

MENGATASI PENCILAN PADA PEMODELAN REGRESI LINEAR BERGANDA DENGAN METODE REGRESI ROBUST PENAKSIR LMS

Solve the Outlier in Multiplication Linear Regression Models with Robust's Regression Method Least Median of Squares (LMS) Estimator

Farida Daniel

*Prodi Pendidikan Matematika, STKIP Soe
Jln. Badak No. 5a, Soe, 85511, Nusa Tenggara Timur, Indonesia*

e-mail: staffaridaniel@stkipsoe.ac.id

Abstrak

Metode Kuadrat Terkecil (OLS) merupakan metode yang sering digunakan untuk menaksir parameter model regresi. Penaksir OLS bukan merupakan prosedur regresi yang robust terhadap adanya pencilan sehingga estimasinya menjadi tidak sesuai. Median Kuadrat Terkecil (LMS) merupakan salah satu penaksir yang robust terhadap adanya pencilan dan memiliki *breakdown value* yang tinggi. LMS menaksir parameter model dengan meminimumkan median kuadrat galat. $LMS = \min_b \text{median}_i |y_i - x_i b|^2$.

Penelitian ini bertujuan untuk mendapatkan suatu persamaan regresi yang lebih baik daripada persamaan regresi yang sebelumnya menggunakan OLS untuk data yang mengandung pencilan. Terlebih dahulu dilakukan pendeteksian keberadaan pencilan dan kemudian mencari persamaan regresi dengan metode LMS. Penelitian ini menggunakan data sekunder berupa data *stackloss* dimana hasil estimasi parameter pada data ini, penaksir LMS menunjukkan hasil yang lebih baik jika dibandingkan dengan penaksir OLS karena persamaan regresi yang dihasilkan mempunyai nilai Rataan Persentase Galat Mutlak (MAPE) yang lebih kecil.

Kata Kunci: *Pencilan, LMS, regresi robust.*

Abstract

*Ordinary Least Squares (OLS) is frequent used method for estimating parameters. OLS estimator is not a robust regression procedure for the presence of outliers, so the estimate becomes inappropriate. Least Median of Squares (LMS) is one of a robust estimator for the presence of outliers and has a high breakdown value. LMS estimate parameters by minimizing the median of squared residuals. Least Median of Squares (LMS) = $\min_b \text{median}_i |y_i - x_i b|^2$. The purpose of this study is getting a regression equation that better than the regression equation before using OLS for the data that having outlier. For the first step, checking if there is outlier at data and then searching regression equation with LMS method. In this study used data *stackloss* and from estimation parameter of this data, LMS estimator showed better results compared to the OLS estimator because the regression equation from LMS method have smaller value of Mean Absolute Percentage Error (MAPE).*

Keywords: *Outlier, LMS, Robust Regression.*

Submitted: 23rd March 2019

Revised: 02nd April 2019

Accepted: 19th July 2019

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



1. PENDAHULUAN

Metode Statistika yang digunakan untuk mengetahui hubungan linear antara variabel terikat (dependen/respon/ y) dengan satu atau lebih variabel bebas (independen/prediktor/ x) disebut regresi linear [11]. Apabila banyaknya variabel bebas hanya ada satu maka disebut regresi linear sederhana, sedangkan regresi linear berganda adalah regresi yang meramalkan hubungan antara satu variabel tak bebas dengan dua atau lebih variabel bebas. Hubungan tersebut dapat dirumuskan dalam bentuk persamaan:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

dengan y adalah variabel tak bebas; x adalah variabel bebas; β_0 adalah intersep atau titik potong antara sumbu tegak y dan garis fungsi linear; $\beta_1, \beta_2, \dots, \beta_k$ adalah koefisien-koefisien regresi atau koefisien kemiringan; ε_i adalah faktor galat dan i adalah pengamatan ke- i .

Koefisien-koefisien regresi dapat ditaksir menggunakan metode *Ordinary Least Squares* (OLS) atau Metode Kuadrat Terkecil/MKT [5]. Metode ini ditemukan oleh Gauss dan Legendre sejak tahun 1800 dengan prinsip meminimumkan jumlah kuadrat residualnya. Pada metode OLS koefisien-koefisien regresi ditaksir dengan meminimumkan jumlah kuadrat galat $\sum_{i=1}^n \varepsilon_i^2$. Taksiran untuk β diperoleh dengan persamaan:

$$\hat{\beta} = [X^T X]^{-1} X^T Y \quad (2)$$

Penggunaan OLS memerlukan beberapa asumsi klasik yang harus dipenuhi oleh komponen sisaan atau galat (ε_i) dalam model yang dihasilkan. Beberapa asumsi itu antara lain bahwa galat harus memenuhi asumsi normalitas, kehomogenan ragam dan tidak terjadi autokorelasi. Apabila asumsi itu terpenuhi, maka penduga parameter yang diperoleh bersifat *Best Linier Unbiased Estimator*/BLUE atau penduga terbaik yang bersifat linear dan tak bias [3].

Seringkali dalam berbagai kasus ditemui hal-hal yang menyebabkan tidak terpenuhinya asumsi klasik tersebut. Data yang diperoleh tidak jarang ditemukan satu atau beberapa yang jauh dari pola kumpulan data keseluruhan yang lazim didefinisikan sebagai pencilan (*outlier*). Pencilan dapat dilihat sebagai pengamatan dengan sisaan yang cukup besar [1]. Pencilan adalah pengamatan yang jauh dari kelompok data yang mungkin berpengaruh besar terhadap koefisien regresi [8]. Adanya Pencilan dapat disebabkan oleh beberapa hal diantaranya adalah kesalahan input data, kekeliruan pada sistem pengukuran ataupun karena terjadinya peristiwa yang luar biasa seperti krisis maupun bencana.

