

ANALISIS REGRESI *ROBUST* ESTIMASI-MM DALAM MENGATASI PENCILAN PADA REGRESI LINEAR BERGANDA

Ariady Zulkarnain, Setyo Wira Rizki, Hendra Perdana

INTISARI

Analisis regresi merupakan suatu analisis statistik yang mempelajari hubungan antara variabel dependen dengan satu atau lebih variabel independen. Model linear yang memuat beberapa variabel independen dan satu variabel dependen disebut model regresi linear berganda. Pada umumnya metode yang digunakan untuk mengestimasi parameter regresi adalah Metode Kuadrat Terkecil (MKT). Namun ketika terdapat pencilan pada data, metode tersebut kurang efektif digunakan. Pencilan dapat dideteksi menggunakan uji DfFITS. Oleh sebab itu diperlukan suatu alternatif terhadap keberadaan pencilan, salah satunya dengan menggunakan metode regresi Robust. Salah satu metode estimasi dalam regresi Robust adalah estimasi Method of Moment (MM). Tujuan dalam penelitian ini adalah untuk mengatasi data pencilan dengan menggunakan regresi Robust estimasi-MM untuk mendapatkan model terbaik. Variabel yang digunakan dalam penelitian ini adalah pengaruh rata-rata lama sekolah (X_1), dan PDRB (X_2) terhadap IPM (Y) di Indonesia pada tahun 2015. Analisis regresi Robust estimasi-MM diawali dengan mengestimasi parameter menggunakan MKT. Setelah mendapatkan model estimasi, selanjutnya menguji apakah data yang digunakan terdapat pencilan atau tidak. Jika terdapat pencilan pada data maka dilanjutkan dengan mencari estimasi parameter menggunakan regresi Robust estimasi-MM. Berdasarkan uji DfFITS, data yang digunakan terdapat pencilan sehingga diperlukan prosedur regresi Robust untuk mengestimasi parameter model matematisnya. Model matematis regresi Robust estimasi-MM yang diperoleh yaitu $\hat{Y} = 44,95 + 2,846X_1 + 0,000002192X_2$, dimana variabel independen berpengaruh signifikan terhadap variabel dependen dengan nilai adjusted-R square sebesar 0,7365.

Kata Kunci: *Estimasi-MM, Regresi Robust, Pencilan.*

PENDAHULUAN

Analisis regresi merupakan suatu analisis statistik yang mempelajari hubungan antara variabel dependen dengan satu atau lebih variabel independen. Pada umumnya metode yang digunakan untuk mengestimasi parameter regresi adalah Metode Kuadrat Terkecil (MKT). Namun apabila ditemukan data pencilan dalam penelitian tersebut, maka penggunaan metode ini kurang tepat karena dapat mempengaruhi hasil estimasi parameter regresi, juga dapat menimbulkan pelanggaran terhadap asumsi kenormalan data. Pencilan adalah suatu pengamatan yang menyimpang cukup jauh dari pengamatan lainnya sehingga menimbulkan dugaan bahwa pengamatan tersebut berasal dari distribusi data yang berbeda [1].

Regresi *Robust* pertama kali dikenalkan oleh Andrews pada tahun 1972. Regresi *Robust* adalah suatu metode regresi yang digunakan ketika distribusi dari *error* tidak normal atau dan adanya beberapa pencilan yang berpengaruh pada model [2]. Metode ini merupakan cara untuk menganalisis data yang dipengaruhi pencilan, sehingga dihasilkan model yang robust atau resisten terhadap pencilan. Terdapat beberapa estimasi yang digunakan dalam mencari parameter regresi *Robust* seperti LTS (*Least Trimmed Square*), Estimasi-S (*Scale*), Estimasi-M (*Maximum Likelihood type*), dan Estimasi-MM (*Method Of Momen*) [3]. Metode estimasi-MM diperkenalkan oleh Yohai pada tahun 1987. Metode ini mempertahankan nilai breakdown point dari metode estimasi-S, serta efisiensi dari metode estimasi-M.

