

# PERBANDING DUA METODE REGRESI *ROBUST* YAKNI *METODE LEAST TRIMMED SQUARES (LTS)* DENGAN METODE ESTIMATOR-MM (ESTMASI-MM) (STUDI KASUS DATA UJIAN TULIS MASUK TERHADAP HASIL IPK MAHASISWA UPGRIS)

Ali Shodiqin<sup>1</sup>, Aurora Nur Aini<sup>2</sup>, Maya Rini Rubowo<sup>3</sup>

Program Studi Pendidikan Matematika, FPMIPATI, Universitas PGRI Semarang  
Jl. Sidodadi Timur No. 24, Dr. Cipto, Semarang

Email: [alishodiqin@upgris.ac.id](mailto:alishodiqin@upgris.ac.id), [auroranuraini@upgris.ac.id](mailto:auroranuraini@upgris.ac.id), [mayarinirubowo@upgris.ac.id](mailto:mayarinirubowo@upgris.ac.id)

## Abstrak

Regresi linear ganda merupakan salah satu metode statistik yang digunakan untuk memodelkan dan menyelidiki hubungan antar satu variabel dependen dengan dua atau lebih variabel independen. *Ordinary Least Squares (OLS)* merupakan metode yang sering digunakan untuk mengestimasi parameter model regresi. Namun metode ini mempunyai kelemahan ketika *outlier* hadir dalam data. Estimator OLS bukan merupakan prosedur regresi yang *robust* terhadap adanya *outlier*, sehingga estimasinya menjadi tidak sesuai meskipun hanya satu kehadiran *outlier*. Regresi robust merupakan alat yang penting untuk menganalisis data yang terdeteksi sebagai data *outlier*. Tujuan dari penelitian ini adalah mengetahui Pencilan (*outlier*) mengganggu persamaan regresi linier, mengetahui hasil penaksir regresi *robust* dengan metode penaksir LTS (*Least Trimmed Squares*), mengetahui hasil penaksir regresi *robust* dengan metode penaksir MM (*MM-Estimator*), serta mengetahui perbandingan antara dua penaksir regresi *robust* tersebut dengan melihat nilai  $R^2$  dan residual masing-masing metode. Data yang digunakan dalam penelitian ini dari nilai ujian penerimaan mahasiswa baru dari Prodi Pendidikan Matematika di Universitas PGRI Semarang. Data ini terdiri merupakan data diskrit yang meliputi 3 (tiga) variabel yaitu nilai Tes (X1), Tes Psikologi (X2) sebagai variabel independen dan IPK (Y) sebagai variabel dependent. Sebelum dilakukan analisis dengan regresi robust, dilakukan pendeteksian *outlier* untuk mengidentifikasi adanya outlier atau tidak. Metode pendeteksian outlier dilakukan dengan beberapa, antara lain metode *boxplot*, *Cook's Distance*, dan metode *DfFIT (Difference In fit Standardized)*.

**Kata Kunci:** Regresi ganda, Outlier, Estimation, LTS, MM-Estimation

## Abstract

*Multiple linear regression is one of the statistical methods used to model and investigate the correlation between one dependent variable with two or more independent variables. Ordinary Least Squares (OLS) is a frequently used method for estimating regression model parameters. However this method has a drawback when the outlier is present in the data. OLS estimators are not a robust regression procedure for outliers, so their estimations are not appropriate even if there is only one outlier presence. Robust regression is an important tool for analyzing detected data as outlier data. The purpose of this research is to know the outliers to interfere with the linear regression equation, to know the result of robust regression estimator with the method of appraiser of LTS (Least Trimmed Squares), to know the result of robust regression estimator with MM (MM-Estimator) estimation method, and to know the comparison between two the robust regression estimator by looking at the values of  $R^2$  and the residuals of each method. The data used in this study from the value of the new admissions exam from the Prodi Mathematics Education at the University of PGRI Semarang. This data consists of discrete data covering 3 (three) variables namely Test value (X1), Psychological Test (X2) as independent variable and GPA (Y) as dependent variable. Prior to analysis with robust regression, detection of outliers is done to identify outlier or not. Outlier detection method is done with several, among others *boxplot* method, *Cook's Distance*, and *DfFIT (Difference In Standardized)* method.*

**Keywords:** double regression, Outlier, Estimation, LTS, MM-Estimation

## 1. PENDAHULUAN

Regresi *robust* diperkenalkan oleh Andrews (1972) dan merupakan metode regresi yang digunakan ketika distribusi *error* tidak normal dan atau adanya beberapa *outlier* yang berpengaruh pada model (Ryan, 1997). Metode ini merupakan alat penting untuk menganalisis data yang dipengaruhi oleh *outlier* (pencilan) sehingga dihasilkan model yang *robust* atau *resistance* terhadap *outlier*. (Barnett dan Lewis, 1994) menyebutkan bahwa *outlier* merupakan objek yang secara numerik berbeda dengan data lainnya. Selain itu, (Hair, et al, 1995) juga menyatakan bahwa *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. Definisi lain dari *outlier* adalah objek yang terletak jauh atau berbeda jauh dari pola distribusinya (Moore dan McCabe, 1999).

