**LAPORAN PENELITIAN**

# PERBANDINGAN METODE ESTIMASI *LTS*, ESTIMASI *M*, DAN ESTIMASI *MM* PADA REGRESI *ROBUST*

**Diusulkan Oleh:**
**Dr. Edy Widodo, S.Si., M.Si**
**Arlinda Amalia Dewayanti**

**PROGRAM STUDI STATISTIKA**
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**
**UNIVERSITAS ISLAM INDONESIA**
**YOGYAKARTA**
**2016**

## HALAMAN PENGESAHAN

1. **Identitas Penelitian**
   a. Judul Penelitian : Perbandingan Metode Estimasi *LTS*, Estimasi *M*, dan Estimasi *MM* pada Regresi *Robust*
   b. Bidang Ilmu : Statistika
   c. Kategori Penelitian : Unggulan

2. **Ketua Peneliti**
   a. Nama Lengkap dan Gelar : Dr.Edy Widodo, M.Si.
   b. Jenis Kelamin : Laki Laki
   c. Golongan dan Pangkat : IIIc/Penata
   d. NIP/NIK : 966110103
   e. Jabatan Fungsional : Lektor
   f. Fakultas/Jurusan : FMIPA/Statistika

3. **Alamat Ketua Peneliti**
   a. Alamat Kantor : FMIPA UII, Jalan Kaliurang Km 14,4 Yogjakarta
   b. Telp/Fax: : 0274 895920 Ext 3042/Fax Ext 3020
   c. E-mail : edywidodo@uii.ac.id
   d. Alamat Rumah : Dusun Mendiro, RT 04 RW 26 Sukoharjo Ngaglik Sleman Yogyakarta
   e. Telp/ Hp : 08224248225

4. **Jumlah Anggota Peneliti**
   a. Anggota Peneliti I : Arlinda Amalia Dewayanti
5. **Lokasi Penelitian** : Laboratorium Statistika FMIPA UII
6. **Jangka Waktu Pelaksanaan** : 3 Bulan

Yogyakarta, 13 April 2016

Mengetahui:
Ketua Jurusan Statistika

( Dr. RB. Fajriya Hakim, M.Si )
NIP/NIK: 986110101

Ketua Peneliti

(Dr. Edy Widodo, M.Si.)
NIP/NIK: 966110103

Menyetujui, Direktur
DPPM UII

UNIVERSITAS ISLAM INDONESIA
DIREKTORAT PENELITIAN DAN PENGABDIAN MASYARAKAT
YOGYAKARTA

( Prof. Akhmad Fauzi, Ph.D )
NIP/NIK: 956110101

# DAFTAR ISI

**DAFTAR TABEL**

# DAFTAR GAMBAR

# DAFTAR LAMPIRAN

# BAB I

# PENDAHULUAN

## 1.1 Latar Belakang Masalah

Analisis regresi merupakan salah satu alat dalam analisis statistik yang memanfaatkan hubungan antara dua variabel atau lebih. Tujuannya adalah untuk membuat perkiraan (prediksi) yang dapat dipercaya untuk nilai variabel dependen, jika nilai variabel lain yang berhubungan dengan diketahui variabel independen (Qudratullah, 2013). Salah satu metode dalam mengestimasi parameter-parameter pada model regresi linear adalah Metode Kuadrat Terkecil (MKT) atau *Ordinary Least Square* (*OLS*).

MKT adalah salah satu metode estimasi parameter pada regresi yang dilakukan dengan meminimumkan jumlah simpangan kuadrat residual. Pada MKT terdapat sifat *BLUE* (*Best Linear Unbias Estimator*) dimana asumsi klasik harus terpenuhi. Asumsi klasik tersebut harus dipenuhi oleh komponen residual pada model yang dihasilkan. Asumsi-asumsi tersebut yaitu asumsi normalitas, asumsi homoskedastisitas, asumsi non-autokorelasi, dan asumsi non-multikolinearitas.

Pada berbagai kasus, tidak jarang ditemukan kondisi dimana asumsi-asumsi tersebut tidak terpenuhi. Jika asumsi tidak terpenuhi akan mengakibatkan hasil estimasi parameter pada MKT kurang baik. Diantara asumsi tersebut, salah satu asumsi yang tidak terpenuhi adalah asumsi normalitas. Hal ini disebabkan adanya *outlier* pada data pengamatan.

*Outlier* adalah kasus atau data yang memiliki karakteristik unik yang terlihat sangat berbeda jauh dari observasi-observasi lainnya dan muncul dalam bentuk nilai ekstrim, baik untuk sebuah variabel (Ghozali, 2005). Ada tiga jenis *outlier* menurut Soemartini (2007) yaitu *outlier* pada variabel dependen atau pada arah $y$ (*outlier* vertikal), *outlier* pada variabel independen atau pada arah $x$ *(good leverage point)*, dan *outlier* pada arah $x$

dan *y (bad leverage point)*. Adanya *outlier* dapat menyebabkan residual yang besar. Oleh karena itu diperlukan metode lain untuk menangani adanya *outlier* yaitu Metode Regresi *Robust* (MRR).

MRR adalah metode yang digunakan dalam mengatasi *outlier* tanpa menghapus data *outlier* tersebut. Suatu estimasi yang *resistance* adalah relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data (Mashitah, dkk, 2013). Terdapat berbagai macam MRR diantaranya estimasi *M* (*Maximum Likelihood Type*), estimasi *S* (*Scale*), estimasi *MM* (*Method Of Moment*), estimasi *LTS* (*Least Terimmed Square*) dan estimasi *LMS* (*Least Median Square*).

Estimasi *LTS* merupakan metode yang mempunyai nilai *breakdown point* yang tinggi yaitu hampir 50%. Pada estimasi *LTS* pertama-tama menghitung *h*, banyak data yang menjadikan estimasi *robust*, dengan sebelumnya menyusun residual kuadrat dari yang terkecil sampai dengan yang terbesar (Nurcahyadi, 2010).

Estimasi *M* merupakan MRR yang baik untuk menduga parameter yang memiliki *breakdown point* 1/n (Bekti, 2011). Estimasi *M* dilakukan dengan cara memberi bobot pada $e_i$ kemudian perhitungan nilai parameter dilakukan dengan *WLS* (*Weighted Least Square*).

Estimasi *MM* dilakukan dengan cara menggabungkan cara estimasi pada metode estimasi *S* dengan cara estimasi pada metode estimasi *M*. Perhitungan nilai parameter dilakukan dengan menggunakan *WLS* (*Weighted Least Square*). Metode ini berusaha untuk mempertahankan sifat *robust* dan *resistance* dari estimasi *S*. Selain itu, metode ini juga mempertahankan sifat efisien dari estimasi *M* (Susanti, dkk, 2013).

Penelitian-penelitian yang pernah dilakukan mengenai metode estimasi *M* dan estimasi *MM* pada regresi *robust*, antara lain skripsi Safitri (2015) mengenai "Perbandingan Metode Estimasi *M* dan Estimasi *MM* Pada Regresi *Robust*". Selain itu ada pula penelitian Nurcahyadi (2010) mengenai "Analisis Regresi Pada Data *Outlier* dengan Menggunakan *LTS* dan Estimasi *MM*".

Oleh karena itu, dilakukan penelitian untuk membandingkan metode estimasi *LTS*, estimasi *M*, dan estimasi *MM* pada Regresi *Robust*. Tujuan lain dari penelitian ini adalah untuk mengetahui metode estimasi tersebut mana yang paling baik digunakan dalam mengestimasi data yang mengandung *outlier*. Selain itu, dilakukan pengelolahan data dengan menggunakan *software* R 2.14.2.

## 1.2. Rumusan Masalah

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

a. Bagaimana cara mengestimasi nilai-nilai parameter model regresi dengan adanya data *outlier* dengan menggunakan metode estimasi *LTS*, metode estimasi *M*, dan metode estimasi *MM*.
b. Bagaimana model regresi yang dihasilkan dengan menggunakan metode estimasi *LTS*, metode estimasi *M*, dan metode estimasi *MM*.
c. Dari ketiga metode tersebut, metode manakah yang paling baik dalam mengestimasi data yang mengandung *outlier*.

## 1.3. Tujuan Penelitian

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

a. Bagaimana cara mengestimasi nilai-nilai parameter model regresi dengan adanya data *outlier* dengan menggunakan metode estimasi *LTS*, metode estimasi *M*, dan metode estimasi *MM*.
b. Bagaimana model regresi yang dihasilkan dengan menggunakan metode estimasi *LTS*, metode estimasi *M*, dan metode estimasi *MM*.

c. Dari ketiga metode tersebut, metode manakah yang paling baik dalam mengestimasi data yang mengandung *outlier*.

### 1.4. Manfaat Penelitian

Jika sudah diketahui cara mengestimasi nilai-nilai parameter model regresi dengan menggunakan metode estimasi *robust* yaitu estimasi *LTS*, estimasi *M*, dan estimasi *MM*, maka diharapkan akan mempermudah dalam pengolahan dan analisis data yang mengandung *outlier*. Kemudian pada pemilihan metode estimasi yang tepat diperlukan dalam melihat hasil estimasi yang baik. Oleh sebab itu, perbandingan beberapa metode diperlukan untuk melihat hasil yang paling baik sehingga metode tersebut dapat digunakan untuk penyelesaian kasus regresi dengan data mengandung *outlier*. Hasil estimasi yang paling baik dari estimasi ini berguna untuk melihat prediksi dari suatu kasus tersebut.

# BAB II
# TINJAUAN PUSTAKA

Tinjauan pustaka atau penelitian terdahulu digunakan sebagai dasar dalam melakukan penelitian, dimana dalam penelititian digunakan sebagai bahan kajian untuk melihat hubungan penelitian sebelumnya dengan penelitian yang dilakukan. Tujuan dari penulisan tinjauan pustaka untuk menghindari duplikasi penulisan karya ilmiah.

Safitri (2015) dalam skripsi yang brjudul "Perbandingan Metode Estimasi *Maximum Likelihood Type* (*M*) dan Estimasi *Method Of Moment* (*MM*) pada Regresi *Robust*". Penelitian ini bertujuan untuk mengetahui cara mengestimasi nilai-nilai parameter model regresi dengan menggunakan estimasi *M* dan estimasi *MM* pada data yang mengandung *outlier*, mengetahui perbandingan model regresi yang dihasilkan dengan menggunakan estimasi *M* dan estimasi *MM*, dan mengetahui hasil estimasi yang paling baik diantara estimasi *M* dan estimasi *MM* dalam mengestimasi data yang mengandung *outlier*. Hasil yang diperoleh dari contoh penerapan menunjukkan bahwa untuk data yang mengandung *outlier* estimasi parameter yang diperoleh pada metode regresi *robust* dengan metode estimasi *M* dan *MM* lebih baik digunakan dibandingkan metode MKT. Sedangkan untuk data tanpa *outlier* estimasi parameter yang diperoleh dengan metode MKT lebih baik dibandingkan metode regresi *robust* dengan metode estimasi *M* dan *MM*.

Nurcahyadi (2010) dalam skripsi yang berjudul "Analisis Regresi pada Data *Outlier* dengan Menggunakan *Least Trimmed Square* (*LTS*) dan *MM*-Estimasi". Penelitian ini bertujuan untuk mengidentifikasi data outlier dengan menggunakan *leverage*, nilai *discrepancy*, dan nilai *influence* dari data regresi, mengetahui cara mengestimasi nilai-nilai parameter regresi dengan menggunakan *LTS* dan *MM* pada data regresi, serta membandingkan model regresi yang dihasilkan pada dua metode tersebut. Hasil dari penelitian ini adalah estimasi *LTS* dapat menghasilkan model regresi yang *fit* terhadap data walaupun setengah dari datanya merupakan data *outlier*, karena mempunyai nilai *breakdown point* yang tinggi yaitu 50%. Metode *robust* lain yang memiliki nilai *breakdown point* 50% adalah estimasi *MM* yang menggunakan iterasi awal estimasi *S*. Model *LTS* sangat baik pada analisis regresi sederhana dibandingkan estimasi *MM* dilihat dari estimasi skala *residual standard error*. Sedangkan pada analisis regresi berganda estimasi *MM* lebih baik jika dibandingkan dengan *LTS* dilihat dari estimasi skala *residual standard error*.

Kurniawati (2011) dalam skripsi yang berjudul "Kekekaran Regresi Linier Ganda dengan Estimasi *MM* dalam Mengatasi Pencilan". Tujuan dari penelitian ini adalah menunjukkan langkah-langkah dalam menduga parameter regresi dengan estimasi *MM* dan menunjukkan penerapan estimasi *MM* dalam regresi linier berganda. Hasil penelitian ini menunjukkan bahwa metode estimasi *MM* dapat mengestimasi parameter pada data yang terdapat *outlier* tanpa menghapus *outlier* tersebut, tetapi hanya menurunkan bobot dari *outlier* tersebut. Berbeda dengan metode kuadrat terkecil, apabila data terdeteksi adanya *outlier*, untuk mendapatkan model regresi yang baik data *outlier* tersebut dihapus.

Wijaya (2009) dalam skripsi yang berjudul "Taksiran Parameter pada Model Regresi *Robust* dengan Menggunakan Fungsi *Huber*", penelitian ini menjelaskan bahwa untuk data ada *outlier* taksiran parameter yang diperoleh dengan metode regresi *robust* dengan fungsi *Huber* lebih effisien dibandingkan Metode Kuadarat Terkecil (MKT), sedangkan untuk data tanpa *outlier* taksiran parameter yang diperoleh dengan metode MKT lebih effisien dibandingkan metode regresi *robust* dengan fungsi *Huber*.

Ardiyanti (2011) dalam jurnal yang berjudul "Perbandingan Keefektifan Metode Regresi *Robust* Estimasi *M* dan Estimasi *MM* karena Pengaruh *Outlier* Dalam Analisis Regresi Linear (Contoh Kasus Data Produksi Padi Di Jawa Tengah Tahun 2007)" menjelaskan bahwa estimasi *M resistant* untuk *outlier* pada variabel dependen, akan tetapi kurang *resistant* terhadap *outlier* pada variabel independen. Estimasi *MM* mempunyai *breakdown point* sebesar 0,5 sehingga estimasi *MM resistant* terhadap *outlier* pada variabel independen maupun respon.

Berdasarkan efek nilai *breakdown point*, estimasi *MM* lebih efektif daripada estimasi *M*. Dalam menilai hasil *residual standard error* regresi *robust* dengan membandingkan hasil *residual standard error* yang diperoleh dengan MKT. Apabila *residual standard error* regresi *robust* lebih kecil daripada MKT, maka regresi *robust* tersebut sebagai alternatif tanpa membuang data *outlier*.

