Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)
Volume 08, No. 4 (2019), hal 869 – 874.

METODE *LEAST MEDIAN SQUARE* (LMS) DALAM ANALISIS REGRESI ROBUST KETIKA TERDAPAT *OUTLIER*

Mimi Kurniati, Yundari, Setyo Wira Rizki

INTISARI

Least median square (LMS) adalah salah satu metode estimasi dalam regresi robust yang digunakan untuk mengatasi outlier. Dalam metode ini, dengan meminimumkan median kuadrat sisaannya, penduga yang dihasilkan akan lebih resisten terhadap outlier. Data yang digunakan pada penelitian ini adalah data sekunder tentang produksi jeruk di Indonesia tahun 2016 yang diambil dari Kementrian Pertanian. Ukuran data yang digunakan adalah sebanyak 34 yaitu banyak provinsi di Indonesia. Dengan variabel dependen adalah produksi jeruk, dan variabel bebas (independen) yang diambil sebanyak tiga yaitu luas panen, curah hujan, dan suhu. Proses pertama adalah mendeteksi apakah ada outlier pada data, dan melakukan uji asumsi klasik. Kemudian mencari model regresi dengan metode Least median square (LMS). Apabila dibandingkan, hasil model regresi produksi jeruk dengan metode MLS lebih akurat daripada hasil model regresi dengan metode MKT. Dengan demikian dapat disimpulkan bahwa metode regresi robust Least median square (LMS) cukup layak untuk digunakan sebagai alternatif dalam mencari model regresi pada data produksi jeruk di Indonesia tahun 2016 yang mengandung outlier.

Kata Kunci : *Least Median Square, metode kuadrat terkecil, outlier.*

PENDAHULUAN

Analisis regresi adalah sebuah analisis statistik yang digunakan untuk menyelidiki dan membangun suatu model matematis untuk menjelaskan bentuk hubungan antara variabel. Analisis regresi yang dilakukan untuk satu variabel independen dan satu variabel dependen disebut analisis regresi sederhana. Jika terdapat beberapa variabel independen maka disebut regresi linear berganda [1]. Salah satu metode yang digunakan untuk mengestimasi parameter regresi adalah Metode Kuadrat Terkecil (MKT). Prinsip MKT adalah meminimumkan jumlah kuadrat sisaan (*error*). Metode kuadrat terkecil harus memenuhi asumsi-asumsi yang ada sehingga hasil estimasinya memenuhi sifat *best linear unbiased estimator* (BLUE). Metode kuadrat terkecil mempunyai asumsi-asumsi tertentu yaitu homoskedastisitas, tidak terjadi autokolerasi, tidak terjadi multikolinearitas, dan *error* berdistibusi normal. Jika tidak memenuhi salah satu asumsi misalnya disebabkan adanya *outlier*, maka penduga MKT yang diperoleh menjadi tidak efisien [2].

Dalam beberapa kasus dimungkinkan adanya data yang jauh dari pola kumpulan dan keseluruhan, yang lazim didefinisikan sebagai data *outlier* [3]. Jadi *outlier* adalah data pengamatan yang sangat berbeda dari data yang lain. *Outlier* dapat secara serius mengganggu tingkat keabsahan metode kuadrat terkecil [1]. Regresi robust merupakan metode regresi yang digunakan ketika distribusi dari sisaan tidak normal atau ada beberapa *outlier* yang berpengaruh pada model. Metode ini merupakan alat penting untuk menganalisis data yang dipengaruhi oleh *outlier* sehingga dihasilkan model yang tahan terhadap *outlier* [4]. *Least median square* (LMS) adalah salah satu metode estimasi dalam regresi robust yang digunakan untuk mengatasi *outlier*. Dalam metode ini, dengan meminimumkan median kuadrat sisaannya, penduga yang dihasilkan akan lebih resisten terhadap *outlier* [5].

Penelitian ini bertujuan untuk mempelajari dan memahami tentang regresi robust. Selain itu tujuan dari penelitian ini adalah untuk mendeteksi *outlier* pada data yang digunakan dan menentukan persamaan regresi linear berganda yang mengandung *outlier* dengan menggunakan metode regresi robust *Least median square* (LMS).

