ISBN:1858-1382

J M S K
Jurnal Matematika, Statistika, & Komputasi
http://journal.unhas.ac.id/index.php/jmsk
Vol. 10, No. 2, 114-123, Januari 2014

PENGGUNAAN REGRESI *ROBUST* PADA DATA YANG MENGANDUNG PENCILAN DENGAN METODE MOMEN

Nurmiati Nurdin, Raupong, Anna Islamiyati

Abstrak

Analisis regresi merupakan sebuah alat statistika yang memberikan tentang pola hubungan antara dua variabel atau lebih. Salah satu metode yang umumnya digunakan dalam mengestimasi parameter pada analisis regresi linear adalah metode kuadrat terkecil (OLS). Namun metode ini mempunyai kelemahan apabila data terdeteksi mengandung *outlier*. Maka regresi *robust* disarankan dapat mengatasi masalah *outlier* dalam data untuk mengestimasi parameter, salah satunya adalah *Metode Momen* (MM) yang digunakan untuk data yang terdeteksi *outlier* pada variabel bebas dan variabel terikat serta memiliki nilai *breakdown point* yang tinggi.

Dalam skripsi ini dikaji tentang penggunaan Metode Momen dengan metode iterasi *Iteratively Reweighted Least Square* (IRLS). Metode Momen merupakan gabungan antara estimasi S dan estimasi M. Pada Metode Momen ini digunakan fungsi pembobot *Tukey Bisquare.*

Kata kunci : Analisis Regresi, OLS, Outlier, Breakdown Point, Regresi Robust, Metode Momen, Tukey Bisquare.

1. Pendahuluan

Regresi linier adalah metode statistika yang digunakan untuk membentuk model hubungan antara variabel terikat (*dependent*; respon; *Y*) dengan satu atau lebih variabel bebas (*independent*; prediktor; *X*). Secara umum regresi linear terdiri dari dua yaitu regresi linear sederhana dimana terdapat satu variabel terikat Y dan satu variabel bebas X sedangkan regresi linear berganda dimana terdapat satu variabel terikat Y dan beberapa variabel bebas X.

Pelanggaran asumsi yang sering terjadi pada data, biasanya disebabkan oleh adanya data pencilan. Menurut Soemartini (2007) pencilan adalah pengamatan yang jauh dari kelompok data yang mungkin berpengaruh besar terhadap koefesien regresi.

Regresi *robust* merupakan alat yang penting untuk menganalisis data yang terkontaminasi oleh pencilan. Regresi *robust* digunakan untuk mendeteksi pencilan dan memberikan hasil yang resisten terhadap adanya pencilan (Chen 2000).

Estimasi MM (*Method of Moment*), dikenalkan oleh Yohai (1987). Metode ini menggabungkan estimasi S (estimasi dengan *high breakdown point*) dan estimasi M. Estimasi-MM mempunyai kelebihan yaitu dapat digunakan untuk data yang terdeteksi pencilan pada variabel bebas dan variabel terikat .

Adapun tujuan penulisan ini adalah untuk mengestimasi parameter regresi *robust* menggunakan estimasi Metode Momen dan mendapatkan model regresi pada data indeks prestasi kumulatif mahasiswa yang mengandung pencilan dengan regresi *robust* melalui estimasi Metode Momen.[1]

[1,2,3] *Jurusan Matematika FMIPA Universitas Hasanuddin Makassar, Jl. Perintis Kemerdekaan Km.10 Makassar*

2. Tinjauan Pustaka

2.1 Regresi Linier

Analisis regresi merupakan sebuah alat statistika yang memberikan penjelasan tentang pola hubungan antara dua variabel atau lebih. Variabel terikat disebut juga variabel dependent yaitu variabel yang keberadaannya diperngaruhi oleh variabel lainnya dan dinotasikan dengan *Y*. Variabel bebas disebut juga variabel independent yaitu variabel yang tidak dipengaruhi oleh variabel lainnya dan dinotasikan dengan *X*.

2.1.1 Model Regresi Linier Sederhana

Bentuk hubungan yang paling sederhana antara variabel X dengan variabel Y adalah berbentuk garis lurus atau berbentuk hubungan linier yang disebut dengan regresi linier sederhana atau sering disebut regresi linier saja dengan persamaan matematikanya sebagaimana yang diungkapkan (Walpole, Ronald E, dkk. 1995) adalah sebagai berikut:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \qquad (1)$$

di mana: Y_i : variabel terikat, X_i : variabel bebas, α : intersep, β : koefisien regresi linier sederhana, ε_i : error atau kesalahan.

