UJM 3 (2) (2014)
UNNES Journal of Mathematics
http://journal.unnes.ac.id/sju/index.php/ujm

DETEKSI *OUTLIER* MENGGUNAKAN DIAGNOSA REGRESI BERBASIS ESTIMATOR PARAMETER *ROBUST*

Suyanti✉, YL Sukestiyarno

Jurusan Matematika, FMIPA, Universitas Negeri Semarang, Indonesia
Gedung D7 lantai 1 Kampus Sekaran, Gunungpati, Semarang, 50229

Info Artikel

Sejarah Artikel:
Diterima Desember 2013
Disetujui Mei 2014
Dipublikasikan Nopember 2014

Keywords :
Least Trimmed Square;
M-estimation;
Pencilan;
Regresi Linier;
Regresi Robust.

Abstrak

Pencilan adalah data yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat data, dapat dideteksi dengan menggunakan *leverage*, nilai *discrepancy*, nilai *influence*. Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi dari residual tidak normal dan atau mengandung beberapa outlier yang berpengaruh pada model (Ryan, 1997). *Least Trimmed Square* (LTS) yaitu metode penaksir regresi robust yang menggunakan konsep pemangkasan untuk meminimalkan jumlah kuadrat residual dengan nilai *breakdown point* sebesar $[(n-p)/2+1]/n$. *M-estimation* merupakan estimasi yang meminimumkan suatu fungsi objektif dengan nilai *breakdown point* 0. Tujuan dari penelitian ini adalah membandingkan tingkat efektifitas metode *Least Trimmed Square* (LTS) dan metode *M-estimation* dalam data yang mengandung *outlier* (pencilan). Perbandingan kedua metode ini dilakukan melalui studi pustaka yang melibatkan dua contoh kasus. Kemudian perbandingan keefektifitasan kedua metode dilihat dari nilai koefisien determinasi (R^2) yang diperoleh dengan menggunakan rumus atau bisa juga dengan menggunakan *software* MINITAB 16.

Abstract

Outliers are data that do not follow most of the pattern and located away from the data center, can be detected by using leverage, the value of a discrepancy, the value of influence. Robust regression is a regression method that is used when the distribution of residuals is not normal and it contains a few outliers or influential on the model (Ryan, 1997). Least Trimmed Square (LTS) is a robust regression estimator method which uses the concept of pruning to minimize the sum of squared residuals with a value breakdown point of $[(n-p)/2+1]/n$. *M-estimation is an estimate that minimizes an objective function value of with breakdown point of 0. The purpose of this study was to compare the effectiveness of the method Least Trimmed Square (LTS) and the method of M-estimation in the data that contain outliers. Comparison of the two methods is conducted through a literature study involving two case examples. Then the comparison of the effectiveness of the two methods seen from the coefficient of determination (R^2) is obtained by using the formula or it could be using the software MINITAB 16.*

✉ Alamat korespondensi:
E-mail: zhuy_yhayhan91@yahoo.com

ISSN 2252-6943

Pendahuluan

Regresi merupakan suatu metode statistika yang digunakan untuk menyelidiki pola hubungan antara dua atau lebih variabel, yaitu variabel *dependent* (variabel terikat atau variabel *respon*) dan variabel *independent* (variabel bebas atau variabel *explanatory*). Bentuk model regresi linier sederhana seperti berikut :

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (1)$$

Sedangkan regresi linear berganda adalah regresi linear yang terdiri dari satu variabel respon dan lebih dari satu variabel prediktor. Bentuk model regresi linier berganda sebagai berikut :

$$y_i = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \varepsilon_i, i = 1, 2, ..., n \quad (2)$$

Dengan adanya *outlier* pada data mengakibatkan model regresi tidak memenuhi asumsinya dan model regresi tidak cocok (*fit*) terhadap data yang akan dimodelkan, karena nilai koefisien dari model regresi tersebut dipengaruhi oleh adanya outlier. Sehingga model yang dihasilkan tidak dapat digunakan untuk memprediksi, dan *outlier* pada regresi harus diatasi.