Soemartini [8] mengemukakan bahwa keberadaan pencilan dapat dideteksi dengan metode sebagai berikut:

1) Metode grafis (*scatter plot*).

Untuk melihat apakah terdapat pencilan pada data, dapat dilakukan dengan membuat plot antara data dengan observasi ke- i ($i = 1, 2, 3, \dots, n$).

2) *Box Plot*.

Metode ini merupakan yang paling umum yakni dengan mempergunakan nilai kuartil dan jangkauan. Kuartil 1, 2, dan 3 akan membagi sebuah urutan data menjadi empat bagian. Jangkauan (*IQR*, *Interquartile Range*) didefinisikan sebagai selisih kuartil 3 terhadap kuartil 1, atau $IQR = Q_3 - Q_1$. Data-data pencilan dapat ditentukan yaitu nilai yang kurang dari $1.5 * IQR$ terhadap kuartil 1 dan nilai yang lebih dari $1.5 * IQR$ terhadap kuartil 3.

Nilai $DFFITs_i$ dan *cook distance* dapat digunakan untuk mengidentifikasi apakah suatu pengamatan berpengaruh atau tidak.

a) $DFFITs_i$

Merupakan suatu ukuran berpengaruh yang ditimbulkan oleh pengamatan ke- i terhadap nilai taksiran \hat{y}

$$DFFITs_i = \frac{\hat{y}_i - \hat{y}_{i-i}}{s_{-i} \sqrt{h_{ii}}} \quad (3)$$

Dimana: \hat{y}_i nilai taksiran uji, $\hat{y}_{i,-i}$ nilai taksiran uji tanpa pengamatan ke- i , s_{-i} taksiran galat baku (standar error) tanpa pengamatan ke- i , h_{ii} unsur ke- i dari diagonal matriks H Suatu pengamatan ke- i akan berpengaruh pada persamaan regresi apabila nilai: $|DFFITs_i| > 1$ untuk $n \leq 30$ dan $|DFFITs_i| > 2\left(\frac{p}{n}\right)^{1/2}$ untuk $n > 30$

Dengan p menyatakan banyaknya parameter termasuk intersep dan n menyatakan banyaknya pengamatan.

b) Cook's Distance

Merupakan suatu ukuran pengaruh pengamatan ke- i terhadap semua koefisien regresi taksiran. Pada *Cook's Distance* pengaruh pengamatan ke- i diukur oleh jarak D , Jarak tersebut diperoleh dari persamaan berikut:

$$D_i = \frac{(b - b_{-i})^T (X^T Y)(b - b_{-i})}{ps^2} \quad (4)$$

$$D_i = \left(\frac{e_i^2}{ps^2} \right) \left(\frac{h_{ii}}{(1 - h_{ii})^2} \right)$$

Dengan: b vektor taksiran koefisien regresi termasuk pengamatan ke- i , b_{-i} vektor taksiran koefisien regresi tanpa pengamatan ke- i , e_i nilai residu pada pengamatan ke- i , h_{ii} unsur ke- i dari diagonal matrik H , p banyaknya parameter termasuk intersep dalam model, dan n banyaknya pengamatan. s^2 diperoleh dari persamaan :

$$s^2 = \frac{\sum_{i=1}^n (e_i)^2}{n - p} \quad (5)$$

Suatu pengamatan ke- i akan berpengaruh pada persamaan regresi apabila $D_i > F_{(\alpha, p, n-p)}; \alpha = 0,05$.

Berbagai kaidah telah diajukan untuk menolak pencilan (dengan kata lain untuk memutuskan menyisihkan amatan tersebut dari data, kemudian menganalisis kembali tanpa amatan tersebut). Penolakan begitu saja pada suatu pencilan bukanlah prosedur yang bijaksana. Adakalanya pencilan memberikan informasi yang tidak bisa diberikan oleh titik data lainnya seperti pencilan timbul karena adanya kombinasi keadaan yang tidak biasa dan mungkin saja sangat penting sehingga perlu diselidiki lebih jauh. Sebagai kaidah umum, pencilan baru akan ditolak jika setelah ditelusuri ternyata merupakan akibat dari kesalahan-kesalahan seperti kesalahan mencatat amatan bersangkutan atau kesalahan ketika menyiapkan peralatan. Bila ternyata bukan akibat dari kesalahan-kesalahan semacam itu, penyelidikan yang seksama harus dilakukan [2].

Identifikasi pencilan dalam data amatan dan melihat bagaimana peranannya terhadap taksiran model merupakan tahapan diagnosis yang perlu ditempuh terutama bila penaksiran modelnya dilakukan dengan OLS. Prosedur analisis yang diharapkan adalah menghasilkan keluaran yang cukup baik meskipun beberapa asumsinya tidak terpenuhi secara sempurna. Metode lain yang dapat digunakan untuk mengatasi pencilan adalah regresi robust [1].

Regresi robust diperkenalkan oleh Andrews pada tahun 1972 dan merupakan metode regresi yang digunakan ketika distribusi dari galat tidak normal dan atau adanya beberapa pencilan yang berpengaruh pada model [7]. Metode ini merupakan alat penting untuk menganalisa data yang dipengaruhi oleh pencilan sehingga dihasilkan model yang *robust* atau *kekar* atau *resistance* terhadap pencilan. Dalam regresi robust terdapat beberapa metode yang dapat digunakan untuk menangani data pencilan yaitu penaksir *Maximum Likelihood* (M), *Least Trimmed Square* (LTS), *Scale* (S), *Method of Moment* (MM) serta *Least Median of Squares* (LMS) atau metode Kuadrat Median Terkecil. Metode LMS merupakan salah satu penaksir regresi robust dengan *breakdown point* yang tinggi. *Breakdown point* adalah ukuran kekekaran suatu estimator atau proporsi minimal dari banyaknya pencilan dibandingkan seluruh data pengamatan [4]. Algoritma LMS meminimalkan median (nilai tengah) dari kuadrat residu terurut.