Di dalam regresi *robust* banyak metode yang bisa digunakan, seperti : *Least Median Squares* (LMS) yaitu metode penduga parameter regresi *robust* dengan meminimumkan median dari kuadrat sisaan. *Least Trimmed Squares* (LTS) yaitu metode pendugaan parameter regresi *robust* untuk meminimumkan jumlah kuadrat  $h$  residual (fungsi objektif), penaksir  $M$  ( $M$ -*Estimator*) adalah penduga parameter regresi *robust* untuk meminimumkan fungsi galat, dsb.

Dalam mengatasi permasalahan data yang mengandung pencilan, metode- metode penduga regresi *robust* tersebut memiliki kelebihan dan kelemahan masing- masing. Regresi *robust* dengan metode pendugaan parameter LTS lebih efisien dibanding LMS, karena LTS memiliki fungsi objektif yang lebih *smooth* (halus) sehingga akan lebih sensitif terhadap efek lokal dan mempunyai nilai breakdown yang paling tinggi (Yaffee, 2002). Sedangkan  $S$ -*Estimator* tidak selalu lebih baik dari LTS dan LMS, terutama untuk data yang sedikit. Akan tetapi jika data banyak, kadang  $S$ -*Estimator* lebih baik dari LTS dan LMS (Rousseeuw dan Leory, 1987). Dari uraian di atas, maka dalam penelitian ini peneliti akan membandingkan metode regresi *robust* yaitu metode *Least Trimmed Squares* dan penaksir MM dalam mengatasi permasalahan data pencilan, dengan studi kasus data penerimaan masuk di Universitas PGRI Semarang. Sehingga tujuan dari penelitian ini yakni mengetahui Pencilan (*outlier*) mengganggu persamaan regresi linier, mengetahui hasil penaksir regresi *robust* dengan metode penaksir LTS (*Least Trimmed Squares*), serta mengetahui hasil penaksir regresi *robust* dengan metode penaksir MM ( $MM$ -*Estimator*).

## 2. METODOLOGI PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah metode kajian pustaka yaitu mempelajari buku-

buku, jurnal-jurnal dan bahan-bahan literatur yang berhubungan dengan penelitian, kemudian dianalisis dan dibandingkan, dengan beberapa tahapan, yaitu menentukan masalah, studi pustaka, analisis pemecahan masalah, tahapan penelitian

Pendeteksian adanya data *outlier* menggunakan metode Cook's Distance dengan kriteria nilai  $cook's > F(0.5;k, n-k)$  dan DfFITS dengan kriteria nilai  $DfFITS > 2\sqrt{\frac{k+1}{n}}$ . Untuk mendapatkan nilai *Cook's Distance* dan *DfFITS* dapat menggunakan bantuan SPSS 21.

Pengambilan data untuk simulasi regresi *robust* diperoleh dari nilai IPK dari Mahasiswa pendidikan Matematika dipengaruhi nilai ujian masuk mahasiswa di Universitas PGRI Semarang.

## 3. HASIL PENELITIAN DAN PEMBAHASAN

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

### 3.1. Model Regresi Linear Berganda

Beberapa permasalahan regresi dapat mencakup lebih dari satu variabel bebas. Model-model regresi yang menggunakan lebih dari satu variabel bebas disebut model regresi linier berganda.

Menurut Supranto (2009) bentuk umum regresi linier berganda dapat dinyatakan secara statistik sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + \varepsilon_i$$

Dengan:

$Y_i$  = variabel terikat

$X_{ki}$  = variabel bebas

$\beta_1, \beta_2, \dots, \beta_k$  = parameter

$\varepsilon_i$  = variabel gangguan / error

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

### 3.2. Regresi Robust

Regresi robust diperkenalkan oleh oleh Andrews (1972). “ Regresi robust merupakan metode regresi yang digunakan ketika berdistribusi dari error tidak normal dan atau adanya beberapa outlier yang berpengaruh pada model ”.