Pengidentifikasian data *outlier* dengan melihat nilai *leverage*, nilai *discrapancy*, dan nilai *influence*. Dan untuk estimasi digunakan metode *Least Trimmed Square* (LTS) dan metode *M-estimation*. Pengidentifikasian outlier serta pengestimasian regresi dapat dilakukan dengan bentuan software MINITAB 16. Data yang termasuk *outlier* adalah yang melebihi dari masing-masing *cutoff* yang telah ditentukan, sebagai berikut:

1. *Leverage* > $\frac{2(k+1)}{n}$
2. *Eksternally studentized residuals* > t_{tabel}
3. *DFFITS* > $2\sqrt{\frac{(k+1)}{n}}$
4. *Cook's Distance* > $F(0.5; p; n-p)$

Ordinary least square (OLS) bukan merupakan prosedur regresi yang *robust* terhadap adanya *outlier*, karena estimasinya menjadi tidak sesuai meskipun hanya dengan kehadiran satu *outlier* dalam data (Rousseeuw dan Leroy, 1987). Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi dari *residual* tidak normal dan atau mengandung beberapa *outlier* yang berpengaruh pada model (Ryan, 1997). Pada regresi *robust*, taksiran yang *robust* terhadap *outlier* (tidak terpengaruh oleh adanya *outlier*) akan dicari sehingga *outlier* yang ada tidak perlu dikeluarkan dari analisis.

Metode *robust* yang akan dipakai pada tugas akhir ini adalah *Least Trimmed Square* (LTS) dan *M-estimation*. Menurut Rousseeuw (1984) metode LTS mampu mengatasi pencilan (*outlier*) yang disebabkan baik oleh variabel bebas maupun variabel terikatnya. LTS diusulkan oleh Rousseeuw (1984) sebagai alternatif *robust* untuk mengatasi kelemahan *ordinary least squares* (OLS), yaitu dengan menggunakan sebanyak h ($h \leq n$) kuadrat residual yang diturunkan nilainya.

$$\min_{\beta} \sum_{i=1}^{h} (\varepsilon^2)_{i;n} \quad (3)$$

dengan $h = \frac{(3n+p+1)}{4}$ (4)

dimana $(\varepsilon^2)_{i;n}$ adalah *residual* kuadrat ke-i, yang kemudian diurutkan dari nilai terkecil hingga paling besar: $(\varepsilon^2)_{1:n} \leq ... \leq (\varepsilon^2)_{n:n}$ dan $\varepsilon_i = y_i - \hat{y}$, untuk $i = 1,2,\ldots, n$. Nilai h adalah kostanta *trimming* yang memenuhi

$$\frac{(3n+p+1)}{4} < h < n.$$

Pada saat ini $h = n$, maka estimator LTS identik dengan estimator OLS. Nilai h akan membangun *breakdown point* yang besar sebanding dengan 50%.

Menurut Rousseuw, "*m-estimation* juga dianggap baik untuk mengestimasi parameter yang disebabkan oleh pencilan dan memiliki *breakdown point* 0".

$$\min_{\beta} \sum_{i=1}^{n} \rho(\varepsilon_i) = \min_{\beta} \sum_{i=1}^{n} \rho\left(y_i - \sum_{i=1}^{k} x_{ij}\beta_j\right)$$

Fungsi ρ merupakan representasi pembobot dari residual. Untuk memperoleh suatu skala invariant dari estimator ini, biasanya dilakukan dengan menyelesaikan persamaan

$$\min_\beta \sum_{i=1}^{n} \rho\left(\frac{\varepsilon_i}{s}\right) = \min_\beta \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{i=1}^{k} x_{ij}\beta_j}{s}\right)$$

dengan $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$ merupakan nilai *m-estimation* $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$ dari yang meminimumkan:

(5)

$$\sum_{i=1}^{n} \rho(\mu_i) = \sum_{i=1}^{n} \rho\left(\frac{\varepsilon_i}{s}\right) = \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{i=1}^{k} x_{ij}\beta_j}{s}\right)$$

dimana $\rho(\mu_i)$ adalah fungsi simetris dari residual atau fungsi yang memberikan kontribusi pada masing-masing residual pada fungsi objektif (Jacob,2003).

