi Robust penduga Least Trimmed Square* (LTS) *dan Penduga MM untuk Pendugaa Model Penilaian Aset Modal.* Tidak Diterbitkan. Skripsi. Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Brawijaya, Malang.

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

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

Soemartini, 2007. *Pencilan (Outlier).* Bandung: UNPAD.

Srinadi I.G.A.M. 2014. Pengaruh Outlier Terhadap Estimator Parameter Regresi dan Metode Regresi Robust. *Prosiding Konferensi Nasional Matematika XVII, ITS*, Surabaya.

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

Tirta, I.M. 2009. *Analysis Regresi dengan R*. Jember: Unej ─ *press*.

Ummah, S. 2016. “Pendugaan Parameter Least Median Square (LMS) Pada Data Mengandung Vertical Outlier”. Tidak Diterbitkan. Skripsi. Jember: Jurusan Matematika FMIPA Universitas Jember.

Yaffe R.A. 2002. *Robust Regression Analysis: some popular statistical package options*. www.nyu.edu/its/socsci/Docs/RobustRegression.pdf.

LAMPIRAN

Lampiran A. Struktur fungsi program R

A.1 Struktur fungsi `rnorm()`

merupakan fungsi untuk membangkitkan variabel bebas yang berdristribusi normal. Struktur fungsi `rnorm()` yang digunakan dalam penelitian ini adalah:

`rnorm(n,` μ, σ^2`)`

Keterangan:

`n` : banyaknya data yang dibangkitkan

μ : rata-rata

σ^2 : standart deviasi

A.2 Struktur fungsi `sort.list()`

Fungsi `sort.list()` merupakan fungsi yang digunakan untuk menghasilkan sebuah vektor yang berukuran sama dengan vektor dari variabel yang diteliti dimana elemennya sudah terurut dengan urutan menaik dengan fasilitas pengurutan yang lebih fleksibel. Struktur fungsi `sort.list()` yang digunakan dalam penelitian ini adalah

```
sort.list(variable, decreasing = TRUE)
```

Keterangan:

`Variable` : variabel yang diteliti (variabel respon atau prediktor)

`decreasing = TRUE` : logika agar elemennya menjadi urutan menurun

A.3 Struktur fungsi `lm()`

Fungsi `lm()` merupakan fungsi yang digunakan untuk menentukan model linier. Struktur fungsi `lm()` yabg digunakan dalam penelitian ini adalah:

```
lm(formula)
```

Keterangan:

`formula` : menyatakan hubungan antara variabel respon (y) dan variabel prediktor (x)

A.4 Struktur fungsi `Boxplot()` pada paket `car`

Fungsi `Boxplot()` merupakan fungsi yang digunakan untuk melihat sebaran data. Pada penelitian ini digunakan untuk mengindentifikasi pencilan. Struktur fungsi `Boxplot()` yang digunakan dalam penelitian ini adalah:

`Boxplot()`

Keterangan:

`Variable` : variabel yang diteliti (variabel respon atau prediktor)

A.5 Struktur fungsi `ltsreg()` pada paket `MASS`

Fungsi `ltsreg()` merupakan fungsi untuk mendapatkan niai parameter (β) dari salah satu metode *robust* yaitu *Least Trimmed Square*. Struktur fungsi `ltsreg()` yang digunakan dalam penelitian ini adalah:

`ltsreg(formula)`

Keterangan:

`formula` : menyatakan hubungan antara variabel respon (y) dan prediktor (x)

Lampiran B. *Syntax* program R

B.1 Tahapan Simulasi

B.1.1 Data sebelum diberi pencilan pada keempat skenario simulasi dengan n = 100

```
> set.seed(01)
> n<-100
> x0<-rep(1,n)
> r=1
> while (abs(r)>0)
+ {
+   x1<-rnorm(n,0,0.5)
+   x2<-rnorm(n,0,0.5)
+   r=round(cor(x1,x2),2)
+ }
> cor(x1,x2)
```

```
[1] -0.0009943199
> r
[1] 0
> X<-cbind(x0,x1,x2)
> eps<-rnorm(n,0,0.8)
> beta<-c(1.4,1.2,2)
> Y<-X%*%beta+eps
```

B.1.2 Data setelah diberi pencilan pada keempat skenario simulasi dengan n = 100

a. Skenario simulasi 1 dengan 5% pencilan yang diberikan pada variabel x_1

```
> stdx1<-sqrt(sum((x1-mean(x1))^2/(length(x1)-1)))
> stdx1
[1] 0.4490997
> x1p<-x1
> m<-sort.list(x1p,decreasing = TRUE)
> for(i in 1:5)
+ {
+   x1p[m[i]]<-x1p[m[i]]+(3*stdx1)
+ }
```

b. Skenario simulasi 1 dengan 10% pencilan yang diberikan pada variabel x_1

```
> stdx1<-sqrt(sum((x1-mean(x1))^2/(length(x1)-1)))
> stdx1
[1] 0.4490997
> x1p<-x1
> m<-sort.list(x1p,decreasing = TRUE)
> for(i in 1:10)
+ {
+   x1p[m[i]]<-x1p[m[i]]+(3*stdx1)
+ }
```

c. Skenario simulasi 2 dengan 5% pencilan yang diberikan pada variabel x_2

```
> stdx2<-sqrt(sum((x2-mean(x2))^2/(length(x2)-1)))
> stdx2
[1] 0.4789395
> x2p<-x2
> m<-sort.list(x2p,decreasing = TRUE)
> for(i in 1:5)
+ {
```

```
+   x2p[m[i]]<-x2p[m[i]]+(3*stdx2)
+ }
```

d. Skenario simulasi 2 dengan 10% pencilan yang diberikan pada variabel x_2

```
> stdx2<-sqrt(sum((x2-mean(x2))^2/(length(x2)-1)))
> stdx2
[1] 0.4789395
> x2p<-x2
> m<-sort.list(x2p,decreasing = TRUE)
> for(i in 1:10)
+ {
+   x2p[m[i]]<-x2p[m[i]]+(3*stdx2)
+ }
```

e. Skenario simulasi 3 dengan 5% pencilan yang diberikan pada variabel x_1 dan x_2

```
> stdx1<-sqrt(sum((x1-mean(x1))^2/(length(x1)-1)))
> stdx1
[1] 0.4490997
> x1p<-x1
> m<-sort.list(x1p,decreasing = TRUE)
> for(i in 1:10)
+ {
+   x1p[m[i]]<-x1p[m[i]]+(3*stdx1)
+ }
> stdx2<-sqrt(sum((x2-mean(x2))^2/(length(x2)-1)))
> stdx2
[1] 0.4789395
> x2p<-x2
> mA<-sort.list(x2p,decreasing = TRUE)
> for(i in 1:10)
+ {
+   x2p[mA[i]]<-x2p[mA[i]]+(3*stdx2)
+ }
```

f. Skenario simulasi 3 dengan 10% pencilan yang diberikan pada variabel x_1 dan x_2

```
> stdx1<-sqrt(sum((x1-mean(x1))^2/(length(x1)-1)))
> stdx1
[1] 0.4490997
> x1p<-x1
> m<-sort.list(x1p,decreasing = TRUE)
> for(i in 1:10)
+ {
+   x1p[m[i]]<-x1p[m[i]]+(3*stdx1)
+ }
> stdx2<-sqrt(sum((x2-mean(x2))^2/(length(x2)-1)))
> stdx2
```

```
[1] 0.4789395
> x2p<-x2
> mA<-sort.list(x2p,decreasing = TRUE)
> for(i in 1:10)
+ {
+   x2p[mA[i]]<-x2p[mA[i]]+(3*stdx2)
+ }
```