Metode regresi *robust* terus berkembang dan banyak digunakan dalam meneliti berbagai permasalahan, seperti : pengoptimalan kekuatan *torque* pada lampu TL yaitu menggunakan metode penduga parameter LTS, dengan alasan terdapat pencilan pada data kekuatan *torque* (Akbar dan Maftukhah, 2007).

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

- (1) Masalah dengan outlier yang terdapat pada peubah  $y$  (*respon*);
- (2) Masalah dengan outlier yang terdapat pada peubah  $x$  (*levergae point*); dan
- (3) Masalah dengan outlier yang terdapat pada keduanya yaitu pada peubah  $y$  (respon dan peubah  $x$  (penjelas).

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

### 3.3. Estimasi –M

Wilcon (2005: 51) menjelaskan “estimasi –M pertama kali diperkenalkan oleh Huber pada tahun 1973 dan merupakan penggambaran dari suatu percobaan yang menggabungkan sifat efisiensi Ols dan ketahanan dari estimasi LAV (LAD)”. LAV merupakan estimasi yang meminimumkan jumlah nilai mutlak dari residu.

### 3.4. Least Trimmed Squares (LTS)

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

### 3.5. Estimator-S

Metode robust S merupakan metode high Breakdown Value yang diperkenalkan pertama oleh Rousseeuw pada tahun 1984. Menurut Wilcon (2005 : 55), “Estimasi-S merupakan solusi dengan kemungkinan terkecil penyebaran residual”. Selain meminimumkan variansi dari residual, Estimasi-S juga meminimumkan skala residual dari estimasi-M.

### 3.6. Estimasi –MM

Wilcox (2005:5) menjelaskan “metode estimasi–MM dikenalkan oleh Yohai (1987) yang menggabungkan suatu *high breakdown point* (50%) dengan efisiensi tinggi (mencapai 95%)”. Estimasi MM dimulai

dengan mencari estimasi S yang sangat robust dan resisten yang meminimumkan suatu skala residual. Selanjutnya skala residual tetap konstan dan di akhiri dengan menetapkan parameter-parameter regresi menggunakan estimasi-M. Estimasi-MM mempunyai *breakdown point* = 0,5, menjelaskan bahwa banyaknya outlier hingga separoh data pengamatan tidak berpengaruh terhadap estimasi-MM.

Estimasi MM didefinisikan sebagai berikut.

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

Estimasi S sebagai permulaan dengan nilai *breakdown* yang tinggi dan di akhiri dengan estimasi-M membuat estimasi mempunyai efisiensi yang tinggi. Pada umumnya di gunakan fungsi Tukey Bisquare baik pada estimasi-S maupun estimasi-M. Prosedure estimasi-MM dapat diuraikan sebagai berikut.

1. Mengestimasi koefisien  $\beta^{(1)}$ , sehingga diperoleh  $\hat{e}^{(1)}$  yang diambil dari regresi robust dengan *high breakdown point*.
2. Residual  $e_i$  pada langkah pertama digunakan untuk menghitung skala residual Estimasi-M,  $\hat{\sigma}$  dan menghitung pula bobot awal  $w_i^{(1)}$ .
3. Residual  $e_i^{(1)}$  dan skala residual  $\hat{\sigma}$  dari langkah (2) digunakan dalam iterasi awal dengan metode WLS untuk menghitung koefisien regresi.

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

Dimana  $w_i$  menggunakan pembobot Huber atau Tukey bisquare.

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

Dari kelima metode di atas, peneliti memilih dua metode robust, yaitu metode LTS dan metode robust estimasi-MM karena kedua metode salah satu metode populer digunakan, dan peneliti ingin membandingkan kedua estimasi robust tersebut. Perbandingan dari kelima metode tersebut dapat dilihat pada Tabel 1 (Wilcox, 2005:58).

**Tabel 1** Perbandingan Estimasi Regresi Robust

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

### 3.7. Regresi Robust Estimasi LTS dan Estimasi-MM untuk Permasalahan Outlier pada OLS

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

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

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

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

### 3.8. Breakdown Point Estimasi-LTS

Sama dengan metode LMS, Metode LTS juga merupakan metode *High Breakdown Value* yang diperkenalkan oleh Rousseeuw pada tahun 1984. Metode LTS adalah suatu metode estimasi parameter

regresi robust dengan untuk meminimumkan jumlah kuadrat  $h$  residual.

$$m \sum_{i=1}^h e_i^2$$

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

Keterangan :  $e_i^2$  = Kuadrat residual yang diurutkan dari terkecil terbesar.