Maharani, Satyahadewi, dan Kusnandar (2014) dalam jurnal yang berjudul "Metode *Ordinary Least Squares* dan *Least Trimmed Sq*uares dalam Mengestimasi Parameter Regresi Ketika Terdapat *Outlier*" menjelaskan bahwa penelitian ini menggunakan 20 kondisi data yang berbeda dalam ukuran sampel dan persentase *outlier*. Tingkat efisiensi dari kedua metode dibandingkan berdasarkan nilai bias dan *Mean Square Error* (*MSE*) dari nilai estimasi yang dihasilkan. Penelitian ini menunjukkan bahwa estimasi *LTS* menghasilkan nilai bias dan *MSE* lebih kecil dibandingkan metode MKT. Sehingga estimasi *LTS* lebih efisien dalam mengestimasi parameter regresi dibandingkan metode MKT ketika terdapat outlier dalam data. Estimasi *LTS* merupakan metode estimasi parameter yang baik ketika terdapat outlier dalam data sebesar 5%, 10% dan 20%. Hal ini ditunjukkan dari nilai bias dan *MSE* yang lebih kecil dibandingkan metode MKT, sehingga model estimasi *LTS* dapat dikatakan sebagai penduga yang tak bias dan efisien ketika terdapat *outlier* dalam data.

Berdasarkan uraian diatas mengenai regresi *robust* estimasi *LTS*, estimasi *M* dan estimasi *MM* serta perbandingan regresi *robust* tersebut dengan metode lain, maka dalam tugas akhir ini peneliti tertarik untuk melakukan penelitian membahas perbandingan antara metode estimasi *LTS*, estimasi *M*, dan metode estimasi *MM* pada regresi *robust*. Perbandingan tersebut mencakup cara estimasi nilai parameter, model yang didapatkan serta keakuratan ketiga metode tersebut.

## 2.1. Landasan Teori

### 2.1.1. Metode Kuadrat Terkecil

Didalam model regresi terdapat parameter-parameter yaitu $\beta_0$, $\beta_1$, $\beta_2, \dots, \beta_k$. Parameter tersebut perlu diestimasi karena nilai belum diketahui. Metode yang sering digunakan dalam menduga parameter regresi adalah metode kuadrat terkecil. Metode ini merupakan salah satu prosedur estimasi garis regresi dimana suatu garis regresi yang dipilih dapat meminumkan jumlah kuadrat residual (Draper & Smith, 1992). Jumlah kuadrat residual ($J$) adalah sebagai berikut:

$$J = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik})^2 \qquad (2.1)$$

Untuk meminimumkan jumlah kuadrat residual pada persamaan (2.1), dapat dicari turunan dari persamaan (3.4) secara parsial terhadap $\beta_j$, dengan $j = 0,1, \dots, k$ dan disama dengan nol, sehingga diperoleh:

$$\frac{\partial J}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik}) = 0$$

$$\frac{\partial J}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik})\, x_{i1} = 0$$

$$\frac{\partial J}{\partial \beta_2} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik})\, x_{i2} = 0$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$\frac{\partial J}{\partial \beta_k} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik})\, x_{ik} = 0 \qquad (2.4)$$

Penjabaran dari persamaan (2.4) tersebut manghasilkan persamaan-persamaan sebagai berikut:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} + \cdots + \hat{\beta}_\mathrm{k} \sum_{i=1}^{n} x_{ik} = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_1 \sum_{i=1}^{n} {x_{i1}}^2 + \hat{\beta}_2 \sum_{i=1}^{n} x_{i1} x_{i2} + \cdots + \hat{\beta}_\mathrm{k} \sum_{i=1}^{n} x_{i1} x_{ik} = \sum_{i=1}^{n} x_{i1} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{i2} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} x_{i2} + \hat{\beta}_2 \sum_{i=1}^{n} {x_{i2}}^2 + \cdots + \hat{\beta}_\mathrm{k} \sum_{i=1}^{n} x_{i2} x_{ik} = \sum_{i=1}^{n} x_{i2} y_i$$

$$\vdots \qquad\qquad \vdots$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{ik} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} x_{ik} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} x_{ik} + \cdots + \hat{\beta}_\mathrm{k} \sum_{i=1}^{n} {x_{ik}}^2 = \sum_{i=1}^{n} x_{ik} y_i$$

(2.5)

Ketika disusun dalam bentuk matrik maka persamaan (2.5) akan menjadi:

$$\boldsymbol{X'X\hat{\beta}} = \boldsymbol{X'Y} \tag{2.6}$$

dengan,

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

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

$$\boldsymbol{X'Y} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_{21} & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & x_{3k} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1}y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ik}y_i \end{bmatrix}$$

Dalam menentukan nilai estimasi $\beta$ , dapat digunakan penyelesaian persamaan (3.7) dimana kedua ruas dikalikan dengan invers dari ($\boldsymbol{X'X}$). Sehingga estimasi kuadrat terkecil dari $\beta$ adalah:

$$\boldsymbol{\hat{\beta}} = (\boldsymbol{X'X})^{-\mathbf{1}}\boldsymbol{X'Y} \tag{2.7}$$

### 2.1.2. Pengeretian *Outlier* dan Identifikasi *Outlier*

*Outlier* menurut Ghozali (2005) adalah kasus atau data yang memiliki karakteristik unik yang terlihat sangat berbeda jauh dari observasi-observasi lainnya dan muncul dalam bentuk nilai ekstrim, baik untuk sebuah variabel tunggal maupun variabel kombinasi. Sedangkan *outlier* menurut Sembiring (2003) adalah pengamatan yang jauh dari pusat data yang mungkin berpengaruh besar terhadap koefesien regresi. Sehingga dapat disimpulkan *outlier* adalah suatu data pengamatan yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat data. Identifikasi outlier dapat dilakukan dengan perhitungan statistik sebagai berikut:

**a. Metode *leverage***

Metode ini mengukur pengaruh suatu observasi terhadap besarnya estimasi parameter antara lain dapat dilihat dari jarak nilai *x* terhadap pusat nilai *x* semua observasi. Menurut skripsi Wijaya (2009), nilai *leverage* untuk linier sederhana dapat ditentukan sebagai berikut:

$$leverage(h_{ii}) = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{(n-1)S_x^2} \tag{2.8}$$

dengan:

$h_{ii}$ : *Leverage* kasus ke-*i*

$n$ : Banyaknya data

$X_i$ : Nilai untuk kasus ke-*i*

$\bar{X}$ : *Mean* dari *X*

$S_x^2$ : Kuadrat *n* kasus dari simpangan $X_i$ terhadap mean

Jika suatu kasus terdiri dari beberapa variabel independen maka perhitungan nilai *leverage* dapat dilakukan dengan menggunakna matrik berikut ini:

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \tag{2.9}$$

dengan ***H*** adalah *hat* matriks. Elemen ke-*i* dari diagonal dari *hat* matriks merupakan *leverage* ($h_{ii}$) dan ***X*** merupakan matriks *X*.

Pendeteksian *outlier* didasarkan pada nilai *cutoff* dan apabila nilai $h_{ii}$ lebih dari *cutoff* dideteksi sebagai *outlier*. Menurut (Kutner, dkk, 2005) nilai *cutoff* dari *leverage* adalah *2k/n*, dimana *k* merupakan banyaknya parameter dari dalam persamaan regresi termasuk *intersept*. *Leverage* digunakan untuk mendeteksi *outlier* pada arah *x*.

**b. Metode *DFFITS (Difference in fit statndarized)***

Metode ini merupakan pengukuran *influence* global atau memberikan informasi mengenai pengaruh kasus ke-*i* terhadap keseluruhan karateristik dari persamaan regresi. Didalam metode ini menampilkan nilai perubahan dalam harga yang diprediksi bilamana kasus tertentu dikeluarkan, yang sudah distandarkan (Kurniawati, 2011). *DFFITS* merupakan metode gabungan antara metode *leverage* ($h_{ii}$) dan *externally studentized residuals* ($t_i$). Nilai *DFFITS* dapat didefinisikan sebagai berikut (Montgomery & Peck, 1982):

$$DFFITS_i = t_i\sqrt{\frac{h_{ii}}{1-h_{ii}}} \tag{2.10}$$

dengan:

$$t_i = e_i\sqrt{\frac{n-k-1}{JKG(1-h_{ii})-e_i^2}} \tag{2.11}$$

$e_i$ adalah residual ke-*i* dan JKG merupakan jumlah kuadrat residual. Data dikatakan *outlier* ketika nilai |*DFFITS*| > $2\sqrt{k/n}$ dengan *k* banyaknya parameter dalam model dan *n* banyaknya pengamatan.

**c. Metode *Cook's Distance***

*Influence* global dapat diukur dengan *Cook's distance*. *Cook's distance* merupakan ukuran pengaruh observasi ke-*i* terhadap semua *estimator* parameter regresi (Wijaya, 2009). Secara matematis ukuran *Cook's Distance* dapat didefinisikan sebagai berikut (Montgomery & Peck, 1982):

$$\begin{aligned} D_i &= \frac{(\hat{\beta}_{(i)} - \hat{\beta})' X' X (\hat{\beta}_{(i)} - \hat{\beta})}{kMSE} \\ &= \frac{(y_i - \hat{y}_i)^2}{kMSE}\left[\frac{h_{ii}}{(1-h_{ii})^2}\right] \end{aligned} \tag{2.12}$$

dengan:

$D_i$ : Ukuran *Cook's Distance*

$\hat{\beta}$ : Vektor estimasi atau *estimator* koefisien regresi

$\hat{\beta}_{(i)}$ : Vektor estimasi koefisien regresi tanpa observasi ke-*i*

$k$ : Banyaknya parameter

*MSE* : *Mean Square Error* ($\sum e_i{}^2 /n$)

Berdasarkan persamaan (3.24), besarnya nilai *Cook's Distance* bergantung pada nilai residual dan nilai *leverage* ($h_{ii}$). Nilai $D_i$ besar atau $D_i > F_{(0.5,k,n-k)}$ mengartikan bahwa data atau observasi ke-*i* sudah dapat dikatakan sebagai *outlier* (Montgomery & Peck, 1982).

**d. Metode *DFBETAS*$_{ij}$ (*Diference in Beta*)**

*DFBETAS*$_{ij}$ merupakan ukuran pengaruh dengan melihat selisih nilai taksiran koefisien regresi ke-*j*, $\hat{\beta}_j$ dengan nilai taksiran koefisien regresi ke-*j* saat observasi ke-*i* dikeluarkan ($\hat{\beta}_{j(i)}$) . *DFBETAS*$_{ij}$ memperlihatkan bahwa observasi ke-*i* cukup mempengaruhi parameter regresi (Wijaya, 2009). Secara matematis, *DFBETAS*$_{ij}$ dapat didefinisikan seperti persamaan berikut ini (Montgomery & Peck, 1982):

$$DFBETAS_{ij} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{S_{(i)}^2 C_{jj}}} \qquad (2.13)$$

dengan:

$\hat{\beta}_j$ : Estimasi koefisien regresi ke-*j*

$\hat{\beta}_{j(i)}$ : Estimasi koefisien regresi ke-*j* saat data ke-*i* dikeluarkan

$S_{(i)}^2$ : Variansi sampel saat observasi ke-*i* dikeluarkan

$C_{jj}$ : Elemen diagonal ke-*j* dari $((X^T X)^{-1}$

Berdasarkan persamaan (2.13), pendeteksian *outlier* menggunakan *DFBETAS* adalah ketika nilai *DFBETAS* suatu observasi > 1 untuk ukuran sampel yang kecil dan nilai *DFBETAS* > $\frac{2}{\sqrt{n}}$ untuk ukuran sampel yang besar.

**e. *R-student***

Menurut Wijaya (2009), metode ini memiliki perhitungan yang hampir sama dengan *stdentized residuals*, tetapi variansi yang digunakan untuk perhitungan *R-student* memperhitungkan saat observasi ke-*i* dikeluarkan dari pengamatan. Variansi residual dapat diestimasi dengan persamaan berikut ini:

$$S_{(i)}^2 = \frac{(n-k)MSE - \frac{e_i^2}{(1-h_{ii})}}{n-k-1} \tag{2.14}$$

Berdasarkan persamaan (2.14), *R-student* didefinisikan seperti berikut:

$$t_i = \frac{e_i}{\sqrt{S_{(i)}^2(1-h_{ii})}} \tag{2.15}$$

dengan $S_{(i)}^2$ merupakan variansi residual dan $t_i$ adalah nilai *R-student*. Suatu observasi dikatakan *outlier* ketika $|t_i| > t_{\frac{\alpha}{2},n-k-1}$ (Montgomery & Peck, 1982).

### 2.1.3. Regresi *Robust*

Regresi *robust* merupakan alat penting untuk menganalisa data yang dipengaruhi oleh *outlier* sehingga dihasilkan model yang *robust* atau *resistance* terhadap *outlier*. Suatu estimasi yang *resistance* adalah relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data.

Regresi *robust* bertujuan untuk mengatasi penyimpangan-penyimpangan sebagai pengganti metode kuadrat terkecil. Kelebihan metode tersebut adalah kurang peka terhadap penyimpangan-penyimpangan yang sering terjadi dari asumsi klasik. Prosedur statsistik yang bersifat kekar ditujukan untuk mengakomodasi adanya keanehan data dan sekaligus meniadakan pengaruhnya terhadap analisis tanpa terlebih dahulu mengadakan identifikasi (Safitri, 2015).

Dua hal yang diperlukan dalam estimasi *robust* adalah *resistance* dan efisiensi. Suatu estimasi dikatakan *resistance* terhadap *outlier* jika sebagaian kecil dari data tidak memberikan efek yang terlalu besar terhadap *estimator*. Estimasi memiliki efisiensi yang cukup baik pada berbagai sebaran jika raagamnya mendekati ragam minimum untuk setiap sebaran (Montgomery & Peck, 1982).

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

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

### 2.1.4. *Breakdown Point*

Salah satu ukuran *robust* yang sering digunakan adalah *breakdown point*. *Breakdown point* merupakan proporsi minimal dari banyaknya *outlier* dibandingkan seluruh data pengamatan (Kurniawati, 2011). Diasumsikan bahwa sebuah pada sebuah sampel *Z* (berdstribusi normal dengan ukuran sampel *n*), *T* merupakan estimasi regresi, nilai *breakdown point* dari sebuah *estimator T=T(Z)* dapat didefinisikan seperti berikut:

$$\varepsilon^*(T,Z) = min\left\{\frac{m}{n}; bias\ (m;T,Z) is\ infinite\right\} \tag{2.16}$$

dengan:

$$bias(m;T,Z) = \ {}^{sup}_{Z\prime}||T(Z') - T(Z)$$

dimana $\varepsilon^*(T,Z)$ merupakan nilai *breakdown point* dari *estimator T*, *supremum* (*sup*) diambil sari semua kemungkinan pada sampel $Z'$ yang diperoleh dengan mengganti observasi *m* dari *Z* dengan nilai sembarang dan $||T||$ adalah normal (Rousseeuw, 1987).

