

SIMULASI PENANGANAN PENCILAN PADA ANALISIS REGRESI MENGUNAKAN *METODE LEAST MEDIAN SQUARE (LMS)*

Tusilowati¹, L. Handayani², dan Rais³

^{1,2,3}Program Studi Statistika Jurusan Statistik FMIPA Universitas Tadulako

Jalan Soekarno-Hatta Km. 09 Tondo, Palu 94118, Indonesia.

¹tusilowati02@gmail.com, ²lilies.stath@gmail.com, ³rais76_untad@yahoo.co.id

ABSTRACT

The simulation of handling of outliers on regression analysis used the method which was commonly used to predict the parameter in regression analysis, namely *Least Median Square (LMS)* due to the simple calculation it had. The data with outliers would result in unbiased parameter estimate. Hence, it was necessary to draw up the robust regression to overcome the outliers. The data used were simulation data of the number of data pairs (X, Y) by 25 and 100 respectively. The result of the simulation was divided into 5 subsets of data cluster of parameter regression prediction by *Ordinary Least Square (OLS)* and *Least Median Square (LMS)* methods. The prediction result of the parameter of each method on each subset of data cluster was tested with both method to discover the which better one. Based on the research findings, it was found that The *Least Median Square (LMS)* method was known better than *Ordinary Least Square (OLS)* method in predicting the regression parameter on the data which had up to 3% of the percentage of the outlier.

Keywords : **Regression Analysis, Least Median Square (LMS), Outlier Data of Robust Regression.**

ABSTRAK

Simulasi Penanganan Pencilan Pada Analisis Regresi Menggunakan Metode yang umum digunakan untuk menduga parameter dalam analisis regresi adalah *Least Median Square (LMS)* karena memiliki perhitungan yang sederhana. Data yang memiliki pencilan akan menghasilkan dugaan parameter yang tak bias. Sehingga perlu disusun metode regresi kekar untuk mengatasi pencilan. Data yang digunakan adalah data simulasi banyaknya pasangan data (X, Y) sebesar masing-masing 25 dan 100. Hasil simulasi diperoleh data yang terbagi menjadi 5 subset gugus data parameter regresi diduga dengan metode *Ordinary Least Square (OLS)* dan *Least Median Square (LMS)*. Hasil dugaan parameter masing-masing metode pada setiap subset gugus data diuji dengan kedua metode untuk mengetahui metode mana yang lebih baik. Dari hasil penelitian diperoleh Metode *Least Median Square (LMS)* lebih baik dari pada metode *Ordinary Least Square (OLS)* dalam penduga parameter regresi pada data yang memiliki pencilan hingga persentase pencilan sebesar 3%.

Kata kunci : **Analisis Regresi, Least Median Square (LMS), data pencilan Regresi Kekar (Robust).**

I. PENDAHULUAN

1.1. Latar Belakang

Analisis regresi adalah metode statistika yang memanfaatkan hubungan antara dua atau lebih peubah kuantitatif sehingga salah satu peubah dapat diramalkan dari peubah lainnya (Draper dan Smith, 1992). Peubah tersebut dibedakan menjadi dua, yaitu peubah bebas dan peubah tak bebas. Analisis regresi adalah metode yang sangat populer untuk mengetahui hubungan linier antara variabel respon dengan satu atau lebih peubah bebas. Metode pendugaan parameter regresi yang paling populer adalah *Ordinary Least Square* (OLS). Metode ini populer karena selain pertama kali dikembangkan juga dianggap paling mudah. Setiap metode selalu memiliki kelebihan dan kekurangan. OLS akan sah jika sejumlah asumsi tertentu dipenuhi.