METODE KUADRAT TERKECIL (MKT)

MKT adalah metode yang bertujuan untuk meminimumkan jumlah kuadrat dari *error*. Misalkan terdapat p parameter dan n pengamatan, maka model yang akan diperoleh adalah sebagai berikut [5].

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_0 \\ \vdots \\ \beta_0 \end{pmatrix} + \begin{pmatrix} \beta_1 X_{11} \\ \beta_1 X_{21} \\ \vdots \\ \beta_1 X_{n1} \end{pmatrix} + \begin{pmatrix} \cdots \\ \cdots \\ \vdots \\ \cdots \end{pmatrix} + \begin{pmatrix} \beta_{p-1} X_{1,p-1} \\ \beta_{p-1} X_{2,p-1} \\ \vdots \\ \beta_{p-1} X_{n,p-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

dalam notasi matriks dapat ditulis sebagai:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1)$$

Estimasi parameter MKT untuk kasus n pengamatan dapat diperoleh dengan meminimumkan:

$$\sum_{i=1}^n \varepsilon_i^2 = \sum \left[\left(Y_i - (\beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{ip-1}) \right) \right]^2$$

Dengan $\sum \varepsilon_i^2$ merupakan jumlah kuadrat galat. Notasi matriks untuk meminimumkan $\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}$ dari Persamaan (1) diperoleh :

$$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \quad (2)$$

oleh karena itu,

$$\begin{aligned} \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} &= \mathbf{Y}^T \mathbf{Y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} \end{aligned} \quad (3)$$

dengan menggunakan sifat-sifat transpos suatu matriks, yaitu $(\mathbf{X}\boldsymbol{\beta})^T = \boldsymbol{\beta}^T \mathbf{X}^T$ dan karena $\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y}$ merupakan suatu skalar maka bentuk itu sama dengan transposnya $\mathbf{Y}^T \mathbf{X} \boldsymbol{\beta}$. Untuk menaksir parameter $\beta_1, \beta_2, \dots, \beta_{p-1}$ maka $\sum_{i=1}^n \varepsilon_i^2$ harus sekecil mungkin. Hal tersebut dicapai dengan menurunkan persamaan $\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}$ terhadap $\boldsymbol{\beta}$ dan membuatnya sama dengan nol. Selanjutnya diperoleh:

$$\begin{aligned} \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} &= \mathbf{Y}^T \mathbf{Y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} \\ \frac{\partial \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta}} &= -2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} + 2\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} \\ \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} &= \mathbf{X}^T \mathbf{Y} \\ \boldsymbol{\beta} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \end{aligned} \quad (4)$$

UJI DfFITS

Metode ini menampilkan nilai perubahan dalam data yang diprediksi bilamana *case* tertentu dikeluarkan, yang sudah distandarkan [6]. Hipotesis uji *DfFITS* adalah sebagai berikut:

H_0 : pencilan ke- i tidak berpengaruh.

H_1 : pencilan ke- i berpengaruh.

Statistik uji:

$$(DfFITS_i) = t_i \left(\frac{h_{ii}}{1-h_{ii}} \right)^{\frac{1}{2}} \quad (5)$$

$$\text{dengan } t_i = e_i \sqrt{\frac{\varepsilon_i}{S_{(i)}^2 (1-h_{ii})}}, \quad i=1,2,\dots,n; \quad S_{(i)}^2 = \frac{(n-p)MSE - \varepsilon_i^2 / (1-h_{ii})}{n-p-1}; \quad MSE = \frac{\mathbf{Y}^T \mathbf{Y} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y}}{n-k-1};$$

dimana h_{ii} adalah elemen diagonal dari matriks \mathbf{H} . Matriks \mathbf{H} adalah matriks berukuran $n \times n$ serta

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T.$$

Kriteria pengujian H_0 ditolak adalah nilai $|DfFITS| > 1$ untuk gugus data kecil atau nilai $|DfFITS| >$

$2\sqrt{\frac{p}{n}}$ untuk gugus data besar, dimana $p = k + 1$ dengan k adalah banyaknya variabel bebas [6].