g. Skenario simulasi 4 dengan 5% pencilan yang berikan pada variabel x_1 , x_2 dan Y

```
> stdx1<-sqrt(sum((x1-mean(x1))^2/(length(x1)-1)))
> stdx1
[1] 0.4490997
> x1p<-x1
> m<-sort.list(x1p,decreasing = TRUE)
> for(i in 1:5)
+ {
+   x1p[m[i]]<-x1p[m[i]]+(3*stdx1)
+ }
> stdx2<-sqrt(sum((x2-mean(x2))^2/(length(x2)-1)))
> stdx2
[1] 0.4789395
> x2p<-x2
> mA<-sort.list(x2p,decreasing = TRUE)
> for(i in 1:5)
+ {
+   x2p[mA[i]]<-x2p[mA[i]]+(3*stdx2)
+ }
> stdY<-sqrt(sum((Y-mean(Y))^2/(length(Y)-1)))
> stdY
[1] 1.352562
> Yp<-Y
> mB<-sort.list(Yp,decreasing = TRUE)
> for(i in 1:5)
+ {
+   Yp[mB[i],]<-Yp[mB[i],]+(3*stdY)
+ }
```

h. Skenario simulasi 4 dengan 10% pencilan diberikan pada variabel x_1 , x_2 dan Y

```
> stdx1<-sqrt(sum((x1-mean(x1))^2/(length(x1)-1)))
> stdx1
[1] 0.4490997
> x1p<-x1
> m<-sort.list(x1p,decreasing = TRUE)
> for(i in 1:10)
+ {
+   x1p[m[i]]<-x1p[m[i]]+(3*stdx1)
+ }
> stdx2<-sqrt(sum((x2-mean(x2))^2/(length(x2)-1)))
```

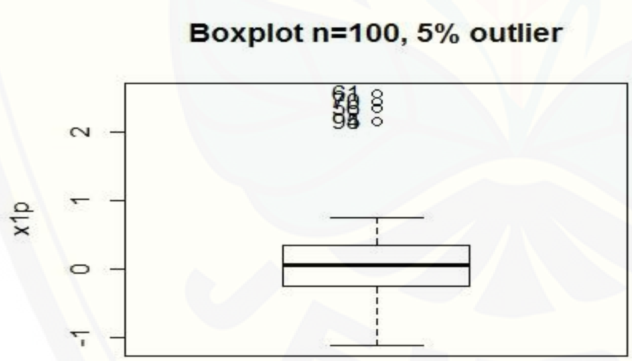
```
> stdx2
[1] 0.4789395
> x2p<-x2
> mA<-sort.list(x2p,decreasing = TRUE)
> for(i in 1:10)
+ {
+   x2p[mA[i]]<-x2p[mA[i]]+(3*stdx2)
+ }
> stdY<-sqrt(sum((Y-mean(Y))^2/(length(Y)-1)))
> stdY
[1] 1.352562
> Yp<-Y
> mB<-sort.list(Yp,decreasing = TRUE)
> for(i in 1:10)
+ {
+   Yp[mB[i],]<-Yp[mB[i],]+(3*stdY)
+ }
```

B.2 Tahapan Estimasi Data

B.2.1 Identifikasi dan Deteksi Pencilan

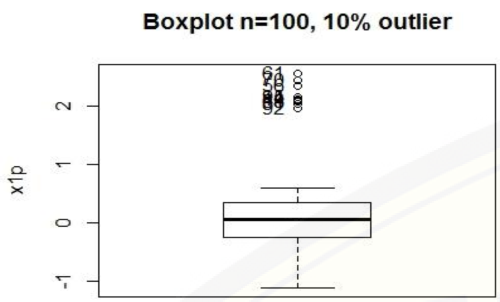
a. Skenario simulasi 1 dengan 5% pencilan yang diberikan pada variabel x_1

```
> library(car)
> Boxplot(x1)
> Boxplot(x1p,main='Boxplot n=100, 5% outlier')
```



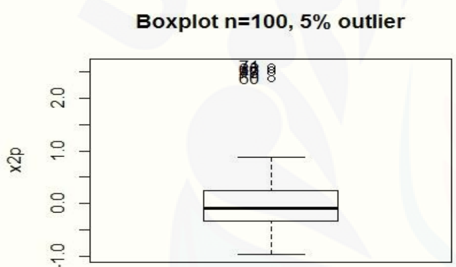
b. Skenario simulasi 1 dengan 10% pencilan yang diberikan pada variabel x_1

```
> library(car)
> Boxplot(x1)
> Boxplot(x1p,main='Boxplot n=100, 10% outlier')
```



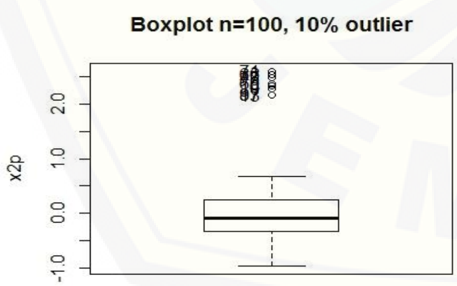
c. Skenario simulasi 2 dengan 5% pencilan yang diberikan pada variabel x_2

```
> library(car)
> Boxplot(x2)
> Boxplot(x2p,main='Boxplot n=100, 5% outlier')
```



d. Skenario simulasi 2 dengan 10% pencilan yang diberikan pada variabel x_2

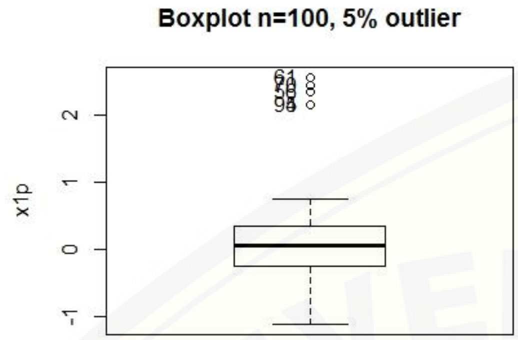
```
> library(car)
> Boxplot(x2)
> Boxplot(x2p,main='Boxplot n=100, 10% outlier')
```

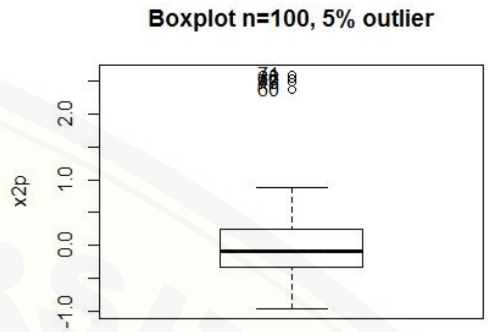


e. Skenario simulasi 3 dengan 5% pencilan yang diberikan pada variabel x_1 dan x_2

```
> library(car)
> Boxplot(x1)
> Boxplot(x1p,main='Boxplot n=100, 5% outlier')
> Boxplot(x2)
```

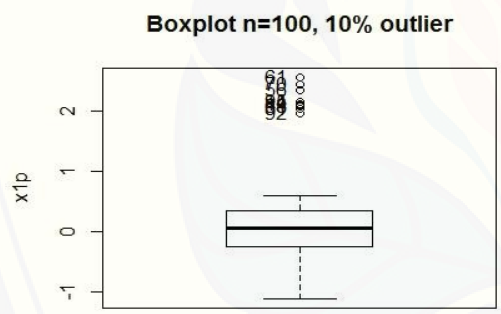
```
> Boxplot(x2p,main='Boxplot n=100, 5% outlier')
```

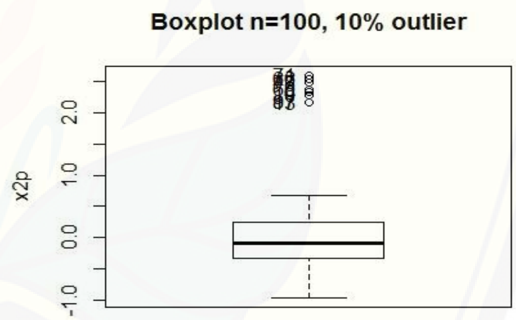




f. Skenario simulasi 3 dengan 10% pencilan yang diberikan pada variabel x_1 dan x_2

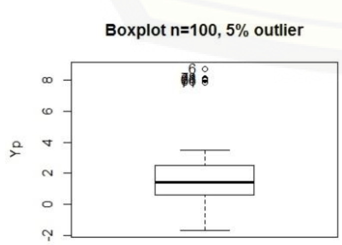
```
> library(car)
> Boxplot(x1)
> Boxplot(x1p,main='Boxplot n=100, 10% outlier')
> Boxplot(x2)
> Boxplot(x2p,main='Boxplot n=100, 10% outlier')
```

