

ESTIMASI-M IRLS MENGGUNAKAN FUNGSI PEMBOBOT HUBER DAN *BISQUARE TUKEY* PADA DATA KEMISKINAN DI INDONESIA

Alda Oktaviani, Shantika Martha

INTISARI

Metode kuadrat terkecil (MKT) adalah metode yang umum digunakan untuk mengestimasi parameter pada regresi linear. Akan tetapi estimasi dengan MKT mempunyai kelemahan ketika outlier atau pencilan terdapat dalam data yang menyebabkan estimator dari parameter bersifat bias. Sebagai alternatif, metode regresi robust dapat digunakan untuk mengestimasi parameter ketika terdapat data outlier. Metode robust yang digunakan dalam penelitian ini adalah estimasi-M IRLS (Iteratively Reweighted Least Square) dengan fungsi pembobotan Huber dan Bisquare Tukey. Studi kasus yang digunakan dalam penelitian ini adalah pengaruh Indeks Pembangunan Manusia (IPM) (Z_1), Produk Domestik Regional Bruto (PDRB) menurut pengeluaran (Z_2), dan Tingkat Pengangguran Terbuka (TPT) (Z_3), terhadap jumlah kemiskinan (Y) di 34 Provinsi Indonesia pada tahun 2018. Tujuan penelitian ini adalah melakukan estimasi parameter dan membandingkan hasil estimasi-M menggunakan pembobotan Huber dan Bisquare Tukey. Berdasarkan uji boxplot data yang digunakan teridentifikasi data outlier, sehingga diperlukan prosedur robust untuk mengestimasi parameter model matematis. Nilai standart error dari pembobotan Bisquare Tukey lebih kecil dari pembobotan Huber ($221 < 229,2$). Berdasarkan nilai standart error dapat disimpulkan bahwa estimasi-M dengan pembobotan Bisquare Tukey merupakan metode yang paling baik digunakan, dengan model matematis $\hat{Y} = 827,85 - 197,98Z_1 + 1183,28Z_2$.

Kata Kunci: *Estimasi-M IRLS, Regresi Robust, Tukey Bisquare, Huber*

PENDAHULUAN

Penyebab kemiskinan dikarenakan adanya keterbelakangan dan ketertinggalan Sumber Daya Manusia (SDM), ketidaksempurnaan pasar, dan kurangnya modal yang menyebabkan rendahnya produktivitas. Rendahnya produktivitas menyebabkan rendahnya pendapatan yang mereka terima (tercermin oleh rendahnya Produk Domestik Regional Bruto (PDRB) per kapita) rendahnya pendapatan akan berimplikasi pada rendahnya tabungan dan investasi. Rendahnya investasi berakibat pada rendahnya akumulasi modal sehingga proses penciptaan lapangan kerja rendah (tercermin oleh tingginya angka pengangguran). Rendahnya akumulasi modal disebabkan oleh keterbelakangan [1], sehingga untuk menganalisa hubungan faktor penyebab kemiskinan diperlukan analisis regresi.

Analisis regresi merupakan analisis ketergantungan dari satu atau lebih variabel bebas terhadap satu variabel terikat, dengan tujuan untuk menduga atau memprediksi nilai rata-rata populasi berdasarkan nilai-nilai variabel bebasnya. Untuk memperoleh taksiran parameter regresi linear maka dapat dilakukan dengan menggunakan *Ordinary Least Square* (OLS). Analisis regresi memiliki asumsi klasik yang harus dipenuhi agar model yang dihasilkan dapat dikatakan model terbaik [2].

Permasalahan yang dapat timbul dari regresi dengan metode OLS adalah adanya *outlier* atau pencilan pada data. Pencilan adalah suatu pengamatan yang menyimpang cukup jauh dari pengamatan lainnya sehingga menimbulkan dugaan bahwa pengamatan tersebut berasal dari distribusi data yang berbeda [3].