ESTIMASI-MM

Estimasi-MM menggabungkan estimasi *high breakdown point* dan efisiensi statistik yang dikenalkan oleh Yohai tahun 1987. Langkah pertama dalam estimasi ini adalah mencari estimator dengan estimasi-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 [8]:

$$\hat{\beta}_{MM} = \arg \min_{\beta} \sum_{i=1}^n \rho \left(\frac{e_i}{\hat{\sigma}_S} \right) = \arg \min_{\beta} \sum_{i=1}^n \rho \left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{\hat{\sigma}_S} \right) \tag{6}$$

Metode-MM juga menggunakan IRLS (*Iteratively Reweighted Least Square*) untuk mencari estimasi parameter regresi.

Prosedur estimasi parameter pada model regresi linier ganda dengan regresi robust estimasi MM [8]:

1. Menghitung estimator awal koefisien $\hat{\beta}_j^{(1)}$ dan residual $e_i^{(1)}$ dari regresi *Robust* dengan *high breakdown poin* dengan pembobotan *Tukey Bisquare*.
2. Residual $e_i^{(1)}$ pada langkah pertama digunakan untuk menghitung skala estimasi $\hat{\sigma}_s$ dan dihitung pula pembobotan awal $w_i^{(1)}$.
3. Residual $e_i^{(1)}$ dengan skala estimasi $\hat{\sigma}_s$ pada langkah kedua digunakan dalam iterasi awal sebagai penaksiran WLS untuk menghitung koefisien regresi.

$$\sum_{i=1}^n w_i^{(1)} \left(\frac{e_i^{(1)}}{\hat{\sigma}_s} \right) x_i = 0, \quad w_i^{(1)} \text{ merupakan pembobot } \textit{Tukey Bisquare}.$$

4. Menghitung bobot baru $w_i^{(2)}$ dengan skala estimasi dari iterasi awal WLS.

Mengulang langkah 2, 3, dan 4 (dengan skala estimasi tetap konstan) sampai mendapatkan $\sum_{i=1}^n |e_i^{(m)}|$ konvergen (selisih $\hat{\beta}_j^{(m+1)}$ dan $\hat{\beta}_j^{(m)}$ mendekati 0, dengan m banyaknya iterasi).

PEMBOBOTAN TUKEY BISQUARE

Fungsi obyektif untuk pembobotan *Tukey Bisquare* dalam Persamaan [7]:

$$\rho(u_i) = \begin{cases} \frac{u_i^2}{2} - \frac{u_i^4}{2c^2} - \frac{u_i^6}{6c^4}, & |u_i| \leq c \\ \frac{c^2}{6}, & |u_i| > c \end{cases} \tag{7}$$

dengan $\rho(u_i)$ adalah fungsi obyektif dari sebuah pembobotan, u_i adalah skala sisaan dan c adalah nilai *tuning constant*.

Kemudian untuk menentukan fungsi pengaruh *Tukey Bisquare* (ψ) yang merupakan turunan dari $\rho(u_i)$ terhadap u , sehingga dapat ditulis $\psi = \rho'$, sehingga fungsi *Tukey Bisquare* menjadi

$$\psi(u_i) = \begin{cases} u_i \left(1 - \left(\frac{u_i}{c} \right)^2 \right)^2, & |u_i| \leq c \\ 0, & |u_i| > c \end{cases} \tag{8}$$

Untuk fungsi pembobotan digunakan rumus sebagai berikut:

$$w(u_i) = \frac{\psi(u_i)}{u_i} \quad (9)$$

sehingga diperoleh,

$$w(u) = \begin{cases} \left(1 - \left(\frac{u}{c}\right)^2\right)^2 & |u| \leq c \\ 0 & |u| > c \end{cases} \quad (10)$$

dengan $u_i = \frac{e_i}{\sigma_s}$, nilai $c = 4,685$ adalah nilai *tuning constant* yang telah ditetapkan.