2.1.2 Model Regresi Linier Berganda

Hubungan fungsional atau hubungan kausal antara dua atau lebih variable yang dinyatakan dalam suatu bentuk fungsi linier pada umumnya dapat dinyatakan dalam bentuk persamaan matematika yang dibahas dalam analisis regresi. Untuk hubungan fungsional yang linier dapat dirumuskan dalam bentuk persamaan regresi sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i \quad ; \quad i = 1,2,\ldots,n \qquad (2)$$

di mana: Y_i : variabel terikat, X_{ik} : variabel bebas, β_0 : intersep, $\beta_1,\ldots,\beta_k$: koefisien-koefesien regresi, ε_i : galat (*error*).

2.2 Pencilan (*Outlier*)

Pencilan adalah pengamatan yang jauh dari pusat data yang mungkin berpengaruh besar terhadap koefesien regresi. Pencilan dapat muncul karena kesalahan dalam memasukkan data, kesalahan pengukuran, analisis, atau kesalahan-kesalahan lain. Pengaruh pencilan dalam analisis data dapat dibedakan berdasarkan asal pencilan tersebut yaitu yang berasal dari peubah respon (*y-outliers*; titik *influence*) atau berasal dari peubah bebasnya (*x-outliers*; titik *leverage*).

2.2.1 Identifikasi Pencilan

Terdapat beberapa metode untuk mengidentifikasi adanya pencilan yang berpengaruh dalam koefisien regresi antara lain :

1. Metode Grafis

Keuntungan dari metode ini yaitu mudah dipahami karena menampilkan data secara grafis (gambar) dan tanpa melibatkan perhitungan yang rumit. Sedangkan menurut Soemartini, kelemahan dari metode ini adalah keputusan bahwa suatu data merupakan pencilan sangat bergantung pada *judgement* peneliti, karena hanya mengandalkan visualisasi grafis, untuk itu dibutuhkan seseorang yang ahli dan berpengalaman dalam menginterpretasikan plot tersebut.

2. Metode *DfFITS* (*Difference fitted value FITS*) atau *Standardized DfFITS*

DfFITS merupakan suatu ukuran berpengaruh yang ditimbulkan oleh pengamatan ke-i terhadap nilai taksiran $\hat{y}_i$. Nilai $DfFITS_i$ diperoleh dari persamaan berikut:

$$(DfFITS)_i = \frac{\hat{y}_i - \hat{y}_{i-1}}{S_{i-1}^2 - \sqrt{h_{ii}}} \quad (3)$$

di mana h_{ii} : elemen diagonal ke-i dari matriks $H = X_i^T (X^T X)^{-1} X_i$
Suatu pengamatan ke-*i* data diidentifikasikan sebagai pencilan apabila nilai :

$$|DfFITS_i| > 1 \qquad untuk\ n \leq 30$$

$$|DfFITS_i| > 2\left(\frac{p}{n}\right)^{1/2} \qquad untuk\ n > 30$$

dengan *p* banyaknya parameter dan *n* banyaknya pengamatan .

3. Nilai Pengaruh (*Leverage Point*)

Metode yang digunakan dalam mengidentifikasi pencilan terhadap variabel X adalah nlai pengaruh (*Leverage Point*). Nilai pengaruh (h_{ii}) dari penamatan (X_i, Y_i) menunjukan besarnya peranan Y_i terhadap $\hat{Y}_i$ dan didefinisikan sebagai:

$$h_{ii} = X_i^T (X^T X)^{-1} X_i \quad ;\ i: 1,2,\ldots,n$$

dengan $X_i = \begin{bmatrix} X_{i1} & X_{i2} & \cdots & X_{ik} \end{bmatrix}$ adalah vektor baris yang berisi nilai-nilai dari peubah variabel bebas dalam pengamatan ke-*i*. Nilai h_{ii} berada diantara 0 dan 1 dengan $\sum_{i=1}^{n} h_{ii} = k$ dimana *k=p-1*. Sehingga dapat dituliskan menjadi

$$2\bar{h}_{ii} = \frac{2\sum_{i=1}^{n} h_{ii}}{n} = \frac{2k}{n} = \frac{2(p-1)}{n}$$

Suatu pengamatan ke-*i* data diidentifikasikan sebagai pencilan apabila nilai $h_{ii} > 2\bar{h}_{ii}$. Sehingga pengamatan ke-*i* dikatakan pencilan terhadap *X*.