Metode *ordinary least square* (OLS) adalah metode pendugaan dengan meminimumkan jumlah kuadrat galat. Metode ini cenderung memiliki perhitungan yang sederhana. Sejumlah asumsi yang harus dipenuhi adalah tidak adanya multikolinieritas antar peubah bebas, galat menyebar mengikuti sebaran normal, galat memiliki ragam yang homogeny (homoskedastisitas), dan antar komponen galat tidak terjadi autokorelasi. Salah satu asumsi yang cukup sulit untuk dipenuhi adalah asumsi tidak adanya multikolinieritas (korelasi kuat antar peubah bebas) dikarenakan jumlah peubah bebas yang cukup banyak. Untuk mengatasi pelanggaran asumsi tersebut dikembangkan beberapa alternatif. Alternatif tersebut adalah regresi komponen utama, membuang salah satu peubah bebas yang berkorelasi, atau dengan regresi.

Myers (1990) menyatakan bahwa apabila asumsi terpenuhi maka penduga parameter yang diperoleh bersifat BLUE (*best linier unbiased estimator*). Padahal permasalahan yang sering dihadapi yaitu adanya asumsi yang dilanggar. Salah satu penyebab dilanggarnya asumsi ialah adanya pencilan (*outlier*) pada data. Adanya pencilan pada regresi menggunakan metode OLS merupakan masalah, karena OLS sensitif terhadap pencilan dimana persamaan regresi yang dihasilkan oleh OLS cenderung mudah berubah-ubah. Salah satu metode untuk mengatasi pencilan yang paling populer digunakan adalah regresi kekar dengan pendekatan *least median square* (LMS).

Metode LMS adalah salah satu metode estimasi regresi yang diperkenalkan oleh Andrews (1972). Metode ini melakukan perhitungan dengan menghilangkan pengaruh-pengaruh sisaan. Dengan menggunakan penduga yang dihasilkan akan lebih baik dalam menghadapi pencilan. Kriteria dari LMS adalah metode analisis standar dari data ketika adanya kemungkinan banyaknya data pencilan, dimana estimasi terhadap nilai pengganggu tetap dimungkinkan (Rousseeuw, 1984). Dalimunthe (2010) menggunakan metode LMS pada analisis regresi dengan pencilan menggunakan PMC. Dalam penelitian ini peneliti akan

menggunkan regresi kekar dengan metode LMS pada analisis regresi dengan pencilan menggunkan OLS.

1.2. Rumusan Masalah

Dari latar belakang di atas peneliti dapat merumuskan permasalahan yaitu bagaimana menentukan pengaruh *Ordinary Least Square* (OLS) terhadap pendugaan parameter model regresi?.

1.3. Batasan Masalah

Dalam peneliti ini, metode digunakan adalah data simulasi banyaknya pasangan data (X,Y) sebesar masing-masing 25 dan 100.

II. TINJAUAN PUSTAKA

2.1. Variabel Penelitian

Metode *Ordinary Least Square* (OLS) adalah estimasi parameter bertujuan mendapatkan model regresi yang akan digunakan dalam analisis regresi. Metode Kuadrat Terkecil ini bertujuan meminimumkan jumlah kuadrat error. (Srinadi, I.G.A.M 2014).

2.2. Pencilan (*Outlier*)

2.2.1. Pengertian Metode pencilan (*Outlier*)

Pencilan adalah suatu data yang menyimpang dari sekumpulan data yang lain ada beberapa definisi pencilan (*outlier*) menurut para ahli:

2.2.2. Perumusan Metode pencilan

diterapkan bila anggapan-anggapan berikut terpenuhi:

- a) Menurut Barnett dan Lewis, T. 1994. pencilan adalah pengamatan yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat data.
- b) Menurut Sembiring, *R.K Sembing* 2005. pencilan adalah pengamatan yang jauh dari pusat data yang mungkin berpengaruh besar terhadap koefisien regresi.
- c) Menurut Barnett dan Lewis, T. 1994. pencilan adalah suatu pengamatan yang tampak bertentangan atau tidak konsisten terhadap pengamatan yang lain.