Mengatasi masalah pencilan tersebut maka diperlukan regresi *robust*. Beberapa metode estimasi dalam regresi *robust* di antaranya adalah estimasi-M, *Least Trimmed Squares* (LTS), *Least Median Squares* (LMS), estimasi S dan estimasi *Method of Moment* (MM) [4]. Estimasi-M merupakan suatu

metode regresi *robust* yang terkenal dan paling luas digunakan dari pada metode regresi *robust* yang lain, karena hasilnya lebih teliti. Metode *Iteratively Reweighted Least Squares* (IRLS) merupakan salah satu metode iterasi pada estimasi-M yang memerlukan 3 fungsi pembobot, yaitu metode kuadrat terkecil, fungsi Huber dan fungsi *Bisquare Tukey*.

Metode Kuadrat Terkecil (MKT)

Kasus k-variabel bebas, penaksir metode kuadrat terkecil (MKT) diperoleh dengan meminimumkan jumlah kuadrat *error* yaitu [5]:

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 \quad (1)$$

Dimana $\sum_{i=1}^n \varepsilon_i^2$ adalah jumlah kuadrat *error*. Dalam notasi skalar, metode kuadrat terkecil tercapai dalam menaksir $\beta_0, \beta_1, \dots, \beta_k$ sehingga $\sum_{i=1}^n \varepsilon_i^2$ sekecil mungkin. Notasi matriks metode kuadrat terkecil sama dengan meminimumkan $\varepsilon' \varepsilon$ dengan $\varepsilon = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, yaitu [6]:

$$\begin{aligned} \text{Oleh karena itu:} \quad \varepsilon' \varepsilon &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}' \mathbf{y} - 2\boldsymbol{\beta}' \mathbf{X}' \mathbf{y} + \boldsymbol{\beta}' \mathbf{X}' \boldsymbol{\beta} \end{aligned} \quad (2)$$

$\boldsymbol{\beta}' \mathbf{X}' \mathbf{y}$ adalah matrik 1×1 , atau suatu skalar. Transposnya $(\boldsymbol{\beta}' \mathbf{X}' \mathbf{y})' = \mathbf{y}' \mathbf{X} \boldsymbol{\beta}$ adalah skalar. Estimator metode kuadrat terkecil harus memenuhi

$$\frac{\partial \sum_{i=1}^n \varepsilon_i^2}{\partial \boldsymbol{\beta}} = \frac{\partial \varepsilon' \varepsilon}{\partial \boldsymbol{\beta}} = -2\mathbf{X}' \mathbf{y} + 2\mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{0} \quad (3)$$

$$\text{Bila disederhanakan menjadi:} \quad \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}' \mathbf{y} \quad (4)$$

Untuk menyelesaikan Persamaan (2) kalikan keduanya dengan invers dari $\mathbf{X}' \mathbf{X}$ Jadi estimator kuadrat terkecil dari $\boldsymbol{\beta}$ adalah

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} \quad (5)$$

Box Plot

Box Plot (juga dikenal sebagai diagram *box-and-whisker*) [7], merupakan suatu *box* (kotak berbentuk bujur sangkar). *Boxplot* adalah salah satu cara dalam statistic deskriptif untuk menggambarkan secara grafik dari data numeris melalui lima ukuran sebagai berikut:

1. Nilai observasi terkecil,
2. Kuartil terendah atau kuartil pertama (Q_1), yang memotong 25% dari data terendah,
3. Median (Q_2) atau nilai pertengahan,
4. Kuartil tertinggi atau kuartil ketiga (Q_3), yang memotong 25% dari data tertinggi,
5. Nilai observasi terbesar.