Contoh penerapan *breakdown point* jika *T* adalah fungsi median $T_{med}$ didefinisikan pada $P_T = P$ dengan $\Theta = R$ dan $D(\theta_1, \theta_2) = |\theta_1 - \theta_2|$, kemudian diberikan nilai *fsbp*($T_{med}$, *x*, *D*) = $\left[\frac{n+1}{2}\right]/n$. Sebuah distribusi *breakdown point* memerlukan sebuah matrik *d* pada *P* dengan $\ {}^{sup}_{P,Q\in P} d(P,Q) = 1$. Maka nilai *breakdown point T* pada distribusi P ∈ $P_T$ adalah:

$$\varepsilon^*(T,P,d,D) = inf\left\{\varepsilon > 0; {}^{sup}_{d(P,Q)<\varepsilon} D(T(P), QT)) = \infty\right\}$$

dimana $D(T(P), QT)) = \infty$ jika $Q \notin P_T$.

### 2.1.5 Fungsi Objektif, Fungsi *Influence*, dan Fungsi Pembobot

Fungsi objektif digunakan untuk representasi pembobot dari residual atau $\rho(u)$. Fungsi *influence* ($\psi(u)$) digunakan untuk mengukur pengaruh dari sebuah data terhadap estimasi parameter. Fungsi *influence* secara matematis didefinisikan seperti berikut:

$$\psi(u) = \frac{\partial\rho(u)}{\partial u} \tag{2.17}$$

dengan $\rho(u)$ adalah representasi pembobot dari residual (fungsi objektif).

Berdasarkan persamaan (3.28), misal $\rho(u) = \frac{u^2}{2}$, maka bdiperoleh nilai $\psi(u) = u$. Hal ini mengartikan bahwa pengaruh estimasi suatu data terhadap estimasi parameter secara linier sejalan dengan naiknya *u* (Wijaya, 2009). Fungsi pembobot dicari dengan menggunakan fungsi objektif.

### 2.1.6 Sifat *Equavariant*

Menurut Nurcahyadi (2010), kata "*equivariant*" dalam statistik menunjuk pada transformasi sebagaimana mestinya, kata lawannya yaitu *invariant* menunjuk pada kuantitas yang tetap tidak berubah. Sifat-sifat *equivariant* yang harus dimiliki oleh suatu *estimator* ada tiga, yaitu regresi *equivariant*, skala *equivariant*, dan *affine equivariant.*

Suatu *estimator T* disebut sebagai regresi *equivariant* jika memenuhi:

$$T(\{(x_i, y_i + x_i v); i = 1, \dots, n\}) = T(\{(x_i, y_i); i = 1, \dots, n\}) + v \quad (2.18)$$

dengan *v* merupakan sebarang vektor kolom. suatu *estimator T* disebut sebagai skala *equivariant* jika memenuhi:

$$T(\{(x_i, cy_i); i = 1, \dots, n\}) = cT(\{(x_i, y_i); i = 1, \dots, n\}) \quad (2.19)$$

Untuk sembarang konstanta *c*. Skala *equivariant* menyebabkan bahwa kecocokan secara esensial independen dari pemilihan satuan pengukuran pada variabel dependen *y*. Sedangkan, suatu *estimator T* adalah *affine equivariant* jika memenuhi:

$$T(\{(x_i, A, y_i); i = 1, \dots, n\}) = A^{-1}T(\{(x_i, y_i); i = 1, \dots, n\}) \quad (2.20)$$

Untuk sembarang matriks persegi *A* yang *nonsingular*. *Affine equivariant* berarti bahwa suatu transformasi linear dengan $x_i$ yang harus mentransformasikan *estimator T*, karena $\hat{y}_i = x_i T = (x_i A)(A^{-1}T)$ . Hal ini memperbolehkan penggunaan sistem koordinat lain dari variabel independen, dengan tanpa mempengaruhi pengestimasian $\hat{y}_i$ . Terdapat suatu teorema yang menyatakan bahwa sebarang regresi *equivariant* dari *estimator T* memenuhi:

$$\varepsilon_n^*(T, Z) \leq \left(\left[\frac{(n - \mathrm{k})}{2}\right] + 1\right)/n \quad (2.21)$$

dengan $\varepsilon_n^*(T, Z)$ merupakan nilai *breakdown point* dari *estimator T*, *n* merupakan banyaknya sampel pada ada seluruh sampel *Z* (berdistribusi normal).

### 2.1.5 *Residual Standard Error*

Dalam keaukaratan sutau model dapat diukur dengan beberapa macam metode. Salah satu ukuran tersebut adalah *residual standard error*. Menurut Buechler (2007), secara matematis *residual standard erro*r dapat didefinisikan seperti berikut ini:

$$Residual\ Standard\ Error = \sqrt{\frac{SSE}{df}} \tag{2.22}$$

dengan *SSE* merupakan jumlah kuadrat residual dan *df* merupakan derajat bebas residual $(n-2)$. Suatu metode dikatakan baik ketika nilai *residual standard error* kecil atau mendekati nilai 0.

# BAB III
# METODOLOGI PENELITIAN

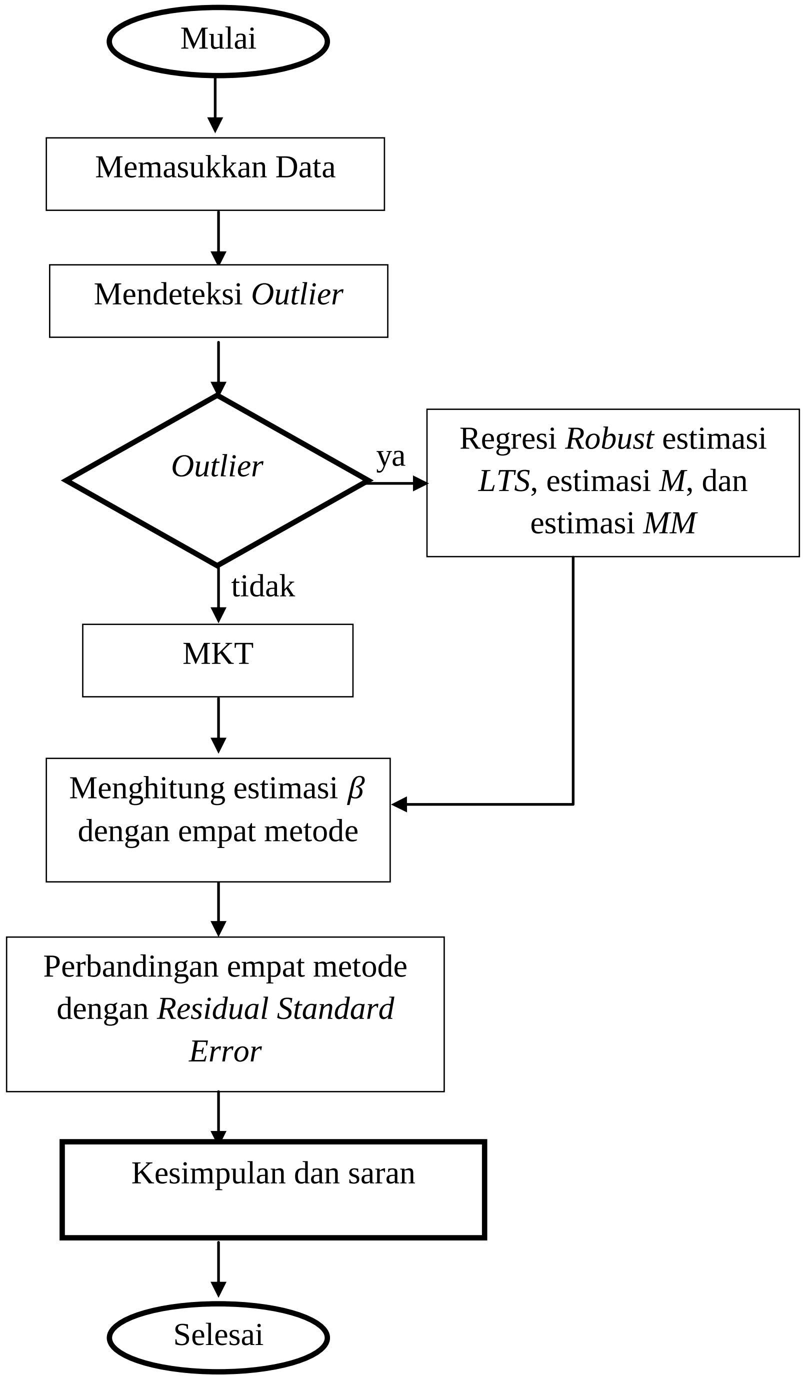
## 3.1. Data

Data yang digunakan pada penelitian ini adalah data nilai tukar petani pada tanaman kedelai tahun 2015. Data tersebut mengandung data *outlier* dan dibuktikan dengan uji *kolmogorov-smirnov*. Kemudian untuk data tanpa outlier dengan menghilangkan data yang mengandung *outlier*. Pengerjaan dilakukan dengan menggunakan *software* R.2.14.2.

## 3.2. Tahapan Analisis Data

Pada penelitian ini dilakukan dengan tahapan estimasi data mengandung *outlier*

a. Memasukkan data nilai tukar petani tanaman kedelai tahun 2015.
b. Melakukan estimasi regresi dengan Metode Kuadrat Terkecil (MKT).
c. Mendeteksi *outlier* dengan menggunakan R-*student*, *leverage*, $DFFITS_i$, $DFBETAS_{ji}$, dan *Cook's distance*.
d. Melakukan estimasi regresi dengan metode estimasi *LTS* pada regresi *robust*.
e. Melakukan estimasi regresi dengan metode estimasi *M* pada regresi *robust*.
f. Melakukan estimasi regresi dengan metode estimasi *MM* pada regresi *robust.*
g. Membandingkan hasil estimasi yaitu nilai $\beta$, *standard error,* dan *residuals standard error* pada keempat metode.
h. Memilih metode estimasi terbaik melalui nilai *residuals standard error* pada keempat metode.




Gambar 3.1 Alur Tahapan Analisis Data

# BAB IV
# HASIL DAN PEMBAHASAN

Metode kuadrat terkecil merupakan suatu cara untuk mengestimasi parameter pada model regresi dengan meminimumkan jumlah kuadrat residual. Pada estimasi ini memiliki sifat *BLUE*, jika asumsi klasik harus terpenuhi. Tidak jarang ditemukan dalam berbagai kasus bahwa terdapat penyimpangan asumsi normalitas. Sehingga, metode ini kurang tepat digunakan ketika residual tidak normal dan teridentifikasi adanya *outlier*. Metode lain yang dapat digunakan ketika terdapat *outlier* di dalam data adalah regresi *robust*. Regresi *robust* merupakan metode yang digunakan untuk mengatasi *outlier* tanpa menghapus data *outlier* tersebut. Pada regresi *robust*, metode yang sering digunakan dalam mengestimasi parameter adalah estimasi *LTS* (*Least Trimmed Square*), estimasi *M* (*Maximum Likelihood Type*), dan estimasi *MM* (*Method of Moment*).

## 4.1. Estimasi *LTS* (*Least Trimmed Square*)

Salah satu metode penduga parameter model regresi terhadap data yang mengandung outlier adalah estimasi *LTS*. *LTS* adalah metode nilai *high breakdown point* yang diperkenalkan oleh Rousseeuw pada tahun 1984. *LTS* merupakan suatu metode pendugaan parameter regresi *robust* untuk meminimumkan jumlah kuadrat $h$ residual (fungsi objektif). Rumus pada estimasi *LTS* sebagai berikut (Chen, 2002):

$$\tilde{\beta}_{LTS} = arg\ min\ Q_{LTS}$$
$$Q_{LTS} = \sum_{i=1}^{h} e_i^2 \tag{4.1}$$

dengan $h = [n/2] + [(k+2)/2]$, $e_i = \left(\hat{y}_i - X_i\hat{\beta}_0\right)$

Keterangan:

$e_i^2$ : Kuadrat residual yang diurutkan dari terkecil ke terbesar. $e_1^2 < e_2^2 < \cdots < e_i^2$.

*n* :Banyaknya pengamatan.

*k* : Parameter

Jumlah *h* menunjukkan sejumlah *subset* data dengan kuadrat fungsi objektif terkecil. Nilai *h* pada persamaan (5.1) akan membangun *breakdown point* yang besar sebanding dengan 50%. *LTS* mempunyai nilai kekonvergenan $n^{-1/2}$. Algoritma *LTS* menurut Rousseeauw dan Van Driessen (1984) adalah gabungan *FAST-LTS* dan *C-steps*. Dapat dikatakan bahwa prosedur *LTS* terdiri dari dua skala *estimator* yaitu $S_{LTS}$ dan $W_{scale}$. Pada $W_{scale}$, *estimator* berdasarkan pada estimasi $S_{LTS}$ disebut juga proses *Final Weighted Scale Estimator* (*FWLS*). Secara sistematis fungsi pembobotnya jika nilai *r* = 3 sebagai berikut:

$$w_i = \begin{cases} 0 \quad , & \frac{|e_i|}{S_{LTS}} > 3 \\ 1 \quad , & lainnya \end{cases} \tag{4.2}$$

dengan:

$$S_{LTS} = d_{h,n}\sqrt{\frac{1}{h}\sum_{i=1}^{h} e_{(i)}^2}$$

$$d_{h,n} = \frac{1}{\sqrt{1-\frac{2n}{hC_{h,n}}\phi(\frac{1}{C_{h,n}})}}$$

$$C_{h,n} = \frac{1}{\Phi^{-1}\left((h+n)/2n\right)}$$

Keterangan:

*n* : Banyaknya pengamatan

*Φ* : Fungsi komulatif normal standar

$\phi$ : Fungsi *density* normal standar

Prosedur estimasi dengan menggunakan estimasi *LTS* atau algoritma *FAST-LTS, C-steps*, dan *FWLS* diuraikan sebagai berikut:

a. Dihitung estimasi dari $\beta_0$, dinotasikan $\hat{\beta}_0$ menggunakan MKT.

b. Ditentukan $n$ residual $e_i^2 = \left(\hat{y}_i - X_i\hat{\beta}_0\right)^2$ yang bersesuaian dengan ($\hat{\beta}_0$), kemudian menghitung jumlah $h_0 = (n + k + 2)/2$ pengamatan dengan nilai $e_i^2$ terkecil.
c. Dihitung $\sum_{i=1}^{h_0} e_i^2$.
d. Dilakukan estimasi parameter $\hat{\beta}_{new}$ dari $\hat{\beta}_0$ pengamatan.
e. Ditentukan $n$ kuadrat residual $e_i^2 = \left(\hat{y}_i - X_i\hat{\beta}_{new}\right)^2$ yang bersesuaian dengan ($\hat{\beta}_{new}$) kemudian menghitung sejumlah $\hat{\beta}_{new}$ pengamatan dengan nilai $e_i^2$ terkecil.
f. Dihitung $\sum_{i=1}^{h_{new}} e_i^2$.
g. Dilakukan *C-steps* yaitu tahap d sampai f untuk mendapatkan fungsi objektif yang kecil dan konvergen.