III. PROSEDUR PENELITIAN

Prosedur dalam penelitian ini adalah:

1. Memulai penelitian
2. Menentukan topik penelitian untuk dianalisis, serta mempelajari metode pendugaan parameter *Ordinary Least Square*
3. Mencari dan menyeleksi penelitian yang berkaitan dengan topik penduga parameter
4. Input data yang akan dianalisis melalui pembangkitan data simulasi

5. Melakukan pemodelan regresi menggunakan metode penduga parameter *Ordinary Least Square* dan *Least Mendian Square*
6. Membandingkan model yang dihasilkan menggunakan MAPE
7. Pemilihan model terbaik
8. Interpretasi atau penarikan kesimpulan dari hasil yang diperoleh
9. Selesai

IV. HASIL DAN PEMBAHASAN

4.1. Data Penelitian

Data yang digunakan dalam penelitian ini adalah data bangkitan dengan 6 gugus data dimana 5 gugus data pertama (n) sebanyak 25 disetiap gugus berdasarkan model regresi $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ (Tabel 1). Dengan kriteria dan data sebagai berikut:

Tabel 1 : Data Pencilan- X

Gugus Data	Pencilan- X	Pencilan- Y	Pencilan- XY
1	-	-	-
2	√	-	√
3	-	√	√
4	√	√	-
5	√	√	√

Ke-5 gugus data di atas akan dibangkitkan dengan menggunakan *software* R yang tertera pada Lampiran 1-5. Hasil bangkitan datanya disajikan pada Lampiran 6. Gugus data 1-5 kemudian diregresikan dengan OLS/LMS. Lalu ditentukan nilai $\underline{\beta}$, \underline{Y} dan $\underline{\varepsilon}$ serta MAPE menggunakan *software* R. (Gugus data selanjutnya ialah gugus data yang di dalamnya terdapat ke-5 kriteria pada data diatas yaitu:

1. Ditentukan ukuran subset 5, dengan jumlah subset 5 dan batas kesalahan $\alpha = 5\%$.
2. Membagi secara acak gugus data yang berukuran 25 kedalam 5 buah subset berukuran 5. Dilakukan pengulangan 2x.

Tabel 2 : Data Menggunakan OLS Sehingga diperoleh Parameter Setiap Subset

Pembagian secara acak					
	x	Y		x	Y
Subset 1	10	100	Subset 1	27	105
	7	19		13	30
	2	13		11	33
	35	50		2	13
	17	56		9	30
Subset 2	15	42	Subset 2	10	30
	5	26		35	50
	27	105		7	19
	8	21		40	92
	1	12		3	18
Subset 3	16	43	Subset 3	25	100
	25	100		8	21
	18	50		12	31
	9	30		5	26
	19	50		6	22
Subset 4	40	92	Subset 4	14	35
	20	52		17	56
	12	31		18	50
	3	18		20	52
	13	30		19	50
Subset 5	11	33	Subset 5	4	17
	14	35		15	42
	4	17		16	43
	10	30		1	12
	6	22		10	100

3. Meregresikan dengan menggunakan OLS sehingga diperoleh parameter setiap subset. Mencari median dari kuadrat error disetiap subset.

Tabel 3 : Data yang diacak ke-1

Pengacakan ke-1	X	y	\hat{y}	e	Mede
Subset 1	10	100	34.2393	65.7607	-31.52139848
	7	19	28.3776	-9.3776	
	2	13	18.608	-5.608	
	35	50	83.0872	-33.087	
	17	56	47.9167	8.08329	
Subset 2	15	42	44.0089	-2.0089	46.01775913
	5	26	24.4697	1.53028	
	27	105	67.4559	37.5441	
	8	21	30.3315	-9.3315	
	1	12	16.6541	-.46541	
Subset 3	16	43	45.9628	-2.9628	0.129373152
	25	100	63.548	36.452	
	18	50	49.8706	0.12937	
	9	30	32.2854	-2.2854	
	19	50	51.8245	-1.8245	
Subset 4	40	92	92.8568	-0.8568	-7.147132282
	20	52	53.7785	-1.7785	
	12	31	38.1471	-7.1471	
	3	18	20.5619	-2.5619	
	13	30	40.101	-10.101	
Subset 5	11	33	36.1932	-3.1932	-5.515806194
	14	35	42.055	-7.055	
	4	17	22.5158	-5.5158	
	10	30	34.2393	-4.2393	
	22	21.775	57.6863	-35.911	