$$\text{Least Median of Squares (LMS)} = \min_b \text{median}_i |y_i - x_i b|^2 \quad (6)$$

Misalkan diberikan sebuah gugus data sampel berukuran N , dan ingin diduga vektor β berdimensi p yang berisi parameter dari gugus data tersebut [12]. Akan diambil berulang kali secara acak M buah subset berukuran n dari sampel berukuran N . Kemudian dicari dugaan parameter $\hat{\beta}_j$ untuk setiap subset. Cari median dari kuadrat galat e_{ij}^2 dari setiap subset. Indeks i adalah indeks untuk sampel, $i = 1, 2, 3, \dots, n$ dan indeks j adalah untuk subset, $j = 1, 2, 3, \dots, M$. Definisikan:

$$m = \arg \min_j \text{med}_i (e_{ij}^2) \quad (7)$$

sehingga solusi LMS adalah $\hat{\beta}_m$. Jumlah maksimum subset yang dapat dipilih adalah C_p^N untuk mendapatkan solusi optimal. Ini adalah komputasi yang *infeasible* karena akan memakan waktu lama jika ukuran N dan p besar. Dalam kasus tersebut M dapat dipilih sedemikian rupa sehingga kemungkinan (probabilitas) bahwa setidaknya satu dari M subset terdiri dari p pengamatan yang baik adalah mendekati 1. Probabilitas bahwa setidaknya satu dari M subset terdiri dari n pengamatan yang baik tersebut diberikan oleh:

$$P = 1 - [1 - (1 - \varepsilon)^p]^M \quad (8)$$

dimana ε adalah bagian dari pencilan (*outlier*) yang mungkin ada dalam data.

Karena efisiensi relatif LMS kecil pada *Gaussian Noise* maka sebuah langkah tunggal dari algoritma WLS digabungkan berdasarkan pada pendugaan LMS. Ukuran sebaran dari galat dapat ditaksir dengan cara menentukan terlebih dahulu nilai awal:

$$s_0 = 1,4826(1 + 5/(n - p)) \sqrt{\text{median}_i e_i^2} \quad (9)$$

Faktor $1,4826 = \frac{1}{\Phi^{-1}(0,75)}$ diusulkan karena $\frac{\text{median}_i |z_i|}{\Phi^{-1}(0,75)}$ merupakan penaksir konsisten untuk σ jika z_i

berdistribusi $N(0, \sigma^2)$ atau menyatakan estimasi yang konsisten dari s_0 pada *Gaussian Noise* dan $5/(n - p)$ menyatakan koreksi sampel yang terbatas untuk meningkatkan penaksiran ketika ukuran sampel kecil.

Selanjutnya nilai awal s_0 digunakan untuk menentukan pembobot w_i untuk setiap pengamatan, yaitu

$$w_i = \begin{cases} 1 & ; \text{ jika } |e_i/s_0| \leq 2,5 \\ 0 & ; \text{ jika } |e_i/s_0| > 2,5 \end{cases} \quad (10)$$

Berdasarkan pembobot awal w_i nilai akhir taksiran robust dihitung berdasarkan:

$$\hat{\sigma} = \sqrt{\left(\sum_{i=1}^n w_i e_i^2 \right) / \left(\sum_{i=1}^n w_i - p \right)} \quad (11)$$

Bobot akhir dihitung dengan menggunakan persamaan: $w_i = \begin{cases} 1 & ; \text{ jika } |e_i/\hat{\sigma}| \leq 2,5 \\ 0 & ; \text{ jika } |e_i/\hat{\sigma}| > 2,5 \end{cases} \quad (12)$

Nilai $\hat{\beta}$ diperoleh dengan menggunakan metode kuadrat terkecil terboboti (WLS). Nilai akhir parameter dinyatakan dalam:

$$\hat{\beta} = (X^T W X)^{-1} X^T W Y \quad (13)$$

dimana:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, W = \begin{bmatrix} w_1 & 0 & 0 & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix}$$

Untuk membandingkan tingkat akurasi penduga antar model regresi digunakan rata-rata persentase galat mutlak atau *Mean Absolute Percentage Error* (MAPE). MAPE didefinisikan oleh:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (14)$$

Dengan y_i adalah nilai aktual dan \hat{y}_i adalah nilai pendugaan. Rentang norma MAPE adalah $[0, 100]$. Semakin kecil nilai MAPE, model dinilai semakin baik. Penelitian ini bertujuan untuk mendapatkan suatu persamaan regresi yang lebih baik daripada persamaan regresi yang sebelumnya menggunakan OLS untuk data yang mengandung pencilan.

2. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah kajian pustaka. Kajian mengenai penggunaan metode regresi robust penaksir LMS dalam mengatasi pencilan pada pemodelan regresi linear berganda ini bersifat penelitian murni atau penelitian dasar, yaitu pencarian terhadap sesuatu karena ada perhatian dan keingintahuan terhadap hasil suatu aktifitas atau masalah. Data yang digunakan dalam penelitian ini adalah data sekunder yaitu data *stackloss*. Tahapan penelitian ini adalah pendeteksian pencilan yang berpengaruh pada data amatan berdasarkan kajian teori yang ada kemudian dilanjutkan dengan mencari model regresi linear dengan metode penaksir LMS serta melihat ketepatan metode LMS dibandingkan dengan metode OLS pada data yang terkontaminasi pencilan menggunakan nilai MAPE dari model regresi yang dihasilkan oleh kedua metode tersebut. Penelitian menggunakan bantuan program Minitab dan Microsoft Excel dalam memudahkan pengujian maupun perhitungan sehingga dapat diperoleh kesimpulan yang akurat.