## 4.2. Estimasi *M* (*Maximum Likelihood Type*)

Metode ini pertama kali diperkenalkan oleh Huber pada tahun 1964. Menurut Montgomery dan Peck (1982), metode ini mengasumsikan bahwa sebagian besar yang terdeteksi *outlier* berada pada variabel independen. Estimasi *M* meminimumkan fungsi objektif ($\rho$) dari fungsi residual ($e_i$). Fungsi tersebut dapat dilihat pada persamaan (4.3) dan berikut ini persamaan fungsi tersebut:

$$\min_{\hat{\beta}} \sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = \min_{\hat{\beta}} \sum_{i=1}^{n} \rho\left(\frac{y_i - x_i'\hat{\beta}}{s}\right) \tag{4.3}$$

dengan $s = \frac{median|e_i - median(e_i)|}{0,6745}$ merupakan skala dari suatu estimasi *robust* dan $\rho\left(\frac{e_i}{s}\right)$ merupakan fungsi yang memberikan kontribusi pada masing-masing residual pada fungsi objektif.

Berdasarkan persamaan (4.3), dapat diperoleh estimasi parameter dari persamaan regresi. Meminimumkan fungsi objektif ($\rho$) dengan mencari turunan parsial pertama dari $\rho$ terhadap $\beta_j$ dengan $j=0,1,\ldots,k$, kemudian disama dengan 0. Turunan pertama dari $\rho$ disebut dengan $\psi$ ($\psi = \rho'$). Perhitungan tersebut dapat dilihat pada persamaan (4.4).

$$\frac{\partial \sum_{i=1}^{n} \rho\left(\frac{\varepsilon_i}{s}\right)}{\partial \beta} = \frac{\sum_{i=1}^{n} \frac{\rho\left(y_i - x_i' \beta\right)}{s}}{\partial \beta} = 0 \tag{4.4}$$

Sehingga diperoleh hasil sebagai berikut:

$$\sum_{i=1}^{n} x_{ij} \psi \left(\frac{y_i - x_i' \hat{\beta}}{s}\right) = 0, j = 0,1, \dots, k \tag{4.5}$$

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

Estimasi parameter dengan metode *M* disebut juga dengan *Iteratively Reweighted Least* Squares *(IRLS)*. Penyelesaian metode *IRLS* pada persamaan (5.4) menghasilkan persamaan (4.5) berikut ini:

$$\sum_{i=1}^{n} x_{ij} \psi \left(\frac{y_i - x_i' \hat{\beta}}{s}\right) = \sum_{i=1}^{n} x_{ij} \frac{\psi\left(\frac{y_i - x_i' \hat{\beta}}{s}\right)}{\left(\frac{y_i - x_i' \hat{\beta}}{s}\right)} \left(\frac{y_i - x_i' \hat{\beta}}{s}\right) = 0, j = 0,1, \dots, k \tag{4.6}$$

atau

$$\sum_{i=1}^{n} x_{ij} w_{i0} \left(\frac{y_i - x_i' \hat{\beta}}{s}\right) = 0\,, j = 0,1, \dots, k \tag{4.7}$$

dengan

$$w_{i0} = \begin{cases} \dfrac{\psi\left(\frac{y_i - x_i' \hat{\beta}}{s}\right)}{\left(\frac{y_i - x_i' \hat{\beta}}{s}\right)} & jika\ y_i \neq x_i' \hat{\beta} \\ 1 & jika\ y_i = x_i' \hat{\beta} \end{cases} \tag{4.8}$$

Jika persamaan tersebut dinotasikan dengan matrik, maka akan menjadi:

$$\boldsymbol{X' W_0 X \hat{\beta}} = \boldsymbol{X' W_0 y} \tag{4.9}$$

dengan $\boldsymbol{W_0}$ merupakan diagonal matriks "*weight*" yang berukuran *n x n*. Regresi terboboti tersebut dapat digunakan sebagai alat untuk mendapatkan estimasi *M*. Sehingga estimasi parameter menjadi:

$$\boldsymbol{\hat{\beta}} = (\boldsymbol{X' W_0 X})^{-1} \boldsymbol{X' W_0 y} \tag{4.10}$$

Prosedur estimasi dengan menggunakan estimasi *M* atau *IRLS* diuraikan sebagai berikut:

a. Dihitung estimasi dari $\beta$, dinotasikan $\hat{\beta}$ menggunakan MKT, sehingga didapatkan $\hat{y}_{i0}$ dan $\varepsilon_{i0} = y_i - \hat{y}_{i0}$, *(i = 1, 2, ... n)* yang diperlakukan sebagai nilai awal ($y_i$ adalah hasil observasi).

b. Dari nilai-nilai residual ini dihitung *s* dan pembobot awal $w_{i0} = \frac{\psi(\varepsilon_{i0}^*)}{(\varepsilon_{i0}^*)}$. Nilai $\psi(\varepsilon_{i0}{}^*)$ dihitung sesuai fungsi *Huber*, dan $\varepsilon_{i0}^* = \frac{\varepsilon_{i0}}{s}$.

c. Disusun matrik pembobot berupa matrik diagonal dengan elemen $w_{10}$, $w_{20}, \ldots, w_{n0}$.

d. Dihitung estimasi koefisien regresi,

$$\hat{\beta}_{Robust\ ke\ 1} = (\boldsymbol{X'W_0X})^{-1}\boldsymbol{X'W_0y}$$

e. Dengan menggunakan nilai $\hat{\beta}_{Robust\ ke\ 1}$ dihitung pula $\sum_{i=1}^{n}|y_i - \hat{y}_{i1}|$ atau $\sum_{i=1}^{n}|\varepsilon_{i1}|$.

f. Selanjutnya langkah b sampai dengan e diulang sampai didapatkan $\sum_{i=1}^{n}|\varepsilon_{im}|$ konvergen. Dengan kata lain, $\sum_{i=1}^{n}|\varepsilon_{im}|$ cukup kecil untuk *j=0,1,...,k.*

Jika diambil nilai terstandarisasi dari *e*, maka berdasarkan simulasi yang dilakukan oleh Huber, dipilih nilai *k=1.345*, sehingga diperoleh persamaan $w_i$ sebagai berikut:

$$w_i = \begin{cases} 1, & jika\ |e_i| \leq 1.345 \\ \frac{1.345}{|e_i|}, & jika\ |e_i| > 1.345 \end{cases} \tag{4.11}$$

dengan $w_i$ merupakan elemen diagonal ke-*i* dari matriks bobot $\boldsymbol{W}$. Dengan metode *IRLS* ini dapat digunakan untuk mengestimasi nilai parameter $\hat{\beta}$.

## 4.3. Estimasi *MM* (*Method of Moment*)

Menurut Chen (2002), estimasi *MM* merupakan merupakan kombinasi antara h*igh breakdown point* dan estimasi *M* dikenalkan oleh Yohai pada tahun 1987. Estimasi ini mempunyai efisiensi yang lebih tinggi dibanding estimasi *S.*

Pada umumnya digunakan fungsi *Tukey Bisquare β* baik pada estimasi *S* maupun estimasi *M*. Persamaan dari estimasi *MM* adalah sebagai berikut:

$$\text{argmin}_{\beta} \sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) \qquad (4.12)$$

Langkah pertama dalam estimasi ini adalah mencari nilai estimasi *S*. Estimasi *S* merupakan estimasi yang dapat digunakan untuk membedakan *good leverage point* dan *bad leverage point*. Estimasi ini memiliki nilai *breakdown point* yang sangat tinggi yaitu lebih dari 50%. Bentuk persamaan estimasi *S* dapat dilihat seperti berikut ini:

$$\tilde{\beta}_S = \underset{\beta}{\text{argmin}}\, s(e_1, e_2, \dots, e_n) \qquad (4.13)$$

dengan *s* adalah estimasi skala *robust* dalam persamaan (4.3) yang memenuhi $\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = K$ . $K$ merupakan konstan yang didefinisikan sebagai $K = E[\Phi(\rho)]$. $\Phi$ adalah distribusi normal standar. Nilai *breakdown* dari estimasi *S* dapat ditulis $\frac{K}{max\,\rho(e)} = 0{,}5$.

Setelah diketahui nilai estimasi *S*, langkah selanjutnya adalah menetapkan parameter-parameter regresi menggunakan estimasi *M*. Residual awal yang digunakan dalam estimasi *MM* adalah residual yang diperoleh dari hasil perhitungan estimasi *S*. Estimasi parameter yang digunakan dalam estimasi *MM* adalah *IRLS* seperti pada sub bab 5.6. Penyelesaian estimasi parameter dengan estimasi *M* dapat dilihat pada sub bab tersebut. Berikut ini merupakan langkah-langkah estimasi parameter pada model linier berganda dengan regresi *robust* estimasi *MM*:

a. Menghitung nilai estimasi awal koefisien $\hat{\beta}_j$ dan residual $e_i$ dari regresi *robust* dengan *high breakdown point* (estimasi *S*) dengan pembobotan *tukey bisquare*.
b. Residual $e_i$ pada langkah pertama dilakukan untuk menghitung skala estimasi *S*, dan dihitung pula pembobot awal $w_i$.
c. Residual $e_i$ dengan skala estimasi *S* pada langkah kedua digunakan dalam iterasi awal sebagai estimasi *WLS* (*Weighted Least Square*) untuk menghitung koefisien regresi.

d. Menghitung bobot baru $w_i$ dengan skala estimasi dari iterasi awal *WLS*. Perhitungan estimasi koefisien regresi menggunakan metode ini menggunakan persamaan $\hat{\boldsymbol{\beta}}^{(\boldsymbol{m}+\mathbf{1})} = (\boldsymbol{X}'\boldsymbol{W}^{\boldsymbol{m}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{W}^{\boldsymbol{m}}\boldsymbol{y}$.

e. Mengulang langkah b sampai d sampai mendapatkan $\sum_{i=1}^{n}\left|e_i^{(m)}\right|$ konvergen (selisih $\hat{\beta}_j^{(m+1)}$ dan $\hat{\beta}_j^{(m)}$ mendekati 0, dengan *m* banyaknya iterasi).

Jika diambil nilai terstandarisasi dari *e*, maka berdasarkan simulasi yang dilakukan oleh Tukey, dipilih nilai *k=4.685*, sehingga diperoleh persamaan $w_i$ sebagai berikut:

$$w_i = \begin{cases} \left[1 - \left(\frac{e_i}{4.685}\right)^2\right]^2, jika\ |e_i| \leq 4.685 \\ 0, \qquad jika\ |e_i| > 4.685 \end{cases} \tag{4.14}$$

dengan $w_i$ merupakan elemen diagonal ke-*i* dari matriks bobot $\boldsymbol{W}$. Dengan metode *IRLS* ini dapat digunakan untuk mengestimasi nilai parameter $\hat{\beta}$.

## 4.4. Pengujian Signifikansi Parameter

Pengujian signifikansi parameter dalam model regresi bertujuan untuk mengetahui ada atau tidak hubungan yang nyata antara variabel independen dan variabel dependen. Menurut Montgomery (1982), terdapat dua tahap pengujian yaitu uji *overall* (serentak) dan uji parsial (individu).

### a. Uji *Overall* (serentak)

Uji *overall* merupakan pengujian yang dilakukan secara bersama semua parameter dalam model regresi. Hipotesis dalam pengujian ini adalah sebagai berikut:

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

$H_1$ : paling tidak ada salah satu $\beta_j \neq 0$ untuk $j=0,1,\ldots,k$

Statistik uji yang digunakan adalah uji *F* yang dapat dilihat pada persamaan berikut ini:

$$F_{hitung} = \frac{MSR_{weight}}{MSE_{weight}}$$

$$F_{hitung} = \frac{\left[\sum_{i=1}^{n} w_i(\hat{y}_i - \bar{y})^2\right]/k}{\left[\sum_{i=1}^{n} w_i(y_i - \hat{y}_i)^2\right]/(n-k-1)} \quad (4.15)$$

dengan *MSR* adalah *mean square regression* dan *MSE* adalah *mean square* residual.

Pengambilan keputusan dengan menggunakan uji statistik tersebut adalah apabila $F_{hitung} > F_{\alpha(k,\, n-k-1)}$ dengan $k$ adalah parameter maka $H_0$ ditolak pada tingkat signifikansi $\alpha$, yang artinya paling sedikit ada satu $\beta_j$ yang tidak sama dengan nol. Selain menggunakan $F_{hitung}$, pengambilan keputusan juga dapat menggunakan *P-value* dimana $H_0$ ditolak jika *P-value* $< \alpha$.

**b. Uji Parsial**

Uji parsial merupakan pengujian secara sendiri (individu) parameter dalam model regresi yang bertujuan untuk mengetahui adanya pengaruh antara variabel independen ke-$j$ dimana $j=1,2,...,k$ dengan variabel dependen. Hipotesis yang digunakan dalam pengujian ini adalah sebagai berikut:

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$ untuk $j=0,1,...,k$

Statistik uji yang digunakan adalah uji $t$ yang dapat dilihat pada persamaan berikut ini:

$$t_{hitung} = \frac{\hat{\beta}_{j(weight)}}{S(\hat{\beta}_{j(weight)})} \quad (4.16)$$

dengan $S(\hat{\beta}_{j(weight)})$ merupakan diagonal matriks kovarians.

Pengambilan keputusan pada uji parsial yaitu apabila $|t_{hitung}| > t_{(1-(\alpha/2),\, n-k-1)}$ dengan $k$ adalah parameter maka $H_0$ ditolak pada tingkat signifikansi $\alpha$, artinya ada pengaruh $x_i$ terhadap model. Selain dengan $t_{hitung}$, pengambilan keputusan juga dapat melalui P-*value*, dimana $H_0$ ditolak jika *P-value* $< \alpha$.

## 4.5. Studi Kasus

Berdasarkan data nilai tukar petani tanaman kedelai di Indonesia tahun 2015 sebanyak 34 data, data terdapat pada lampiran 1, teridentifikasi adanya *outlier*. Dimana pada perhitungan, residual tidak berdistribusi normal atau nilai $D_{perhitungan} > D_{tabel}$ ditunjukkan dengan nilai 0.3814 > 0.209. Maka tolak $H_0$ sehingga residual tidak berdistribusi normal. $X_1$ adalah produksi tanaman kedelai, $X_2$ adalah luas panen tanaman kedelai, dan $Y$ adalah nilai tukar petani tanaman kedelai. Selanjutnya akan dilihat model yang terbentuk dari beberapa metode serta keakuratan masing-masing metode tersebut. Ada dua analisis yang dilakukan pada data produksi kedelai yaitu analisis data yang mengandung *outlier* dan analisis data tanpa *outlier* (menghilangkan data *outlier*).

