

**PENERAPAN REGRESI ROBUST ESTIMASI-S DENGAN FUNGSI  
PEMBOBOT TUKEY BISQUARE DAN WELSCH PADA KASUS DATA  
IPM DI PROVINSI PAPUA TAHUN 2021**

**Nita Nurcahyani**

Jurusan Matematika, Universitas Jenderal Soedirman  
nita.nurcahyani@mhs.unsoed.ac.id

**Budi Pratikno\***

Jurusan Matematika, Universitas Jenderal Soedirman  
budi.pratikno@unsoed.ac.id

**Supriyanto**

Jurusan Matematika, Universitas Jenderal Soedirman  
Supriyanto2505@unsoed.ac.id

**ABSTRACT.** *The research studied the robust regression estimation-S on Tukey-Bisquare and Welsch weighting function. Due to the data contains outliers, we therefore used robust regression to handle outliers. The purpose is to find the best model between the Tukey Bisquare or Welsch. The mean square error (MSE) and adjusted  $R^2$  is then used. The simulation data is a human development index (HDI) in Papua in 2021. The result showed that the robust regression model on the S-estimation between Tukey Bisquare and Welsch are similar (close). However, the adjusted  $R^2$  weighting function Welsch is greater than the Tukey Bisquare, and the MSE of the Welsch is smaller than the Tukey Bisquare, so we recommended the robust S-estimation Welsch is a little better.*

**Keywords:** *Outlier, robust, estimate-S, Tukey Bisquare weighting function, Welsch weighting function.*

**ABSTRAK.** Penelitian ini membahas penerapan model estimasi parameter regresi *robust* estimasi S fungsi pembobot *Tukey-Bisquare* dan *Welsch* pada kasus data IPM tahun 2021. Karena data memuat *outlier*, maka model regresi *robust* dipertimbangkan sebagai model yang dapat mengatasi *outlier*. Tujuan penelitian ini adalah mencari model terbaik antara fungsi pembobot *Tukey Bisquare* dengan *Welsch* berdasarkan kriteria *mean square error (MSE)* dan *adjusted  $R^2$* . Hasil riset menunjukkan bahwa estimasi regresi *robust* estimasi-S dengan fungsi pembobot *Tukey Bisquare* dan *Welsch* serupa atau cenderung sama. Namun jika dilihat dari nilainya, *adjusted  $R^2$*  fungsi pembobot *Welsch* lebih besar dari pembobot *Tukey Bisquare*, dan *MSE* fungsi pembobot *Welsch* lebih kecil dari pembobot *Tukey Bisquare*, karena itu regresi *robust* estimasi-S dengan fungsi pembobot *Welsch* lebih di rekomendasikan

**Kata Kunci:** *Outlier, regresi robust, Estimasi-S, fungsi pembobot Tukey Bisquare, fungsi pembobot Welsch.*

---

\*Penulis Korespondensi

Info Artikel : dikirim 20 Agt. 2022; direvisi 16 Nov. 2022; diterima 27 Des. 2022.

## 1. PENDAHULUAN

Analisis regresi linier adalah analisis yang dilakukan untuk mengetahui hubungan yang linier antara satu atau beberapa variabel bebas dengan variabel tak bebas serta digunakan untuk tujuan prediksi atau peramalan. Analisis regresi menggunakan dua variabel disebut regresi linier sederhana sedangkan analisis regresi dengan lebih dari dua variabel disebut regresi linier berganda (Mardiana dkk., 2021).

Salah satu metode pendugaan parameter model regresi linier adalah Metode Kuadrat Terkecil (MKT). Metode ini merupakan metode penaksir tak bias terbaik untuk parameter model regresi jika data yang digunakan memenuhi asumsi klasik. Dalam kasus model regresi, terdapat kemungkinan adanya data *outlier* (pencilan) yaitu pengamatan dengan nilai mutlak residual jauh lebih besar daripada residual-residual lain sehingga akan mempengaruhi model regresi yang terbentuk. Metode kuadrat terkecil dalam analisis regresi kurang tepat dilakukan untuk data yang mengandung *outlier* karena akan mempengaruhi model regresi. Oleh karena itu, dibutuhkan suatu metode yang kekar (*robust*) terhadap kondisi data yang *outlier* (Lainun dan Tinungki, 2018). Regresi *robust* merupakan metode regresi yang digunakan ketika adanya *outlier* yang berpengaruh pada model (Olive, 2005:3). Menurut Chen (2002:13) regresi *robust* terdiri dari 5 metode estimasi, yaitu (1) estimasi *M* (*Maximum Likelihood type*), (2) estimasi *LMS* (*Least Median Squares*), (3) estimasi *LTS* (*Least Trimmed Squares*), (4) estimasi *MM* (*Method of Moment*), dan (5) estimasi *S* (*Scale*).