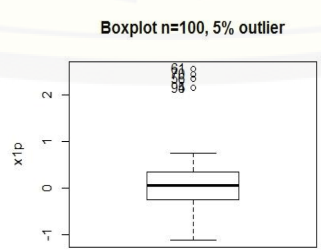


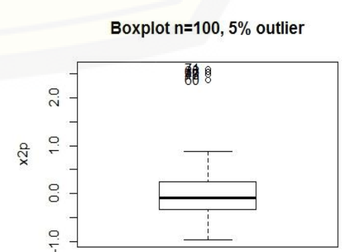


g. Skenario simulasi 4 dengan 5% pencilan yang berikan pada variabel x_1 , x_2 dan Y

```
> library(car)
> Boxplot(x1p,main='Boxplot n=100, 5% outlier')
> Boxplot(x2p,main='Boxplot n=100, 5% outlier')
> Boxplot(Yp,main='Boxplot n=100, 5% outlier')
```

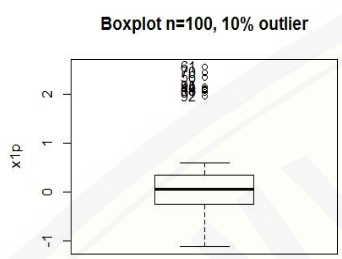


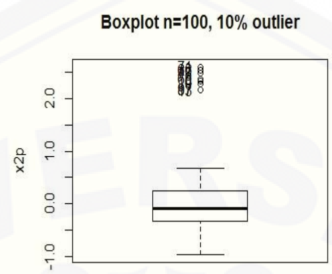


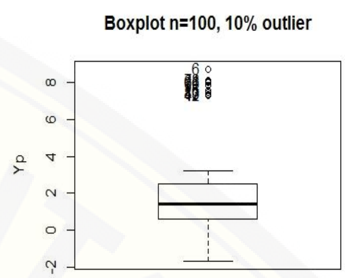


h. Skenario simulasi 4 dengan 10% pencilan diberikan pada variabel x_1 , x_2 dan Y

```
> library(car)
> Boxplot(x1p,main='Boxplot n=100, 10% outlier')
> Boxplot(x2p,main='Boxplot n=100, 10% outlier')
> Boxplot(Yp,main='Boxplot n=100, 10% outlier')
```







B.2.2 Estimasi dengan Metode *Ordinary Least Square*

a. Skenario simulasi 1 tanpa pencilan

```
> OLS<-lm(Y~x1+x2)
> summary(OLS)

Call:
lm(formula = Y ~ x1 + x2)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.35487 -0.34916  0.00161  0.50953  2.11153 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.42028    0.08416  16.877  < 2e-16 ***
x1           1.23378    0.18680   6.605 2.14e-09 ***
x2           1.91445    0.17517  10.929  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8347 on 97 degrees of freedom
Multiple R-squared:  0.6268,	Adjusted R-squared:  0.6191 
F-statistic: 81.46 on 2 and 97 DF,  p-value: < 2.2e-16
```

b. Skenario simulasi 1 dengan 5% pencilan

```
> OLSp<-lm(Y~x1p+x2)
> summary(OLSp)

Call:
lm(formula = Y ~ x1p + x2)
```

```
Residuals:
     Min       1Q   Median       3Q      Max 
-2.32999 -0.40944  0.09675  0.54201  1.96698 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   1.3850     0.0857  16.161  < 2e-16 ***
x1p           0.8461     0.1315   6.434 4.72e-09 ***
x2            1.9468     0.1767  11.021  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8414 on 97 degrees of freedom
Multiple R-squared:  0.6208,	Adjusted R-squared:  0.613 
F-statistic: 79.41 on 2 and 97 DF,  p-value: < 2.2e-16
```

c. Skenario simulasi 1 dengan 10% pencilan

```
> OLSp<-lm(Y~x1p+x2)
> summary(OLSp)

Call:
lm(formula = Y ~ x1p + x2)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.30164 -0.45076  0.07134  0.56216  1.90889 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   1.3624     0.0891  15.291  < 2e-16 ***
x1p           0.6692     0.1145   5.846 6.78e-08 ***
x2            1.9973     0.1819  10.978  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8643 on 97 degrees of freedom
Multiple R-squared:  0.5999,	Adjusted R-squared:  0.5917 
F-statistic: 72.73 on 2 and 97 DF,  p-value: < 2.2e-16
```

d. Skenario simulasi 2 tanpa pencilan

```
> OLS<-lm(Y~x1+x2)
> summary(OLS)

Call:
lm(formula = Y ~ x1 + x2)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.35487 -0.34916  0.00161  0.50953  2.11153 
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.42028    0.08416  16.877  < 2e-16 ***
x1           1.23378    0.18680   6.605 2.14e-09 ***
x2           1.91445    0.17517  10.929  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8347 on 97 degrees of freedom
Multiple R-squared:  0.6268,	Adjusted R-squared:  0.6191 
F-statistic: 81.46 on 2 and 97 DF,  p-value: < 2.2e-16
```

e. Skenario simulasi 2 dengan 5% pencilan

```
> OLSp<-lm(Y~x1+x2p)
> summary(OLSp)

Call:
lm(formula = Y ~ x1 + x2p)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.46037 -0.52029 -0.02748  0.53004  2.77444 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.32040    0.09146  14.437  < 2e-16 ***
x1           1.21766    0.20260   6.010 3.25e-08 ***
x2p          1.21974    0.13076   9.328 3.78e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9053 on 97 degrees of freedom
Multiple R-squared:  0.561,	Adjusted R-squared:  0.552 
F-statistic: 61.99 on 2 and 97 DF,  p-value: < 2.2e-16
```

f. Skenario simulasi 2 dengan 10% pencilan

```
> OLSp<-lm(Y~x1+x2p)
> summary(OLSp)

Call:
lm(formula = Y ~ x1 + x2p)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.4244 -0.4975 -0.0675  0.5082  2.2918 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.24590    0.08824  14.120  < 2e-16 ***
```

```
x1           1.31596    0.19387   6.788 9.09e-10 ***
x2p          1.07166    0.10492  10.214  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8655 on 97 degrees of freedom
Multiple R-squared:  0.5988,   Adjusted R-squared:  0.5905
F-statistic: 72.38 on 2 and 97 DF,  p-value: < 2.2e-16
```

g. Skenario simulasi 3 tanpa pencilan

```
> OLS<-lm(Y~x1+x2)
> summary(OLS)

Call:
lm(formula = Y ~ x1 + x2)

Residuals:
     Min       1Q   Median       3Q      Max
-2.35487 -0.34916  0.00161  0.50953  2.11153

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.42028    0.08416  16.877  < 2e-16 ***
x1           1.23378    0.18680   6.605 2.14e-09 ***
x2           1.91445    0.17517  10.929  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8347 on 97 degrees of freedom
Multiple R-squared:  0.6268,   Adjusted R-squared:  0.6191
F-statistic: 81.46 on 2 and 97 DF,  p-value: < 2.2e-16
```

h. Skenario simulasi 3 dengan 5% pencilan

```
> OLSp<-lm(Y~x1p+x2p)
> summary(OLSp)

Call:
lm(formula = Y ~ x1p + x2p)

Residuals:
     Min       1Q   Median       3Q      Max
-2.43524 -0.58415  0.00004  0.46915  2.63305

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.28329    0.09295  13.806  < 2e-16 ***
x1p          0.83775    0.14224   5.889 5.58e-08 ***
x2p          1.24537    0.13151   9.470 1.87e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 0.9102 on 97 degrees of freedom
Multiple R-squared:  0.5563,	Adjusted R-squared:  0.5471 
F-statistic:  60.8 on 2 and 97 DF,  p-value: < 2.2e-16
```