Pilihan estimasi yang populer untuk s adalah

$$s = \frac{median\{|\varepsilon_i - median(\varepsilon_i)|\}}{0,6745}$$

Pemilihan konstanta 0,6745 membuat sedemikian hingga *s* merupakan suatu estimator yang mendekati tak bias dari σ, jika *n* besar dan error berdistribusi normal (Montgomery, 1992). Untuk meminimumkan persamaan (5), turunan parsial pertama dari σ terhadap $\beta_j, j = 0, 1, ..., k$, harus disamakan dengan 0 yang merupakan kondisi minimum yang menghasilkan parameter sebanyak $p = (k + 1)$ dari sistem persamaan:

(6)

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{y_i - \sum_{i=1}^{k} x_{ij}\beta_j}{s}\right) = 0, j = 0, 1, ..., k$$

(Jacob,2003)

dengan $\psi = \rho'$ dan x_{ij} adalah observasi ke-i pada regresor ke-j dan x_{i0}=1.

Estimasi koefisien regresi dengan *m-estimation* dilakukan dengan estimasi kuadrat terkecil dengan pembobot iteratif. Prosedur tersebut dinamakan *Iteratively Reweighted Least Squares* (IRLS).

Sedangkan untuk fungsi pembobot dalam *m-estimation* yang digunakan adalah fungsi Huber:

$$W_H(u_i) = \begin{cases} 1 & \text{untuk } |u_i| \le r, \\ \frac{r}{|u_i|} & \text{when } |u_i| > r. \end{cases}$$

dengan r = 1,345.

Solusi dari menggunakan metode ini adalah melakukan *weigthed least squares* secara iterasi, sehingga diperoleh persamaan

(7)

$$\sum_{i=1}^{n} \frac{\psi(u_i)}{u_i}(u_i)x_i = 0$$

Dengan u_i adalah residual yang telah diskalakan, dimana $u_i = \varepsilon_i/s$, maka persamaan (6) dapat ditulis sebagai:

(8)

$$\sum_{i=1}^{n} W_i(u_i)x_i = 0$$

dengan $_i = \frac{\psi(u_i)}{u_i}$.

Dengan demikian persamaan (7) juga merupakan solusi jumlah kuadrat *error* terboboti (WLS), yaitu:

(9)

$$\sum_{i=1}^{n} W_i(y_i - \hat{y}_i)^2$$

Untuk kasus regresi berganda perhitungan parameternya dapat diperoleh dari persamaan matriks

(10)

$$X'W^0X\hat{\beta} = X'W^0Y$$

Pengidentifikasian adanya data pencilan serta untuk mengestimasi persamaan regresi robust dengan menggunakan bantuan *software* MINITAB 16. Paket program MINITAB merupakan perangkat lunak statistika yang dapat digunakan sebagai media pengolahan data yang menyediakan berbagai jenis perintah yang memungkinkan proses pemasukkan data, manipulasi data, pembuatan grafik, peringkasan nilai-nilai numerik, dan analisis statistika lainnya. Tujuannya adalah untuk mendapatkan hasil yang lebih akurat serta untuk mempermudah dan mempercepat proses identifikasi dan estimasi.

Sedangkan untuk melihat metode mana yang lebih efektif yaitu dengan melihat nilai dari koefisien determinasi (R^2) yang didapatkan dari hasil estimasi menggunakan *software* MINITAB 16. Dalam uji regresi linier sederhana maupun berganda, koefisien determinasi (R^2) digunakan untuk mengetahui presentase sumbangan pengaruh serentak variabel-variabel bebas terhadap variabel terikat. Kisaran nilai R^2 adalah 0 hingga 1. Makin dekat R^2 dengan 1 makin baik kecocokan data dengan model, dan sebaliknya, makin dekat R^2 dengan 0 makin jelek kecocokan tersebut (Sembiring, 1995: 47).