3. HASIL DAN PEMBAHASAN

Data yang diambil adalah data sekunder yaitu data *stackloss* atau data pertumbuhan oksidasi amonia yang dilihat berdasarkan jumlah amonia yang hilang akibat pengaruh beberapa variabel seperti pada Tabel 1. Diketahui data tiga variabel yang diteliti terhadap responden untuk mengukur pertumbuhan oksidasi amonia ke nitrat acid pada tanaman selama 21 hari yaitu x_1 : pergerakan udara ke tanaman (*air flow to the plant*), x_2 : kadar temperatur air (*cooling water inlet temperature*), x_3 : konsentrasi asam (*acid concentration*) dan y_i : amonia yang hilang/ml (*the permillage of ammonia lost/stackloss*).

Tabel 1. Data *Stackloss*

Obs	x_1	x_2	x_3	y	Obs	x_1	x_2	x_3	y
1	80	27	89	42	12	58	17	88	13
2	80	27	88	37	13	58	18	82	11
3	75	25	90	37	14	58	19	93	12
4	62	24	87	28	15	50	18	89	8
5	62	22	87	18	16	50	18	86	7
6	62	23	87	18	17	50	19	72	8
7	62	24	93	19	18	50	19	79	8
8	62	24	93	20	19	50	20	80	9
9	58	23	87	15	20	56	20	82	15
10	58	18	80	14	21	70	20	91	15
11	58	18	89	14					

Sumber Data: Peter J. Rousseeuw, Annick M. Leroy, 1987, *Robust Regression And Outlier Detection*, Canada, John Willey & Sons [6]

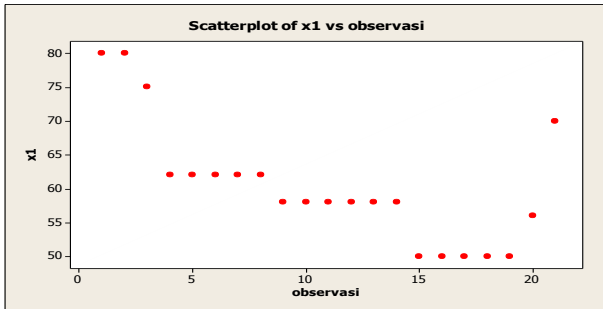
Terlebih dahulu data diuji apakah berdistribusi normal atau tidak dengan uji *Kolmogorov-Smirnov* dengan hipotesis sebagai berikut:

- i. H_0 : Data berdistribusi normal
- ii. H_1 : Data berdistribusi tidak normal

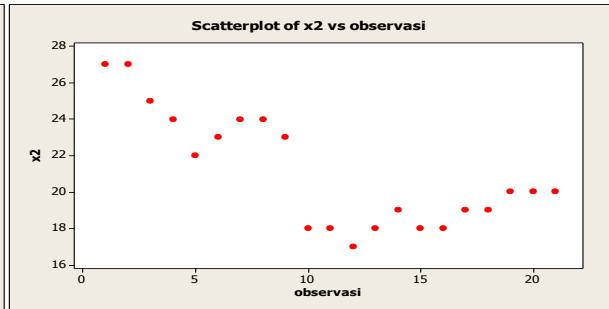
dengan *software* Minitab 16 diperoleh P value = 0,011 < 0,05 sehingga keputusan tolak H_0 .

Jadi data tidak berdistribusi normal sehingga diduga terdapat pencilan. Selanjutnya dilakukan pendeteksian keberadaan pencilan dengan metode grafis dan *box plot*.

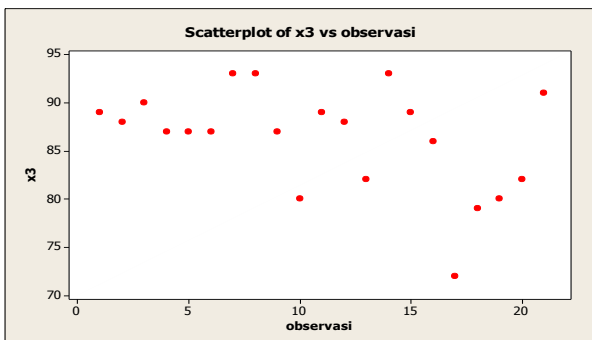
Dari *scatter plot* yang ada dapat dilihat bahwa pada data x_1 terdapat data yang agak jauh dari sebaran yakni data ke 1 dan 2. Pada x_2 dan x_3 sebaran data tidak ada yang menyimpang, sedangkan pada y data ke 1,2,3 dan 4 agak menjauh dari sebaran.



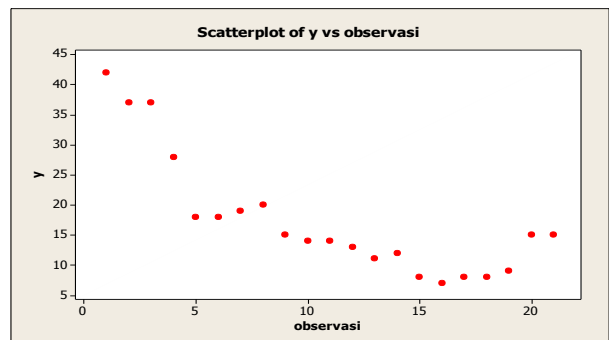
Gambar 1. Scatter Plot dari x_1 vs Observasi



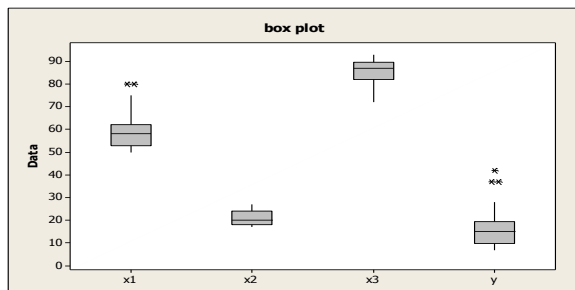
Gambar 2. Scatter Plot dari x_2 vs Observasi