i. Skenario simulasi 3 dengan 10% pencilan

```
> OLSp<-lm(Y~x1p+x2p)
> summary(OLSp)

Call:
lm(formula = Y ~ x1p + x2p)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.38296 -0.55585 -0.01011  0.52434  2.06520 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.18239    0.09498  12.449  < 2e-16 ***
x1p          0.69204    0.12053   5.742 1.07e-07 ***
x2p          1.10567    0.11055  10.001  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9081 on 97 degrees of freedom
Multiple R-squared:  0.5583,	Adjusted R-squared:  0.5492 
F-statistic: 61.31 on 2 and 97 DF,  p-value: < 2.2e-16
```

j. Skenario simulasi 4 tanpa pencilan

```
> OLS<-lm(Y~x1+x2)
> summary(OLS)

Call:
lm(formula = Y ~ x1 + x2)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.35487 -0.34916  0.00161  0.50953  2.11153 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.42028    0.08416  16.877  < 2e-16 ***
x1           1.23378    0.18680   6.605 2.14e-09 ***
x2           1.91445    0.17517  10.929  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8347 on 97 degrees of freedom
Multiple R-squared:  0.6268,	Adjusted R-squared:  0.6191 
F-statistic: 81.46 on 2 and 97 DF,  p-value: < 2.2e-16
```

k. Skenario simulasi 4 dengan 5% pencilan

```
> OLSp<-lm(Yp~x1p+x2p)
> summary(OLSp)

Call:
lm(formula = Yp ~ x1p + x2p)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.2153 -0.4987  0.0251  0.5624  5.9993 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   1.4232     0.1235  11.523  < 2e-16 ***
x1p           1.0419     0.1890   5.513 2.92e-07 ***
x2p           1.9643     0.1747  11.241  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.209 on 97 degrees of freedom
Multiple R-squared:  0.6131,	Adjusted R-squared:  0.6051 
F-statistic: 76.84 on 2 and 97 DF,  p-value: < 2.2e-16
```

l. Skenario simulasi 4 dengan 10% pencilan

```
> OLSp<-lm(Yp~x1p+x2p)
> summary(OLSp)

Call:
lm(formula = Yp ~ x1p + x2p)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.3356 -0.6902  0.0256  0.5978  5.4475 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   1.4444     0.1578   9.153 9.05e-15 ***
x1p           0.8642     0.2003   4.315 3.84e-05 ***
x2p           1.9968     0.1837  10.871  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.509 on 97 degrees of freedom
Multiple R-squared:  0.5701,	Adjusted R-squared:  0.5612 
F-statistic: 64.31 on 2 and 97 DF,  p-value: < 2.2e-16
```

B.2.3 Estimasi dengan Metode *Least Trimmed Square*

a. Skenario simulasi 1 tanpa pencilan

```
> library(MASS)
> LTS<-ltsreg(Y~x1+x2)
> LTS
Call:
lqs.formula(formula = Y ~ x1 + x2, method = "lts")

Coefficients:
(Intercept)           x1           x2
      1.357        1.376        1.564

Scale estimates 0.6628 0.7188
```

b. Skenario simulasi 1 dengan 5% pencilan

```
> LTSp<-ltsreg(Y~x1p+x2)
> LTSp
Call:
lqs.formula(formula = Y ~ x1p + x2, method = "lts")

Coefficients:
(Intercept)          x1p           x2
      1.329        1.074        1.611

Scale estimates 0.6673 0.7060
```

c. Skenario simulasi 1 dengan 10% pencilan

```
> LTSp<-ltsreg(Y~x1p+x2)
> LTSp
Call:
lqs.formula(formula = Y ~ x1p + x2, method = "lts")

Coefficients:
(Intercept)          x1p           x2
      1.326        1.058        1.622

Scale estimates 0.7041 0.7639
```

d. Skenario simulasi 2 tanpa pencilan

```
> library(MASS)
> LTS<-ltsreg(Y~x1+x2)
> LTS
Call:
lqs.formula(formula = Y ~ x1 + x2, method = "lts")
```

```
Coefficients:
(Intercept)           x1           x2
      1.345        1.197        1.629

Scale estimates 0.6585 0.7017
```

e. Skenario simulasi 2 dengan 5% pencilan

```
> LTSp<-ltsreg(Y~x1+x2p)
> LTSp
Call:
lqs.formula(formula = Y ~ x1 + x2p, method = "lts")

Coefficients:
(Intercept)           x1          x2p
     1.3319       0.8242       1.0389

Scale estimates 0.7217 0.7751
```

f. Skenario simulasi 2 dengan 10% pencilan

```
> LTSp<-ltsreg(Y~x1+x2p)
> LTSp
Call:
lqs.formula(formula = Y ~ x1 + x2p, method = "lts")

Coefficients:
(Intercept)           x1          x2p
     1.2397       0.8072       0.9028

Scale estimates 0.6957 0.7914
```

g. Skenario simulasi 3 tanpa pencilan

```
> library(MASS)
> LTS<-ltsreg(Y~x1+x2)
> LTS
Call:
lqs.formula(formula = Y ~ x1 + x2, method = "lts")

Coefficients:
(Intercept)           x1           x2
      1.358        1.297        1.552

Scale estimates 0.6563 0.7023
```

h. Skenario simulasi 3 dengan 5% pencilan

```
> LTSp<-ltsreg(Y~x1p+x2p)
> LTSp
Call:
```

```
lqs.formula(formula = Y ~ x1p + x2p, method = "lts")

Coefficients:
(Intercept)          x1p          x2p  
     1.3438       0.8343       1.0695  

Scale estimates 0.7117 0.7620
```

i. Skenario simulasi 3 dengan 10% pencilan

```
> LTSp<-ltsreg(Y~x1p+x2p)
> LTSp
Call:
lqs.formula(formula = Y ~ x1p + x2p, method = "lts")

Coefficients:
(Intercept)          x1p          x2p  
     1.2797       0.8430       0.9993  

Scale estimates 0.7249 0.7831
```

j. Skenario simulasi 4 tanpa pencilan

```
> library(MASS)
> LTS<-ltsreg(Y~x1+x2)
> LTS
Call:
lqs.formula(formula = Y ~ x1 + x2, method = "lts")

Coefficients:
(Intercept)           x1           x2  
      1.358        1.356        1.577  

Scale estimates 0.6591 0.7178
```

k. Skenario simulasi 4 dengan 5% pencilan

```
> LTSp<-ltsreg(Yp~x1p+x2p)
> LTSp
Call:
lqs.formula(formula = Yp ~ x1p + x2p, method = "lts")

Coefficients:
(Intercept)          x1p          x2p  
      1.348        1.054        1.447  

Scale estimates 0.7582 0.8029
```

l. Skenario simulasi 4 dengan 10% pencilan

```
> LTSp<-ltsreg(Yp~x1p+x2p)
```

```
> LTSp
Call:
lqs.formula(formula = Yp ~ x1p + x2p, method = "lts")

Coefficients:
(Intercept)          x1p          x2p
     1.1427       0.3884       0.8197

Scale estimates 0.8039 0.8318
```

B.3 Tahapan Analisis Data

B.3.1 Uji hipotesis β dengan menggunakan Uji T

a. Hasil Uji T pada metode OLS telah terangkum dalam formula dari OLS itu sendiri yang ada pada pembahasan sebelumnya (B.2.2) baik pada OLS tanpa pencilan maupun OLS dengan pencilan (OLSp) pada semua skenario simulasi.