Metode

Langkah-langkah dalam penelitian ialah sebagai berikut:

1. Menggunakan contoh data yang mengandung permasalahan pencilan.
2. Membuat regresi awal dengan OLS
3. Mendeteksi *outlier*
4. Pengujian asumsi analisis regresi
5. Pendugaan parameter regresi robust dengan penduga LTS
6. Pendugaan parameter regresi robust dengan penduga *M-estimation*
7. Membandingkan nilai koefisien determinasi (R^2)

Hasil dan Pembahasan

Pendeteksian Pencilan

Pada kasus 1 diuraikan data dengan kehadiran pencilan. Data yang diambil adalah Data Ketahanan Pangan di Jawa Tengah Tahun 2007, dengan x: rata-rata produksi (ton/ha) dan y: rasio ketahanan pangan di Jawa Tengah seperti pada tabel 1.

Tabel 1. Contoh Data Regresi Sederhana

No	x	Y	No	x	Y
1	6,7	2,59	21	5,8	3,95
2	5,1	1,7	22	3,4	1,12
3	7,4	2,16	23	6,7	2,1
4	6,5	2,01	24	5,8	3,4
5	7,8	4,3	25	6,3	2,95
6	5,8	1,42	26	5,8	3,5
7	5,7	1,91	27	5,2	2,45
8	3,6	2,57	28	11,2	5,59
9	6	2,5	29	5,2	2,71
10	3,7	2,4	30	5,8	2,58
11	6,3	4,13	31	3,2	0,74
12	6,7	1,86	32	8,3	2,52
13	5,8	3,95	33	5	3,5
14	7,7	3,4	34	5,8	3,3
15	7,4	2,4	35	5,4	2,64
16	6	2,98	36	5,3	2,6
17	3,7	1,55	37	2,6	2,05
18	7,3	3,56	38	4,3	2,85
19	5,6	3,02	39	4,8	2,45
20	5,2	2,85	40	5,4	1,81

Sumber: www.bps.go.id

Pada kasus 2 data yang diambil adalah data yang diperoleh dari (Soemartini, 2007) yang terdiri dari dua variabel independen yaitu: *prognostic index* (x_1), *enzyme function test* (x_2) dan variabel dependennya adalah *survival time* dengan data berjumlah 30 buah seperti pada tabel 2.

Tabel 2. Data regresi berganda

No	y	x_1	x_2	No	y	x_1	x_2
1	43	88	6	16	215	59	52
2	127	86	73	17	70	4	52
3	191	78	59	18	58	23	45
4	71	46	73	19	168	106	4
5	172	56	61	20	136	76	86
6	144	101	85	21	184	38	59
7	200	73	83	22	34	41	51
8	220	85	72	23	58	93	38
9	153	72	49	24	178	72	46
10	204	63	85	25	101	45	40
11	217	84	84	26	101	83	28
12	65	99	58	27	65	70	67
13	203	77	52	28	151	86	5
14	118	120	59	29	87	100	78
15	70	72	90	30	95	81	68

Sumber: Soemartini 2007

Pendeteksian pencilan pada kasus 1 dilakukan dengan metode *Leverage*, nilai discrepancy dengan menggunakan metode *externally studentized residuals*, dan nilai influence dengan menggunakan metode DFFITS dan Cook's Distance. Hasil perhitungan menggunakan *software* MINITAB 16 dapat dilihat pada tabel 3.