Gambar 3. Scatter Plot dari x_3 vs Observasi



Gambar 4. Scatter Plot dari y vs Observasi



Gambar 5. Boxplot Data Stackloss

Tabel 2. Kuartil Data Stackloss

	x_1	x_2	x_3	y
Q ₁	56	18	82	11
Q ₂	58	20	87	15
Q ₃	62	24	89	19
IQR	6	6	7	8
1,5*IQR	9	9	10,5	12

Gambar 5, menunjukkan nilai kuartil untuk masing-masing variabel dan juga jangkauan untuk data *stackloss* yang juga disajikan dalam Tabel 2. Tabel 2 menunjukkan bahwa pada data x_1 terdapat dua data yang agak jauh dari sebaran, pada x_2 dan x_3 sebaran data tidak ada yang menyimpang, sedangkan pada y ada dua yang agak menjauh dari sebaran. Data-data tersebut dianggap sebagai pencilan. Selanjutnya dilakukan pendeteksian apakah pengamatan pencilan tersebut berpengaruh atau tidak:

$$Cook's Distance = F(0.5; p, n - p) = F(0.5; 5, 16) = 2,85$$

Dengan Minitab 16 diperoleh hasil seperti pada Tabel 3.

Tabel 3. Nilai Cook's Distance dan Dffits Data Stackloss

Observasi	COOK1	DFIT1	Observasi	COOK1	DFIT1
1	0,153710	0,794720	12	0,065066	0,509180
2	0,059683	-0,481320	13	0,010765	-0,202690
3	0,126414	0,744160	14	0,000020	-0,008630
4	0,130542	0,787880	15	0,038516	0,388340
5	0,004048	-0,124520	16	0,003379	0,113090
6	0,019565	-0,279160	17	0,065473	-0,502020
7	0,048802	-0,437670	18	0,001122	-0,065030
8	0,016502	-0,250990	19	0,002179	-0,090680
9	0,044556	-0,423400	20	0,004492	0,130830
10	0,011930	0,213120	21	0,692000	-2,100300
11	0,035866	0,376210			

Untuk nilai DFFITS karena $n < 30$ maka pengamatan akan berpengaruh jika $|DFFITS_i| > 1$. Tabel 3 menunjukkan bahwa hanya ada 1 pengamatan yang memiliki nilai $|DFFITS_i|$ melebihi nilai kritisnya (1) yaitu pengamatan ke 21. Hal ini berarti bila pengamatan 21 dikeluarkan dari kumpulan datanya maka akan berpengaruh pada nilai taksiran \hat{y}_i sedangkan untuk nilai Cook's tidak ada pengamatan yang melebihi nilai kritisnya ($F = 2,85$). Ini berarti taksiran koefisien regresi sudah stabil.

Dalam metode LMS, estimasi yang dilakukan menghasilkan persamaan regresi yang berbeda pada setiap pengacakan. Dalam hal ini peneliti menggunakan metode perulangan dengan mencari nilai MAPE yang lebih kecil dari metode OLS. Jika nilai MAPE lebih kecil dari OLS maka nilai persamaan regresi disimpan.

1. Diketahui $N = 21$, maka ditentukan $M = 3$ dan $n = 7$.
2. Secara acak diambil 3 buah subset berukuran 7 dari sampel berukuran 21, seperti pada Tabel 4.

Tabel 4. Pengelompokkan Subset Data Stackloss

	Observasi	x_1	x_2	x_3	Y
SUBSET 1	1	58	19	93	12
	2	56	20	82	15
	3	58	17	88	13
	4	58	18	82	11
	5	50	18	89	8
	6	62	24	93	19
	7	50	19	79	8
SUBSET 2	1	75	25	90	37
	2	62	24	87	28
	3	58	18	89	14
	4	50	19	72	8
	5	50	20	80	9
	6	80	27	88	37
	7	50	18	86	7
SUBSET 3	1	62	22	87	18
	2	62	23	87	18
	3	58	18	80	14
	4	62	24	93	20
	5	70	20	91	15
	6	58	23	87	15
	7	80	27	89	42

3. Kemudian dicari dugaan \hat{y} dari ketiga subset data diatas. Dengan bantuan minitab diperoleh secara berturut-turut subset 1, 2 dan 3 adalah:

$$\hat{y} = -31,8882 + 0,5826x_1 + 0,7611x_2 - 0,0362x_3$$

$$\hat{y} = -72,4633 + 0,4623x_1 + 1,8770x_2 + 0,2827x_3$$

$$\hat{y} = -3,35701 + 0,79174x_1 + 2,19236x_2 - 0,87389x_3$$

4. Setelah itu dicari median dari kuadrat galat dari setiap subset seperti pada Tabel 5.

Tabel 5. Median Tiap Subset

Subset	Obs	x_1	x_2	x_3	y	\hat{y}	$e = y - \hat{y}$	e^2	Median
SUBSET 1	1	58	19	93	12	12,99690	-0,99690	0,99381	0,99381 (Minimum)
	2	56	20	82	15	12,99100	2,00900	4,03608	
	3	58	17	88	13	11,65570	1,34430	1,80714	
	4	58	18	82	11	12,63400	-1,63400	2,66996	
	5	50	18	89	8	7,71980	0,28020	0,07851	
	6	62	24	93	19	19,13280	-0,13280	0,01764	
	7	50	19	79	8	8,84290	-0,84290	0,71048	
SUBSET 2	1	75	25	90	37	34,57720	2,42280	5,86996	3,267779
	2	62	24	87	28	25,84220	2,15780	4,65610	
	3	58	18	89	14	13,29640	0,70360	0,49505	
	4	50	19	72	8	6,66910	1,33090	1,77130	
	5	50	20	80	9	10,80770	-1,80770	3,26778	
	6	80	27	88	37	40,07730	-3,07730	9,46978	
	7	50	18	86	7	8,74990	-1,74990	3,06215	
SUBSET 3	1	62	22	87	18	17,93436	0,06564	0,00431	3,552509
	2	62	23	87	18	20,12672	-2,12672	4,52294	
	3	58	18	80	14	12,11519	1,88481	3,55251	
	4	62	24	93	20	17,07574	2,92426	8,55130	
	5	70	20	91	15	16,38800	-1,38800	1,92654	
	6	58	23	87	15	16,95976	-1,95976	3,84066	
	7	80	27	89	42	41,39970	0,60030	0,36036	