b. Skenario simulasi 1 tanpa pencilan

```
> summary(LTS)
              Length Class      Mode
crit            1    -none-     numeric
sing            1    -none-     character
coefficients    3    -none-     numeric
bestone         3    -none-     numeric
fitted.values 100    -none-     numeric
residuals     100    -none-     numeric
scale           2    -none-     numeric
terms           3    terms      call
call            3    -none-     call
xlevels         0    -none-     list
model           3    data.frame list
> LTS$coefficients[2]
      x1
1.376302
> LTS$coefficients[3]
      x2
1.564157
> T1 <- LTS$coefficient[2]/LTS$scale[1]
> T1
      x1
2.076515
> T2 <- LTS$coefficient[3]/LTS$scale[2]
> T2
      x2
2.175959
> p.value1 = dt(2.076515 , df=100-3)
> p.value1
[1] 0.04723241
> p.value2 = dt(2.175959 , df=100-3)
```

```
> p.value2
[1] 0.03851226
```

c. Skenario simulasi 1 dengan 5% pencilan

```
> summary(LTSp)
              Length Class      Mode
crit            1    -none-     numeric
sing            1    -none-     character
coefficients    3    -none-     numeric
bestone         3    -none-     numeric
fitted.values 100    -none-     numeric
residuals     100    -none-     numeric
scale           2    -none-     numeric
terms           3    terms      call
call            3    -none-     call
xlevels         0    -none-     list
model           3    data.frame list
> LTSp$coefficients[2]
     x1p
1.073621
> LTSp$coefficients[3]
      x2
1.611291
> T1p <- LTSp$coefficient[2]/LTSp$scale[1]
> T1p
     x1p
1.608819
> T2p <- LTSp$coefficient[3]/LTSp$scale[2]
> T2p
      x2
2.282313
> p.value1p = dt(1.608819 , df=100-3)
> p.value1p
[1] 0.109496
> p.value2p = dt(2.282313 , df=100-3)
> p.value2p
[1] 0.03066534
```

d. Skenario simulasi 1 dengan 10% pencilan

```
> summary(LTSp)
              Length Class      Mode
crit            1    -none-     numeric
sing            1    -none-     character
coefficients    3    -none-     numeric
bestone         3    -none-     numeric
fitted.values 100    -none-     numeric
residuals     100    -none-     numeric
scale           2    -none-     numeric
terms           3    terms      call
call            3    -none-     call
```

```
xlevels         0    -none-     list
model           3    data.frame list
> LTSp$coefficients[2]
     x1p
1.057575
> LTSp$coefficients[3]
      x2
1.621651
> T1p <- LTSp$coefficient[2]/LTSp$scale[1]
> T1p
     x1p
1.502009
> T2p <- LTSp$coefficient[3]/LTSp$scale[2]
> T2p
      x2
2.122849
> p.value1p = dt(1.502009 , df=100-3)
> p.value1p
[1] 0.1289785
> p.value2p = dt(2.122849 , df=100-3)
> p.value2p
[1] 0.04299416
```

e. Skenario simulasi 2 tanpa pencilan

```
> summary(LTS)
              Length Class      Mode
crit            1    -none-     numeric
sing            1    -none-     character
coefficients    3    -none-     numeric
bestone         3    -none-     numeric
fitted.values 100    -none-     numeric
residuals     100    -none-     numeric
scale           2    -none-     numeric
terms           3    terms      call
call            3    -none-     call
xlevels         0    -none-     list
model           3    data.frame list
> LTS$coefficients[2]
      x1
1.197353
> LTS$coefficients[3]
      x2
1.629323
> T1 <- LTS$coefficient[2]/LTS$scale[1]
> T1
     x1
1.81837
> T2 <- LTS$coefficient[3]/LTS$scale[2]
> T2
      x2
2.322048
```

```
> p.value1 = dt(1.81837 , df=100-3)
> p.value1
[1] 0.07699942
> p.value2 = dt(2.322048 , df=100-3)
> p.value2
[1] 0.02809246
```

f. Skenario simulasi 2 dengan 5% pencilan

```
> summary(LTSp)
              Length Class      Mode
crit            1    -none-     numeric
sing            1    -none-     character
coefficients    3    -none-     numeric
bestone         3    -none-     numeric
fitted.values 100    -none-     numeric
residuals     100    -none-     numeric
scale           2    -none-     numeric
terms           3    terms      call
call            3    -none-     call
xlevels         0    -none-     list
model           3    data.frame list
> LTSp$coefficients[2]
       x1
0.8242254
> LTSp$coefficients[3]
     x2p
1.038857
> T1p <- LTSp$coefficient[2]/LTSp$scale[1]
> T1p
      x1
1.142069
> T2p <- LTSp$coefficient[3]/LTSp$scale[2]
> T2p
     x2p
1.340224
> p.value1p = dt(1.142069 , df=100-3)
> p.value1p
[1] 0.2067987
> p.value2p = dt(1.340224 , df=100-3)
> p.value2p
[1] 0.1619324
```

g. Skenario simulasi 2 dengan 10% pencilan

```
> summary(LTSp)
              Length Class      Mode
crit            1    -none-     numeric
sing            1    -none-     character
coefficients    3    -none-     numeric
bestone         3    -none-     numeric
fitted.values 100    -none-     numeric
```

```
residuals     100    -none-     numeric
scale           2    -none-     numeric
terms           3    terms      call
call            3    -none-     call
xlevels         0    -none-     list
model           3    data.frame list
> LTSp$coefficients[2]
       x1
0.8072488
> LTSp$coefficients[3]
      x2p
0.9027599
> T1p <- LTSp$coefficient[2]/LTSp$scale[1]
> T1p
      x1
1.160379
> T2p <- LTSp$coefficient[3]/LTSp$scale[2]
> T2p
     x2p
1.140652
> p.value1p = dt(1.160379 , df=100-3)
> p.value1p
[1] 0.2024994
> p.value2p = dt(1.140652 , df=100-3)
> p.value2p
[1] 0.2071324
```

h. Skenario simulasi 3 tanpa pencilan

```
> summary(LTS)
              Length Class      Mode
crit            1    -none-     numeric
sing            1    -none-     character
coefficients    3    -none-     numeric
bestone         3    -none-     numeric
fitted.values 100    -none-     numeric
residuals     100    -none-     numeric
scale           2    -none-     numeric
terms           3    terms      call
call            3    -none-     call
xlevels         0    -none-     list
model           3    data.frame list
> LTS$coefficients[2]
      x1
1.297399
> LTS$coefficients[3]
      x2
1.551784
> T1 <- LTS$coefficient[2]/LTS$scale[1]
> T1
      x1
1.976717
```

```
> T2 <- LTS$coefficient[3]/LTS$scale[2]
> T2
      x2
2.209491
> p.value1 = dt(1.976717 , df=100-3)
> p.value1
[1] 0.05746135
> p.value2 = dt(2.209491 , df=100-3)
> p.value2
[1] 0.03588074
```

i. Skenario simulasi 3 dengan 5% pencilan

```
> summary(LTSp)
              Length Class      Mode
crit            1    -none-     numeric
sing            1    -none-     character
coefficients    3    -none-     numeric
bestone         3    -none-     numeric
fitted.values 100    -none-     numeric
residuals     100    -none-     numeric
scale           2    -none-     numeric
terms           3    terms      call
call            3    -none-     call
xlevels         0    -none-     list
model           3    data.frame list
> LTSp$coefficients[2]
      x1p
0.8342546
> LTSp$coefficients[3]
     x2p
1.069544
> T1p <- LTSp$coefficient[2]/LTSp$scale[1]
> T1p
     x1p
1.172263
> T2p <- LTSp$coefficient[3]/LTSp$scale[2]
> T2p
     x2p
1.403644
> p.value1p = dt(1.172263 , df=100-3)
> p.value1p
[1] 0.1997221
> p.value2p = dt(1.502883 , df=100-3)
> p.value2p
[1] 0.1288114
```

j. Skenario simulasi 3 dengan 10% pencilan

```
> summary(LTSp)
              Length Class      Mode
crit            1    -none-     numeric
```

```
sing              1     -none-     character
coefficients      3     -none-     numeric
bestone           3     -none-     numeric
fitted.values   100     -none-     numeric
residuals       100     -none-     numeric
scale             2     -none-     numeric
terms             3     terms      call
call              3     -none-     call
xlevels           0     -none-     list
model             3     data.frame list
> LTSp$coefficients[2]
      x1p
0.8429939
> LTSp$coefficients[3]
      x2p
0.9992638
> T1p <- LTSp$coefficient[2]/LTSp$scale[1]
> T1p
     x1p
1.162971
> T2p <- LTSp$coefficient[3]/LTSp$scale[2]
> T2p
     x2p
1.275967
> p.value1p = dt(1.162971 , df=100-3)
> p.value1p
[1] 0.2018927
> p.value2p = dt(1.275967 , df=100-3)
> p.value2p
[1] 0.1760252
```