2.3 Regresi *Robust*

Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi dari galat tidak normal dan atau adanya beberapa pencilan yang berpengaruh pada model (Ryan, 1997).

Menurut Chen (2002:1) metode-metode estimasi dalam regresi robust diantaranya adalah:

1. Estimasi M (*Maximum likelihood type*) yang dikenalkan oleh Huber (1973) adalah metode yang sederhana baik dalam penghitungan maupun secara teoritis. Estimasi ini menganalisis data dengan mengasumsikan bahwa sebagian besar yang terdeteksi pencilan pada variabel independen.
2. Estimasi LTS (*Least Trimmed Squares*) adalah metode dengan *high breakdown point* yang dikenalkan oleh Rousseeuw (1984). *Breakdown point* adalah ukuran proporsi minimal dari banyaknya data yang terkontaminasi pencilan dibandingkan seluruh data pengamatan.
3. Estimasi S (*Scale*) juga merupakan metode dengan *high breakdown point* yang dikenalkan oleh Rousseeuw and Yohai (1984). Dengan nilai *breakdown* yang sama, metode ini mempunyai efisiensi yang lebih tinggi dibanding estimasi LTS.
4. Estimasi MM (*Method of Moment*), dikenalkan oleh Yohai (1987). Metode ini menggabungkan estimasi S (estimasi dengan *high breakdown point*) dan estimasi M.

2.3.1 *Breakdown Point*

Breakdown point adalah salah satu cara yang digunakan untuk mengukur ke-*robust*-an (kekekaran) suatu estimator. *Breakdown point* merupakan proporsi minimal dari banyaknya pencilan dibandingkan seluruh data pengamatan.

Dengan kata lain, *breakdown point* sebagai suatu ukuran ke-r*obuts*-an dari suatu penaksir. Semakin besar nilai persen dari *breakdown point* pada suatu penaksir, maka penaksir tersebut semakin *robust*.

Regresi *robust* yang mempunyai *breakdown point* adalah regresi *robust* dengan metode estimasi S, LTS, LMS, dan MM. Metode estimasi MM mempunyai *breakdown point* 50%. *Breakdown point* 50% adalah *breakdown point* yang tinggi.

2.3.2 Fungsi Obyektif

Fungsi obyektif adalah fungsi yang digunakan untuk mencari fungsi pembobot pada regresi *robust*. Fungsi pembobot yang digunakan yaitu fungsi pembobot *Tukey Bisquare*.
Diberikan suatu fungsi obyektif sebagai berikut:

$$\rho(e_i^*) = \begin{cases} \frac{r^2}{6}\left[1-\left[1-\left(\frac{e_i^*}{r}\right)^2\right]^3\right], & |e_i^*| < r \\ \frac{r^2}{6}, & |e_i^*| \geq r \end{cases} \quad (4)$$

dengan fungsi *influence* yaitu:

$$\psi(e_i^*) = \rho'(e_i^*) = \frac{\partial\left(\rho(e_i^*)\right)}{\partial e_i^*} = \begin{cases} e_i^*\left[1-\left(\frac{e_i^*}{r}\right)^2\right]^2, & |e_i^*| < r \\ 0, & |e_i^*| \geq r \end{cases} \quad (5)$$

Sehingga diperoleh fungsi pembobot

$$w_i = w(e_i^*) = \frac{\psi(e_i^*)}{e_i^*} = \begin{cases} \left[1-\left(\frac{e_i^*}{r}\right)^2\right]^2, & |e_i^*| < r \\ 0, & |e_i^*| \geq r \end{cases} \quad (6)$$

di mana e_i^* : galat yang distandarisasi dan r : 4,685
Nilai r pada fungsi objektif, *influence* dan pembobot adalah *tunning constant*. Kelly (2006) menyatakan permasalahan dalam estimasi regresi robust adalah perlu dilakukan pemilihan *tunning constant* agar estimasi yang diperoleh lebih spesifik dan memimimumkan jumlah kuadrat galat. Menurunkan *tunning constant* akan menaikan pembobot terhadap galat yang besar. Menaikkan *tunning constant* akan menurunkan pembobot terhadap galat yang besar. Semakin besar r maka estimasi robust akan mendekati *least square*.