### a. Analisis data mengandung *outlier*

Sebelum mencari nilai estimasi dari parameter $\beta_0$, $\beta_1$ dan $\beta_2$, akan diperiksa terlebih dahulu keberadaan *outlier* dalam data. Pendeteksian *outlier* dalam data dilihat berdasarkan jenis *outlier*, yaitu *outlier* pada arah *x*, *outlier* pada arah *y*, dan *outlier* pada arah *x* dan *y (influence)*. Metode yang digunakan untuk mendeteksi *outlier* adalah $DFBETAS_{ji}$ dan $DFFITS_i$ untuk melihat *influence, R-student* untuk melihat *outlier* pada arah *y*, dan *leverage* untuk melihat *outlier* pada arah *x*. Data diolah dengan menggunakan *software R 2.14.2*. Dpengelolahan data nilai tukar petani tanaman kedelai di Indonesia dapat dilihat pada lampiran 2.

Pada kasus ini, nilai *cutoff* untuk masing-masing metode ditentukan berdasarkan jumlah sampel (*n*) yaitu 34 dan banyaknya parameter (*k*) adalah 3. Oleh sebab itu, observasi dikatakan *outlier* jika:

- $|R - student| > 2$
- *Leverage* $(h_{ii}) > \frac{2k}{n} = \frac{2.3}{34} = 0.17647$
- $|DFFITS_i| > 2\sqrt{\frac{k}{n}} = 2\sqrt{\frac{3}{34}} = 0.10189$
- $|DFBETAS_{ji}| > \frac{2}{\sqrt{n}} = \frac{2}{\sqrt{34}} = 0.51449$

Berdasarkan kriteria tersebut, dapat diketahui observasi-observasi yang merupakan *outlier*. Hasil perhitungan nilai-nilai tersebut dapat dilihat pada lampiran 5 mengeni deteksi *outlier*. Pada lampiran 5 dapat dilihat bahwa terdapat data observasi yang merupakan data *outlier* pada arah *y* dan pada kedua arah sekaligus. Terdapat data observasi yang *outlier* pada arah *x* yaitu data observasi ke-11 dan ke-15. Dimana nilai kedua observasi tersebut sebagai berikut:

1. Pada observasi ke-11 yaitu $|Leverage\ (hii)\ _{11}| = 0.5448 > 0.17647.$
2. Pada observasi ke-15 yaitu $|Leverage\ (hii)\ _{15}| = 0.7274 > 0.17647.$

Data yang *outlier* pada arah *y* adalah data observasi ke-24. Hal ini dikarenakan pada pendeteksian *outlier* menggunakan *R-student* diperoleh hasil bahwa nilai $|R - student| = 25.6301$ lebih dari 2. Data yang *outlier* pada kedua arah sekaligus yaitu pada arah *x* dan pada arah *y* dilihat berdasarkan nilai *influence*. Terdapat beberapa data yang merupakan *outlier* pada dua arah sekaligus yaitu data observasi ke-11 dan 24. Berikut ini merupakan hasil nilai *DFFITS* dan *DFBETAS* pada dua observasi tersebut:

1. Pada observasi ke-11
   - $|DFFITS_{11}| \quad = 1.187700 > 0.10189.$
   - $\left|DFBETAS_{2,11}\right| = 1.187736 > 0.51449.$
2. Pada observasi ke-24
   - $|DFFITS_{24}| \quad = 6.6422 > 0.10189.$
   - $|DFBETAS_{1}| \quad = 5.6082 > 0.51449.$
   - $\left|DFBETAS_{2,24}\right| = 4.5149 > 0.51449.$

Pada penjabaran nilai tersebut, terlihat bahwa berdasarkan nilai *DFFITS* dan *DFBETAS* dapat dikatakan data observasi ke-11 dan 24 merupakan data *outlier* pada dua arah sekaligus. Selain itu, terdapat beberapa data yang memiliki nilai *DFBETAS*$_{ji}$ lebih dari 0.514490. Namun, hal ini bukan merupakan indikasi kuat bahwa data-data tersebut merupakan *outlier* pada dua arah sekaligus. Hal ini dikarenakan *DFBETAS*$_{ji}$ melihat secara khusus nilai *influence* setiap parameter tanpa mengetahui nilai *influence* secara global pada semua parameter. Nilai *influence* global dapat mengetahui bahwa data tersebut merupakan data *outlier* dilihat secara keseluruhan pada semua parameter. Oleh sebab itu, meskipun data tersebut memiliki nilai *DFBETAS*$_{ji}$ sesuai dengan kriteria tetapi perlu dilihat pula nilai *DFFITS* memenuhi kriteria sebagai *outlier*. Pada kasus ini, *outlier* yang digunakan adalah *outlier* pada arah *y*. Sehingga, data observasi yang dikatakan merupakan *outlier* berdasarkan kriteria tersebut adalah data observasi ke-24.

Estimasi untuk mencari nilai parameter pada data yang mengandung *outlier* dilakukan dengan beberapa metode, yaitu metode MKT, metode regresi *robust* estimasi *LTS*, estimasi *M* dan metode estimasi *MM*. Hasil dari nilai estimasi parameter pada metode MKT tersebut adalah sebagai berikut:

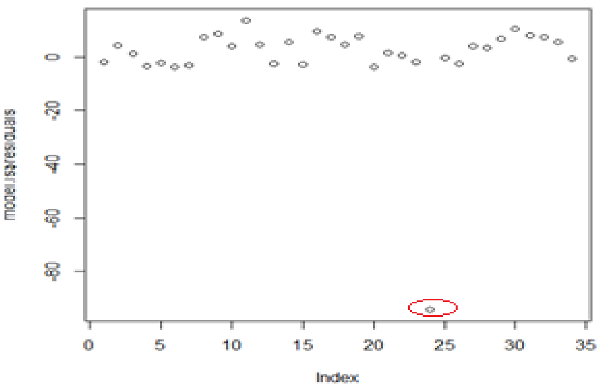
Tabel 4.1. Hasil Estimasi Parameter Menggunakan MKT Pada Data Mengandung *Outlier*

| Parameter | Nilai Estimasi | *Standard Error* |
|---|---|---|
| (*Intercept*) | 85.30000 | 13.18000 |
| $x_1$ | 0.00004 | 0.00008 |
| $x_2$ | 0.90450 | 1.01900 |

Model persamaan regresi yang terbentuk dari estimasi nilai parameter metode MKT adalah:

$$\hat{y} = 85.30000 + 0.00004x_1 + 0.90450x_2 + \varepsilon \quad (5.17)$$

Estimasi pada metode ini menghasilkan *residual standard error* sebesar 17.85000. Plot residual pada metode MKT mayoritas residual terletak disekitar 0. Namun, ada beberapa residual yang terletak jauh dari 0. Hal ini mengidentifikasi bahwa data terdapat *outlier*. Berikut ini merupakan plot residual metode MKT:




Gambar 4.1. Plot Residual Metode MKT Mengandung *Outlier*

Pada gambar 4.1. tersebut dapat dilihat bahwa terdapat residual yang nilainya paling besar diantara nilai residual yang pada observasi lain. Hal ini mengindikasikan adanya *outlier* sehingga terdapat metode lain yang dapat digunakan menangani *outlier* yaitu regresi *robust*.

Regresi *robust* yang pertama digunakan adalah metode estimasi *LTS*. Berikut ini merupakan hasil estimasi nilai parameter menggunakan metode estimasi *LTS*:

Tabel 4.2. Hasil Estimasi Parameter Menggunakan Estimasi *LTS* Pada Data Mengandung *Outlier*

| Parameter | Nilai Estimasi | *Standard Error* |
|---|---|---|
| (*Intercept*) | 111.70000 | 5.49500 |
| $x_1$ | 0.00030 | 0.00007 |
| $x_2$ | -1.01000 | 0.42930 |

Model persamaan regresi yang terbentuk dari estimasi nilai parameter metode estimasi *LTS* adalah:

$$\hat{y} = 111.70000 + 0.00030x_1 - 1.01000x_2 + \varepsilon \qquad (4.18)$$

Estimasi pada metode ini menghasilkan *residual standard error* sebesar 4.26400. Plot residual pada metode *LTS* memiliki hasil yang hampir sama dengan metode MKT yaitu tedapat residual observasi yang terletak jauh dari 0. Oleh sebab itu metode ini, digunakan untuk mengatasi hal tersebut tanpa perlu mengeluarkan observasi yang diindikasi sebagai *outlier*.

Regresi robust selanjutnya adalah estimasi *M*. Berikut ini merupakan hasil estimasi nilai parameter menggunakan metode estimasi *M*:

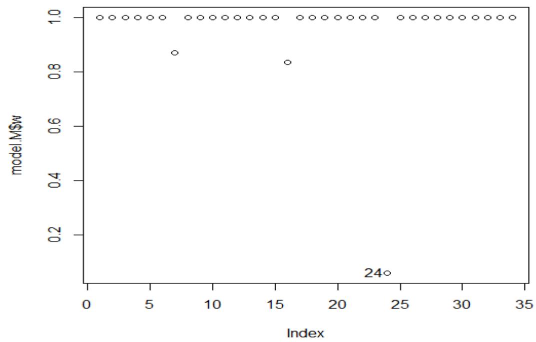
Tabel 4.3. Hasil Estimasi Parameter Menggunakan Estimasi *M* Pada Data Mengandung *Outlier*

| Parameter | Nilai Estimasi | *Standard Error* |
|---|---|---|
| (*Intercept*) | 100.03710 | 3.08800 |
| $x_1$ | 0.00001 | 0.00000 |
| $x_2$ | -0.01520 | 0.23890 |

Pada estimasi *M*, model model persamaan regresi yang terbentuk dari estimasi nilai parameter adalah:

$$\hat{y} = 100.03710 + 0.00001 - 0.01520x_2 + \varepsilon \qquad (4.17)$$

Estimasi pada metode ini menghasilkan *residual standard error* sebesar 4.59700. Plot residual pada metode *M* memiliki hasil yang hampir sama dengan metode MKT dan metode *LTS* yaitu tedapat residual observasi yang terletak jauh dari 0. Oleh sebab itu metode ini, digunakan untuk mengatasi hal tersebut tanpa perlu mengeluarkan observasi yang diindikasi sebagai *outlier*. Pada metode *M* sebelum mengestimasi nilai parameter terlebih dahulu dilakukan pembobotan pada data tersebut. Dalam hal ini pembobotan yang digunakan adalah fungsi pemobot *Huber*. Berikut ini hasil plot pembobotan tersebut:




Gambar 4.2. Pembobotan Pada Metode *M* Menggunakan *Huber*

Pada plot tersebut terlihat bahwa ada data yang memiliki residual besar diberikan bobot yang kecil dan plot ini juga dapat digunakan untuk mendeteksi *outlier*. Hal ini terlihat pada data observasi ke-24 dan teidentifikasi adanya *outlier*. Regresi *robust* yang lain yang dapat digunakan adalah metode estimasi *MM*. Berikut ini hasil estimasi nilai parameter menggunakan metode estimasi *MM*:

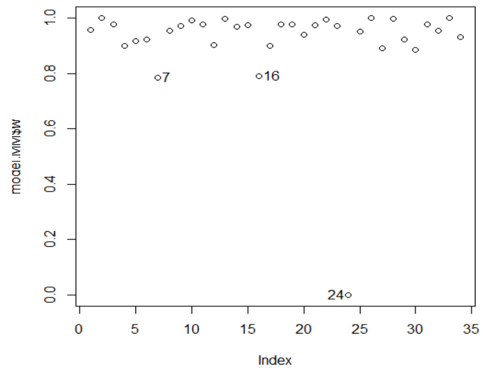
Tabel 4.4. Hasil Estimasi Parameter Menggunakan Estimasi *MM* Pada Data Mengandung *Outlier*

| Parameter | Nilai Estimasi | *Standard Error* |
|---|---|---|
| (*Intercept*) | 101.08150 | 3.03370 |
| $x_1$ | 0.00001 | 0.00000 |
| $x_2$ | -0.08060 | 0.23460 |

Pada estimasi MM, model model persamaan regresi yang terbentuk dari estimasi nilai parameter adalah:

$$\hat{y} = 101.08150 + 0.00001x_1 - 0.08060x_2 + \varepsilon \qquad (4.19)$$

Estimasi pada metode ini menghasilkan *residual standard error* sebesar 4.65400. Plot residual pada metode *MM* memiliki hasil yang hampir sama dengan metode MKT ,metode *LTS*, dan metode *M* yaitu tedapat residual observasi yang terletak jauh dari 0. Dalam hal ini pembobotan yang digunakan adalah fungsi pemobot *Tukey Bisquare*. Berikut ini hasil plot pembobotan tersebut:




Gambar 4.3. Pembobotan Pada Metode *MM* Menggunakan *Tukey Bisquare*

Pada plot tersebut terlihat bahwa beberapa data yang memiliki residual besar diberikan bobot yang kecil dan plot ini juga dapat digunakan untuk mendeteksi *outlier*. Hal ini terlihat pada data observasi ke-7, ke-16, dan ke-24. Pembobotan metode *MM* dengan *Tukey Bisquare* memberikan bobot yang kecil pada nilai residual yang tinggi. Bobot pada setiap observasi dapat dilihat pada lampiran 7.

Berdasarkan hasil estimasi parameter menggunakan empat metode tersebut diperoleh *residual standard error* secara keseluruhan seperti berikut:

Tabel 4.5. *Residual Standard Error* Pada Empat Metode Pada Data Mengandung *Outlier*

| Metode | *Residual Standard Error* |
|---|---:|
| MKT | 17.85000 |
| Estimasi *LTS* | 4.26400 |
| Estimasi *M* | 4.59700 |
| Estimasi *MM* | 4.65400 |

Pada tabel 4.5. dapat dilihat bahwa metode estimasi *LTS* memiliki *residual standard error* paling kecil diantara metode yang lain yaitu sebesar 4.2640. Oleh sebab itu, metode estimasi *LTS* lebih baik digunakan untuk mengatasi *outlier* dari pada dua metode yang lain pada kasus nilai tukar petani tanaman kedelai tahun 2015. Metode MKT merupakan metode yang paling tidak bagus digunakan ketika terdapat data *outlier*. Metode ini memiliki *residual standard error* terbesar diantara yang lain. Padahal model dan estimasi parameter yang baik jika nilai *standard error* mendekati 0 (kecil).