5. Subset pertama merupakan subset dengan median kuadrat galat terkecil. Nilai s_0 adalah: 1,912711.

6. Kemudian bobot awal (w) tiap observasi disajikan pada Tabel 6.

7. Berdasarkan pembobot w_i dihitung nilai akhir taksiran Robust seperti pada Tabel 7.

8. Berdasarkan nilai akhir taksiran Robust diperoleh bobot final seperti pada Tabel 8.

Tabel 6. Bobot Awal Tiap Observasi

Obs	x_1	x_2	x_3	y	\hat{y}	$e = y - \hat{y}$	s_0	e/s_0	$ e/s_0 $	w_i
1	80	27	89	42	32,04770	9,95230	1,91271	5,20324	5,20324	0
2	80	27	88	37	32,08390	4,91610	1,91271	2,57023	2,57023	0
3	75	25	90	37	27,57630	9,42370	1,91271	4,92688	4,92688	0
4	62	24	87	28	19,35000	8,65000	1,91271	4,52238	4,52238	0
5	62	22	87	18	17,82780	0,17220	1,91271	0,09003	0,09003	1
6	62	23	87	18	18,58890	-0,58890	1,91271	-0,30789	0,30789	1
7	62	24	93	19	19,13280	-0,13280	1,91271	-0,06943	0,06943	1
8	62	24	93	20	19,13280	0,86720	1,91271	0,45339	0,45339	1
9	58	23	87	15	16,25850	-1,25850	1,91271	-0,65797	0,65797	1
10	58	18	80	14	12,70640	1,29360	1,91271	0,67632	0,67632	1
11	58	18	89	14	12,38060	1,61940	1,91271	0,84665	0,84665	1
12	58	17	88	13	11,65570	1,34430	1,91271	0,70282	0,70282	1
13	58	18	82	11	12,63400	-1,63400	1,91271	-0,85428	0,85429	1
14	58	19	93	12	12,99690	-0,99690	1,91271	-0,52120	0,52120	1
15	50	18	89	8	7,71980	0,28020	1,91271	0,14649	0,14649	1
16	50	18	86	7	7,82840	-0,82840	1,91271	-0,43310	0,43310	1
17	50	19	72	8	9,09630	-1,09630	1,91271	-0,57317	0,57317	1
18	50	19	79	8	8,84290	-0,84290	1,91271	-0,44068	0,44068	1
19	50	20	80	9	9,56780	-0,56780	1,91271	-0,29686	0,29686	1
20	56	20	82	15	12,99100	2,00900	1,91271	1,05034	1,05034	1
21	70	20	91	15	20,82160	-5,82160	1,91271	-3,04364	3,04364	0

Tabel 7. Nilai Akhir Taksiran Robust

Obs	x_1	x_2	x_3	y	\hat{y}	$e = y - \hat{y}$	s_0	w_i	e^2	$w_i e_i^2$	$\hat{\sigma}$
1	80	27	89	42	32,04770	9,95230	1,91271	0	99,04828	0	1,275811
2	80	27	88	37	32,08390	4,91610	1,91271	0	24,16804	0	1,275811
3	75	25	90	37	27,57630	9,42370	1,91271	0	88,80612	0	1,275811
4	62	24	87	28	19,35000	8,65000	1,91271	0	74,82250	0	1,275811
5	62	22	87	18	17,82780	0,17220	1,91271	1	0,02965	0,02965	1,275811
6	62	23	87	18	18,58890	-0,58890	1,91271	1	0,34680	0,34680	1,275811
7	62	24	93	19	19,13280	-0,13280	1,91271	1	0,01764	0,01764	1,275811
8	62	24	93	20	19,13280	0,86720	1,91271	1	0,75204	0,75204	1,275811
9	58	23	87	15	16,25850	-1,25850	1,91271	1	1,58382	1,58382	1,275811
10	58	18	80	14	12,70640	1,29360	1,91271	1	1,67340	1,67340	1,275811
11	58	18	89	14	12,38060	1,61940	1,91271	1	2,62246	2,62246	1,275811
12	58	17	88	13	11,65570	1,34430	1,91271	1	1,80714	1,80714	1,275811
13	58	18	82	11	12,63400	-1,63400	1,91271	1	2,66996	2,66996	1,275811
14	58	19	93	12	12,99690	-0,99690	1,91271	1	0,99381	0,99381	1,275811
15	50	18	89	8	7,71980	0,28020	1,91271	1	0,07851	0,07851	1,275811
16	50	18	86	7	7,82840	-0,82840	1,91271	1	0,68625	0,68625	1,275811
17	50	19	72	8	9,09630	-1,09630	1,91271	1	1,20187	1,20187	1,275811
18	50	19	79	8	8,84290	-0,84290	1,91271	1	0,71048	0,71048	1,275811
19	50	20	80	9	9,56780	-0,56780	1,91271	1	0,32240	0,32240	1,275811
20	56	20	82	15	12,99100	2,00900	1,91271	1	4,03608	4,03608	1,275811
21	70	20	91	15	20,82160	-5,82160	1,91271	0	33,89103	0	1,275811
Jumlah								16		19,53231	