Interquartile Range (IQR) diperoleh dari $IQR = Q_3 - Q_1$, suatu nilai dikatakan *outlier* atau pencilan jika $Q_3 + (1,5 \times IQR) < outlier \leq Q_3 + (3 \times IQR)$ atau $Q_1 - (1,5 \times IQR) > outlier \geq Q_1 - (3 \times IQR)$, selanjutnya suatu nilai dikatakan ekstrim jika lebih besar dari $Q_3 + (3 \times IQR)$ atau lebih dari $Q_1 - (3 \times IQR)$

PROSEDUR ESTIMASI-M

Prosedur estimasi-M sebagai berikut [8]:

1. Dihitung penaksir $\boldsymbol{\beta}$, dinotasikan \mathbf{b} menggunakan metode kuadrat terkecil, sehingga didapatkan $\hat{y}_{i,0}$ dan $\varepsilon_{i,0}$, dimana $\varepsilon_{i,0} = y_i - \hat{y}_{i,0}$ ($i = 1, 2, \dots, n$) yang diperlakukan sebagai nilai awal
2. Menghitung nilai $\hat{\sigma}$

$$\hat{\sigma} = \frac{MAD}{0,6745} = \frac{\text{median}\left\{\left|e_i - \text{median}(e_i)\right|\right\}}{0,6745} \quad (6)$$

3. Mencari nilai skala residual (u_i)

$$u_i = \frac{(y_i - \hat{y}_i)}{\hat{\sigma}} = \frac{e_i}{\hat{\sigma}} \quad (7)$$

4. Mendefinisikan pembobot berdasarkan fungsi pembobot $w_i = w(u_i)$, dengan konstanta untuk pembobot Huber sebesar 1,345 dan *Bisquare Tukey* sebesar 4,685
5. Memperbaiki estimasi $\hat{\beta}$ berdasarkan metode *Weighted Least Square* (WLS) dengan pembobot w_i sehingga diperoleh $\hat{\beta}$ yang baru pada iterasi ke-1
6. Selanjutnya ulangi Langkah 2 sampai Langkah 5 sehingga nilai akan berubah pada tiap iterasinya sehingga diperoleh $\hat{\beta}_m$ yang konvergen.

HASIL DAN PEMBAHASAN

Analisis yang dilakukan pada penelitian ini membahas tentang analisis pengaruh Indeks Pembangunan Manusia (IPM), Produk Domestik Regional Bruto (PDRB) menurut pengeluaran, dan Tingkat Pengangguran Terbuka (TPT) terhadap jumlah Kemiskinan. Data yang digunakan merupakan data sekunder dari (BPS) dan dilakukan standarisasi data, analisis dilakukan dengan *software R*.

Tabel 1. Estimasi Parameter Metode Kuadrat Terkecil

Parameter	Estimator Regresi	<i>p-value</i>
$\hat{\beta}_0$	755,14	0,00
$\hat{\beta}_1$	-374,85	0,01
$\hat{\beta}_2$	904,20	0,00
$\hat{\beta}_3$	-91,69	0,47

Berdasarkan Tabel 1 bisa dilihat pada nilai *p-value* menunjukkan tingkat signifikan masing-masing variabel. Signifikan menunjukkan ada pengaruh atau tidak ada pengaruh antara variabel bebas dan variabel terikat dari ketiga variabel bebas. Suatu variabel dikatakan signifikan apabila *p-value* $\leq \alpha$ dengan ($\alpha = 0,05$). Dari hasil estimasi parameter MKT ada satu variabel bebas yang tidak signifikan yaitu variabel tingkat pengangguran terbuka ($\hat{\beta}_3$). Sehingga variabel tingkat pengangguran terbuka tidak ada pengaruh yang signifikan terhadap variabel jumlah kemiskinan, oleh sebab itu dilakukan estimasi tanpa variabel tingkat pengangguran terbuka. Hasil estimasi parameter MKT tanpa variabel tingkat pengangguran terbuka ($\hat{\beta}_3$) dapat dilihat pada Tabel 2.