Metode estimasi *M* dan estimasi *MM* memiliki nilai *residual standard error* sebesar 4.59700 dan 4.65400. Hal ini mengartikan bahwa metode estimasi *M* dan estimasi *MM* lebih baik digunakan dari pada MKT jika data mengandung *outlier*. Metode estimasi *LTS* lebih baik digunakan dari pada metode estimasi *M* dan estimasi *M*M. Hal ini disebabkan karena data mengandung *outlier* pada arah *y*. Oleh sebab itu, motode yang cocok digunakan untuk mengatasi *outlier* adalah regresi *robust* baik menggunakan estimasi *LTS*, estimasi *M* dan estimasi *MM*. Namun, ketika terdapat data *outlier* pada arah *y* metode yang cocok digunakan adalah metode estimasi *LTS*.

### b. Analisis data tanpa *outlier*

Data yang digunakan pada kasus ini sama dengan data yang digunakan pada kasus pertama. Hanya saja data-data yang mengandung *outlier* pada arah *y* tidak diikutkan dalam proses pengolahan data. Setelah mengeluarkan data ke-24 setelah dicek masih terdapat *outlier* pada data ke-16, dan data ke-7 pada arah *y*. Sehingga data tanpa *outlier* dengan menghilangkan ketiga data tersebut. Maka diperoleh 31 data yang tidak mengandung *outlier*. Data inilah yang yang digunakan untuk analisis data selanjutnya. Data dapat dilihat pada lampiran 4.

Langkah awal sebelum mencari nilai estimasi dari parameter $\beta_0$, $\beta_1$ dan $\beta_2$, akan diperiksa terlebih dahulu keberadaan *outlier* dalam data. Hal ini seperti yang dilakukan pada kasus pertama. Nilai *cutoff* untuk masing-masing metode ditentukan berdasarkan jumlah sampel (*n*) yang dalam hal ini ketika empat observasi telah dihilangkan yaitu 31 dan banyaknya parameter (*k*) adalah 3. Oleh sebab itu, observasi dikatakan *outlier* jika:

- $|R - student| > 2$
- *Leverage* ($h_{ii}$) > $\frac{2k}{n} = \frac{2.3}{31} = 0.19355$

- $|DFFITS_i| > 2\sqrt{\frac{k}{n}} = 2\sqrt{\frac{3}{31}} = 0.11175$
- $|DFBETAS_{ji}| > \frac{2}{\sqrt{n}} = \frac{2}{\sqrt{31}} = 0.35921$

Berdasarkan kriteria tersebut, dapat diketahui observasi-observasi yang merupakan *outlier*. Hasil perhitungan nilai-nilai tersebut dapat dilihat pada lampiran 6 mengeni deteksi *outlier*. Berdasarkan tabel yang terdapat pada lampiran 6, tidak terdapat observasi yang diduga kuat merupakan *outlier* ke arah *y* yaitu dilihat dari nilai *R-student* pada masing-masing observasi dibandingkan dengan kriteria yang ada.

Estimasi untuk mencari nilai parameter dilakukan dengan beberapa metode, yaitu metode MKT, metode regresi *robust* estimasi *LTS*, metode estimasi *M* dan metode estimasi *MM*. Hal ini digunakan untuk mengetahui metode yang baik untuk digunakan ketika data tidak ada yang *outlier*. Hasil dari nilai estimasi parameter pada metode MKT adalah sebagai berikut:

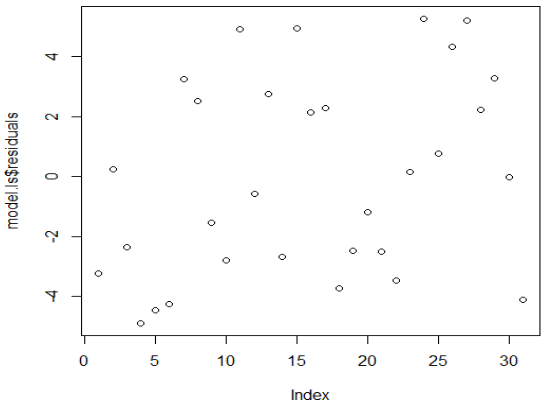
Tabel 4.6. Hasil Estimasi Parameter Menggunakan MKT Pada Data Tanpa *Outlier*

| Parameter | Nilai Estimasi | *Standard Error* |
|---|---|---|
| (*Intercept*) | 101.60000 | 2.58300 |
| $x_1$ | 0.00004 | 0.00002 |
| $x_2$ | -0.11740 | 0.19870 |

Model persamaan regresi yang terbentuk dari estimasi nilai parameter metode MKT adalah:

$$\hat{y} = 101.60000 + 0.00004x_1 - 0.11740x_2 + \varepsilon \qquad (5.20)$$

Estimasi pada metode ini menghasilkan *residual standard error* sebesar 3.40300. Plot residual pada metode MKT dapat dilihat bahwa semua residual terletak disekitar 0. residual menyebar dan terletak disekitar 0. Hal ini mengartikan bahwa data memenuhi asumsi. Berikut ini merupakan plot residual metode MKT:




Gambar 4.4. Plot Residual Metode MKT Tanpa *Outlier*

Pada gambar 4.4 tersebut dapat dilihat bahwa tidak ada residual yang nilainya paling besar diantara nilai residual yang pada observasi lain. Hal ini mengartikan bahwa metode ini sudah baik untuk digunakan. Namun, untuk mengetahui secara pasti dilakukan perbandingan dengan metode estimasi *LTS*, metode estimasi *M*, dan metode estimasi *MM*. Berikut ini merupakan hasil estimasi menggunakan metode estimasi *LTS*:

Tabel 4.7. Hasil Estimasi Parameter Menggunakan Estimasi *LTS* Pada Data Tanpa *Outlier*

| Parameter | Nilai Estimasi | *Standard error* |
|---|---|---|
| (*Intercept*) | 115.60000 | 4.73900 |
| $x_1$ | 0.00004 | 0.00007 |
| $x_2$ | -1.3560 | 0.43750 |

Model persamaan regresi yang terbentuk dari estimasi nilai parameter metode estimasi *M* adalah:

$$\hat{y} = 115.60000 + 0.00004 - 1.3560x_2 + \varepsilon \tag{4.21}$$

Estimasi pada metode ini menghasilkan *residual standard error* sebesar 3.62400. Plot residual pada metode *LTS* memiliki hasil yang hampir sama dengan metode MKT yaitu tidak tedapat residual observasi yang terletak jauh dari 0. Metode selanjutnya yang digunakan untuk membandingkan adalah metode estimasi *M*. Berikut ini merupakan hasil estimasi nilai parameter:

Tabel 4.8. Hasil Estimasi Parameter Menggunakan Estimasi *M* Pada Data Tanpa *Outlier*

| Parameter | Nilai Estimasi | *Standard Error* |
|---|---|---|
| (*Intercept*) | 101.56470 | 2.58310 |
| $x_1$ | 0.00001 | 0.00000 |
| $x_2$ | -0.11740 | 0.19780 |

Pada estimasi *MM*, model model persamaan regresi yang terbentuk dari estimasi nilai parameter adalah:

$$\hat{y} = 101.56470 + 0.00001x_1 - 0.11740x_2 + \varepsilon \qquad (4.22)$$

Estimasi pada metode ini menghasilkan *residual standard error* sebesar 4.05800. Plot residual pada metode *M* memiliki hasil yang hampir sama dengan metode MKT dan metode *LTS* yaitu tidak tedapat residual observasi yang terletak jauh dari 0.

Metode selanjutnya yang digunakan untuk membandingkan adalah metode estimasi *MM*. Berikut ini merupakan hasil estimasi nilai parameter menggunakan metode estimasi *MM*:

Tabel 4.9. Hasil Estimasi Parameter Menggunakan Estimasi *MM* Pada Data Tanpa *Outlier*

| Parameter | Nilai Estimasi | *Standard Error* |
|---|---|---|
| (*Intercept*) | 101.57340 | 2.81590 |
| x1 | 0.00001 | 0.00000 |
| x2 | -0.12010 | 0.21560 |

Pada estimasi *MM*, model model persamaan regresi yang terbentuk dari estimasi nilai parameter adalah:

$$\hat{y} = 101.57340 + 0.00001x_1 - 0.12010x_2 + \varepsilon \qquad (4.23)$$

Estimasi pada metode ini menghasilkan *residual standard error* sebesar 4.06900. Plot residual pada metode *MM* memiliki hasil yang hampir sama dengan metode MKT, *LTS*, dan *M* yaitu tidak tedapat residual observasi yang terletak jauh dari 0.

Oleh sebab itu, sebenarnya tidak diperlukan analisis data menggunakan metode estimasi *LTS*, *M*, dan *MM* karena tidak terdapat *outlier* dalam data.

Berdasarkan hasil estimasi parameter menggunakan empat metode tersebut diperoleh *residual standard error* secara keseluruhan seperti berikut:

Tabel 4.10. *Residual Standard Error* Pada Empat Metode Pada Data Tanpa *Outlier*

| Metode | *Residual Standard Error* |
|---|---|
| MKT | 3.40300 |
| Estimasi *LTS* | 3.62400 |
| Estimasi *M* | 4.05800 |
| Estimasi *MM* | 4.06900 |

Pada tabel 4.10. dapat dilihat bahwa MKT memiliki *residual standard error* paling kecil diantara metode yang lain yaitu sebesar 3.40300. Oleh sebab itu, MKT lebih baik digunakan untuk data tanpa mengandung *outlier* dari pada dua metode yang lain. Metode *MM* merupakan metode yang paling tidak bagus digunakan ketika tidak terdapat data *outlier*. Metode ini memiliki *residual standard error* terbesar diantara yang lain. Hal ini mengartikan bahwa motode yang cocok digunakan ketika tidak terdapat *outlier* adalah metode MKT.

# BAB V
# KESIMPULAN DAN SARAN

## 5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan maka dapat diperoleh kesimpulan sebagai berikut:

1. Estimasi parameter menggunakan metode estimasi *LTS* dilakukan dengan cara meminimumkan jumlah kuadrat *h error* (fungsi objektif) dan menggunakn *FWLS*. Metode estimasi *M* dilakukan dengan cara memberi bobot pada $e_i$ kemudian perhitungan nilai parameter dilakukan dengan *WLS*. Pada estimasi parameter menggunkan metode estimasi *MM* dilakukan dengan cara menggabungkan cara estimasi pada metode estimasi *S* dengan cara estimasi pada metode estimasi *M*. Perhitungan nilai parameter dilakukan dengan menggunakan *WLS*.
2. Estimasi parameter untuk $\hat{\beta}$ pada model regresi *robust* berganda dengan estimasi *LTS*, diperoleh sebagai berikut:

   $$\tilde{\beta}_{LTS} = arg\ min \sum_{i=1}^{h} e_i^2$$

   Estimasi parameter untuk $\hat{\beta}$ pada model regresi *robust* berganda dengan estimasi *M*, diperoleh sebagai berikut:

   $$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{W_0}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{W_0}\boldsymbol{y}$$

   sedangkan estimasi parameter untuk $\hat{\beta}$ pada model regresi *robust* berganda dengan estimasi *MM*, diperoleh sebagai berikut:

   $$\widehat{\boldsymbol{\beta}}^{(\boldsymbol{m}+\mathbf{1})} = (\boldsymbol{X}'\boldsymbol{W}^{\boldsymbol{m}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{W}^{\boldsymbol{m}}\boldsymbol{y}$$

3. Pada kasus data nilai tukar petani 2015 yang mengandung *outlier* pada arah *y*, metode yang paling baik digunakan adalah metode estimasi *LTS* pada regresi *robust* dibandingkan dengan MKT, metode estimasi *M*, dan metode estimasi *MM*. Pada kasus data nilai tukar petani 2015 tanpa *outlier*¸ metode yang paling baik digunakan adalah MKT. Perbandingan dilakukan dengan melihat nilai *residual standard error*.

## 5.2 Saran

Berdasarkan kesimpulan yang diperoleh dari analisis maka diketahui bagaimana cara estimasi menggunakan metode estimasi *LTS*, metode estimasi *M* dan metode estimasi *MM* serta persamaan estimasi $\beta$ pada dua metode tersebut, sehingga dapat digunakan sebagai referensi dalam perhitungan estimasi $\beta$ dalam mengatasi *outlier*.

Berdasarkan analisis metode yang paling baik digunakan dalam kasus data nilai tukar petani 2015 data yang mengandung *outlier* khususnya *outlier* pada arah *y* adalah metode estimasi *LTS* pada regresi *robust*, sehingga dapat digunakan sebagai referensi ketika data mengandung *outlier* pada arah *y*, metode regresi *robust* yang dapat digunakan dalam mengatasi *outlier* adalah metode estimasi *LTS*.

## DAFTAR PUSTAKA


Andersen, R. (2008). *Modern Methods For Robust Regression*. Thousand Oaks: SAGE Publications.

Algifari. (1997). *Analisis Regresi Teori, Kasus dan Solusi*. Yogyakarta: BPFE.

Ardiyanti, H. (2011). *Perbandingan Keefektifan Metode Regresi Robust Estimasi-M Dan Estimasi-MM Karena Pengaruh Outlier Dalam Analisis Regresi Linear (Contoh Kasus Data Produksi Padi Di Jawa Tengah Tahun 2007)*. Skripsi. Semarang: Universitas Negeri Semarang.

Barrera, M. S., & Yohai, V. J. (2008). High Breakdown Point Robust Regression with Cencored Data. The Annals of Statistics, Vol. 36, No. 1, 118-146.

Bekti, D. R. (2011). *Materi Statistik*. http://www.statisticsanalyst.files.wordpress.

com/2011/10/11.doc. Diakses Tanggal 6 Februari 2016.

BPS. (2015). *Indikator Pertanian Nasional Tahun 2015*. http://www.bps.go.id/site

/resultTab. Diakses Tanggal 6 Februari 2016.

BPS. (2015). *Luas Panen Tanaman Kedelai Tahun 2015*. Jakarta: BPS RI.

BPS. (2015). *Nilai Tukar Petani Tanaman Kedelai Tahun 2015*. Jakarta: BPS RI.

BPS. (2015). *Produksi Tanaman Kedelai Tahun 2015*. Jakarta: BPS RI.

Buechler, S. (2007). *Statistical Models in R Some Examples*. http://www.3.nd.edu

/~steve/Rcourse/Lecture8v1.pdf. Diakses Tanggal 7 Februari 2016.

Draper, N., & Smith, H. (1992). *Analisis Regresi Terapan Edisi Kedua*. Jakarta: Gramedia Pustaka Utama.

Ghozali, I. (2005). *Analisis Multivariate dengan Program SPSS Ed 3*. Semarang: Badan Penerbit Universitas Diponegoro.

Gujarati, D. N. (2004). *Basic Econometrics Fourth editon*. New York: McGraw-Hill.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1992). *Multivariate Data Analysis*. New York: Macmillan Publishing Company.

Kutner, M. H., Nachtshein, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Model*. New York: McGraw-Hill.

Kurniawati, L. D. (2011). *Kekekaran Regresi Linier Ganda Dengan Estimasi MM (Method Of Moment) Dalam Mengatasi Pencilan*. Skripsi. Yogyakarta: Universitas Negeri Yogyakarta.