Tabel 8. Bobot Final

Obs	x_1	x_2	x_3	y	$e = y - \hat{y}$	$\hat{\sigma}$	$e/\hat{\sigma}$	$ e/\hat{\sigma} $	w_i final
1	80	27	89	42	9,95230	1,275811	7,80077	7,80077	0
2	80	27	88	37	4,91610	1,275811	3,85332	3,85332	0
3	75	25	90	37	9,42370	1,275811	7,38644	7,38644	0
4	62	24	87	28	8,65000	1,275811	6,78000	6,78000	0
5	62	22	87	18	0,17220	1,275811	0,13497	0,13497	1
6	62	23	87	18	-0,58890	1,275811	-0,46159	0,46159	1
7	62	24	93	19	-0,13280	1,275811	-0,10409	0,10409	1
8	62	24	93	20	0,86720	1,275811	0,67972	0,67972	1
9	58	23	87	15	-1,25850	1,275811	-0,98643	0,98643	1
10	58	18	80	14	1,29360	1,275811	1,01394	1,01394	1
11	58	18	89	14	1,61940	1,275811	1,26931	1,26931	1
12	58	17	88	13	1,34430	1,275811	1,05368	1,05368	1
13	58	18	82	11	-1,63400	1,275811	1,28075	1,28075	1
14	58	19	93	12	-0,99690	1,275811	-0,78139	0,78139	1
15	50	18	89	8	0,28020	1,275811	0,21963	0,21963	1
16	50	18	86	7	-0,82840	1,275811	-0,64931	0,64931	1
17	50	19	72	8	-1,09630	1,275811	-0,85930	0,85930	1
18	50	19	79	8	-0,84290	1,275811	-0,66068	0,66068	1
19	50	20	80	9	-0,56780	1,275811	-0,44505	0,44505	1
20	56	20	82	15	2,00900	1,275811	1,57469	1,57469	1
21	70	20	91	15	-5,82160	1,275811	-4,56306	4,56306	0

9. Kemudian di cari \hat{y} final

Dengan minitab diperoleh: $\hat{y} = -35,4842 + 0,6861x_1 + 0,5671x_2 - 0,0173x_3$

Untuk melihat ketepatan metode Regresi Robust penaksir LMS dibandingkan dengan metode OLS pada data yang mengandung pencilan yaitu *stackloss* maka dilihat nilai MAPE dari model yang dihasilkan oleh metode OLS dan LMS pada data tersebut.

a) Metode OLS :

Dengan Minitab 16 diperoleh model regresinya adalah: $\hat{y} = -39,9197 + 0,7156x_1 + 1,2953x_2 - 0,1521x_3$ maka perhitungan MAPE-nya pada Tabel 9.

Tabel 9. MAPE OLS

Obs	x_1	x_2	x_3	Y	\hat{y}	$e = y - \hat{y}$	e/y	$ e/y $
1	80	27	89	42	38,7645	3,23550	0,07704	0,07704
2	80	27	88	37	38,9166	-1,91660	-0,05180	0,05180
3	75	25	90	37	32,4438	4,55620	0,12314	0,12314
4	62	24	87	28	22,302	5,69800	0,20350	0,20350
5	62	22	87	18	19,7114	-1,71140	-0,09508	0,09508
6	62	23	87	18	21,0067	-3,00670	-0,16704	0,16704
7	62	24	93	19	21,3894	-2,38940	-0,12576	0,12576
8	62	24	93	20	21,3894	-1,38940	-0,06947	0,06947
9	58	23	87	15	18,1443	-3,14430	-0,20962	0,20962
10	58	18	80	14	12,7325	1,26750	0,09054	0,09054
11	58	18	89	14	11,3636	2,63640	0,18831	0,18831
12	58	17	88	13	10,2204	2,77960	0,21382	0,21382
13	58	18	82	11	12,4283	-1,42830	-0,12985	0,12985
14	58	19	93	12	12,0505	-0,05050	-0,00421	0,00421
15	50	18	89	8	5,6388	2,36120	0,29515	0,29515
16	50	18	86	7	6,0951	0,90490	0,12927	0,12927
17	50	19	72	8	9,5198	-1,51980	-0,18998	0,18998
18	50	19	79	8	8,4551	-0,45510	-0,05689	0,05689
19	50	20	80	9	9,5983	-0,59830	-0,06648	0,06648
20	56	20	82	15	13,5877	1,41230	0,09415	0,09415
21	70	20	91	15	22,2372	-7,23720	-0,48248	0,48248
Jumlah								3,06356

$$MAPE = \frac{1}{21}(3,06356) \times 100\%$$

$$= 14,58836\%$$

b) Metode LMS :

Dari hasil sebelumnya diketahui bahwa model regresi yang dihasilkan oleh metode LMS adalah:

$$\hat{y} = -35,4842 + 0,6861x_1 + 0,5671x_2 - 0,0173x_3 \quad \text{maka perhitungan MAPE-nya pada Tabel 10.}$$