Tabel 2. Estimasi Parameter Metode Kuadrat Terkecil

Parameter	Estimator Regresi	<i>p-value</i>
$\hat{\beta}_0$	755,1	0,00
$\hat{\beta}_1$	-389,2	0,00
$\hat{\beta}_2$	881,1	0,00

Berdasarkan Tabel 2 dapat kita ketahui untuk nilai *p-value* $\leq \alpha$ dari masing-masing variabel bebas telah signifikan. Jadi untuk model persamaan regresi Metode Kuadrat Terkecil yaitu $\hat{Y} = 755,1 - 389,2Z_1 + 881,1Z_2$. Nilai *adj R-Square* untuk MKT adalah 0,5524, yang artinya variabel

bebas menjelaskan sebesar 55,24%, dan sisanya sebesar 44,76% dijelaskan oleh variabel lain. Nilai *Standard Error* pada model regresi tersebut sebesar 704,9, hal ini berarti banyaknya kesalahan dalam memprediksi variabel terikat sebesar 704,9% jumlah kemiskinan.

ASUMSI KLASIK

Uji Normalitas

Uji normalitas bertujuan untuk mengetahui apakah dalam model regresi, variabel pengganggu atau residual memiliki distribusi normal. Uji *Kolmogorov-Smirnov* (KS) merupakan salah satu uji untuk menguji kenormalan. KS_{hitung} sebesar 0,5 dan nilai *p-value* adalah 2,196e-08. Nilai tabel Kolmogorov-Smirnov untuk KS_{tabel} adalah 0,231. Dari nilai tersebut dapat disimpulkan bahwa $KS_{tabel} < KS_{hitung}$ dan $p - value < \alpha$ ($\alpha = 0,05$), sehingga H_0 ditolak dan dapat disimpulkan bahwa *error* tidak berdistribusi normal.

Uji Multikolinearitas

Uji multikolinearitas bertujuan untuk menguji apakah ditemukan adanya korelasi antar variabel bebas (independen). Multikolinearitas di dalam model regresi, dapat dilihat pada nilai VIF (*Variance Inflation Factors*). Nilai VIF variabel bebas Z_1 dan Z_2 adalah 1,207816 lebih kecil dari 10 ($VIF < 10$). Sehingga dapat disimpulkan bahwa tidak terjadi multikolinearitas.

Uji Autokorelasi

Uji autokorelasi bertujuan untuk menguji apakah dalam model regresi linier ada korelasi antara kesalahan pengganggu pada periode t dengan periode $t - 1$ (sebelumnya). Uji autokorelasi dapat dilakukan menggunakan uji Durbin-Watson (DW). Nilai Durbin-Watson sebesar 1,4892, sedangkan untuk nilai DL dan DU masing-masing adalah 1,3325 dan 1,5805. Nilai $DW > DL$ dan $p - value < \alpha$ ($\alpha = 0,05$) sehingga dapat disimpulkan tidak terjadi autokorelasi.

Uji Heterokedastisitas

Uji heterokedastisitas yaitu untuk menguji apakah terjadi ketidaksamaan varians dan residual satu pengamatan ke pengamatan yang lain. Uji heterokedastisitas dapat dilakukan dengan menggunakan uji *Breusch Pagan* (BP). Nilai BP sebesar 61,561 dan nilai *p-value* 4,287e-14, sedangkan untuk nilai $\chi^2_{(p_j+1)}$ tabel adalah sebesar 5,99. Nilai $BP > \chi^2_{(p_j+1)}$ dan $p - value < \alpha$ ($\alpha = 0,05$) sehingga disimpulkan bahwa H_0 ditolak yang artinya tidak terjadi homoskedastisitas.

Pendeteksi Pencilan

Suatu data diduga dan dinyatakan sebagai suatu pencilan dapat dideteksi menggunakan metode *boxplot*. Dengan metode *boxplot*, suatu data dikatakan pencilan jika nilai data pengamatan lebih kecil dari $Q1 - (1,5 \times IQR)$ atau lebih besar dari $Q3 + (1,5 \times IQR)$. Pendeteksian pencilan yang dilakukan menggunakan diagram *boxplot* menduga masing-masing variabel. Variabel jumlah kemiskinan (Y) memiliki 3 pencilan yang terdiri provinsi Jawa Barat, Jawa Tengah, Jawa Timur. Variabel IPM (Z_1) memiliki 9 pencilan yang terdiri dari provinsi DKI Jakarta, Yogyakarta, Kalimantan Timur, Bali, Kepulauan Riau, Sulawesi Barat, NTB, Papua Barat dan Papua. Variabel PDRB (Z_2) memiliki 4 pencilan yang terdiri dari Provinsi DKI Jakarta, Jawa Barat, Jawa Tengah, dan Jawa Timur.