2.3.3 Estimasi MM

Estimasi MM menggabungkan estimasi *high breakdown point* dan efisiensi statistik yang dikenalkan oleh Yohai (1987). Langkah pertama dalam estimasi ini adalah mencari estimator S, kemudian menetapkan parameter-parameter regresi menggunakan estimasi M. Estimasi S menjamin nilai *breakdown point* tinggi dan estimasi M membuat estimator mempunyai efisiensi tinggi. Pada umumnya digunakan fungsi *Tukey Bisquare* baik pada estimasi S maupun estimasi M.
Bentuk dari metode estimasi MM:

$$\hat{\beta}_{MM} = \underset{\beta}{min} \sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}}\right) = \underset{\beta}{min} \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}}\right) \quad (7)$$

di mana ρ : fungsi obyektif dan $\hat{\sigma}$: estimator skala *robust*.

3. Metodologi Analisis

3.1 Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang dikumpulkan melalui pengambilan sampel nilai Ujian Nasional dan Indeks Prestasi Kumulatif selama tiga semester pada 150 orang mahasiswa jurusan matematika Fakultas Matematika dan Ilmu Pengetahuan Universitas Hasanuddin angkatan tahun 2010, 2011 dan 2012. Yang dimana setiap angkatannya mewakili 50 orang mahasiswa sebagai sampel. Dalam penelitian ini diasumsikan bahwa kondisi yang terjadi antara tiga tahun angkatan itu dianggap sama.

Variabel yang digunakan dalam penelitian ini adalah : Variabel terikat yaitu indeks prestasi kumulatif (y_i). Variabel bebas yaitu terdiri dari nilai bahasa Indonesia(x_{i1}), bahasa Inggris(x_{i2}), Matematika(x_{i3}), Fisika(x_{i4}), Kimia(x_{i5}) dan Biologi(x_{i6}).

3.2 Metode dan Analisis

Adapun langkah-langkah yang dilakukan berdasarkan tujuan penelitian adalah sebagai berikut :

a. Melakukan pengambilan data sekunder.
b. Mengidentifikasi adanya pencilan pada data menggunakan metode *DfFITS* dan *Leverage Point.*
c. Melakukan estimasi MM yaitu langkahnya sebagai berikut:
   1. Menghitung estimator awal koefisien $\hat{\beta}_j^{(1)}$ dan galat $e_i^{(1)}$ dari regresi *robust* dengan *high breakdown point* (estimasi S) dengan bobot *Tukey Bisquare* .
   2. Menghitung skala estimasi $\hat{\sigma}$ dan dihitung pula pembobot awal $w_i^{(1)}$ menggunakan galat $e_i^{(1)}$ pada langkah pertama.
   3. Menghitung koefisien regresi menggunakan galat $e_i^{(1)}$dengan skala estimasi $w_i^{(1)}$ pada langkah kedua.
   4. Menghitung bobot baru $w_i^{(2)}$ dengan skala estimasi dari iterasi awal.
   5. Mengulang langkah 2, 3, 4 (dengan skala estimasi tetap konstan) sampai mendapatkan $\sum_{i=1}^{n}\left|e_i^{(m)}\right|$ konvergen (selisih $\hat{\beta}_j^{(m+1)}$ dan $\hat{\beta}_j^{(m)}$ mendekati 0, dengan m banyaknya iterasi).
d. Uji signifikansi parameter antara variabel IPK dengan nilai-nilai Ujian Akhir Nasional.

4. Hasil dan Pembahasan

4.1 Estimasi Parameter Regresi Robust Menggunakan Metode Momen

Secara umum persamaan (2) merupakan model regresi linier berganda untuk data ke-i dan data n pengamatan yang dapat dituliskan sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i = \boldsymbol{X\beta} + \varepsilon_i$$

Taksiran modelnya yaitu:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \cdots + \hat{\beta}_k X_{ik} = \boldsymbol{X\hat{\beta}} \qquad (8)$$

Estimasi parameter menggunakan *OLS* menjadi kurang baik apabila distribusi galatnya tidak normal dan mengandung pencilan. Salah satu cara untuk mengatasinya adalah menggunakan regresi *robust.* Salah satu metode regresi *robust* yang digunakan adalah Metode Momen, yang diperkenalkan oleh Yohai pada tahun 1978 (Chen,2002).