k. Skenario simulasi 4 tanpa pencilan

```
> summary(LTS)
              Length Class      Mode
crit              1     -none-     numeric
sing              1     -none-     character
coefficients      3     -none-     numeric
bestone           3     -none-     numeric
fitted.values   100     -none-     numeric
residuals       100     -none-     numeric
scale             2     -none-     numeric
terms             3     terms      call
call              3     -none-     call
xlevels           0     -none-     list
model             3     data.frame list
> LTS$coefficients[2]
      x1
1.355773
> LTS$coefficients[3]
      x2
1.576652
```

```
> T1 <- LTS$coefficient[2]/LTS$scale[1]
> T1
      x1
2.056963
> T2 <- LTS$coefficient[3]/LTS$scale[2]
> T2
      x2
2.196534
> p.value1 = dt(2.056963 , df=100-3)
> p.value1
[1] 0.04911595
> p.value2 = dt(2.196534 , df=100-3)
> p.value2
```

l. Skenario simulasi 4 dengan 5% pencilan

```
> summary(LTSp)
              Length Class      Mode
crit            1    -none-     numeric
sing            1    -none-     character
coefficients    3    -none-     numeric
bestone         3    -none-     numeric
fitted.values 100    -none-     numeric
residuals     100    -none-     numeric
scale           2    -none-     numeric
terms           3    terms      call
call            3    -none-     call
xlevels         0    -none-     list
model           3    data.frame list
> LTSp$coefficients[2]
    x1p
1.05366
> LTSp$coefficients[3]
     x2p
1.446729
> T1p <- LTSp$coefficient[2]/LTSp$scale[1]
> T1p
     x1p
1.389693
> T2p <- LTSp$coefficient[3]/LTSp$scale[2]
> T2p
     x2p
1.801812
> p.value1p = dt(1.389693 , df=100-3)
> p.value1p
[1] 0.1514486
> p.value2p = dt(1.801812 , df=100-3)
> p.value2p
[1] 0.07928817
```

m. Skenario simulasi 4 dengan 10% pencilan

```
> summary(LTSp)
              Length Class      Mode
crit            1    -none-     numeric
sing            1    -none-     character
coefficients    3    -none-     numeric
bestone         3    -none-     numeric
fitted.values 100    -none-     numeric
residuals     100    -none-     numeric
scale           2    -none-     numeric
terms           3    terms      call
call            3    -none-     call
xlevels         0    -none-     list
model           3    data.frame list
> LTSp$coefficients[2]
      x1p
0.3883969
> LTSp$coefficients[3]
      x2p
0.8196658
> T1p <- LTSp$coefficient[2]/LTSp$scale[1]
> T1p
      x1p
0.4831435
> T2p <- LTSp$coefficient[3]/LTSp$scale[2]
> T2p
      x2p
0.9854345
> p.value1p = dt(0.4831435 , df=100-3)
> p.value1p
[1] 0.3537053
> p.value2p = dt(0.9854345 , df=100-3)
> p.value2p
[1] 0.2442354
```

B.3.2 Analisis *Mean Square Error* (MSE)

a. Skenario simulasi 1 dengan 5% pencilan

```
> YpredOLS<-fitted(OLS)
> YpredLTS<-fitted(LTS)
> YpredOLSp<-fitted(OLSp)
> YpredLTSp<-fitted(LTSp)
> MSE1<-sum((YpredOLS-Y)^2)/n
> MSE1
[1] 0.6758742
> MSE2<-sum((YpredLTS-Y)^2)/n
> MSE2
[1] 0.7102358
> MSE3<-sum((YpredOLSp-Y)^2)/n
> MSE3
```

```
[1] 0.6867527
> MSE4<-sum((YpredLTSp-Y)^2)/n
> MSE4
[1] 0.7353899
```

b. Skenario simulasi 1 dengan 10% pencilan

```
> YpredOLS<-fitted(OLS)
> YpredLTS<-fitted(LTS)
> YpredOLSp<-fitted(OLSp)
> YpredLTSp<-fitted(LTSp)
> MSE1<-sum((YpredOLS-Y)^2)/n
> MSE1
[1] 0.6758742
> MSE2<-sum((YpredLTS-Y)^2)/n
> MSE2
[1] 0.7020874
> MSE3<-sum((YpredOLSp-Y)^2)/n
> MSE3
[1] 0.7245708
> MSE4<-sum((YpredLTSp-Y)^2)/n
> MSE4
[1] 0.8533732
```

c. Skenario simulasi 2 dengan 5% pencilan

```
> YpredOLS<-fitted(OLS)
> YpredLTS<-fitted(LTS)
> YpredOLSp<-fitted(OLSp)
> YpredLTSp<-fitted(LTSp)
> MSE1<-sum((YpredOLS-Y)^2)/n
> MSE1
[1] 0.6758742
> MSE2<-sum((YpredLTS-Y)^2)/n
> MSE2
[1] 0.6997916
> MSE3<-sum((YpredOLSp-Y)^2)/n
> MSE3
[1] 0.7950003
> MSE4<-sum((YpredLTSp-Y)^2)/n
> MSE4
[1] 0.8423025
```

d. Skenario simulasi 2 dengan 10% pencilan

```
> YpredOLS<-fitted(OLS)
> YpredLTS<-fitted(LTS)
> YpredOLSp<-fitted(OLSp)
> YpredLTSp<-fitted(LTSp)
> MSE1<-sum((YpredOLS-Y)^2)/n
> MSE1
```

```
[1] 0.6758742
> MSE2<-sum((YpredLTS-Y)^2)/n
> MSE2
[1] 0.6933536
> MSE3<-sum((YpredOLSp-Y)^2)/n
> MSE3
[1] 0.7266618
> MSE4<-sum((YpredLTSp-Y)^2)/n
> MSE4
[1] 0.7981071
```

e. Skenario simulasi 3 dengan 5% pencilan

```
> YpredOLS<-fitted(OLS)
> YpredLTS<-fitted(LTS)
> YpredOLSp<-fitted(OLSp)
> YpredLTSp<-fitted(LTSp)
> MSE1<-sum((YpredOLS-Y)^2)/n
> MSE1
[1] 0.6758742
> MSE2<-sum((YpredLTS-Y)^2)/n
> MSE2
[1] 0.709271
> MSE3<-sum((YpredOLSp-Y)^2)/n
> MSE3
[1] 0.8036589
> MSE4<-sum((YpredLTSp-Y)^2)/n
> MSE4
[1] 0.8210486
```

f. Skenario simulasi 3 dengan 10% pencilan

```
> YpredOLS<-fitted(OLS)
> YpredLTS<-fitted(LTS)
> YpredOLSp<-fitted(OLSp)
> YpredLTSp<-fitted(LTSp)
> MSE1<-sum((YpredOLS-Y)^2)/n
> MSE1
[1] 0.6758742
> MSE2<-sum((YpredLTS-Y)^2)/n
> MSE2
[1] 0.6954044
> MSE3<-sum((YpredOLSp-Y)^2)/n
> MSE3
[1] 0.7999541
> MSE4<-sum((YpredLTSp-Y)^2)/n
> MSE4
[1] 0.8354568
```