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

$n$  = Banyaknya pengamatan

$k$  = Parameter regresi

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

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

Prosedur estimasi LTS dapat diuraikan sebagai berikut.

- (1) Menghitung estimasi parameter  $\beta$ .
- (2) Menentukan  $n$  residual  $e_i^2 = (y_i - \sum_{j=0}^k x_{ij}\beta_j)^2$  yang sesuai dengan  $\hat{\beta}$ , kemudian menghitung  $h_0 = \left[ \frac{n}{2} \right] + \left[ \frac{k+1}{2} \right]$  pengamatan dengan nilai  $e_i^2$  terkecil.
- (3) Menghitung  $\sum_{i=1}^{h_0} e_i^2$ .
- (4) Melakukan estimasi parameter  $\beta_{new}$  dari  $h$  pengamatan.
- (5) Menentukan  $n$  kuadrat residual yang baru yang bersesuaian dengan  $\hat{\beta}_{new}$  kemudian menghitung sejumlah  $h_{new}$  pengamatan dengan nilai  $e_i^2$  terkecil.
- (6) Menghitung  $\sum_{i=1}^{h_{new}} e_i^2$  dan mengulang langkah 4 sampai 6 untuk mendapatkan fungsi obyektif yang kecil dan konvergen.

### 3.9. Analisis Regresi Robus dalam studi kasus

Dalam penelitian ini mengambil simulasi pada suatu kasus dengan menggunakan data dari Tes Masuk Penerimaan Mahasiswa Baru Pendidikan Matematika UPGRIS tahun 2017 yakni dari tes tulis dan tes psikologi sebagai variabel independen, sedangkan

variabel dependennya adalah nilai IPK setelah masuk perkuliahan dengan banyaknya data  $n = 52$ . Proses analisis regresi robust dimulai dengan regresi kuadrat terkecil terlebih dahulu, kemudian pengidentifikasi outlier dan selanjutnya dengan metode regresi robust estimasi LTS dan estimasi-MM. Pengolahan data komputasi yang digunakan sebagai alat batu adalah program SPSS.

### 3.10. Metode Kuadrat Terkecil

Analisis dimulai dengan menganalisis regresi bisa menggunakan kuadrat terkecil. Berdasarkan hasil output pada data lampiran 1, diperoleh model regresi antara variabel independen dan variabel dependen data PMB Universitas PGRI Semarang tahun 2015 sebagai berikut

$$\hat{Y} = 2,9338 + 0,0062 X_1 + 0,00518 X_2$$

Dengan  $Y$  = Indeks Prestasi Kumulatif (IPK)

$X_1$  = Nilai Tes

$X_2$  = Nilai Psikologi

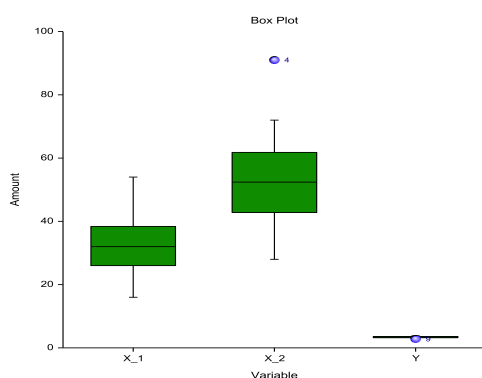
Model regresi tersebut mempunyai nilai  $R^2$  sebesar  $0,1709 = 17,09\%$ , tahap selanjutnya adalah melakukan pendeteksian outlier untuk mengetahui ada atau tidaknya outlier dalam data observasi.

### 3.11. Pendeteksian Outlier

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

### 3.12. Metode Boxplot

Metode *boxplot* ini disajikan dalam bentuk perhitungan manual maupun dengan bantuan *software* komputer. Hasil *Boxplot* dengan program SPSS dapat dilihat pada Gambar 4.1 dengan perhitungan manual dapat dilihat pada Tabel 4.2. Dengan metode *boxplot*, suatu data dikatakan outlier jika nilai data pengamatan lebih kecil dari  $Q1 - (1,5 * IQR)$  atau lebih besar dari  $Q3 + (1,5 * IQR)$ .