Tabel 4 : Data yang di acak ke-2

Pengacakan ke-2	X	y	\hat{y}	e	mede
Subset 1	27	105	70.5515	35.4485325	-4.049660471
	13	30	41.2374	-11.23739625	
	11	33	37.0497	-4.049660471	
	2	13	18.2048	-5.204849441	
	9	30	32.8619	-2.861924686	
Subset 2	10	30	34.9558	-4.95579578	-6.74188902
	35	50	87.3025	-37.30248988	
	7	19	28.6742	-9.674188902	
	40	92	97.7718	-5.771829344	
	3	18	20.2987	-2.298717333	
Subset 3	25	100	66.3638	33.63618904	-8.143528363
	8	21	30.7681	-9.768056794	
	12	31	39.1435	-8.143528363	
	5	26	24.4865	1.513546883	
	6	22	26.5803	-4.58032101	
Subset 4	14	35	43.3313	-8.331264147	-1.706735716
	17	56	49.6129	6.387132176	
	18	50	51.7067	-1.706735716	
	20	52	55.8945	-3.8944715	
	19	50	53.8006	-3.800603608	
Subset 5	4	17	8.37547	8.624528431	-4.518999931
	15	42	45.4251	-3.425132039	
	16	43	47.519	-4.518999931	
	1	12	16.111	-4.110981549	
	10	100	34.9558	65.04420742	

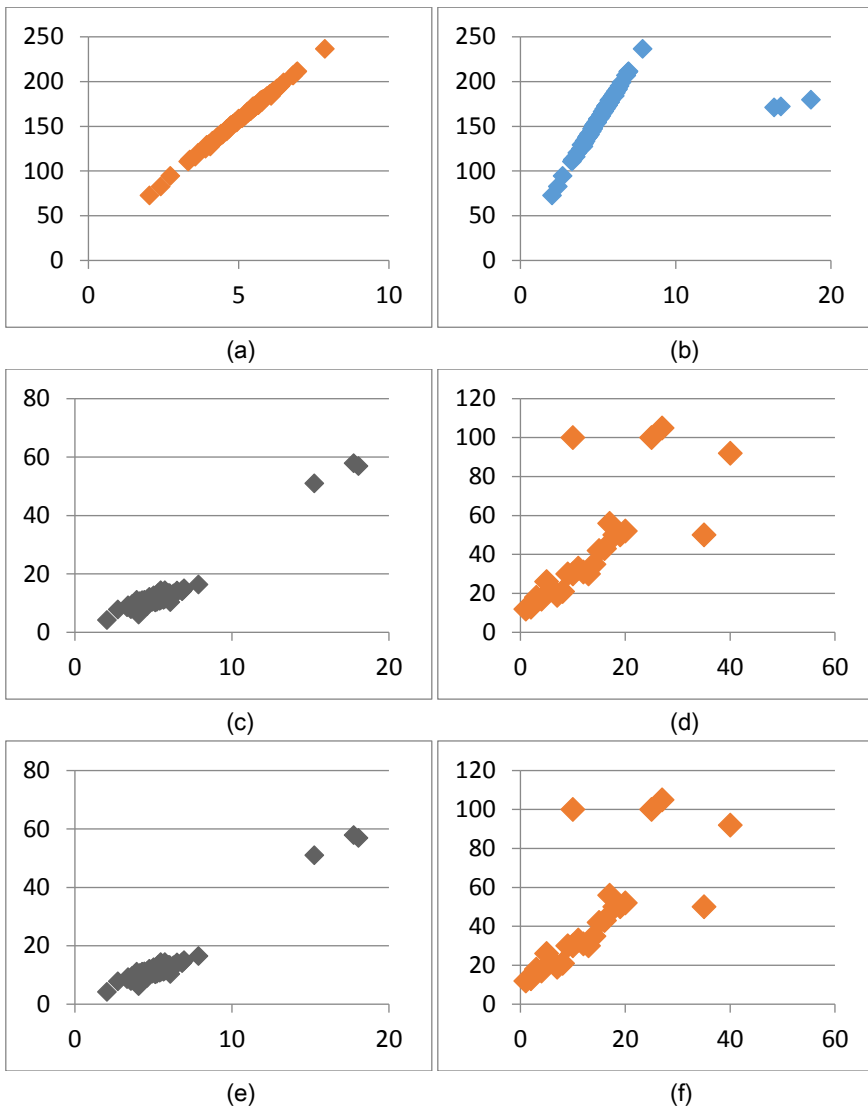
4. Dengan melihat data diatas, maka diperoleh minimum median: -8.143528363
5. Berdasarkan minimum median, maka diperoleh $b = \begin{bmatrix} 10.525 \\ 1.875 \end{bmatrix}$ dari hasil regresi anak gugus yang mempunyai median terkecil pada langkah 4.
6. Menentukan kuadrat galat:

Tabel 5 : Langkah 3 dan 4 diperoleh sebagai berikut:

<i>obs</i>	<i>x</i>	<i>y</i>	\hat{y}	e_i	e_i^2
1	1	12	16.111	-4.111	16.9002
2	2	13	18.2048	-5.2048	27.0905
3	3	18	20.2987	-2.2987	5.2841
4	4	17	22.3926	-5.3926	29.08
5	5	26	24.4865	1.51355	2.29082
6	6	22	26.5803	-4.5803	20.9793
7	7	19	28.6742	-9.6742	93.5899
8	8	21	30.7681	-9.7681	95.4149
9	9	30	32.8619	-2.8619	8.19061
10	10	30	34.9558	-4.9558	24.5599
11	11	33	37.0497	-4.0497	16.3997
12	12	31	39.1435	-8.1435	66.3171
13	13	30	41.2374	-11.237	126.279
14	14	35	43.3313	-8.3313	69.41
15	15	42	45.4251	-3.4251	11.7315
16	16	43	47.519	-4.519	20.4214
17	17	56	49.6129	6.38713	40.7955
18	18	50	51.7067	-1.7067	2.91295
19	19	50	53.8006	-3.8006	14.4446
20	20	52	55.8945	-3.8945	15.1669
21	10	100	34.9558	65.0442	4230.75
22	35	50	87.3025	-37.302	1391.48
23	40	92	97.7718	-5.7718	33.341
24	25	100	66.3638	33.6362	1131.39
25	27	105	70.5515	34.4487	1186.7

Menentukan $S_0 = 1.4826 \left(1 + \frac{5}{N-p}\right) \sqrt{\text{Med } e_{ij}^2} = 6.948882$

7. Menghitung bobot w_i , dengan menggunakan $w_i = 1$ jika $\left|\frac{s_0}{e_i}\right| \leq \alpha$ dan $w_i = \frac{s_0}{|e_i|}$ selainnya.



Gambar 1 : Gugus data yang (a) tanpa pencilan, sedangkan gugus data (b) sampe (f) semua data menggunakan pencilan.

V. KESIMPULAN

Berdasarkan cara membangkitkan data untuk menentukan ada tidaknya suatu pencilan dengan menggunakan metode OLS dan LMS, untuk menentukan seberapa besar kekekaran metode *Least Median Square* (LMS) dalam menduga parameter regresi dari gugus data yang memiliki pencilan.

DAFTAR PUSTAKA

- [1] Andrew, The most destructive hurricane in United States history ,1972, jakarta.
- [2] Daliunthe, *Metode Least Median Of Square (LMS) pada Analisi Regresi Dengan Pencilan*. Jurusan Departemen Statistika FMIPA ITB, 2010, Bogor.
- [3] Draper, *Analisis Regresi Terapan* Edisi Ke-2, Jakarta PT Gramedia Pusaka Utama. Terjemahan dari Applied Regression Analysis,1992, Jakarta.
- [4] Myers, R, *Classical And Modern Regression With Applications*. Boston: PWS-KENT Publishing Company, 1990, Surabaya.
- [5] Rousseeuw, P, J, *Least Median of Squares*. Journal of The American Statistician Association Vol.76, No 388: 871-880,1984, Jakarta.