Jika dilihat dari nilai *breakdown point*nya, estimasi-*S* merupakan estimasi *robust* yang memiliki nilai *breakdown point* yang paling tinggi hingga 50%, *breakdown point* digunakan untuk menjelaskan ukuran kekekaran (*robustness*) dari teknik *robust*. Menurut Huber (1981: 13), *breakdown point* adalah persentase dari *outlier* yang dapat menyebabkan nilai estimator menjadi besar. Di antara beberapa estimasi untuk mengatasi hal tersebut adalah estimasi-*S*. Estimasi-*S* tersebut memiliki fungsi pembobot di antaranya adalah *Tukey Bisquare* dan *Welsch*. Fungsi pembobot *Tukey Bisquare* dan *Welsch* memiliki nilai skala *robust* yang lebih kecil dibandingkan dengan fungsi pembobot lain sehingga hasilnya

lebih efektif (Yaziz dan Rizki, 2019). Fungsi pembobot ini digunakan untuk menghasilkan nilai skala pembobot yang diperoleh dengan cara melakukan iterasi hingga estimasi yang diperoleh konvergen.

Pada penelitian sebelumnya, Setiawan dkk. (2019) telah mengkaji metode regresi *robust* estimasi-S dengan fungsi pembobot *tukey bisquare* pada kasus data kemiskinan di Indonesia. Hasil penelitian tersebut menyatakan bahwa metode regresi *robust* lebih baik digunakan dibandingkan dengan metode kuadrat terkecil pada data yang mengandung *outlier*. Penelitian lain juga telah dilakukan oleh Cahyawati dkk. (2009) tentang efektivitas metode regresi *robust* penduga *Welsch* dalam mengatasi pencilan pada pemodelan regresi linier berganda. Diperoleh hasil bahwa pembobot *welsch* menghasilkan model regresi yang lebih baik dari metode kuadrat terkecil untuk berbagai ukuran sampel dan banyaknya *outlier*.

Berdasarkan uraian sebelumnya, pada penelitian ini dilakukan penerapan estimasi-S pada kasus data Indeks Pembangunan Manusia di Provinsi Papua. Perbedaan penelitian ini dengan penelitian sebelumnya terletak pada kasus data dan fungsi pembobot yang digunakan.

## 2. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah studi kasus. Dalam hal ini, peneliti menerapkan metode regresi *robust* untuk mengidentifikasi faktor-faktor yang berpengaruh pada nilai IPM di Provinsi Papua tahun 2021. Variabel bebas yang digunakan adalah angka harapan hidup ( $X_1$ ), rata-rata lama sekolah ( $X_2$ ), dan pengeluaran perkapita disesuaikan ( $X_3$ ). Adapun langkah-langkah yang dilakukan dalam penelitian ini adalah:

1. mengumpulkan data dan menentukan variabel bebas dan tak bebas;
2. mencari estimasi parameter regresi dengan menggunakan MKT;
3. melakukan pendeteksian *outlier* dengan metode *DFFITs*;
4. mencari estimasi parameter regresi *robust* estimasi-S menggunakan fungsi pembobot *Tukey Bisquare* dan *Welsch*;
5. melakukan uji signifikansi parameter;
6. menghitung adjusted  $R^2$  dan *MSE*;

7. memilih model terbaik.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Estimasi Regresi Linier Berganda