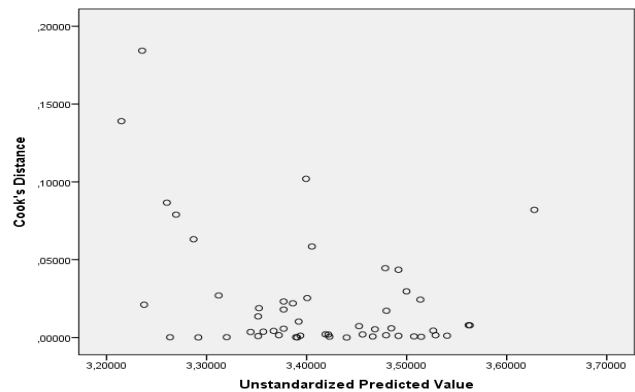
Gambar 1. Gambar Boxplot

Dari boxplot di atas dapat terlihat bahwa data ke 4

dan data ke-9 merupakan outlier.

### 3.13. Metode Cook's Distance

Hasil diagnosis metode *cook's distance* pada tiap observasi terhadap data juga dapat dilihat Tabel 1 maupun Gambar 2. Suatu observasi diduga sebagai outlier apabila nilai *Cook's distance*  $> (4/n)$ , dengan  $n$  adalah banyaknya data. Selanjutnya disajikan gambar scatter plot yang menyajikan *Cook's Distance* dan variabel *Unstandaized Predicted Value*.

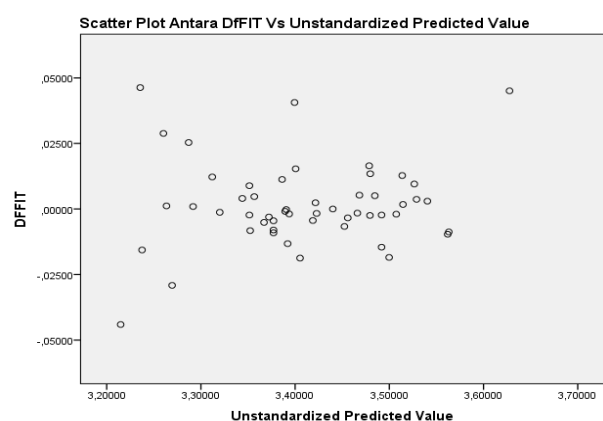


Gambar 2. Scatter Plot antara *Cook's Distance* Vs *Unstandaized Predicted Value*

Berdasarkan hasil yang diperoleh nilai *Cook's distance* pada data ke-1, 4 dan 9 diduga sebagai outlier karena memiliki nilai *Cook's distance* yang lebih besar dari nilai  $(4/n) = 0,07692$ .

### 3.14. Metode DfFITS

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



Gambar 3 Scatter Plot antara *DfFITS* vs *Unstandaized Predicted Value*

Pada plot di atas dapat dilihat bahwa ada titik yang menjauh dari titik lainnya, hal ini menunjukkan bahwa ada data yang teridentifikasi sebagai outlier. Sehingga perlu pengecekan data tersebut.

Pada tahap selanjutnya pada hasil pengolahan data menggunakan metode DfFIS untuk masing-masing data. Dengan ketentuan jika nilai DfFIT masing-masing data yang lebih dari  $2\sqrt{\frac{k+1}{n}}$  maka dikategorikan sebagai outlier. Batas nilai penentuan berdasarkan DfFIT > 0,4803 merupakan data outlier. Dari data terlihat bahwa data yang mempunyai nilai DfFIT > 0,4803, data yang mendekati outlier data ke-4. Selanjutnya dilakukan analisis regresi menggunakan metode robust untuk data yang mengandung outlier, agar hasil regresi yang dihasilkan lebih tepat.

### 3.15. Analisis Regresi Robus dengan metode *robust Least Trimmed Square (LTS)* dan metode *MM-Estimation*.

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

### 3.16. Regresi Robust *Estimasi Least Trimmed Square*

Penerapan metode *Least Trimmed Square (LTS)* memerlukan beberapa iterasi untuk mendapatkan model terbaik. Pada iterasi 1 diperoleh persamaan model

$$Y = 2,9338 + 0,0062 X_1 + 0,00518 X_2$$

Karena  $h_0 = \left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{k+1}{2} \right\rfloor = 26 + 2 = 28$ , maka pada iterasi selanjutnya digunakan data sebanyak 28 dengan mengambil  $\varepsilon_i^2$  yang terkecil. Estimasi dilakukan dengan menggunakan bantuan *software Microsoft Excel* dan SPSS, dapat dilihat pada lampiran 4. Dalam literasi dengan 28 data, diperoleh persamaan model

$$\hat{Y} = 3,015 + 0,007 X_1 + 0,003 X_2 \text{ dan } \sum_{i=0}^{h_0} \varepsilon_i^2 = 0,127,$$