Maharani, I.F., Satyahadewi, N., & Kusnandar, D. (2014). Metode Ordinary Least Squares Dan Least Trimmed Squares Dalam Mengestimasi Parameter Regresi Ketika Terdapat Outlier. Buletin Ilmiah Mat, Stat, dan Terapannya (Bimaster), Vol. 03, No. 3, 163-168.

Mashitah, Wibowo, A., & Indriani, D. (2013). Metode Robust Regression on Ordered Statistics (ROS) pada Data Tersensor Kiri dengan Outlier. Jurnal Biometrika dan Kependudukan, Vol. 2, No. 2 , 148–157.

Montgomery, D. C., & Peck, E. A. (1982). Introdution to Linear Regression Analysis. New York: John WIley and Sons.

Nurcahyadi, H. (2010). *Analisis Regresi pada Data Outlier dengan Menggunakan Least Trimmed Square (LTS) dan MM-Estimasi*. Skripsi. Jakarta: Universitas Islam Negeri Syarif Hidayatullah.

Putri, N. A. (2013). Studi Komparatif Metode Kuadrat Terkecil dengan Metode Regresi Robust Pembobot Welsch Pada Data yang Mengandung Outlier. Jurnal Matematika UNAND Vol. 02, 18-26.

Qudratullah, M. F. (2013). *Analisis Regresi Terapan Teori, Contoh Kasus, dan Aplikasi dengan SPSS*. Yogyakarta: ANDI.

Rousseeuw, P. J. (1987). *Robust Regression and Outlier Detection*. New York: Wiley and Sons.

Safitri, D. A . (2015). *Perbandingan Metode Estimasi M Dan Estimasi MM (Method Of Moment) Pada Regresi Robust*. Skripsi. Yogyakarta: Universitas Islam Indonesia.

Sembiring, R. K. (2003). *Analisis Regresi Edisi Kedua*. Bandung: ITB.

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

Susanti, Y., Pratiwi, H., & H, Sri Sulistiowati. (2013). Optimasi Model Regresi Robust Untuk Memprediksi Produksi Kedelai Di Indonesia. Surakarta: Universitas Sebelas Maret Surakarta.

Widodo, E., Guritno, S., & Haryatmi, S. (2015). *Estimasi Model Permukaan Respon Multivariat Dengan Data Outlier*. Disertasi Doktor. Yogyakarta: Universitas Gajah Mada.

Wijaya, S. (2009). *Taksiran Parameter pada Model Regresi Robust dengan Menggunakan Fungsi Huber*. Skripsi. Jakarta: Universitas Indonesia.

Lampiran 1 Data Nilai Tukar Petani Tanaman Kedelai Tahun 2015

| **No.** | **Provinsi** | $X_1$ | $X_2$ | $Y$ |
|---|---|---|---|---|
| 1 | Aceh | 34826 | 14.65 | 98.13 |
| 2 | Sumatra Utara | 5481 | 12.01 | 100.62 |
| 3 | Sumatra Barat | 347 | 12.59 | 97.75 |
| 4 | Riau | 1309 | 14.51 | 95.03 |
| 5 | Jambi | 5148 | 13.80 | 95.72 |
| 6 | Sumatera Selatan | 12421 | 15.46 | 96.03 |
| 7 | BENGKULU | 4702 | 11.90 | 92.96 |
| 8 | Lampung | 10071 | 12.02 | 103.84 |
| 9 | Kep.Bangka Belitung | 1 | 10 | 102.92 |
| 10 | Kep. Riau | 15 | 10.67 | 98.78 |
| 11 | DKI Jakarta | 0 | 0 | 98.77 |
| 12 | Jawa Barat | 61677 | 16.51 | 107.24 |
| 13 | Jawa Tengah | 72694 | 18.21 | 102.03 |
| 14 | DIY | 13948 | 13.37 | 103.34 |
| 15 | Jawa Timur | 210761 | 16.61 | 106.13 |
| 16 | Banten | 5143 | 13.72 | 107.45 |
| 17 | Bali | 5115 | 13.59 | 105.13 |
| 18 | Nusa Tenggara Barat | 94760 | 13.78 | 106.22 |
| 19 | Nusa Tenggara Timur | 2507 | 10.82 | 102.69 |
| 20 | Kalimantan Barat | 1697 | 16.10 | 96.03 |
| 21 | Kalimantan Timur | 1141 | 12 | 97.74 |
| 22 | Kalimantan Tengah | 7604 | 14.15 | 99.03 |
| 23 | Kalimantan Selatan | 1079 | 15.39 | 97.31 |
| 24 | Kalimantan Utara | 2824 | 9.60 | 0 |
| 25 | Sulawesi Utara | 6330 | 12.98 | 96.85 |
| 26 | Sulawesi Tengah | 6940 | 18.75 | 99.82 |
| 27 | Sulawesi Selatan | 37128 | 17.46 | 106.39 |
| 28 | Sulawesi Tenggara | 6035 | 13.48 | 101.01 |
| 29 | gorontalo | 2384 | 13.47 | 104.41 |
| 30 | Sulawesi barat | 5497 | 10.93 | 105.71 |
| No. | Provinsi | $X_1$ | $X_2$ | $Y$ |
| 31 | Maluku | 962 | 10.34 | 102.61 |
| 32 | Maluku Utara | 532 | 11.97 | 103.46 |
| 33 | Papua Barat | 1343 | 10.61 | 100.35 |
| 34 | Papua | 2426 | 12.72 | 96.08 |

Sumber: BPS

Keterangan:
*X1* : Produksi Kedelai Tahun 2015 (Kg)
*X2* : Luas Panen Kedelai Tahun 2015 (Ha)
Y : Nilai Tukar Petani Tanaman Kedelai Tahun 2015

Lampiran 2 Perintah Pada *Software* R untuk Estimasi Data Mengandung *Outlier*

```
> #input data
> data=read.delim("clipboard")
> data=data.frame(x1=data$x1,x2=data$x2,y=data$y)
> #OLS
> model.ls<-lm(y~x1+x2,data=data)
> summary(model.ls)

Call:
lm(formula = y ~ x1 + x2, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-94.097   -2.192   3.522   7.150  13.472

Coefficients:
            Estimate    Std. Error t value Pr(>|t|)
((Intercept) 8.530e+01  1.318e+01  6.473   3.21e-07 ***
x1           4.114e-05  8.264e-05   0.498  0.622
x2           9.045e-01  1.019e+00   0.888  0.382
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
1

Residual standard error: 17.85 on 31 degrees of freedom
Multiple R-squared: 0.049,     Adjusted R-squared:-0.01236
F-statistic:  0.7986 on 2 and 31 DF,  p-value: 0.459

> plot(model.ls$residuals)
> #uji normalitas pada residual
> library(nortest)
> lillie.test(model.ls$residuals)

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  model.ls$residuals
D = 0.3814, p-value = 5.673e-14
> #deteksi outlier
> (im<-influence.measures(model.ls))
> rstudent(model.ls)
> #LTS
> library(robustbase)
> model.LTS<-ltsReg(y~x1+x2,data=data)
> summary(model.LTS)
```

```
Call:
ltsReg.formula(formula = y ~ x1 + x2, data = data)
Residuals (from reweighted LS):
   Min      1Q      Median     3Q         Max
-9.784    -2.192     0.000      1.081    7.978
Coefficients:
            Estimate   Std. Error  t value   Pr(>|t|)
Intercept  1.117e+02   5.495e+00   20.329    < 2e-16 ***
x1         3.161e-04   6.913e-05   4.572      0.000104 ***
x2        -1.010e+00   4.293e-01  -2.354      0.026431 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
1

Residual standard error: 4.264 on 26 degrees of freedom
Multiple R-Squared: 0.4457,     Adjusted R-squared: 0.4031
F-statistic: 10.45 on 2 and 26 DF,  p-value: 0.0004662
> plot(model.LTS$residuals)
> #model M
> library(MASS)
> model.M<-rlm(y~x1+x2,data=data)
> summary(model.M)

Call: rlm(formula = y ~ x1 + x2, data = data)
Residuals:
     Min      1Q       Median    3Q       Max
   -100.0113  -3.0311  -0.5257   3.0044   7.4027

Coefficients:
             Value      Std. Error     t value
(Intercept)  100.0371    3.0889        32.3861
x1             0.0001    0.0000         2.1969
x2            -0.0152    0.2389        -0.0637
Residual standard error: 4.597 on 31 degrees of freedom
> plot(model.M$residuals)
> plot(model.M$w)
> smallweights<-which(model.M$w<0.8)
> library(car)
>
showLabels(1:50,model.M$w,rownames(data),id.method=smallweig
hts,cex.=.6)
 24
 24
> model.M$w
> #model MM
> model.MM<-rlm(y~x1+x2,data=data,method="MM")
> summary(model.MM)
Call: rlm(formula = y ~ x1 + x2, data = data, method = "MM")
Residuals:
      Min      1Q          Median    3Q       Max
   -100.4271   -3.0739  -0.3463    2.7218   7.2570
```

```
Coefficients:
               Value       Std. Error    t value
(Intercept)   101.0815     3.0337        33.3191
x1              0.0001     0.0000         2.2157
x2             -0.0806     0.2346        -0.3434
Residual standard error: 4.654 on 31 degrees of freedom
> plot(model.MM$residuals)
> plot(model.MM$w)
> smallweights<-which(model.MM$w<0.8)
>
showLabels(1:50,model.MM$w,rownames(data),id.method=smallwei
ghts,cex.=.6)
> model.MM$w
```

Lampiran 3 Perintah Pada *Software* R untuk Estimasi Data Tanpa *Outlier*

```
> #data tanpa outlier
> newdata<-data[-c(24,16,7), ]
> #OLS
> model.ls<-lm(y~x1+x2,data=newdata)
> summary(model.ls)
Call:
lm(formula = y ~ x1 + x2, data = newdata)

Residuals:
    Min      1Q  Median      3Q     Max
-4.8876 -2.7347 -0.0271  2.6336  5.2569

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.016e+02  2.583e+00  39.319   <2e-16 ***
x1           4.361e-05  1.581e-05   2.758   0.0101 *
x2          -1.174e-01  1.978e-01  -0.594   0.5574
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
1

Residual standard error: 3.403 on 28 degrees of freedom
Multiple R-squared: 0.2184,     Adjusted R-squared: 0.1625
F-statistic: 3.911 on 2 and 28 DF,  p-value: 0.03178
> plot(model.ls$residuals)
> #deteksi outlier
> (im<-influence.measures(model.ls))
> rstudent(model.ls)
> #Model LTS
> model.LTS<-ltsReg(y~x1+x2,data=newdata)
> summary(model.LTS)
Call:
ltsReg.formula(formula = y ~ x1 + x2, data = newdata)
Residuals (from reweighted LS):
   Min     1Q Median     3Q    Max
-9.185 -1.382  0.000  1.144  7.031
Coefficients:
```

```
             Estimate Std. Error t value Pr(>|t|)
Intercept  1.156e+02  4.739e+00  24.394  < 2e-16 ***
x1         3.763e-04  6.333e-05   5.943 4.66e-06 ***
x2        -1.356e+00  3.696e-01  -3.669  0.00127 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
1
Residual standard error: 3.624 on 23 degrees of freedom
Multiple R-Squared: 0.6092,     Adjusted R-squared: 0.5752
F-statistic: 17.93 on 2 and 23 DF,  p-value: 2.029e-05
>  #model M hubber
>  library(MASS)
>  model.M<-rlm(y~x1+x2,data=newdata)
>  summary(model.M)


Call: rlm(formula = y ~ x1 + x2, data = newdata)
Residuals:
     Min       1Q   Median       3Q      Max
-4.88758 -2.73467 -0.02711  2.63357  5.25691
Coefficients:
            Value    Std. Error t value
(Intercept) 101.5647   2.5831    39.3194
x1            0.0001   0.0000     2.7576
x2           -0.1174   0.1978    -0.5938
Residual standard error: 4.058 on 28 degrees of freedom>
model.MM$w
>  model.M$w
>  #model MM bisquare
>  model.MM<-rlm(y~x1+x2,data=newdata,method="MM")
>  summary(model.MM)
Call: rlm(formula = y ~ x1 + x2, data = newdata, method =
"MM")
Residuals:
     Min       1Q   Median       3Q      Max
-4.85722 -2.67056 -0.00706  2.65918  5.31234
Coefficients:
            Value    Std. Error t value
(Intercept) 101.5734   2.8159    36.0715
x1            0.0001   0.0000     2.5016
x2           -0.1201   0.2156    -0.5570
Residual standard error: 4.069 on 28 degrees of freedom>
> model.MM$w
```

Lampiran 4 Data Nilai Tukar Petani Tanaman Kedelai 2015 Tanpa *Outlier*

| No. | Provinsi | $X_1$ | $X_2$ | $Y$ |
|---|---|---|---|---|
| 1 | Aceh | 34826 | 14.65 | 98.13 |
| 2 | Sumatra Utara | 5481 | 12.01 | 100.62 |
| 3 | Sumatra Barat | 347 | 12.59 | 97.75 |
| 4 | Riau | 1309 | 14.51 | 95.03 |
| 5 | Jambi | 5148 | 13.80 | 95.72 |
| 6 | Sumatera  Selatan | 12421 | 15.46 | 96.03 |
| 8 | Lampung | 10071 | 12.02 | 103.84 |
| 9 | Kep.Bangka Belitung | 1 | 10 | 102.92 |
| 10 | Kep. Riau | 15 | 10.67 | 98.78 |
| 11 | DKI Jakarta | 0 | 0 | 98.77 |
| 12 | Jawa Barat | 61677 | 16.51 | 107.24 |
| 13 | Jawa Tengah | 72694 | 18.21 | 102.03 |
| 14 | DIY | 13948 | 13.37 | 103.34 |
| 15 | Jawa Timur | 210761 | 16.61 | 106.13 |
| 17 | Bali | 5115 | 13.59 | 105.13 |
| 18 | Nusa Tenggara Barat | 94760 | 13.78 | 106.22 |
| 19 | Nusa Tenggara Timur | 2507 | 10.82 | 102.69 |
| 20 | Kalimantan Barat | 1697 | 16.10 | 96.03 |
| 21 | Kalimantan Timur | 1141 | 12 | 97.74 |
| 22 | Kalimantan Tengah | 7604 | 14.15 | 99.03 |
| 23 | Kalimantan Selatan | 1079 | 15.39 | 97.31 |
| 25 | Kalimantan Utara | 6330 | 12.98 | 96.85 |
| 26 | Sulawesi Utara | 6940 | 18.75 | 99.82 |
| 27 | Sulawesi Tengah | 37128 | 17.46 | 106.39 |
| 28 | Sulawesi Selatan | 6035 | 13.48 | 101.01 |
| 29 | Sulawesi Tenggara | 2384 | 13.47 | 104.41 |
| 30 | gorontalo | 5497 | 10.93 | 105.71 |
| 31 | Sulawesi barat | 962 | 10.34 | 102.61 |
| 32 | Maluku | 532 | 11.97 | 103.46 |
| 33 | Maluku Utara | 1343 | 10.61 | 100.35 |
| No. | Provinsi | $X_1$ | $X_2$ | $Y$ |
| 34 | Papua Barat | 2426 | 12.72 | 96.08 |