Estimasi-M dengan Pembobotan *Bisquare Tukey*

Proses iteratif dimulai dengan menentukan estimasi awal koefisien regresi yang diperoleh dengan metode kuadrat terkecil, diperoleh taksiran model regresi linier sebagai berikut:

$\hat{Y} = 755,1 - 389,2Z_1 + 881,1Z_2$. Selanjutnya, dihitung nilai residual dan dihitung nilai skala estimasi serta nilai skala residual. Selanjutnya dilakukan pembobotan *Bisquare Tukey*, sehingga diperoleh taksiran model regresi linier hingga konvergen atau sama dengan hasil iterasi sebelumnya. Pada pembobotan *Bisquare Tukey* untuk setiap iterasi ternyata berhenti pada iterasi ke-17 karena nilai $\hat{\beta}$ yang baru sama dengan nilai $\hat{\beta}$ sebelumnya. Hasil iterasi estimasi menggunakan pembobot *Bisquare Tukey* dapat dilihat pada Tabel 3.

Tabel 3. Hasil Iterasi Estimasi *Bisquare Tukey*

Iterasi	<i>Bisquare Tukey</i>		
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
1	779,57	-266,94	1073,56
2	804,29	-240,53	1149,84
3	813,65	-228,85	1181,21
⋮	⋮	⋮	⋮
15	827,82	-198,04	1183,29
16	827,85	-197,98	1183,28
17	827,85	-197,98	1183,28

Estimasi parameter regresi *robust* estimasi-M dengan menggunakan pembobot *Bisquare Tukey* menggunakan *software* R dapat diketahui yaitu $\hat{Y} = 827,85 - 197,98Z_1 + 1183,28Z_2$.

Estimasi-M dengan Pembobotan Huber

Proses iteratif dimulai dengan menentukan estimasi awal koefisien regresi yang diperoleh dengan metode kuadrat terkecil, diperoleh taksiran model regresi linier sebagai berikut: $\hat{Y} = 755,1 - 389,2Z_1 + 881,1Z_2$. Selanjutnya, dihitung nilai residual dan dihitung nilai skala estimasi serta nilai skala residual. Selanjutnya dilakukan pembobotan Huber, sehingga diperoleh taksiran model regresi linier hingga konvergen atau sama dengan hasil iterasi sebelumnya. Pembobotan Huber untuk setiap iterasi ternyata berhenti pada iterasi ke-16 karena nilai $\hat{\beta}$ yang baru sama dengan nilai $\hat{\beta}$ sebelumnya. Hasil iterasi estimasi menggunakan pembobot Huber dapat dilihat pada Tabel 4.

Tabel 4. Hasil Iterasi Estimasi Huber

Iterasi	Huber		
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
1	753,00	-307,04	980,26
2	776,12	-285,27	1048,51
3	789,39	-275,17	1084,86
⋮	⋮	⋮	⋮
14	803,52	-251,47	1131,73
15	803,54	-251,46	1131,77
16	803,54	-251,46	1131,77

Estimasi parameter regresi *robust* estimasi-M dengan menggunakan pembobot Huber menggunakan *software* R dapat diketahui yaitu $\hat{Y} = 803,54 - 251,46Z_1 + 1131,77Z_2$.