Data yang digunakan dalam penelitian ini berupa data sekunder tentang produksi jeruk di Indonesia tahun 2016 yang diambil dari Kementrian Pertanian. Ukuran data yang digunakan adalah sebanyak 34 yaitu banyak provinsi di Indonesia. Variabel bebas (independen) yang diambil sebanyak tiga yaitu luas panen, curah hujan, dan suhu. Langkah yang dilakukan dalam mengestimasi parameter pada regresi robust dengan metode *Least Median Square* (LMS) adalah melakukan estimasi koefisien regresi pada data dengan metode kuadrat terkecil (MKT). Kemudian menguji asumsi klasik dari model regresi dan mendeteksi adanya *outlier* pada data.

REGRESI ROBUST

Regresi robust merupakan metode regresi yang digunakan ketika distribusi dari sisaan tidak normal atau ada beberapa *outlier* yang berpengaruh pada model. Metode ini merupakan alat penting untuk menganalisis data yang dipengaruhi oleh *outlier* sehingga dihasilkan model yang tahan terhadap *outlier* [4]. Ketika menyusun model regresi dan melakukan uji asumsi sering ditemui bahwa asumsi regresi dilanggar, transformasi yang dilakukan tidak akan menghilangkan atau melemahkan pengaruh dari *outlier* yang akhirnya prediksi menjadi bias. Dalam keadaan ini, regresi robust yang tahan terhadap pengaruh *outlier* adalah metode yang terbaik. Regresi robust adalah metode yang penting untuk menganalisis data yang terkontaminasi oleh *outlier*. Regresi robust terdiri dari 5 metode estimasi, antara lain: M-estimator, *Least Median Square* (LMS)-estimator, *Least Trimmed Square* (LTS)-estimator, S-estimator, dan MM-estimator [6].

METODE *LEAST MEDIAN SQUARE* (LMS)

Least median square (LMS) adalah salah satu metode estimasi dalam regresi robust yang digunakan untuk mengatasi *outlier*. Pada metode MLS, dengan meminimumkan median kuadrat sisaannya, penduga yang dihasilkan akan lebih resisten terhadap *outlier* [5]. Metode ini adalah salah satu metode estimasi dalam regresi robust, dimana data *outlier* yang ada tidak dibuang begitu saja, tetapi diproses dan dieliminasi melalui sebuah iterasi. Jika pada metode kuadrat terkecil hal yang perlu dilakukan adalah meminimumkan jumlah kuadrat sisaan $\left(\sum_{i=1}^{n} e_i^2\right)$, maka pada LMS hal yang perlu dilakukan adalah mencari nilai median kuadrat sisaan pada setiap iterasi, yaitu:

$$M_j = med\ e_i^2 \tag{1}$$

sehingga terbentuk $M_1, M_2, \dots, M_s$ yang merupakan median kuadrat sisaan dari setiap h_i pengamatan. Untuk mendapatkan nilai M_1, dicari himpunan bagian data sejumlah h_i pengamatan, yaitu:

$$h_i = \left[\frac{n}{2}\right] + \left[\frac{p+1}{2}\right] \tag{2}$$

dengan n adalah banyaknya data, dan p adalah banyaknya parameter. Dalam perhitungan nilai h_i harus selalu dalam bentuk bilangan bulat. Oleh karena itu, jika nilai h_i bukan dalam bentuk bilangan bulat maka dilakukan pembulatan ke atas. Demikian seterusnya sampai iterasi berakhir pada iterasi ke-i yaitu saat $h_i = h_{i+1}$. Setelah itu dicari nilai minimum dari $M_1, M_2, \dots, M_s$ [7].

Karena LMS merupakan penduga pada regresi robust, maka sama halnya dengan penduga lain pada regresi robust, prinsip dasar dari LMS adalah dengan memberikan bobot w_{ii} pada sehingga data *outlier* tidak mempengaruhi model parameter taksiran. Bobot w_{ii} dirumuskan dengan ketentuan sebagai berikut:

$$w_{ii} = \begin{cases} 1 & \text{jika } |e_i/\hat{\sigma}| \leq 2,5 \\ 0 & \text{lainnya} \end{cases} \tag{3}$$

dengan;

$\hat{\sigma} = 1,4826\left[1 + \frac{5}{(n-p)}\right]\sqrt{M_j min}$

Setelah bobot w_{ii} dihitung, dapat dibentuk matrik W sebagai berikut:

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & 0 & \dots & 0 \\ 0 & w_{22} & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & w_{nn} \end{bmatrix} \quad (4)$$