Sehingga fungsi pembobot menjadi seperti berikut:

$$w(u) = \begin{cases} \left[1 - \left(\frac{u}{4,685}\right)^2\right]^2, & |u| \leq 4,685 \\ 0, & |u| > 4,685 \end{cases} \quad (11)$$

STUDI KASUS

Analisis yang dilakukan pada penelitian ini adalah menganalisis pengaruh rata-rata lama sekolah dan PDRB terhadap IPM di Indonesia. Estimasi parameter dengan MKT untuk data IPM (Y), rata-rata lama sekolah (X_1) dan PDRB (X_2) dilakukan dengan *software* R 3.5.1. Hasil estimasi parameter dengan MKT dapat dilihat pada Tabel 1 dibawah ini.

Tabel 1 Estimasi Parameter Metode Kuadrat Terkecil

Variabel Bebas	Estimator regresi
Konstanta	41,590
X_1	3,281
X_2	$1,986 \times 10^{-6}$

dari Tabel 1 dihasilkan model matematis:

$$\hat{Y} = 41,59 + 3,281X_1 + 0,000001986X_2$$

Nilai *adj R-square* untuk MKT adalah 0,6932. Yang artinya 69,32% variasi pada variabel dependen (Y) dapat dijelaskan oleh variabel independen (X), sedangkan sisanya dapat dijelaskan oleh variabel lain.

Uji normalitas bertujuan untuk mengetahui asumsi bahwa *error* (ε_i) berdistribusi normal. Uji kenormalan dilakukan menggunakan uji *Kolmogorov-Smirnov*. Berdasarkan nilai *p-value* yaitu 0,2894 lebih besar dari $\alpha = 0,05$ ($p - value > \alpha$), hal ini berarti *error* berdistribusi normal.

Uji non autokorelasi dapat dilakukan menggunakan uji *Durbin-Watson*. Hasil uji *Durbin-Watson* dapat dilihat pada Tabel 2 dibawah ini.

Tabel 2 Hasil Uji *Durbin-Watson*

DW	P-Value
1,0274	0,0007338

Diperoleh nilai *Durbin-Watson* adalah 1,0274, serta untuk $k = 2$, $n = 34$ dan $\alpha = 5\%$ diperoleh nilai $DL = 1,3325$ dan $DU = 1,5805$. Nilai $DW < DL$ dan $p - value < \alpha$ ($\alpha=0,05$) sehingga dapat disimpulkan bahwa terdapat autokorelasi.

Uji non multikolinearitas adalah uji untuk mengetahui adanya hubungan diantara variabel-variabel independen dalam model regresi. Uji multikolinearitas dapat dilakukan dengan melihat nilai *VIF*. Multikolinearitas tidak terjadi apabila nilai *VIF* lebih besar dari 10 pada masing-masing variabel bebas. Tabel 3 menyajikan hasil *output* nilai *VIF* pada masing-masing variabel independen.

Tabel 3 Nilai *VIF*

Variabel	<i>VIF</i>
Rata-rata Lama Sekolah	1,0579
PDRB	1,0579

Berdasarkan nilai *VIF* pada masing-masing variabel independen dapat diartikan bahwa variabel independen tidak terjadi multikolinearitas.

Uji homoskedastisitas dapat dilakukan menggunakan uji Breusch Pagan. Pasangan hipotesis untuk uji Breusch Pagan adalah sebagai berikut.

H_0 : Terjadi homoskedastisitas

H_1 : Tidak terjadi homoskedastisitas

Hasil pengujian Breusch Pagan dapat dilihat pada Tabel 4.

Tabel 4 Hasil Uji *Breusch Pagan*

BP	<i>P-Value</i>
1,1356	0,5668

dari Tabel 4, diperoleh nilai BP sebesar 1,1356, sedangkan untuk nilai $\chi^2_{(p_j+1)}$ adalah sebesar 5,99. Nilai BP < $\chi^2_{(p_j+1)}$ dan *p-value* > α ($\alpha=0,05$) sehingga disimpulkan bahwa H_0 diterima yang artinya terjadi homoskedastisitas.

Pendeteksi pencilan menggunakan uji *DfFITS*, diperoleh nilai *DfFITS* pada pengamatan ke-11 dengan nilai *DfFITS* 1,0058 yang lebih besar dari satu ($|DfFITS| > 1$), sehingga dapat disimpulkan bahwa pengamatan tersebut merupakan pencilan.