Pada umumnya, estimasi MM meminimumkan fungsi obyektif dengan persamaan:

$$\sum_{i=1}^{n}\rho(e_i) = \sum_{i=1}^{n}\rho(Y_i - \boldsymbol{X\beta}) \qquad (9)$$

Selanjutnya dari persamaan (9) dicari turunan parsial pertama dari ρ terhadap β_j, $j = 0,1,\dots,k$ dan disamakan dengan 0, diperoleh:

$$\sum_{i=1}^{n}\psi(Y_i - \boldsymbol{X\beta})\boldsymbol{X^T} = 0 \qquad (10)$$

dengan $\psi = \rho'$ dan ψ merupakan fungsi *influence* yang digunakan untuk memperoleh pembobot. Kemudian galatnya distansarisasi, sehingga pers.(10) menjadi

$$\sum_{i=1}^{n}\psi\left(\frac{Y_i - \boldsymbol{X\beta}}{\hat{\sigma}}\right)\boldsymbol{X^T} = 0 \qquad (11)$$

Didefenisikan suatu fungsi pembobot $w_i = \frac{\psi(e_i^*)}{e_i^*}$ dengan e_i^* adalah galat yang distandarisasi sehingga $e_i^* = \frac{e_i}{\hat{\sigma}}$. Maka pers. (11) dapat ditulis menjadi:

$$\sum_{i=1}^{n} w_i\left(\frac{Y_i - \boldsymbol{X\beta}}{\hat{\sigma}}\right)\boldsymbol{X^T} = 0$$
$$\sum_{i=1}^{n}\left(\boldsymbol{X^T} w_i Y_i - \boldsymbol{X^T} w_i \boldsymbol{X\beta}\right) = 0 \qquad (12)$$
$$\sum_{i=1}^{n}\boldsymbol{X^T} w_i Y_i - \sum_{i=1}^{n}\boldsymbol{X^T} w_i \boldsymbol{X\beta} = 0$$

Dalam bentuk matriks persamaan (12) dapat dituliskan menjadi :

$$\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} \tag{13}$$

dimana W adalah matriks diagonal yang berukuran *n x n* dengan elemen diagonalnya $w_1, w_2, \dots, w_n$ (*n* banyaknya pengamatan). Pers. (13) dikalikan dengan $\left(\boldsymbol{X}^T\,\boldsymbol{W}^{(\mathbf{1})}\,\boldsymbol{X}\right)^{-\mathbf{1}}$ pada kedua ruas maka didapatkan estimasi parameter sebagai berikut:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}\,\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} \tag{14}$$

Pada langkah selanjutnya dihitung kembali bobot w_i yang baru menggunakan $\hat{\beta}_j$ dari hasil sebelumnya dan skala parameter $\hat{\sigma}_s$ konstan. Untuk $w_i^{(m)}$ bobot yang diberikan, dapat diperoleh estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}^{(\boldsymbol{m+1})} = \left(\boldsymbol{X}^T\,\boldsymbol{W}^{(\boldsymbol{m})}\boldsymbol{X}\right)^{-1}\ X'\boldsymbol{W}^{(\boldsymbol{m})}\boldsymbol{Y}$ sampai $\sum_{i=0}^{n}\left|e_i^{(m)}\right|$ konvergen (selisih nilai $\hat{\beta}_j^m$ dan $\hat{\beta}_j^{m+1}$ mendekati 0) dengan *m* banyaknya iterasi.

4.2 Pengolahan Data

4.2.1 Identifikasi Pencilan

Untuk mendeteksi pencilan dapat menggunakan metode *DfFITS* untuk mengidentifikasi pencilan di variabel Y dan *Leverage Value* (h_{ii}) untuk mengidentifikasi pencilan di variabel X.

a. Metode *DfFITS*

Suatu data dikatakan terdeteksi adanya pencilan pada variabel terikat apabila nilai $|DfFITS| > 2\sqrt{p/n} = 0.432$. Data yang terdeteksi pencilan yaitu data ke-1, 6, 24, 86, dan 122.

b. *Leverage Value* (h_{ii})

Suatu data dikatakan terdeteksi adanya pencilan pada variabel bebas apabila nilai $h_{ii} > 2k/n = 0.080$. Data yang terdeteksi pencilan yaitu data ke- 5, 7, 25, 27, 40, 90, 100, 111, 122, 124, 131, 133, 138, 139, 142, 143 dan 144.