Setelah terbentuk matriks W, maka penaksiran parameter regresi LMS dapat dihitung dengan menggunakan rumus [7]:

$$\hat{\boldsymbol{\theta}}_{LMS} = \left(\boldsymbol{X^T W X}\right)^{-1}\left(\boldsymbol{X^T W Y}\right) \quad (5)$$

MODEL REGRESI DENGAN METODE *LEAST MEDIAN SQUARE* (LMS)

Sebelum melakukan pemodelan dilakukan terlebih dahulu pendeteksian *outlier*. Pendeteksian *outlier* dilakukan dengan menggunakan metode DfFITS (*Difference fitted value FITS*) atau *Standardized* DfFITS. Suatu data dikatakan terdeteksi outlier apabila nilai $|DfFITS| > 2\sqrt{\frac{p}{n}}$. Hasil diagnosis metode DfFITS terhadap data dengan p adalah banyak parameter dana n adalah banyaknya data diperoleh nilai $|\text{DfFITS}| = 2\sqrt{\frac{p}{n}} = 2\sqrt{\frac{4}{34}} = 0{,}686$. Berdasarkan hasil perhitungan diperoleh bahwa data ke- 2, 15 dan 17 terdeteksi outlier. Sedangkan dengan menggunakan software MINITAB terdeteksi adanya *outlier* sebanyak 4 pengamatan, yaitu data ke- 2, 12, 15, dan 17.

Selanjutnya dilakukan estimasi koefisien regresi pada data dengan metode kuadrat terkecil (MKT). Didapat model sebagai berikut:

$$Y_i = -309 + 31{,}7\, X_{1i} + 17\, X_{2i} - 45\, X_{3i}$$

Adapun langkah-langkah pemodelan regresi dengan LMS diawali dengan penentuan M_j. Untuk menentukan nilai M_j yaitu median dari error kuadrat pada iterasi ke-i, terlebih dahulu dicari nilai error kuadrat dari MKT dengan jumlah pengamatan pada iterasi ke-i.

Diberikan $i = 1, 2, \dots, l$ dan berhenti saat $h_i = h_{i+1}$, diperoleh iterasi sebagai berikut:

(i) Iterasi ke-1 dengan $n = 34$ diperoleh
$M_1 = median\left\{e_i^2 : i = 1,2, \dots, 34\right\} = 8214671{,}83$, dan
$h_1 = \left[\frac{n}{2}\right] + \left[\frac{p+1}{2}\right] = \left[\frac{34}{2}\right] + \left[\frac{4+1}{2}\right] = \frac{39}{2} = 20$
Artinya pada iterasi ke-2 akan diambil 20 pengamatan yang jarak nilai e_i^2 ke M_1 nya minimum. Dengan kata lain dihilangkan sebanyak 14 pengamatan.

(ii) Iterasi ke-2 dengan $n = 20$, melalui MKT diperoleh model sebagai berikut:
$Y_i = 18987 + 35{,}2\, X_{1i} - 6{,}75\, X_{2i} - 661\, X_{3i}$
dengan demikian diperoleh;
$M_2 = median\left\{e_i^2 : i = 1,2, \dots, 20\right\} = 425581{,}94$, dan
$h_2 = \left[\frac{n}{2}\right] + \left[\frac{p+1}{2}\right] = \left[\frac{20}{2}\right] + \left[\frac{4+1}{2}\right] = 13$
Artinya pada iterasi ke-3 akan diambil 13 pengamatan yang jarak nilai e_i^2 ke M_2 nya minimum. Dengan kata lain dihilangkan sebanyak 7 pengamatan.

(iii) Iterasi ke-3 dengan $n = 13$, melalui MKT diperoleh model sebagai berikut:
$Y_i = 15891 + 35{,}9\, X_{1i} - 3{,}47\, X_{2i} - 562\, X_{3i}$
dengan demikian diperoleh;
$M_3 = median\left\{e_i^2 : i = 1,2, \dots, 13\right\} = 85703{,}62$, dan
$h_3 = \left[\frac{n}{2}\right] + \left[\frac{p+1}{2}\right] = \left[\frac{13}{2}\right] + \left[\frac{4+1}{2}\right] = 9$

Artinya pada iterasi ke-4 akan diambil 9 pengamatan yang jarak nilai e_i^2 ke M_1 nya minimum. Dengan kata lain dihilangkan sebanyak 4 pengamatan.