Sebelum melakukan analisis regresi *robust*, terlebih dahulu dilakukan analisis menggunakan metode kuadrat terkecil (MKT) untuk mendapatkan estimasi parameter model regresi linier berganda. Pada analisis tersebut IPM merupakan variabel tak bebas sedangkan variabel bebas adalah angka harapan hidup, rata-rata lama sekolah, serta pengeluaran perkapita disesuaikan. Diperoleh model dari hasil perhitungan menggunakan R-Studio sebagai berikut:

$$\hat{Y} = 3,7965 + 0,4788X_1 + 2,5611X_2 + 0,0009X_3 \quad (1)$$

Analisis regresi persamaan (1) yang menggunakan metode MKT memerlukan uji asumsi klasik untuk menentukan apakah model tersebut layak atau tidak. Uji asumsi klasik dalam hal ini meliputi uji normalitas, uji heteroskedastisitas, uji autokorelasi, dan uji multikolinieritas. Berikut ini adalah uji asumsi klasik tersebut untuk menguji perlu tidaknya metode regresi *robust* karena adanya *outlier*.

#### 3.2 Uji Asumsi Klasik

##### 3.2.1 Uji Normalitas

Model regresi yang baik adalah model regresi yang memiliki nilai residual berdistribusi normal. Dari hasil *output* R-Studio diperoleh nilai  $p - value = 0,7692 > \alpha$ , untuk nilai  $D_{tabel}$  dengan  $n = 29$  dan  $\alpha = 0,05$  diperoleh  $D_{tabel} = 0,246$ . Jika dibandingkan dengan nilai  $D$  yang diperoleh bahwa  $D = 0,11822 < D_{tabel} = 0,246$ . Berdasarkan kriteria keputusan maka  $H_0$  diterima. Hal ini berarti bahwa asumsi normalitas residual terpenuhi.

##### 3.2.2 Uji Heteroskedastisitas

Uji heteroskedastisitas dilakukan untuk mengetahui adanya gejala heteroskedastisitas pada model regresi. Untuk mengetahui terdapat heteroskedastisitas atau tidak dilakukan dengan uji *Glejser*, yaitu dengan meregresikan residual model dengan variabel bebas. Berdasarkan *output* R-Studio diperoleh hasil untuk uji *Glejser* pada Tabel 1.

**Tabel 1** Hasil Uji *Glejser*

	Variabel	$p - value$
1	$X_1$	0,008
2	$X_2$	0,014
3	$X_3$	0,007

Berdasarkan Tabel 1, diperoleh  $p - value$  masing-masing variabel bebas kurang dari  $\alpha = 0,05$ , artinya terdapat gejala heteroskedastisitas pada model tersebut. Oleh sebab itu, asumsi heteroskedastisitas dalam hal ini tidak terpenuhi.

### 3.2.3 Uji Autokorelasi

Uji autokorelasi bertujuan untuk menguji apakah dalam model regresi linier terdapat korelasi antara residual satu dengan yang lainnya. Uji yang dilakukan yaitu menggunakan uji *Durbin-Watson*. Hasil perhitungan dengan menggunakan R-Studio diperoleh nilai *Durbin-Watson* yaitu  $d_{hitung} = 2,0523$  dan  $p - value = 0,4737$ . Pada tabel *Durbin-Watson* diperoleh nilai  $d_L = 1,2138$ , karena nilai  $d_{hitung} > d_L$  maka  $H_0$  gagal ditolak, artinya tidak terdapat autokorelasi.

### 3.2.4 Uji Multikolinieritas

Uji ini dilakukan untuk mengetahui hubungan antara masing-masing variabel bebas pada model regresi. Ada tidaknya suatu gejala multikolinieritas pada model dapat berdasarkan nilai *VIF*. Model regresi dikatakan tidak terjadi multikolinieritas jika  $VIF < 10$ . Adapun nilai *VIF* berdasarkan *output R-Studio* disajikan pada Tabel 2.