g. Skenario simulasi 4 dengan 5% pencilan

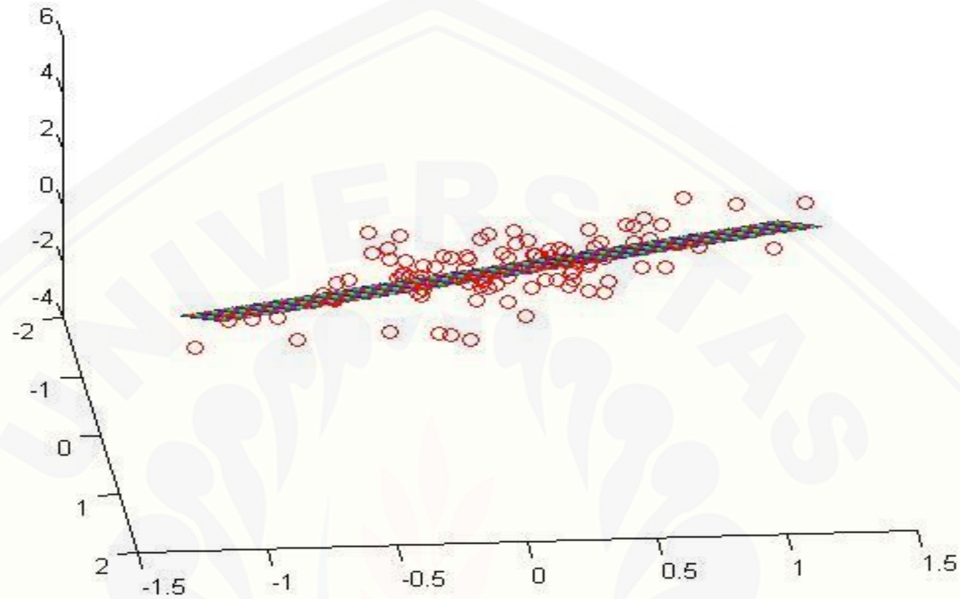
```
> YpredOLS<-fitted(OLS)
> YpredLTS<-fitted(LTS)
> YpredOLSp<-fitted(OLSp)
> YpredLTSp<-fitted(LTSp)
> MSE1<-sum((YpredOLS-Y)^2)/n
> MSE1
[1] 0.6758742
> MSE2<-sum((YpredLTS-Y)^2)/n
> MSE2
[1] 0.7071867
> MSE3<-sum((YpredOLSp-Y)^2)/n
> MSE3
[1] 1.106354
> MSE4<-sum((YpredLTSp-Y)^2)/n
> MSE4
[1] 0.8515324
```

h. Skenario simulasi 4 dengan 10% pencilan

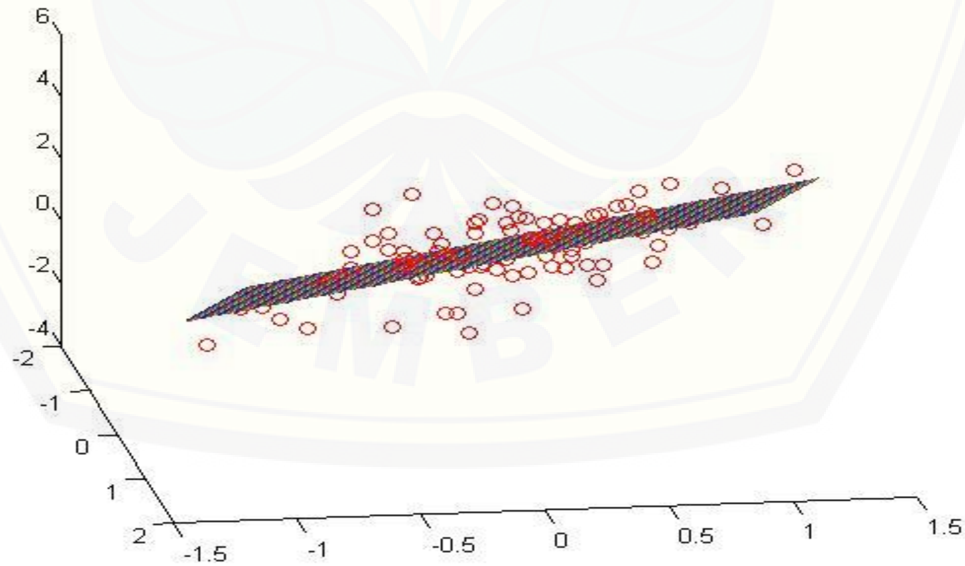
```
> YpredOLS<-fitted(OLS)
> YpredLTS<-fitted(LTS)
> YpredOLSp<-fitted(OLSp)
> YpredLTSp<-fitted(LTSp)
> MSE1<-sum((YpredOLS-Y)^2)/n
> MSE1
[1] 0.6758742
> MSE2<-sum((YpredLTS-Y)^2)/n
> MSE2
[1] 0.6965125
> MSE3<-sum((YpredOLSp-Y)^2)/n
> MSE3
[1] 1.503547
> MSE4<-sum((YpredLTSp-Y)^2)/n
> MSE4
[1] 0.9152311
```

B.3.3 Hasil Plot 3D Menggunakan *Matlab*

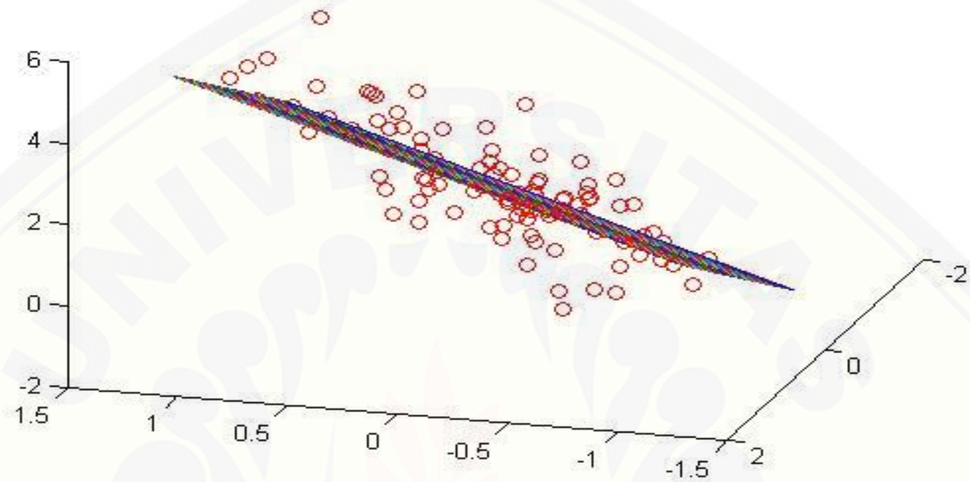
a. Skenario simulasi dengan parameter sebenarnya $\beta_0 = 1{,}4, \beta_1 = 1{,}2, \beta_2 = 2.0$



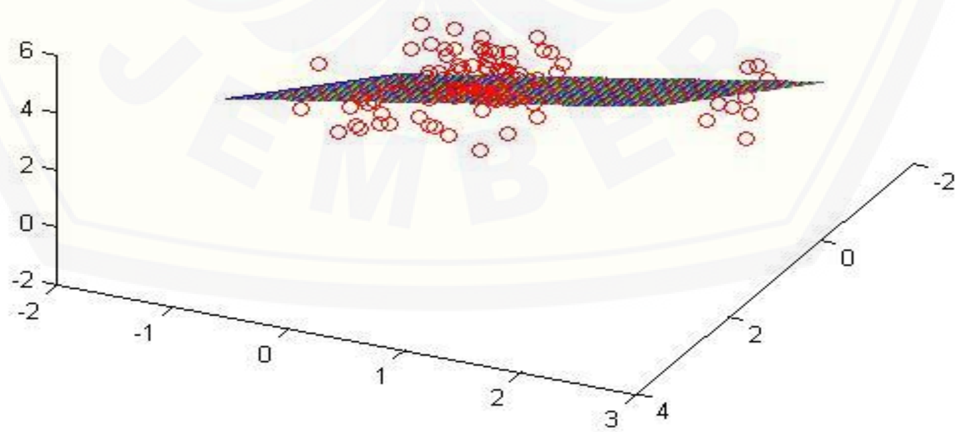
b. Hasil Estimasi metode OLS tanpa pencilan