Penaksiran Robust dengan menggunakan metode LTS dapat dilihat pada Tabel 2

**Tabel 2.** Hasil iterasi *Least Trimmed Square (LTS)*

Tahap	N	H	$\hat{\beta}_0^{(n)}$	$\hat{\beta}_1^{(n)}$	$\hat{\beta}_2^{(n)}$	$\sum_{i=0}^{h_0} \varepsilon_i^2$
1	52	28	2,9338	0,0062	0,005	2,220
2	28	16	3,015	0,007	0,003	0,127

### 3.17. Regresi Robust Estimasi-MM

Pada metode *MM-Estimator*, peneliti menggunakan

pembobot tukey, maka  $c = 4,685$ . Karena metode *MM-Estimation* merupakan gabungan dari metode *M-Estimation* dan *S-Estimation*, maka untuk menyelesaikan langkah pertama yaitu mencari estimator S, kemudian menetapkan parameter-parameter regresi menggunakan metode *M-estimation*.

Pada interasi 1, diperoleh parameter dari *S-Estimation* seperti pada Tabel 3.

**Tabel 3.** Parameter *S-Estimator*

Parameter	Nilai
$\hat{\beta}_0^{(n)}$	2,934
$\hat{\beta}_1^{(n)}$	0,006
$\hat{\beta}_2^{(n)}$	0,005

Dari perhitungan parameter dari *S-estimator* tersebut di gunakan mencari nilai residual  $\varepsilon_i^{(1)}$  selanjutnya digunakan untuk memperoleh nilai pembobot  $w_i^{(1)}$ . Besarnya  $\varepsilon_i^{(1)}$  pada interasi 2 digunakan sebagai nilai WLS pada iterasi 3, untuk mendapatkan persamaan regresi robusnya.

Dari tabel perhitungan di peroleh persamaan

$$\hat{y} = 2,932 + 0,006 X_1 + 0,005 X_2. \text{ Dengan persamaan tersebut diperoleh nilai } \sum_{i=0}^n \varepsilon_i^{(1)} = 0,78904. \text{ Selanjutnya lakukan interasi dengan } w_i^{(2)} \text{ sebagai WLS dan diperoleh output.}$$

Hasil perhitungan dengan SPSS diperoleh persamaan yang sama dengan interasi sebelumnya, ini berarti iterasinya cukup sebanyak dua kali. Dengan persamaan yang dipakai adalah persamaan regresi yang terakhir, yaitu  $\hat{y} = 2,932 + 0,006 X_1 + 0,005 X_2$  dan  $\sum_{i=0}^n \varepsilon_i^{(1)} = 0,89304$ .

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

**Tabel 4.** Hasil iterasi *MM-Estimation*

Tahap	N	$\hat{\beta}_0^{(n)}$	$\hat{\beta}_1^{(n)}$	$\hat{\beta}_2^{(n)}$	$\sum_{i=0}^{h_0} \varepsilon_i^2$
1	52	2,934	0,0062	0,00518	0,7890
2	52	2,932	0,006	0,005	0,78904
3	52	2,932	0,006	0,005	0,89304

Berdasarkan Tabel 4 jelas terlihat  $\sum_{i=0}^n \varepsilon_i^{(m)}$  sudah konvergen. Sehingga persamaan yang paling terakhir yang paling baik yang diperoleh menggunakan metode *MM-Estimation* adalah  $\hat{y} = 2,932 + 0,006 X_1 + 0,005 X_2$

### 3.18. Pemilihan Model Regresi Terbaik

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

**Tabel 5** Perbandingan Nilai  $R^2$  dari metode LTS dan metode MM-Estimation

N	Motode Regresi	$\hat{\beta}_0^{(n)}$	$\hat{\beta}_1^{(n)}$	$\hat{\beta}_2^{(n)}$	$\sum_{i=0}^{h_0} \varepsilon_i^{(n)}$	$R^2$
1	LTS	3,01 5	0,00 7	0,00 3	0,127	0,56 0
2	MM-Estimatio n	2,93 2	0,00 6	0,00 5	0,89304	0,17 2

Dari Tabel 5 dapat dilihat bahwa pada model regresi pada model *Least Trimmed Square* nilai  $R^2$  nya adalah 0,560. Sedangkan pada metode *MM-Estimation* nilai  $R^2$  nya adalah 0,172. Ini berarti model regresi pada model *Least Trimmed Square* memberikan pengaruh yang lebih besar yaitu sebesar 56,0% dibandingkan dengan metode *MM-Estimation* yang hanya memberikan pengaruh sebanyak 17,2,. Dilihat dari nilai residualnya model *Least Trimmed Square* lebih kecil sebesar 0,127 sedangkan *MM-Estimation* nilai residualnya sebesar 0,89304. Dengan demikian, model *Least Trimmed Square* (LTS) merupakan metode terbaik untuk mengestimasi parameter pada saat data terdeteksi mengandung outlier karena  $R^2$  yang lebih banyak dan residual terkecil.