(iv) Iterasi ke-4 dengan $n = 9$, melalui MKT diperoleh model sebagai berikut:

$Y_i = 16168 + 36{,}4\, X_{1i} - 3{,}78\, X_{2i} - 574\, X_{3i}$

dengan demikian diperoleh;

$M_4 = median\left\{e_i^2 : i = 1,2,\dots,9\right\} = 54019{,}43,$ dan

$h_4 = \left[\frac{n}{2}\right] + \left[\frac{p+1}{2}\right] = \left[\frac{9}{2}\right] + \left[\frac{4+1}{2}\right] = 7$

Artinya pada iterasi ke-5 akan diambil 7 pengamatan yang jarak nilai e_i^2 ke M_1 nya minimum. Dengan kata lain dihilangkan sebanyak 2 pengamatan.

(v) Iterasi ke-5 dengan $n = 7$, melalui MKT diperoleh model sebagai berikut:

$Y_i = 10838 + 34{,}9\, X_{1i} - 2{,}70\, X_{2i} - 385\, X_{3i}$

dengan demikian diperoleh;

$M_5 = median\left\{e_i^2 : i = 1,2,\dots,7\right\} = 22645{,}43,$ dan

$h_5 = \left[\frac{n}{2}\right] + \left[\frac{p+1}{2}\right] = \left[\frac{7}{2}\right] + \left[\frac{4+1}{2}\right] = 6$

Artinya pada iterasi ke-6 akan diambil 6 pengamatan yang jarak nilai e_i^2 ke M_1 nya minimum. Dengan kata lain dihilangkan sebanyak 1 pengamatan.

(vi) Iterasi ke-6 dengan $n = 6$, melalui MKT diperoleh model sebagai berikut:

$Y_i = 12263 + 35\, X_{1i} - 2{,}8\, X_{2i} - 436\, X_{3i}$

dengan demikian diperoleh;

$M_6 = median\left\{e_i^2 : i = 1,2,\dots,6\right\} = 20680{,}14,$ dan

$h_6 = \left[\frac{n}{2}\right] + \left[\frac{p+1}{2}\right] = \left[\frac{6}{2}\right] + \left[\frac{4+1}{2}\right] = 6$

Iterasi berakhir pada iterasi ke-5, karena $h_5 = h_6$. Dengan berakhirnya proses iterasi sebanyak 5 kali, maka diperoleh 5 buah nilai M_j yaitu;

Tabel 1. Nilai M_j

Iterasi	M_j
1	8214671,83
2	425581,94
3	85703,62
4	54019,43
5	22645,43
Min	22645,43

Penentuan parameter regresi LMS didapat berdasarkan hasil perhitungan M_j dan $\hat{\sigma}$ yang kemudian dihitung bobot w_{ii}. Dengan mensubsitusikan nilai $n = 34$, $p = 4$, dan $M_j = 22645{,}43$ ke dalam persamaan:

$$\begin{aligned}\hat{\sigma} &= 1{,}4826\left[1 + \frac{5}{(n-p)}\right]\sqrt{M_j min} \\ &= 1{,}4826\left[1 + \frac{5}{(34-4)}\right]\sqrt{22645{,}43} \\ &= 260{,}29\end{aligned}$$

Selanjutnya dengan menggunakan rumus di bawah ini diperoleh nilai penaksiran parameter

$$\hat{\boldsymbol{\theta}}_{LMS} = \left(\boldsymbol{X^T W X}\right)^{-1}\left(\boldsymbol{X^T W Y}\right)$$

$$= \left(\begin{bmatrix} 1 & 198 & 127,28 & 27,34 \\ 1 & 6663 & 155,93 & 27,52 \\ \vdots & \vdots & \vdots & \ddots\ \vdots \\ 1 & 205 & 167,28 & 27,29 \end{bmatrix}^{\boldsymbol{T}} \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} 1 & 198 & 127,28 & 27,34 \\ 1 & 6663 & 155,93 & 27,52 \\ \vdots & \vdots & \vdots & \ddots\ \vdots \\ 1 & 205 & 167,28 & 27,29 \end{bmatrix}\right)^{-1}$$