4.2.2 Regresi *Robust* Estimasi MM (*Method of Momen*)

Hasil identifikasi pencilan dapat disimpulkan bahwa terdapat pencilan pada data. Selanjutnya, untuk mengatasi permasalahan tersebut digunakan regresi *robust* dengan estimasi MM (*Method of Moment*). Adapun langkah-langkah penyelesaiannya adalah sebagai berikut :

1. Menghitung estimator awal koefisien $\hat{\beta}_j^{(1)}$ dan galat $e_i^{(1)}$ dari metode estimasi S. Dengan menggunakan program SAS 9.1 maka didapatkan parameter - parameter regresi awalnya yaitu $\hat{\beta}_0^{(1)} = 1.1744$, $\hat{\beta}_1^{(1)} = 0.0272$, $\hat{\beta}_2^{(1)} = 0.0988$, $\hat{\beta}_4^{(1)} = -0.0564$, $\hat{\beta}_5^{(1)} = 0.1152$, $\hat{\beta}_6^{(1)} = 0.0419$ Selanjutnya estimator dari metode S ini akan digunakan mencari nilai galat $e_i^{(1)}$ dengan $e_i^{(1)} = Y_i - \hat{Y}_i^{(1)}$.
2. Galat $e_i^{(1)}$ pada langkah pertama digunakan untuk menghitung skala estimasi $\hat{\sigma}_s$ dan dihitung pula pembobot awal $w_i^{(1)}$ dengan bobot *Tukey*. Berdasarkan lampiran 3 didapatkan nilai *scale* yang dijadikan sebagai nilai $\hat{\sigma}$ yaitu 0.2827 dan $\psi(e_i^*)$ dihitung sesuai fungsi pembobot *Tukey Bisquare*
3. Galat $e_i^{(1)}$ dengan skala estimasi $w_i^{(1)}$ pada langkah kedua digunakan dalam iterasi awal untuk menghitung koefisien regresi. Nilai $w_i^{(1)}$ akan dijadikan menjadi matriks diagonal dengan ukuran nxn di mana $n = 150$ dengan w_i merupakan elemen diagonalnya kemudian dimasukkan kedalam persamaan berikut:

$$\widehat{\boldsymbol{\beta}}_j^{(2)} = \left(\boldsymbol{X}^{\boldsymbol{T}}\,\boldsymbol{W}^{(1)}\,\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\,\boldsymbol{W}^{(1)}\,\boldsymbol{y} \tag{15}$$

Untuk mendapatkan nilai estimasi parameter kedua.

4. Selanjutnya mengulang langkah 2 dan 3 untuk menghitung bobot baru $w_i^{(2)}$ dengan skala estimasi $\hat{\sigma}_s$tetap konstan) sampai mendapatkan $\sum_{i=1}^{n}\left|e_i^{(m)}\right|$ konvergen (selisih $\widehat{\boldsymbol{\beta}}_j^{(m+1)}$ dan $\widehat{\boldsymbol{\beta}}_j^{(m)}$ mendekati 0, dengan m banyaknya iterasi).
   Hasil iterasi selengkapnya dapat dilihat pada tabel berikut:

Tabel 1 Hasil Iterasi Memperoleh Nilai Koefesien Parameter *Robust*

		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$
Iterasi (*m*)	0	1.1744	0.0272	0.0988	0.0475	-0.0564	0.1152	0.0419
	1	1.30703	0.03256	0.09229	0.04262	-0.05097	0.10811	0.03359
	2	1.32601	0.03311	0.09204	0.04235	-0.05068	0.10710	0.03202
	3	1.32914	0.03318	0.09210	0.04235	-0.05069	0.10696	0.03168
	4	1.32973	0.03318	0.09213	0.04236	-0.05071	0.10693	0.03161
	5	1.32985	0.03319	0.09213	0.04236	-0.05071	0.10693	0.03159
	6	1.32988	0.03319	0.09214	0.04236	-0.05071	0.10693	0.03159
	7	1.32989	0.03319	0.09214	0.04237	-0.05071	0.10693	0.03158
	8	1.32989	0.03319	0.09214	0.04237	-0.05071	0.10693	0.03158
	9	1.32989	0.03319	0.09214	0.04237	-0.05071	0.10693	0.03158

Sumber : Data diolah

Berdasarkan Tabel 1 diatas, terlihat bahwa selisih estimasi parameter pada iterasi ke-9 dan ke-8 sama dengan nol. Hal ini menunjukkan bahwa estimasi parameter telah konvergen, sehingga diperoleh model regresi *robust* sebagai berikut:

$$\hat{y} = 1.32989 + 0.03319X_1 + 0.09214X_2 + 0.04237X_3 - 0.05071X_4 + 0.10693X_5 + 0.03158X_6 \qquad (16)$$