Data di deteksi sebagai outlier apabila:

1. *Leverage* > $\dfrac{2(k+1)}{n} = 0,1$

2. *Eksternally studentized residuals* > $t_{tabel} = 1,686$

3. *DFFITS* > $2\sqrt{\dfrac{(k+1)}{n}} = 0,45$

4. *Cook's Distance* > $F(0.5 ; p ; n\text{-}p) = 0,706$

Tabel 3. pemeriksaan *outlier* pada data sederhana

No	h_{ii}	*Externally studentized residuals*(t_i)	*DFFITS*	*cook's*
1	0,033734	-0,56423	-0,10542	0,005659
2	0,029833	-1,00912	-0,17696	0,015649
3	0,052146	-1,49758	-0,35126	0,059739
4	0,030342	-1,25138	-0,22136	0,02414
5	0,067233	1,214812	0,326147	0,052528
6	0,025003	-1,73977	-0,2786	0,036845
7	0,025071	-1,00458	-0,16109	0,012973
8	0,074423	0,843476	0,239178	0,028822
9	0,025491	-0,36094	-0,05838	0,001744
10	0,069998	0,563596	0,154621	0,012172
11	0,027779	1,68494	0,284812	0,038687
12	0,033734	-1,56347	-0,29213	0,041108
13	0,025003	1,673317	0,267963	0,034278
14	0,06315	0,042216	0,01096	0,000062
15	0,052146	-1,15845	-0,27172	0,036586
16	0,025491	0,263539	0,042623	0,000931
17	0,069998	-0,57173	-0,15685	0,012523
18	0,048894	0,436233	0,098908	0,004998
19	0,025346	0,497478	0,080223	0,003283
20	0,028521	0,457601	0,078406	0,003139
21	0,025003	1,673317	0,267963	0,034278
22	0,083896	-1,02421	-0,30995	0,047972
23	0,033734	-1,22524	-0,22893	0,025864
24	0,025003	0,909002	0,145566	0,010643
25	0,027779	0,088981	0,015041	0,000116
26	0,025003	1,043946	0,167176	0,013941
27	0,028521	-0,06417	-0,01099	0,000062
28	**0,329524**	1,630487	**1,14306**	0,625973
29	0,028521	0,274423	0,04702	0,001133
30	0,025003	-0,1661	-0,0266	0,000363
31	0,0942	-1,47573	-0,40655	0,109835
32	0,09076	-1,46165	-0,43145	0,103532
33	0,031353	1,432725	0,257763	0,032325
34	0,025003	0,775358	0,124165	0,00779
35	0,026518	0,092535	0,015273	0,00012
36	0,027416	0,085752	0,014397	0,000106
37	**0,13009**	0,627963	0,242839	0,029963
38	0,047804	0,881094	0,19742	0,019603
39	0,035016	0,117129	0,022312	0,000256
40	0,026518	-0,99958	-0,16498	0,013609

Pendeteksian pencilan pada kasus 2 dilakukan dengan metode *Leverage*, nilai discrepancy dengan menggunakan metode *externally studentized residuals*, dan nilai influence dengan menggunakan metode DFFITS dan Cook's Distance. Hasil perhitungan menggunakan *software* MINITAB 16 dapat dilihat pada tabel 4.

Data di deteksi sebagai outlier apabila:

1. *Leverage* > $\frac{2(k+1)}{n} = 0,2$

2. *Eksternally studentized residuals* > t_{tabel} = 1,703

3. *DFFITS* > $2\sqrt{\frac{(k+1)}{n}} = 0,632$

4. *Cook' s Distance* > *F*(0.5 ; *p* ; *n-p*) = 0,81

Tabel 4. pemeriksaan *outlier* data berganda

No	h_{ii}	*Externally studentized residuals*	DFFITS	Cook's
1	**0,201018**	-1,27815	-0,64111	0,133865
2	0,059945	-0,27137	-0,06853	0,001621
3	0,035245	0,985086	0,188285	0,01183
4	0,085143	-0,99252	-0,30279	0,030577
5	0,04884	0,787691	0,17849	0,010771
6	0,129159	-0,1875	-0,07221	0,001802
7	0,074969	0,988962	0,281541	0,026443
8	0,056501	1,359882	0,332782	0,035789
9	0,037223	0,455919	0,089646	0,00276
10	0,085141	1,118915	0,341341	0,038479
11	0,086749	1,222279	0,376711	0,046454
12	0,071293	-1,3278	-0,36789	0,043874
13	0,035776	1,269672	0,244567	0,019496
14	0,155639	-0,55637	-0,23887	0,019518
15	0,100151	-1,36188	-0,45434	0,066698
16	0,045098	1,63166	0,354592	0,039481
17	**0,290505**	-0,61738	-0,39505	0,053242
18	0,177715	-0,86624	-0,40271	0,054562
19	**0,256377**	0,998889	0,586516	0,114676
20	0,086153	-0,16591	-0,05094	0,000897
21	0,097796	1,171205	0,385604	0,04889
22	0,090251	-1,44808	-0,4561	0,066635
23	0,075905	-1,23437	-0,35377	0,040924
24	0,040695	0,912539	0,18795	0,011848
25	0,094245	-0,17429	-0,05622	0,001093
26	0,088934	-0,32002	-0,09999	0,003447
27	0,039761	-1,18479	-0,24109	0,019089
28	**0,204631**	0,774596	0,392895	0,052229
29	0,104009	-1,13225	-0,38577	0,049093
30	0,045131	-0,74002	-0,16088	0,008775

Dengan melihat nilai-nilai *Leverage, Externally studentized residuals, DFFITS, Cook'D* dengan bantuan *software* MINITAB 16. Data yang termasuk pencilan pada kasus 1 adalah observasi ke-28 dan ke-37. Sedangkan pada kasus 2, yang termasuk pencilan yaitu observasi ke-1, ke-17, ke-19, dan ke-28.

Penaksiran Parameter Berdasarkan Metode OLS, LTS, dan M-estimation

Hasil penaksiran parameter berdasarkan metode kuadrat terkecil untuk regresi linier sederhana dan berganda dengan model penaksir

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_i X_i$$

ialah sebagai berikut:

Data regresi sederhana:

$$\hat{Y} = 0,6922 + 0,34753X$$

Data regresi berganda:

$$\hat{Y} = 74,21 + 0,3797X_1 + 0,4954X_2$$

Selain itu, penaksiran parameter berdasarkan *least trimmed squares* (LTS) ialah dengan mengurutkan nilai sisaan kuadrat $\varepsilon^2_{(i)}$ dari terkecil hingga terbesar menjadi sebanyak *h*. Hasil penaksiran untuk data dengan model penaksir

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_i X_i$$

ialah sebagai berikut:

Data regresi sederhana:

$$\hat{Y} = 0,4671 + 0,42454X$$

Data regresi berganda:

$$\hat{Y} = 51,6 + 0,671X_1 + 0,5847X_2$$

Selanjutnya, penaksiran parameter berdasarkan *m-estimation* dapat diolah dengan bantuan *software* MINITAB 16. Hasil penaksiran untuk data dengan model penaksir

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_i X_i$$

ialah sebagai berikut:

Data regresi sederhana:

$$\hat{Y} = 0,693 + 0,347X$$

Data regresi berganda:

$$\hat{Y} = 74,2 + 0,3801X_1 + 0,496X_2$$

Secara ringkas, nilai koefisien regresi dan nilai koefisien determinasi (R^2) data dari ketiga metode dapat dilihat pada tabel 5 berikut:

Simpulan dan Saran

Kesimpulan dari hasil penelitian ialah:

Pendeteksian *outlier* yang didilakukan terdiri dari pendeteksian *leverage*, pendeteksian *discrepancy*, dan nilai *influence*. Untuk pendeteksian *leverage* dapat digunakan deteksi menggunakan nilai h_{ii}, pendeteksian nilai *discrepancy* menggunakan *externally studentized residual* (t_i), pendeteksian nilai *influence* dapat digunakan *DFFITS* dan *cook's distance*. Ketiga metode ini dihitung dengan menggunakan bantuan *software* MINITAB 16 yang kemudian dibandingkan dengan suatu nilai *cutoff* . Jika ketiga nilai tersebut melebihi masing-masing nilai *cutoff*-nya maka suatu data dideteksi sebagai *outlier.* Kehadiran data *outlier* pada regresi akan memberikan nilai-nilai konstanta dan koefisien pada model regresi membesar jika menggunakan metode *least square*.