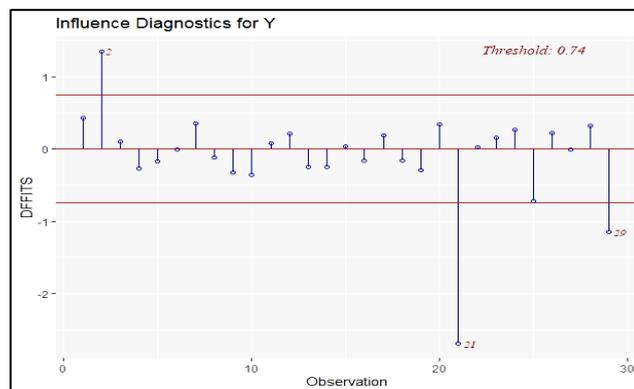
**Tabel 2** Nilai *VIF* masing-masing variabel bebas

	Variabel	$R^2$	<i>VIF</i>
1	$X_1$	0,312733	1,455039
2	$X_2$	0,686013	3,184846
3	$X_3$	0,707281	3,416243

Berdasarkan Tabel 2 didapatkan nilai  $VIF < 10$ , sehingga dapat dikatakan tidak terdapat masalah multikolinieritas pada model.

### 3.3 Identifikasi *Outlier*

Identifikasi *outlier* yaitu menggunakan metode *DFFITs*. Pada identifikasi *outlier* menggunakan metode *DFFITs* berlaku ketentuan jika nilai  $|DFFITs|$  masing-masing data lebih dari  $2\sqrt{\frac{k+1}{n}}$  maka data tersebut dikategorikan sebagai data *outlier*. Nilai  $k + 1$  merupakan jumlah variabel bebas ditambah satu, sedangkan  $n$  merupakan banyaknya pengamatan (Mardiana dkk., 2021). Pada penelitian ini, jumlah variabel bebas yang digunakan ( $k$ ) meliputi angka harapan hidup, rata-rata lama sekolah, dan pengeluaran perkapita disesuaikan.  $k + 1 = 3 + 1 = 4$ , sedangkan banyaknya pengamatan ( $n$ ) adalah 29 pengamatan. Maka diperoleh nilai  $2\sqrt{\frac{k+1}{n}} = 2\sqrt{\frac{4}{29}} = 0,74278$ . Oleh karena itu, batas nilai penentuan berdasarkan  $|DFFITs| > 0,74278$  merupakan data *outlier*. Berdasarkan hasil perhitungan diperoleh bahwa dari 29 data pengamatan terdapat 3 nilai  $DFFITs > 0,74278$ . Dengan menggunakan *R-Studio* dan diperoleh plot *DFFITs* pada Gambar 1.



**Gambar 1** Plot *DFFITs*

Gambar 1 menunjukkan bahwa garis-garis yang keluar pada batas signifikansi menunjukkan adanya pencilan dari setiap variabel. Dari hasil identifikasi *outlier* disimpulkan bahwa terdapat tiga *outlier* yaitu data ke-2, 21, dan 29, hal itu tentu berpengaruh pada model yang diperoleh menggunakan MKT. Selanjutnya untuk mengatasi *outlier* tersebut digunakan regresi *robust* estimasi-S

dengan fungsi pembobot *Tukey Bisquare* dan *Welsch* dengan tujuan memperoleh hasil yang lebih efektif.

### 3.4 Regresi *Robust Estimasi-S*

#### 3.4.1 Estimasi-S dengan Fungsi Pembobot *Tukey Bisquare*

Berdasarkan perhitungan *R-Studio* diperoleh estimasi parameter yang konvergen setelah dilakukan sebanyak 38 iterasi dengan model sebagai berikut:

$$\hat{Y} = 8,824 + 0,429X_1 + 2,281X_2 + 0,001X_3 \quad (2)$$

Persamaan (2) adalah persamaan model regresi robust *Tukey Bisquare* yang menunjukkan bahwa variabel  $X_1$  memberikan pengaruh yang positif terhadap IPM Papua sebesar 0,429 artinya jika nilai  $X_1$  meningkat sebesar 1 satuan, maka nilai IPM Papua akan meningkat sebanyak 0,429. Variabel  $X_2$  memberikan pengaruh positif sebesar 2,281 artinya jika nilai  $X_2$  meningkat sebesar 1 satuan, maka nilai IPM Papua akan meningkat sebanyak 2,281. Variabel  $X_3$  juga memiliki pengaruh yang positif sebesar 0,001 artinya jika nilai  $X_3$  meningkat sebesar 1 satuan, maka nilai IPM Papua akan meningkat sebanyak 0,001.