$$\left(\begin{bmatrix} 1 & 198 & 127,28 & 27,34 \\ 1 & 6663 & 155,93 & 27,52 \\ \vdots & \vdots & \vdots & \ddots\ \vdots \\ 1 & 205 & 167,28 & 27,29 \end{bmatrix}^{\boldsymbol{T}} \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} 6896 \\ 459149 \\ \vdots \\ 18352 \end{bmatrix}\right)$$

$$= \begin{pmatrix} 856,341999 & -0,092568 & -0,167197 & -29,950513 \\ -0,092568 & 0,000062 & -0,000036 & 0,00333 \\ -0,167197 & -0,000036 & 0,000202 & 0,005174 \\ -29,950513 & 0,00333 & 0,005174 & 1,050896 \end{pmatrix} \begin{pmatrix} 18913 \\ 2890437 \\ 3058068,86 \\ 514346,95 \end{pmatrix}$$

$$= \begin{pmatrix} 12176,37512 \\ 31,18644 \\ 12,469467 \\ -480,128192 \end{pmatrix}$$

Dengan demikian model regresi dengan metode regresi robust *Least median square* (LMS) adalah

$$\hat{y}_i = 12176,37512 + 31,18644\, x_{1i} + 12,469467\, x_{2i} - 480,128192\, x_{3i}$$

Dalam penelitian ini digunakan RMSE untuk mengetahui seberapa akurat model yang dibentuk.

Tabel 2. Tabel Perbandingan Keakuratan Model

Metode	**RMSE**
MKT	55616,97
LMS	55679

Berdasarkan Tabel 2, dapat disimpulkan bahwa model yang didapatkan dari metode MKT lebih akurat.

PENUTUP

Dalam pendeteksian *outlier*, dengan menggunakan software MINITAB terdeteksi adanya *outlier* sebanyak 2 pengamatan, yaitu data ke- 2 dan 17. Sedangkan dengan menggunakan metode $DfFITS$, terdeteksi adanya *outlier* sebanyak 3 pengamatan, yaitu data ke- 2, 15 dan 17. Model regresi yang diperoleh menurut metode MKT dan LMS berturut-turut:

$$\hat{y}_i = -309 + 31,7\, x_{1i} + 17\, x_{2i} - 45\, x_{3i}$$

dan,

$$\hat{y}_i = 12176,37512 + 31,18644\, x_{1i} + 12,469467\, x_{2i} - 480,128192\, x_{3i}$$

Apabila dibandingkan, hasil model regresi dengan metode MKT lebih akurat daripada hasil model regresi dengan metode LMS. Dengan demikian dapat disimpulkan bahwa metode regresi robust *Least median square* (LMS) tidak cukup layak untuk digunakan sebagai alternatif dalam mencari model regresi pada data produksi jeruk di Indonesia tahun 2016 yang mengandung *outlier*.

DAFTAR PUSTAKA

[1]. Montgomery DC. and Peck. *Introduction to Linear Regression Analysis*. New York: John Wiley & Sons, Inc; 1991.

[2] Gujarati DN dan Porter DC. *Dasar-dasar Ekonometrika Buku 1 Edisi Kelima*. Jakarta: Salemba Empat; 2010.

[3]. Rousseeuw PJ And Zomeren BC. Unmasking Multivariate Outliers and Leverage Points. *Journal of the American Statistical Association*. **85**: 633-639; 1990.

[4]. Draper NR and Smith H. *Analisis Regresi Terapan Edisi kedua*. Bambang Sumantri, penerjemah. Jakarta: Gramedia Pustaka Utama. Terjemahan dari: *Applied Regression Analysis* 2^{nd} *edition*; 1992.

[5]. Rousseeuw PJ. Least Median of Squares Regression. *Journal of the American Statistical Association*. **79**: 871-880; 1984.

[6]. Chen, C,. Robust Regression and Outlier Detection with the ROBUSTREG Procedure. Paper 265-27. North Carolina: SAS Institute; 2002.

[7]. Parmikanti K. Model regresi kandungan batubara menggunakan metode Least Median of Squares. *Prosiding Seminar Nasional Sains dan Teknologi Nuklir PTNBR*; 2013.

MIMI KURNIATI : Jurusan Matematika FMIPA UNTAN, Pontianak,
mimi.kurniati07@gmail.com

YUNDARI : Jurusan Matematika FMIPA UNTAN, Pontianak,
yundari@math.untan.ac.id

SETYO WIRA RIZKI : Matematika FMIPA UNTAN, Pontianak,
setyo.wirarizki@math.untan.ac.id