### 3.19. Pembahasan

Berdasarkan hasil penelitian, diketahui bahwa data yang diperoleh dari nilai Ujian PMB UPGRIS dengan nilai IPK Mahasiswa Universitas PGRI Semarang dari Prodi Pendidikan Matematika, merupakan data diskrit yang meliputi 3 (tiga) variabel yaitu Nilai Tes (X1), Tes Psikologi (X2) dan IPK (Y).

Setelah memenuhi semua asumsi dalam regresi ganda, dilakukan pengecekan outlier didata tersebut. yakni dilakukan dengan berbagai macam metode. Beberapa metode diantaranya sebagai berikut:

Metode *boxplot*, metode *Cook's Distance*, metode DfFIT untuk masing-masing data seperti pada lampiran. Dengan ketentuan jika nilai DfFIT masing-masing data yang lebih dari  $2\sqrt{\frac{k+1}{n}}$  maka dikategorikan sebagai outlier. Batas nilai penentuan berdasarkan DfFIT > 0,4803 merupakan data outlier. Terlihat data yang mempunyai nilai DfFIT > 0,4803, data yang mendekati outlier data ke-4.

Selanjutnya dilakukan analisis regresi menggunakan metode robust untuk data yang mengandung outlier, agar hasil regresi yang dihasilkan lebih tepat. Langkah selanjutnya adalah melakukan analisis regresi untuk

mendapatkan nilai estimasi parameter dari data tersebut menggunakan metode *robust Least Trimmed Square* (LTS) dan metode MM-Estimation. Selanjutnya dapat dibandingkan metode mana yang lebih efektif digunakan untuk menyelesaikan masalah regresi, perbandingan metode dengan melihat  $R^2$  dan residunya. Pada metode yang pertama dalam regresi robust *Least Trimmed square* (LTS) dihasilkan model regresi  $\hat{y} = 3,015 + 0,007 X_1 + 0,003 X_2$  dan  $\sum_{i=0}^{h_0} \varepsilon_i^2 = 0,127$ . Persamaan tersebut diperoleh dari beberapa iterasi, yakni terjadi pada iterasi yang ke-2. Hal ini terjadi karena pada iterasi ke-3, data outlier tidak termasuk didalamnya, hal ini tidak sesuai dengan konsep regresi *robust* yaitu tetap mengikuti sertakan data outlier dalam menemukan model persamaan regresi. Pada model *Least Trimmed Square* (LTS) juga terjadi pemangkasan sejumlah data sebesar h, dimana nilai h didapat dari rumusan  $h_0 = \left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{k+1}{2} \right\rfloor$ . Inilah yang menyebabkan nilai jumlah kuadrat residualnya pada metode ini semakin kecil dari itersi 1 sampai iterasi terakhir. Nilai  $R^2$  yang didapatkan dalam metode ini adalah 0,560. Hal ini menunjukkan bahwa variabel independent akan memberikan pengaruh yang cukup besar terhadap variabel dependent.

Selanjutnya dalam perhitungan regresi yang menggunakan metode *MM-Estimation* diperoleh hasil model  $\hat{y} = 2,932 + 0,006 X_1 + 0,005 X_2$ .

Pada model ini yang menggunakan metode *MM-Estimation* juga mengalami 2 iterasi untuk sampai pada model regresi terbaik. Nilai residual yang didapat dari metode ini didapat yakni  $\sum_{i=0}^{h_0} \varepsilon_i^2 = 0,89304$ . Sedangkan nilai  $R^2$  yang didapat dalam metode ini adalah 0,172. Ini menunjukkan bahwa model regresi yang diperoleh tadi lemah atau kurang baik digunakan untuk memprediksi.