4.2.3 Koefesien Determinasi ($\boldsymbol{R^2}$)

Menggunakan nilai R^2 dapat diketahui tingkat signifikansi atau kesesuaian hubungan antara variabel bebas dengan variabel terikat dalam model regresi yang dihasilkan. Maka diperoleh nilai R^2 dari regresi *robust* sebesar $0.1429 = 14.29\%$ yang lebih besar dibandingkan dengan nilai R^2 yang diperoleh dari metode OLS yaitu sebesar $0.075 = 7.5\ \%$. Hal ini berarti pengaruh nilai Ujian Nasional terhadap Indeks Prestasi Kumulatif mahasiswa jurusan Matematika FMIPA UNHAS sebesar 14.29 %. Sisanya yaitu 85.71% dipengaruhi oleh variabel-variabel lain.

4.2.4 Uji Signifikansi Parameter

a. Uji Simultan (Uji F)

Uji simultan merupakan pengujian secara bersama semua parameter dalam model regresi. Hipotesis yang digunakan adalah sebagai berikut :

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_6 = 0$$
$$H_1: \beta_1 \neq \beta_2 \neq \cdots \neq \beta_6 \neq 0$$

untuk model regresi *robust* didapatkan nilai $F_{hitung} = 23.8$ dan nilai $F_{tabel} = 2.162$. Karena nilai F_{hitung} lebih besar dari nilai F_{tabel}, maka H_0 ditolak artinya terdapat pengaruh nilai Indeks Prestasi Kumulatif mahasiswa jurusan Matematika FMIPA UNHAS terhadap nilai-nilai Ujian

Nasional yang meliputi nilai bahasa Indonesia(X_1), bahasa Inggris(X_2), Matematika(X_3), Fisika(X_4), Kimia(X_5) dan Biologi(X_6).

b. Uji Parsial (Uji T)

Uji parsial merupakan pengujian parameter secara individu dalam model regresi yang bertujuan untuk mengetahui variabel-variabel mana yang berpengaruh terhadap nilai IPK mahasiswa jurusan Matematika selama tiga semester untuk angkatan 2010, 2011 dan 2012. Hipotesis yang digunakan adalah sebagai berikut :

$$H_0: \beta_j = 0 \quad ; j = 1,2,\dots,6$$
$$H_1: \beta_j \neq 0$$

Tabel 2 Nilai T_{hitung} Model Regresi *Robust*

Variabel	**Nilai t_{hitung}**
X_1	1.1366
X_2	**2.7836**
X_3	1.5350
X_4	-1.5229
X_5	**2.5219**
X_6	1.0288

Sumber : Data diolah

dan diperoleh nilai $t_{tabel} = 1.977$
Berdasarkan Tabel 2 dapat dilihat bahwa variabel nilai bahasa Indonesia(X_1), Matematika(X_3), Fisika(X_4), dan Biologi(X_6) mempunyai nilai t_{hitung} kurang dari t_{tabel} maka H_0 diterima, artinya variabel nilai bahasa Indonesia(X_1), Matematika(X_3), Fisika(X_4), dan Biologi(X_6) tidak signifikan terhadap nilai IPK. Sebaliknya variabel nilai bahasa Inggris(X_2) dan Kimia(X_5) mempunyai nilai t_{hitung} lebih besar dari t_{tabel} maka H_0 ditolak, artinya nilai bahasa Inggris(X_2) dan Kimia(X_5) signifikan terhadap nilai IPK. Maka kita mendapatkan suatu model baru yaitu :

$$\hat{y} = 1.32989 + 0.09214X_2 + 0.10693X_5 \qquad (17)$$

Makna yang dapat diterangkan dari pers. (4.9) adalah sebagai berikut :

a. $\beta_0 = 1.32989$

Apabila nilai bahasa Inggris dan Kimia sama dengan 0. Maka nilai IPK adalah sebesar 1.32989.

b. $\beta_2 = 0.09214$

Angka tersebut menunjukkan koefesien untuk nilai bahasa Inggris. Angka sebesar 0.09214 menunjukkan tanda positif yang berarti apabila nilai bahasa Inggris meningkat satu satuan sedangkan variabel lainnya dianggap konstan, maka mengakibatkan nilai IPK naik sebesar 0.09214 satuan.

c. $\beta_5 = 0.10693$

Angka tersebut menunjukkan koefesien untuk nilai Kimia. Angka sebesar 0.10693 menunjukkan tanda positif yang berarti apabila nilai Kimia meningkat satu satuan sedangkan variabel lainnya dianggap konstan, maka mengakibatkan nilai IPK naik sebesar 0.10693 satuan.