#### 3.4.2 Estimasi-S dengan Fungsi Pembobot *Welsch*

Berdasarkan perhitungan *R-Studio* diperoleh estimasi parameter yang konvergen setelah dilakukan sebanyak 62 iterasi dengan model sebagai berikut:

$$\hat{Y} = 9,409 + 0,418X_1 + 2,295X_2 + 0,001X_3 \quad (3)$$

Persamaan (3) adalah model regresi robust menggunakan fungsi pembobot *Welsch* menunjukkan bahwa variabel  $X_1$  memberikan pengaruh yang positif terhadap IPM Papua sebesar 0,418 artinya jika nilai  $X_1$  meningkat sebesar 1 satuan, maka nilai IPM Papua akan meningkat sebesar 0,418. Variabel  $X_2$  memberikan pengaruh positif sebesar 2,295 artinya jika nilai  $X_2$  meningkat sebesar 1 satuan, maka nilai IPM Papua akan meningkat sebesar 2,295. Variabel  $X_3$  juga memiliki pengaruh yang positif sebesar 0,001 artinya jika nilai  $X_3$  meningkat sebesar 1 satuan, maka nilai IPM Papua akan meningkat sebanyak 0,001.

### 3.5 Uji Signifikansi Parameter

Uji signifikansi dalam hal ini adalah Uji Simultan (Uji  $F$ ) dan Uji Parsial (Uji  $t$ ). Uji  $F$  digunakan untuk pengujian secara simultan (bersamaan) untuk mengetahui ada tidaknya pengaruh simultan dari keseluruhan variabel bebas terhadap variabel tak bebas. Sementara, uji  $t$  digunakan untuk menguji pengaruh masing-masing variabel bebas terhadap variabel tak bebas.

#### 3.5.1 Uji $F$

Pada persamaan estimasi parameter model yang dihasilkan dari fungsi pembobot *Tukey Bisquare* dan *Welsch*, dilakukan uji secara simultan dan diperoleh hasil yang disajikan pada Tabel 3.

**Tabel 3** Hasil Uji  $F$

Model	$F_{Hitung}$	$p$ -value
<i>Tukey bisquare</i>	194,0118	0,0000
<i>Welsch</i>	195,9576	0,0000

Berdasarkan Tabel 3 menunjukkan bahwa masing-masing  $p$  – value untuk pembobot *Tukey Bisquare* dan *Welsch* kurang dari 0,05. Hal ini menunjukkan bahwa kedua model tersebut menunjukkan bahwa secara simultan variabel  $X_1$ ,  $X_2$ , dan  $X_3$  berpengaruh terhadap variabel  $Y$ .

#### 3.5.2 Uji $t$

Pada persamaan estimasi parameter model yang dihasilkan dari fungsi pembobot *Tukey Bisquare* dan *Welsch*, dilakukan uji secara parsial dan diperoleh hasil pada Tabel 4.

**Tabel 4** Uji  $t$  pembobot *Tukey Bisquare*

Variabel	Parameter	$t_{hitung}$	$p$ -value
$X_1$	$\beta_1$	2,957	0,00669
$X_2$	$\beta_2$	9,572	0,0000
$X_3$	$\beta_3$	3,569	0,00148

Berdasarkan Tabel 4 menunjukkan bahwa masing-masing  $p - value$  untuk uji  $t$  pada pembobot *Tukey Bisquare* kurang dari 0,05. Hal ini menunjukkan bahwa setiap variabel bebas bahwa secara parsial berpengaruh terhadap variabel  $Y$ . Selanjutnya untuk uji  $t$  pada pembobot *Welsch* diperoleh hasil pada Tabel 5.

**Tabel 5** Uji  $t$  pembobot *Welsch*

Variabel	Parameter	$t_{hitung}$	$p-value$
$X_1$	$\beta_1$	2,746	0,01101
$X_2$	$\beta_2$	9,225	0,0000
$X_3$	$\beta_3$	3,460	0,00195

Berdasarkan Tabel 5 menunjukkan bahwa masing-masing  $p - value$  untuk uji  $t$  pada pembobot *Welsch* kurang dari 0,05. Hal ini menunjukkan bahwa setiap variabel bebas bahwa secara parsial berpengaruh terhadap variabel  $Y$ .