Sumber: BPS

Keterangan:

$X_1$ : Produksi Kedelai Tahun 2015 (Kg)

$X_2$ : Luas Panen Kedelai Tahun 2015 (Ha)

$Y$ : Nilai Tukar Petani Tanaman Kedelai Tahun 2015

Lampiran 5 Deteksi *Outlier* untuk Data Mengandung *Outlier*

| No. | $DFBETAS_{0,i}$ | $DFBETAS_{1,i}$ | $DFBETAS_{2,i}$ | $DFFITS_{i}$ | $h_{ii}$ | *R-student* |
|---|---|---|---|---|---|---|
| 1 | 0.002907 | -0.00445 | -0.006554 | -0.02073 | 0.0382 | -0.10403996 |
| 2 | 0.019845 | -0.00894 | -0.009089 | 0.04449 | 0.0339 | 0.23757004 |
| 3 | 0.002627 | -0.00440 | 0.000268 | 0.01129 | 0.0354 | 0.05893606 |
| 4 | 0.012164 | 0.02198 | -0.022414 | 0.0474 | 0.0474 | -0.19465820 |
| 5 | 0.002220 | 0.01002 | -0.008469 | -0.02496 | 0.0369 | -0.12757488 |
| 6 | 0.021816 | 0.01733 | -0.032201 | -0.04976 | 0.0517 | -0.21304515 |
| 7 | -0.016411 | 0.00719 | 0.008002 | -0.03503 | 0.0347 | -0.18488157 |
| 8 | 0.036147 | -0.00661 | -0.018676 | 0.07496 | 0.0327 | 0.40764559 |
| 9 | 0.093626 | -0.00957 | -0.071755 | 0.11894 | 0.0560 | 0.48837119 |
| 10 | 0.033080 | -0.00740 | -0.023160 | 0.04779 | 0.0465 | 0.21633592 |
| 11 | 1.228579 | 0.32710 | -1.187736 | 1.22858 | 0.5448 | 1.12310960 |
| 12 | -0.025952 | 0.03377 | 0.032393 | -0.07500 | 0.0785 | 0.25695978 |
| 13 | 0.030401 | -0.02411 | -0.034109 | -0.06021 | 0.1231 | -0.16067118 |
| 14 | 0.005352 | -0.00829 | 0.007716 | 0.05335 | 0.0304 | 0.30125830 |
| 15 | -0.047093 | -0.47175 | 0.077581 | -0.49419 | 0.7274 | -0.30255777 |
| 16 | -0.006938 | -0.04129 | 0.033178 | 0.10433 | 0.0363 | 0.53755535 |
| 17 | -0.002323 | -0.03054 | 0.022348 | 0.07909 | 0.0355 | 0.41241412 |
| 18 | 0.027800 | 0.09951 | -0.027263 | 0.11213 | 0.1452 | 0.27208389 |
| No. | $DFBETAS_{0,i}$ | $DFBETAS_{1,i}$ | $DFBETAS_{2,i}$ | $DFFITS_{i}$ | $h_{ii}$ | *R-student* |
| 19 | 0.062509 | -0.01089 | -0.043521 | 0.09110 | 0.0441 | 0.42409910 |
| 20 | 0.034488 | 0.03319 | -0.047100 | -0.06393 | 0.0755 | -0.22367968 |
| 21 | 0.006811 | -0.00501 | -0.002695 | 0.01667 | 0.0358 | 0.08650691 |
| 22 | -0.001187 | -0.00261 | 0.002875 | 0.00694 | 0.0381 | 0.03488260 |
| 23 | 0.012591 | 0.01493 | -0.018688 | -0.02843 | 0.0615 | -0.11106228 |
| 24 | -5.608263 | -0.07038 | 4.514869 | -6.64217 | 0.0629 | -25.63014672 |
| 25 | 0.000757 | 0.00138 | -0.000416 | -0.00459 | 0.0323 | -0.02512319 |
| 26 | 0.050884 | 0.03101 | -0.060872 | -0.06855 | 0.1505 | -0.16282965 |
| 27 | -0.040771 | -0.00181 | 0.049515 | 0.06557 | 0.0835 | 0.21729709 |
| 28 | 0.000312 | -0.01235 | 0.008466 | 0.03457 | 0.0343 | 0.18353625 |
| 29 | -0.000892 | -0.03243 | 0.020032 | 0.07507 | 0.0367 | 0.38443928 |
| 30 | 0.084463 | -0.00813 | -0.059169 | 0.12260 | 0.0423 | 0.58323482 |
| 31 | 0.077805 | -0.01011 | -0.057667 | 0.10377 | 0.0506 | 0.44940844 |
| 32 | 0.032797 | -0.02476 | -0.013099 | 0.07983 | 0.0362 | 0.41166065 |
| 33 | 0.048237 | -0.00812 | -0.034485 | 0.06772 | 0.0469 | 0.30527585 |
| 34 | -0.001838 | 0.00312 | -0.000388 | -0.00867 | 0.0342 | -0.04612985 |

Lampiran 6 Deteksi *Outlier* untuk Data Tanpa *Outlier*

| No. | $DFBETAS_{0,i}$ | $DFBETAS_{1,i}$ | $DFBETAS_{2,i}$ | $DFFITS_i$ | $h_{ii}$ | *R-Student* |
|---|---|---|---|---|---|---|
| 1 | 0.02116 | -0.038069 | -0.057971 | -0.19748 | 0.0399 | -0.968390380 |
| 2 | 0.00625 | -0.002771 | -0.002995 | 0.01326 | 0.0380 | 0.066748893 |
| 3 | -0.03795 | 0.054952 | 0.001013 | -0.14117 | 0.0393 | -0.698358770 |
| 4 | 0.08240 | 0.176580 | -0.167769 | -0.34698 | 0.0504 | -1.506830306 |
| 5 | 0.01232 | 0.111900 | -0.083714 | -0.27594 | 0.0399 | -1.353689751 |
| 6 | 0.12516 | 0.111376 | -0.193615 | -0.31177 | 0.0541 | -1.303220677 |
| 8 | 0.09576 | -0.019407 | -0.050840 | 0.18943 | 0.0366 | 0.971367261 |
| 9 | 0.15590 | -0.017860 | -0.119094 | 0.19659 | 0.0624 | 0.761928934 |
| 10 | -0.07518 | 0.017384 | 0.052643 | -0.10691 | 0.0521 | -0.455839256 |
| 11 | -1.48676 | -0.379335 | 1.434869 | -1.48676 | 0.5761 | -1.275320486 |
| 12 | -0.15122 | 0.198160 | 0.193748 | 0.45389 | 0.0794 | 1.545179758 |
| 13 | 0.03288 | -0.025652 | -0.037333 | -0.06583 | 0.1244 | -0.174623461 |
| 14 | 0.02120 | -0.025474 | 0.016856 | 0.15091 | 0.0333 | 0.813129551 |
| 15 | -0.22887 | -2.410726 | 0.390449 | -2.52848 | 0.7283 | -1.544527210 |
| 17 | 0.00436 | -0.118263 | 0.074792 | 0.30325 | 0.0386 | 1.513689299 |
| 18 | 0.07184 | 0.245055 | -0.069746 | 0.27886 | 0.1461 | 0.674288794 |
| 19 | 0.10859 | -0.020160 | -0.075635 | 0.15570 | 0.0495 | 0.682614978 |
| No. | $DFBETAS_{0,i}$ | $DFBETAS_{1,i}$ | $DFBETAS_{2,i}$ | $DFFITS_i$ | $h_{ii}$ | *R-Student* |
| 20 | 0.16927 | 0.174968 | -0.238664 | -0.33376 | 0.0784 | -1.144229371 |
| 21 | -0.06532 | 0.045402 | 0.027752 | -0.14980 | 0.0401 | -0.733169521 |
| 22 | 0.00931 | 0.028022 | -0.027851 | -0.07343 | 0.0409 | -0.355677785 |
| 23 | 0.07999 | 0.104383 | -0.124385 | -0.19716 | 0.0644 | -0.751780177 |
| 25 | -0.04070 | 0.061099 | -0.011523 | -0.19993 | 0.0357 | -1.038681916 |
| 26 | -0.01503 | -0.009500 | 0.018241 | 0.02077 | 0.1545 | 0.048592411 |
| 27 | -0.30703 | -0.020171 | 0.379646 | 0.50894 | 0.0854 | 1.665595780 |
| 28 | 0.00232 | -0.016064 | 0.009292 | 0.04441 | 0.0374 | 0.225303675 |
| 29 | 0.00847 | -0.116210 | 0.061921 | 0.26807 | 0.0400 | 1.313268763 |
| 30 | 0.25087 | -0.028431 | -0.175729 | 0.35837 | 0.0474 | 1.605798087 |
| 31 | 0.12322 | -0.017498 | -0.091122 | 0.16266 | 0.0566 | 0.664212039 |
| 32 | 0.08855 | -0.063061 | -0.037901 | 0.20204 | 0.0406 | 0.982759130 |
| 33 | -0.00137 | 0.000245 | 0.000976 | -0.00189 | 0.0525 | -0.008035039 |
| 34 | -0.06117 | 0.088760 | -0.003204 | -0.24570 | 0.0378 | -1.238840903 |

Lampiran 7 Hasil Pembobotan Menggunakan Metode Estimasi *M* dengan Huber dan Metode Estimasi *MM* dengan *Tukey Bisquare* Ketika Data Mengandung *Outlier*

| No. | Pembobotan Estimasi *M* | Pembobotan Estimasi *MM* | No. | Pembobotan Estimasi *M* | Pembobotan Estimasi *MM* |
|---|---|---|---|---|---|
| 1 | 1.00000000 | 0.9563506 | 18 | 1.00000000 | 0.9787472 |
| 2 | 1.00000000 | 0.9996814 | 19 | 1.00000000 | 0.9764200 |
| 3 | 1.00000000 | 0.9772629 | 20 | 1.00000000 | 0.9393864 |
| 4 | 1.00000000 | 0.9000843 | 21 | 1.00000000 | 0.9754638 |
| 5 | 1.00000000 | 0.9178437 | 22 | 1.00000000 | 0.9936277 |
| 6 | 1.00000000 | 0.9227130 | 23 | 1.00000000 | 0.9722661 |
| 7 | 0.87141704 | 0.7850692 | 24 | 0.06182647 | 0.0000000 |
| 8 | 1.00000000 | 0.9546479 | 25 | 1.00000000 | 0.9504907 |
| 9 | 1.00000000 | 0.9708049 | 26 | 1.00000000 | 0.9999922 |
| 10 | 1.00000000 | 0.9912677 | 27 | 1.00000000 | 0.8915317 |
| 11 | 1.00000000 | 0.9776418 | 28 | 1.00000000 | 0.9975695 |
| 12 | 1.00000000 | 0.9019936 | 29 | 1.00000000 | 0.9232640 |
| 13 | 1.00000000 | 0.9982303 | 30 | 1.00000000 | 0.8862727 |
| 14 | 1.00000000 | 0.9684896 | 31 | 1.00000000 | 0.9774657 |
| 15 | 1.00000000 | 0.9739055 | 32 | 1.00000000 | 0.9541539 |
| 16 | 0.83522262 | 0.7907187 | 33 | 1.00000000 | 0.9999812 |
| 17 | 1.00000000 | 0.9004518 | 34 | 1.00000000 | 0.9312434 |

Lampiran 8 Hasil Pembobotan Menggunakan Metode Estimasi *M* dengan Huber dan Metode Estimasi *MM* dengan *Tukey Bisquare* Ketika Data Tanpa *Outlier*

| No. | Pembobotan Estimasi *M* | Pembobotan Estimasi *MM* | No. | Pembobotan Estimasi *M* | Pembobotan Estimasi *MM* |
|---|---|---|---|---|---|
| 1 | 1.0000000 | 0.9449313 | 19 | 1.0000000 | 0.9709022 |
| 2 | 1.0000000 | 0.9996487 | 20 | 1.0000000 | 0.9267566 |
| 3 | 1.0000000 | 0.9704429 | 21 | 1.0000000 | 0.9674698 |
| 4 | 1.0000000 | 0.8743923 | 22 | 1.0000000 | 0.9924581 |
| 5 | 1.0000000 | 0.8954767 | 23 | 1.0000000 | 0.9669375 |
| 6 | 1.0000000 | 0.9043066 | 25 | 1.0000000 | 0.9360349 |
| 8 | 1.0000000 | 0.9418159 | 26 | 1.0000000 | 0.9997810 |
| 9 | 1.0000000 | 0.9646052 | 27 | 1.0000000 | 0.8507163 |
| 10 | 1.0000000 | 0.9874492 | 28 | 1.0000000 | 0.9965217 |
| 11 | 1.0000000 | 0.9572069 | 29 | 1.0000000 | 0.8985066 |
| 12 | 1.0000000 | 0.8676818 | 30 | 1.0000000 | 0.8560865 |
| 13 | 1.0000000 | 0.9986731 | 31 | 1.0000000 | 0.9726536 |
| 14 | 1.0000000 | 0.9581949 | 32 | 1.0000000 | 0.9409248 |
| 15 | 1.0000000 | 0.9648755 | 33 | 1.0000000 | 0.9999997 |
| 17 | 1.0000000 | 0.8687788 | 34 | 1.0000000 | 0.9109090 |
| 18 | 1.0000000 | 0.9731827 | | | |

Lampiran 9 Sertifikat Makalah Tugas Akhir dalam Konferensi Nasional Penelitian Matematika dan Pembelajarannya.


PROGRAM STUDI PENDIDIKAN MATEMATIKA
FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN
UNIVERSITAS MUHAMMADIYAH SURAKARTA

SERTIFIKAT

diberikan kepada

Arlinda Amalia Dewayanti

sebagai

PEMAKALAH

Judul

PERBANDINGAN METODE ESTIMASI M DAN ESTIMASI MM (METHODE OF MOMENT) PADA REGRESI ROBUST

pada Konferensi Nasional Penelitian Matematika dan Pembelajarannya I dengan Tema:
"Tren dan Isu-isu Baru dalam Penelitian Matematika dan Pembelajarannya"
yang diselenggarakan oleh Program Studi Pendidikan Matematika FKIP UMS
di Auditorium Mohamad Djazman UMS, 12 Maret 2016

Rektor UMS,

Prof. Dr. Bambang Setiaji

Surakarta, 12 Maret 2016
Ketua Panitia,

Rita P. Khotimah, M.Sc