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

**Tabel 6** Nilai residual metode LTS dan metode MM-Estimation

No	Motode Regresi	$\sum_{i=0}^{h_0}  \varepsilon_i $	$R^2$
1	LTS	1,58062	0,560
2	MM-Estimation	8,403	0,172

Dari Tabel 6 terlihat bahwa metode *Least Trimmed Square* (LTS) mempunyai nilai residual yang lebih kecil, hal ini disebabkan adanya pemangkasan (*trimmed*) data. Jadi, sama halnya dengan nilai  $R^2$ , nilai metode Least Trimmed Square (LTS) juga lebih baik jika dibandingkan dengan metode MM-Estimation.

#### 4. SIMPULAN

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan sebagai berikut.

1. Sebelum dilakukan analisis dengan regresi robust, dilakukan pendeteksian *outlier* untuk mengidentifikasi adanya outlier atau tidak. Metode pendeteksian outlier dilakukan dengan beberapa, antara lain metode *boxplot*, *Cook's Distance*, dan metode DfFIT (*Difference In fit Standardized*). Berdasarkan hasil perhitungan diperoleh nilai *Cook's distance* pada data ke-1, 4 dan 9 diduga sebagai *outlier* karena memiliki nilai *Cook's distance* yang lebih besar dari nilai  $(4/n) = 0,07692$ . Sedangkan untuk nilai penentuan berdasarkan  $DfFIT > 0,4803$ , data yang mendekati outlier data ke-4. Sehingga data terlihat terdapat data outlier, maka dapat digunakan regresi robust.
2. Pada metode yang pertama dalam regresi robust *Least Trimmed square* (LTS) dihasilkan model regresi  $\hat{y} = 3,015 + 0,007 X_1 + 0,003 X_2$  dan  $\sum_{i=0}^{h_0} \varepsilon_i^2 = 0,127$ .
3. Untuk persamaan regresi robust dengan metode MM-Estimation diperoleh persamaan, yaitu  $\hat{y} = 2,932 + 0,006 X_1 + 0,005 X_2$  dan  $\sum_{i=0}^n \varepsilon_i^{(1)} = 0,89304$ .
4. Regresi robust merupakan metode yang sesuai untuk pendugaan parameter Penduga *Least Trimmed Square* (LTS) lebih baik daripada Penduga MM. Hal ini didasarkan pada kriteria nilai  $R^2$  dan residualnya, hal ini disebabkan adanya pemangkasan (*trimmed*) terhadap data yang mempunyai residual besar.

#### DAFTAR PUSTAKA

- Chen, C., (2002), Robust Regression and Outlier Detection with the ROBUSTREG Procedure, *SAS Institute Inc. Cary NC Paper*, 265(27), hal 1-13.
- Draper, N.R dan Smith, H., (1992), *Analisis Regresi Terapan*, Diterjemahkan oleh Bambang Sumantri, Gramedia, Jakarta.
- Fox, J. 2002. Robust Regression. *Appendix to An R and S-PLUS Companion to Applied Regression*.
- Hogg, R.V,dkk. 2005. *Introduction to Mathematical Statistics*, USA: Pearson Prentice Hall.
- Hubert, M, Rousseeuw, P.J, Aelst, S.V. 2008. High-Breakdown Robust Multivariate Methods *Statistical Science*. Vol.23.No.1.92-119.
- Iswarini, T.H., (2011), Perbandingan Robust Least Median Square (LMS) dan Robust-MM sebagai Metode Pendugaan Parameter pada Regresi Robust Linier Berganda, *Skripsi*, Universitas Brawijaya Malang. Indonesia.
- Rousseeuw, P. J and Yohai, V. J., (1984), *Robust Regression by Means of S Estimator in Robust and Nonlinear Time Series Analysis*, edited by J. Franke, W, Hardle, and R.D. Martin, Lecture Notes in Statistics 26, Springer Verlag, New York, hal. 256-272.
- Ryan, T.P., (1997), *Modern Regression Models*, John Wiley & Sons, New York.
- Yohai, V. J., (1987), High Breakdown Point and High Efficiency Robust Estimates for Regression, *Annals of Statistics*, 15(20), hal 642-656.
- Soemartini. 2007. Pencilan (*Outlier*). Universitas Padjadjaran, Bandung. Melalui [http://resources.unpad.ac.id/unpad-content/uploads/publikasi\\_dosen/OUTLIER\(PENCILAN\).pdf](http://resources.unpad.ac.id/unpad-content/uploads/publikasi_dosen/OUTLIER(PENCILAN).pdf)
- Subali, S.B.W. 2004. Estimasi Robust dengan Least Median Squares and Least Trimmed Squares Pada Model Regresi Linier Parametrik. *Tesis*. 2004. ITS. Surabaya.