5. Kesimpulan dan Saran

5.1. Kesimpulan

Berdasarkan pembahasan hasil yang telah diselesaikan maka dapat ditarik kesimpulan sebagai berikut :

a. Estimasi MM (*Method of Moment*) merupakan salah satu estimasi pada regresi *robust* yang dapat digunakan untuk data yang terdeteksi pencilan pada variabel bebas dan variabel terikat.
b. Model regresi *robust* yang didapatkan dari estimasi MM yaitu:
$$\hat{y} = 1.32989 + 0.03319X_1 + 0.09214X_2 + 0.04237X_3 - 0.05071X_4 + 0.10693X_5 + 0.03158X_6$$
c. Hasil perhitungan koefisien determinasi model regresi menggunakan estimasi MM lebih besar yaitu 14.29 % dibandingkan dengan koefisien determinasi model regresi menggunakan metode kuadrat terkecil yaitu 7.5 % sehingga model regresi *robust* dikatakan lebih baik dibandingkan dengan model regresi menggunakan metode kuadrat terkecil.
d. Berdasarkan uji signifikansi parameter yang dilakukan yaitu uji-F dan uji-T didapatkan bahwa terdapat pengaruh nilai Indeks Prestasi Kumulatif mahasiswa jurusan Matematika FMIPA UNHAS terhadap nilai Ujian Nasional yang meliputi nilai bahasa Indonesia(X_1), nilai bahasa Inggris(X_2), nilai Matematika(X_3), nilai Fisika(X_4), nilai Kimia(X_5) dan nilai Biologi(X_6). Serta dengan $\alpha = 0.05$ diperoleh koefisien regresi $X_2\ dan\ X_5$ signifikan. Maka kita mendapatkan suatu model baru yaitu :
$$\hat{y} = 1.32989 + 0.09214X_2 + 0.10693X_5$$

5.2. Saran

Penelitian ini membahas tentang penggunaan regresi *robust* pada data yang mengandung pencilan dengan Metode Momen. Untuk penelitian selanjutnya dapat dilakukan penelitian atau kajian lebih dalam mengenai sifat-sifat teoritis pada beberapa estimator *robust*, salah satunya *least median square* (LMS) ataupun melakukan pendekatan lain untuk mengatasi data yang mengandung pencilan.

DAFTAR PUSTAKA

[1] Chen, Colin .2002. *Robust Regression and Outlier Detection with the Robustreg Procedure*. SUGI paper 265-27. SAS Institute : Cary, NC

[2] Drapper, N. R.,& Smith, H. 1996. *Applied Regression Analysis, 2nd edition*. New York: John Wiley & Sons. Chapman and Hall.

[3] Hasanah, Isma. Regresi *Robust* Untuk Mengatasi *Outlier* Pada Regresi Linier Berganda. Jurnal Universitas Jenderal Soedirman.

[4] Kelly, M. 2006. "*A Tour Around PROC ROBUSTREG*". *Paper ST01*. Dublin: Quintiles Ireland Ltd.

[5] Kurniawati, Lina Dewi. 2012. Kekekaran Regresi Linier Ganda Dengan Estimasi MM (*Method Of Moment*) Dalam Mengatasi Pencilan. Yogyakarta : UNY.

[6] Montgomery, D. C., & Peck, E. A. 1992. *Introduction to Linear Regression Analysis*. New York : A Wiley-Interscience Publication.

[7] Rousseeuw, R. J. & A. M. Leroy. 1987. *Robust Regression and Outlier Detection.* New York: By John Wiley and Sons, Inc.

[8] Soemartini. 2007. *Outlier* (Pencilan). Bandung: UNPAD.

[9] Puput,Nuraidah. 2011. Estimasi Parameter Dalam Regresi Linear Berganda Dengan Metode *Least Median Square* (LMS). Makassar : Universitas Hasanuddin.

[10] Walpole, Ronald. E and Myres, Raymond. H. 1995. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuan Edisi ke-4.* Bandung : Institut Teknologi Bandung.

[11] Yohai, V. J. (1987). High Breakdown Point and High Efficiency Robust Estimates for Regression. *Annals of Statistics,* Vol. 15, No.20, 642-656.[2]

[1,2,3] *Jurusan Matematika FMIPA Universitas Hasanuddin Makassar, Jl. Perintis Kemerdekaan Km.10 Makassar*