Perbandingan Nilai *Standard Error* antara *Bisquare Tukey* dan Huber

Kriteria yang dipakai untuk melihat baik atau tidaknya suatu estimasi dapat dilihat dari nilai *Standard Error*, jika semakin kecil *Standard Error* maka semakin bagus estimasi yang digunakan. Hasil perbandingan dapat dilihat pada Tabel 5 berikut ini.

Tabel 5. Perbandingan Nilai *Standard Error*

Model	<i>Standard Error</i>
<i>Bisquare Tukey</i>	221
Huber	229,2

Berdasarkan hasil dari 5 dapat dikatakan bahwa estimasi yang paling baik digunakan adalah metode pembobotan *Bisquare Tukey*. *Standard Error* dari pembobotan *Bisquare Tukey* lebih kecil jika dibandingkan dengan pembobotan Huber ($221 < 229,2$), yang artinya tingkat kesalahan memprediksi variabel terikat lebih kecil jika menggunakan pembobotan *Bisquare Tukey*.

KESIMPULAN

Berdasarkan pembahasan yang telah dilakukan, dapat disimpulkan bahwa:

- 1) Pendeteksian pencilan yang dilakukan menggunakan diagram *boxplot* menduga masing-masing variabel. Variabel jumlah kemiskinan (Y) memiliki 3 pencilan yang terdiri dari provinsi Jawa Barat, Jawa Tengah, dan Jawa Timur. Variabel IPM (Z_1) memiliki 9 pencilan yang terdiri dari provinsi DKI Jakarta, Yogyakarta, Kalimantan Timur, Bali, Kep Riau, Sulawesi Barat, NTB, Papua Barat dan Papua Variabel PDRB (Z_2) memiliki 4 pencilan yang terdiri dari Provinsi DKI Jakarta, Jawa Barat, Jawa Tengah, dan Jawa Timur.
- 2) Taksiran model regresi pada data jumlah kemiskinan di 34 provinsi Indonesia sebagai berikut:
 - a. Menggunakan fungsi pembobot Huber
$$\hat{Y} = 803,54 - 251,46Z_1 + 1131,77Z_2$$
 - b. Menggunakan fungsi pembobot *Bisquare Tukey*

$$\hat{Y} = 827,85 - 197,98Z_1 + 1183,28Z_2$$
 - c. Berdasarkan nilai *Standard Error Bisquare Tukey* ($221 < 229,2$). Sehingga dapat dikatakan bahwa metode estimasi-M dengan pembobotan *Bisquare Tukey* lebih baik dipilih dari pada pembobotan Huber.

DAFTAR PUSTAKA

- [1]. Kuncoro M. *Ekonomi Pembangunan: Teori, Masalah dan Kebijakan*. Yogyakarta: UPP AMP YKPN; 1997.
- [2]. Widarjono A. *Ekonometrika Teori dan Aplikasi untuk Ekonomi dan Bisnis*. Yogyakarta: Fakultas Ekonomi UII; 2007.
- [3]. Soemartini. *Outlier (Pencilan)*. Jatinagor: Universitas Padjajaran; 2007.
- [4]. Chen C. *Robust Regression and Outlier Detection with the Robustreg Procedure*. North Carolina: SAS Inc. 2002; 265-27.
- [5]. Gujarati D. *Ekonometrika Dasar*. [Zain, S]. Jakarta: Erlangga; 1997.
- [6]. Montgomery DC, Peck EA. *Introduction to Linear Regression Analysis*. John Wiley & Sons. New York; 1992.
- [7]. Junaidi J. *Deskripsi Data Melalui Boxplot*. Jambi: Fakultas Ekonomi dan Bisnis Universitas Jambi; 2014.
- [8]. Fox J. *Robust Regression. Appendix to An R and S-Plus Companion to Applied Regression*. 2002; 1-8.

ALDA OKTAVIANI : Jurusan Matematika, FMIPA UNTAN, Pontianak
aldaoktaviani@student.untan.ac.id

SHANTIKA MARTHA : Jurusan Matematika, FMIPA UNTAN, Pontianak
shantika.martha@math.untan.ac.id