Tabel 10. MAPE LMS

Obs	x_1	x_2	x_3	y	\hat{y} final	$e = y - \hat{y}$	e/y	$ e/y $
1	80	27	89	42	33,17580	8,82420	0,21010	0,21010
2	80	27	88	37	33,19310	3,80690	0,10289	0,10289
3	75	25	90	37	28,59380	8,40620	0,22720	0,22720
4	62	24	87	28	19,15930	8,84070	0,31574	0,31574
5	62	22	87	18	18,02510	-0,02510	-0,00139	0,00139
6	62	23	87	18	18,59220	-0,59220	-0,03290	0,03290
7	62	24	93	19	19,05550	-0,05550	-0,00292	0,00292
8	62	24	93	20	19,05550	0,94450	0,04723	0,04723
9	58	23	87	15	15,84780	-0,84780	-0,05652	0,05652
10	58	18	80	14	13,13340	0,86660	0,06190	0,06190
11	58	18	89	14	12,97770	1,02230	0,07302	0,07302
12	58	17	88	13	12,42790	0,57210	0,04401	0,04401
13	58	18	82	11	13,09880	-2,09880	-0,19080	0,19080
14	58	19	93	12	13,47560	-1,47560	-0,12297	0,12297
15	50	18	89	8	7,48890	0,51110	0,06389	0,06389
16	50	18	86	7	7,54080	-0,54080	-0,07726	0,07726
17	50	19	72	8	8,35010	-0,35010	-0,04376	0,04376
18	50	19	79	8	8,22900	-0,22900	-0,02862	0,02862
19	50	20	80	9	8,77880	0,22120	0,02458	0,02458
20	56	20	82	15	12,86080	2,13920	0,14261	0,14261
21	70	20	91	15	22,31050	-7,31050	-0,48737	0,48737
Jumlah								2,35767

$$MAPE = \frac{1}{21}(2,35767) \times 100\%$$

$$= 11,227\%$$

Berdasarkan perhitungan nilai MAPE pada model regresi yang dihasilkan oleh kedua metode pada data *stackloss* maka dapat dilihat bahwa metode LMS menghasilkan nilai yang lebih kecil sehingga penggunaan metode ini lebih tepat karena tidak rentan akan pengaruh pencilan. Pada metode OLS estimasi sangat mudah dilakukan akan tetapi pendugaan model regresi terpengaruh oleh data pencilan sehingga persamaan regresi menghasilkan nilai MAPE yang lebih besar. Metode LMS merupakan salah satu penaksir regresi robust yang kekar terhadap pencilan sehingga dapat menghasilkan model regresi yang lebih baik. Hal ini sejalan dengan penelitian Tarno [10] yang menyimpulkan bahwa pada metode LMS, estimasi model yang diperoleh adalah suatu model yang memiliki median kuadrat sesatan terkecil walaupun dalam penggunaannya baik metode OLS maupun LMS diperoleh estimasi model regresi yang tidak memiliki perbedaan yang mencolok atau kedua model yang diperoleh mempunyai keakuratan yang hampir sama. Penelitian Sugiarti dan Megawarni [9] juga menyimpulkan bahwa metode LMS sangat efisien dibanding metode M dalam menaksir koefisien garis regresi jika data mengandung pencilan.

Metode LMS akan menghasilkan estimasi yang lebih baik dibandingkan metode OLS ketika data mengandung pencilan namun perlu diperhatikan bahwa metode LMS kurang stabil karena setiap perulangan pada metode LMS menghasilkan estimasi regresi berbeda. Perulangan bertujuan untuk mencari model regresi dengan nilai MAPE yang terkecil sehingga memerlukan waktu lebih lama.

4. KESIMPULAN

Berdasarkan perhitungan nilai MAPE pada model regresi yang dihasilkan oleh metode OLS dan LMS pada data *stackloss* maka dapat disimpulkan metode LMS menghasilkan nilai yang lebih kecil sehingga penggunaan metode ini lebih tepat karena tidak rentan akan pengaruh pencilan. Pada metode OLS estimasi sangat mudah dilakukan akan tetapi pendugaan terpengaruh oleh data pencilan sehingga persamaan regresi menghasilkan nilai MAPE yang lebih besar. Pada metode LMS estimasi akan lebih baik, tapi perlu diketahui bahwa metode ini kurang stabil karena menggunakan metode perulangan.

DAFTAR PUSTAKA

- [1] Aunuddin, Analisa Data, *Bogor: Institut Pertanian Bogor*, 1989.
- [2] Draper, N. R. dan Smith, H., Analisis Regresi Terapan, *Edisi Kedua. Jakarta: PT Gramedia Pustaka Utama*, 1992.
- [3] Myers, R. H., Classical and Modern Regression With Applications, (2nd Ed). *Boston: PWS- Kent*, 1990.
- [4] Nurdin, N., Raupong dan Islamiyati, A. "Penggunaan Regresi *Robust* pada Data yang Mengandung Pencilan dengan Metode Momen," *Jurnal Matematika, Statistika dan Komputasi*, vol. 10, no. 2, hal. 114-123, 2008.
- [5] Rousseeuw, P. J. 1984. "Least Median of Squares Regression," *Journal of the American Statistical Association*, vol. 79, no. 388, Hal. 871-880, 1984.
- [6] Rousseeuw, P. J. and Leroy, A. M., Robust Regression and Outlier Detection, *New York: Wiley Interscience*, 1987.
- [7] Ryan, T. P., Modern Regression Methods, *Canada: John Wiley & Sons, Inc.*, 1997.
- [8] Soemartini, Pencilan (Outlier). *Bandung: Universitas Padjadjaran*, 2007.
- [9] Sugiarti, H. dan Megawarni, A. "Tingkat Efisiensi Penaksir M terhadap Penaksir LMS dalam Menaksir Koefisien Regresi," *Jurnal Matematika, Sains dan Teknologi*, Vol. 11, No. 2, Hal. 90-98, 2010.
- [10] Tarno. "Estimasi Model Regresi Linier dengan Metode Median Kuadrat Terkecil," *Jurnal Sains dan Matematika*, Vol. 15, No. 2, Hal 69-72, 2007.
- [11] Walpole, R. E. dan Myers, R. H., Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan, *Edisi Keempat, Bandung: ITB*, 1995.
- [12] Yingying, C. et all., Securing Emerging Wireless Systems, Lower Layer Approaches, *New York: Springer Science Bussiness Media*, 2009.