Pada metode *least trimmed square* dan *M-estimation* tidak terpengaruh oleh kehadiran data *outlier.* Nilai *breakdown point* untuk data yang mengandung pencilan pada regresi *robust* dengan menggunakan metode LTS adalah

$$\frac{\left(\left[\frac{n-v}{2}\right]+1\right)}{n}.$$

Sedangkan nilai *breakdown point* pada regresi *robust* menggunakan *m-estimation* adalah 0. Dan nilai koefisien determinasi dari metode *least trimmed square* lebih besar dibandingkan dengan metode *m-estimation* dan *metode least square*.

Berdasarkan nilai koefesien determinasi (R^2) pada contoh kasus 1 dapat dikatakan bahwa metode *least trimmed square* (LTS) lebih efektif dari pada metode *M-estimation*, dan

Tabel 5. Hasil Estimasi Koefisien Regresi dan Rata-rata Kuadrat Sisa

Metode		OLS	LTS	M-est
Data regresi sederhana	b_0	0,6922	0,4671	0,693
	b_1	0,34753	0,42454	0,347
Koef. Determinasi (R^2)		**34,1%**	**85,2%**	**34,3%**
Data regresi berganda	b_0	74,21	51,6	74,2
	b_1	0,3797	0,671	0,3801
	b_2	0,4959	0,585	0,496
Koef. Determinasi (R^2)		**6,2%**	**21,6%**	**6,2%**

metode *M-estimation* lebih efektif dari pada metode *Least Square*. Keefektifan terjadi apabila nilai dari koefesien determinasi (R^2) besar. Sedangkan untuk contoh kasus 2 metode LTS juga menunjukkan koefisien determinasi (R^2) yang lebih besar dari metode *M-estimation* dan metode *least square*. Dan nilai koefisien determinasi (R^2) dari metode *M-estimation* sama dengan metode *least square*. Jadi pada regresi *robust*, metode *Least Trimmed Square* (LTS) lebih efektif dibandingkan metode *M-estimation* dan metode OLS dilihat dari nilai koefisien determinasi (R^2). Hal ini disebabkan karena adanya pemangkasan terhadap data yang mempunyai residual besar, sehingga berpengaruh pada nilai koeffisien determinasi (R^2) dan membuat variabel bebas menjadi lebih kuat memprediksikan variabel terikatnya.

Berdasarkan kesimpulan diatas, maka dalam pengestimasian parameter model dalam regresi linier pada data yang mengandung pencilan dapat digunakan metode *Least Trimmed Squares* (LTS) dan *m-estimation*. Untuk penelitian selanjutnya dapat dikembangkan beberapa metode pengestimasian metode *robust* yang lain seperti *Least Median Square* (LMS), *Least Absolute Value* (LAV), *MM-estimation, S-estimation* dan lainnya.

Ucapan Terima Kasih

Terima kasih kepada bapak, ibu, teman-teman yang telah membantu dan pihak-pihak yang terkait dalam penulisan artikel ilmiah ini.

DAFTAR PUSTAKA

Rousseeuw, P.J. 1984. Least Median Squares Regression. *Journal of the American Statistical Association*. Vol. 79. Number 388.

Rousseeuw, Peter J., & Annick M. Leroy. 1987. *Robust Regression and Outlier Detection*. Canada: John Wiley & Sons, Inc.

Ryan, T.P. 1997. *Modern Regression Methods*. Canada: John Wiley & Sons, Inc.

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

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