Estimasi parameter regresi *Robust* estimasi-MM untuk data IPM (Y), rata-rata lama sekolah (X_1) dan PDRB (X_2) dilakukan dengan *software* R. Estimasi parameter regresi *Robust* estimasi-MM dengan pembobotan *Tukey Bisquare* dapat dilihat pada Tabel 5 berikut ini.

Tabel 5 Estimasi parameter regresi *Robust* dengan pembobotan *Tukey Bisquare*

Variabel Bebas	Estimator Regresi
Konstanta	44,950
X_1	2,846
X_2	$2,192 \times 10^{-6}$

dari Tabel 5 dihasilkan model matematis:

$$\hat{Y} = 44,95 + 2,846X_1 + 0,000002192X_2$$

model tersebut konvergen pada iterasi ke-15. Uji pengaruh variabel independen terhadap variabel dependen dilihat dari nilai *p-value* masing-masing variabel independen, diperoleh nilai *p-value* masing-masing variabel independen lebih besar daripada nilai α ($\alpha=0,05$) sehingga dapat disimpulkan variabel independen berpengaruh signifikan terhadap variabel dependen dengan nilai *adjusted-R square* sebesar 0,7365. Artinya variabel independen mempengaruhi variabel dependen sebesar 73,65%, sedangkan 26,35% dijelaskan oleh variabel lain.

KESIMPULAN

Berdasarkan studi kasus yang telah dilakukan pada data pengaruh rata-rata lama sekolah dan PDRB terhadap Indeks Pempangan Manusia (IPM) setiap provinsi di Indonesia pada tahun 2015, dapat ditarik beberapa kesimpulan bahwa regresi *Robust* estimasi-MM mampu mengatasi data yang mengandung pencilan. Hal tersebut dapat dilihat berdasarkan nilai *adjusted-R square* metode *Robust* estimasi-MM lebih besar dari Metode Kuadrat Terkecil yaitu $0,7365 > 0,6932$. Dengan demikian

dapat disimpulkan bahwa menggunakan regresi *Robust* estimasi-MM memberikan hasil yang lebih efektif daripada Metode Kuadrat Tekecil. Model regresi *Robust* estimasi-MM pada data Indeks Pembangunan Manusia (IPM) setiap provinsi di Indonesia adalah $\hat{Y} = 44,95 + 2,846X_1 + 0,000002192X_2$.

DAFTAR PUSTAKA

- [1]. Soemartini, Outlier (Pencilan), UNPAD, Bandung, 2007.
- [2]. Olive, DJ, Applied Robust Statistics. Southern Illinois University, Carbondale, 2005.
- [3]. Chen, C., Robust Regression and Outlier Detection with the ROBUSTREG Procedure, SUGI Paper 265-27, SAS Institute, North Carolina, 2002.
- [4]. Montgomery, D.C., dan Peck, E.A., Introduction to Linear Regression Analysis second edition, Jhon Wiley and Sons Inc, New York, 1992.
- [5]. Widyaningsing A, Susilawati M, Sumarjaya IW., Estimasi Model Seemingly Unrelated Regression (SUR) dengan Metode Generalized Least Square (GLS). Jurnal Matematika.2014; 4(2):102-110.
- [6]. Romdi, Wahyuningsih S., Yuniarti, D., Regresi Robust Linear Sederhana dengan Menggunakan Estimasi MM (Method of Moment). Jurnal EKSPONENSIAL.2015;2(4):179-186.
- [7]. Rousseew, P., dan Yohai, Robust Regression by Means of S Estimator. Lecture in Statistics, 26:256-272, New York, 1987.
- [8]. Yohai, V.J., High Breakdown Point and High Efficiency Robust Estimates for Regression. Annals of Statistics.1987;15(20):642-656.

ARIADY ZULKARNAIN : Jurusan Matematika FMIPA Untan, Pontianak
ariadyzulkarnain28@gmail.com

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

HENDRA PERDANA : Jurusan Matematika FMIPA Untan, Pontianak
hendra.perdana@math.untan.ac.id