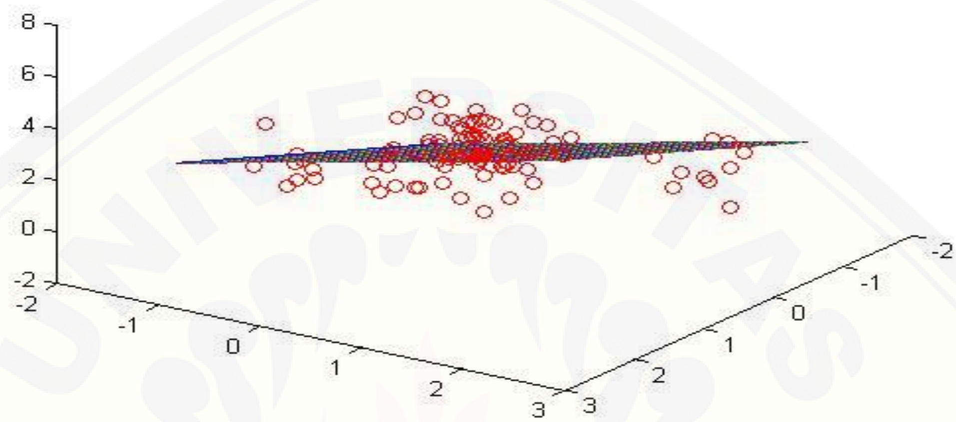
c. Hasil Estimasi metode LTS tanpa pencilan



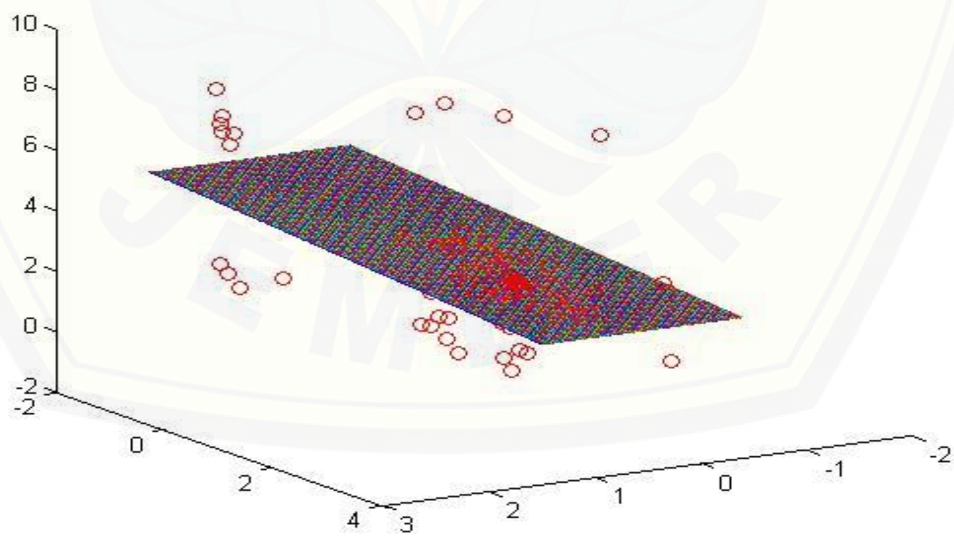
d. Hasil Estimasi metode OLS dengan pencilan x_1 dan x_2 pada presentase 10%



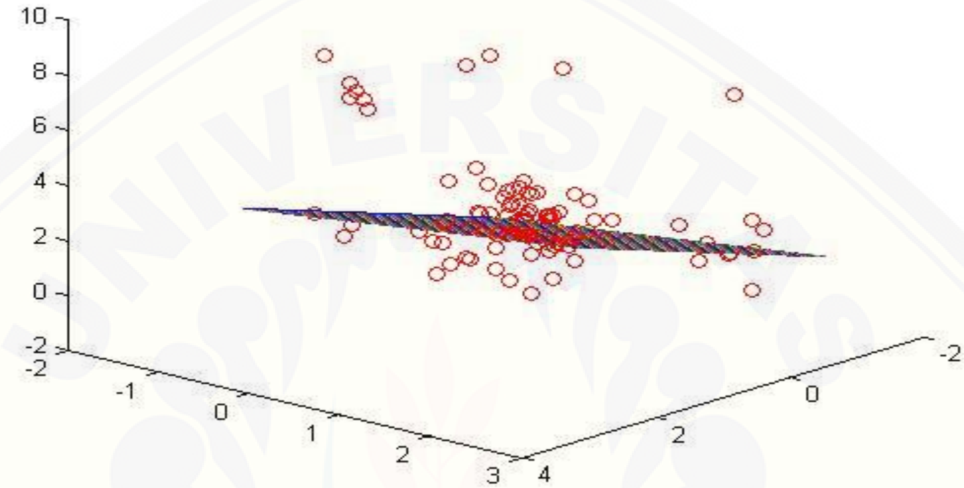
e. Hasil Estimasi metode LTS dengan pencilan x_1 dan x_2 pada presentase 10%



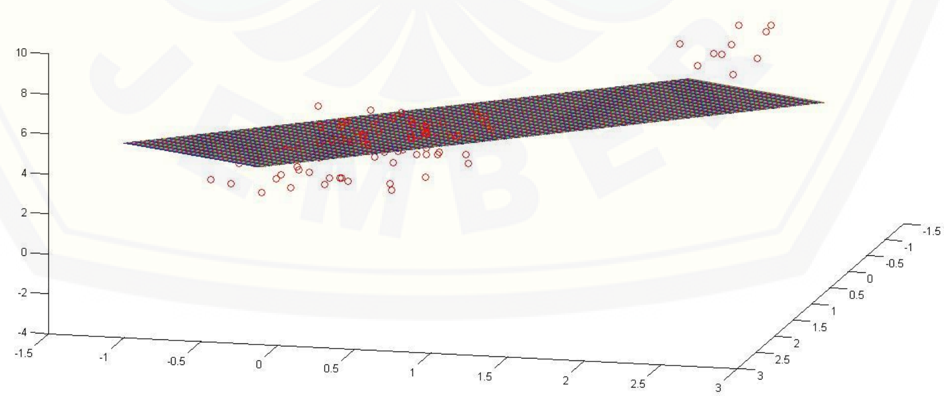
f. Hasil Estimasi metode OLS dengan pencilan x_1, x_2 dan y pada presentase pencilan 10%



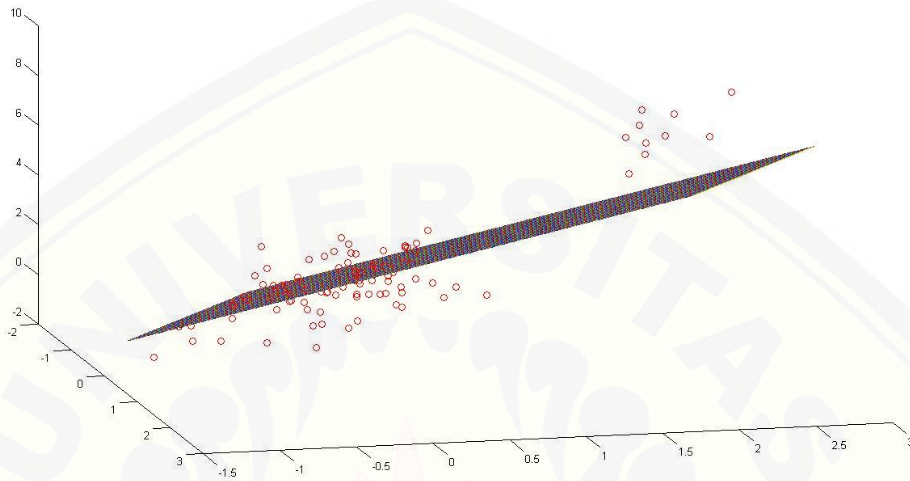
g. Hasil Estimasi metode LTS dengan pencilan x_1, x_2 dan y pada presentase pencilan 10%



h. Skenario simulasi tambahan metode OLS dengan data pencilan x_1, x_2 dan nilai y didapat dari $Y = X_{1\,pencilan}\beta + X_{2\,pencilan}\beta + \varepsilon$



i. Skenario simulasi tambahan metode LTS dengan data pencilan x_1, x_2 dan nilai y didapat dari $Y = X_{1\,pencilan}\beta + X_{2\,pencilan}\beta + \varepsilon$



Rangkuman Nilai MSE pada skenario tambahan point h dan i diatas

MSE	
OLS	LTS
0,6761782	0,7245434