### 3.6 Memilih Model Terbaik

Dalam menentukan metode terbaik, digunakan dua kriteria, yaitu nilai *adjusted R<sup>2</sup>* dan *mean square error* (Mardiana dkk., 2021). Nilai *adjusted R<sup>2</sup>* dan *MSE* selanjutnya disajikan pada Tabel 6.

**Tabel 6** Nilai *adjusted R<sup>2</sup>* dan *MSE*

Fungsi pembobot	<i>Adjusted R<sup>2</sup></i>	<i>MSE</i>
<i>Tukey Bisquare</i>	0,9925	4,210451
<i>Welsch</i>	0,9931	4,172332

Berdasarkan Tabel 6 terlihat bahwa nilai *adjusted R<sup>2</sup>* dan *MSE* untuk fungsi pembobot *Tukey Bisquare* dan *Welsch* memiliki selisih yang tidak terlalu besar, sehingga dapat dikatakan hasil estimasi regresi *robust* dengan fungsi pembobot *Tukey Bisquare* dan *Welsch* serupa atau cenderung sama.

## 4. KESIMPULAN

Berdasarkan analisis yang telah dilakukan diperoleh bahwa analisis regresi dengan menggunakan MKT pada kasus data IPM di Provinsi Papua terjadi adanya pelanggaran asumsi pada model dan diketahui terdapat tiga data *outlier*. Untuk

mengatasi masalah tersebut diperlukan metode yang mampu menangani data yang mengandung *outlier*, yaitu menggunakan metode regresi *robust* dengan fungsi pembobot *Tukey Bisquare* dan *Welsch*. Hasil penelitian ini adalah bahwa estimasi regresi *robust* estimasi-*S* dengan fungsi pembobot *Tukey Bisquare* dan *Welsch* serupa atau cenderung sama. Namun jika dilihat dari nilainya, *adjusted R<sup>2</sup>* fungsi pembobot *Welsch* lebih besar dari pembobot *Tukey Bisquare* dan *MSE* fungsi pembobot *Welsch* lebih kecil dari pembobot *Tukey Bisquare*. Oleh karena itu, model yang direkomendasikan untuk mengestimasi data IPM di Provinsi Papua tahun 2021 adalah regresi *robust* estimasi-*S* dengan fungsi pembobot *Welsch* dengan persamaan model sebagai berikut:

$$\hat{Y} = 9,409 + 0,418X_1 + 2,295X_2 + 0,001X_3$$

dengan nilai *adjusted R<sup>2</sup>* sebesar 0,9931 dan nilai *MSE* sebesar 4,172332.

#### DAFTAR PUSTAKA

- Cahyawati, D., Tanuji, H., dan Abdiati, R. *Efektivitas metode regresi robust penduga Welsch dalam mengatasi pencilan pada pemodelan regresi linear berganda*, Jurnal Penelitian Sains, **12**(1) (2009).
- Chen, C., *Robust Regression and Outlier Detection with the Robustreg Procedure. Statistics and Data Analysis*, SAS Institute: Cary, NC., 2002.
- Huber, P.J., *Robust Statistics*, A John Wiley & Sons, Inc., Canada, 1981.
- Lainun, H., Tinungki, G. M., dan Amran, A., *Perbandingan Penduga M, S, dan MM pada Regresi Linier dalam Menangani Keberadaan Outlier*. Jurnal Matematika, Statistika dan Komputasi, **15**(1) (2018), 88-96.
- Mardiana, M., Wibowo, A., Mahmudah, M., dan Festi, P., *A comparison of m-estimation and s-estimation on the factors affecting IR DHF in East Java in 2017*, The Indonesian Journal of Public Health, **16**(3) (2021), 349-362.
- Olive, D.J., *Applied Robust Statistics*, Southern Illinois University, Carbondale, 2005.
- Setiawan, W., Debataraja, N. N., dan Sulistianingsih, E., *Metode Estimasi-S Pada Analisis Regresi Robust Dengan Pembobotan Tukey Bisquare*, Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya, **8**(2) (2019).

---

Yaziz D. K., dan Rizki, S. W., *Analisis Regresi Robust Estimasi-M dengan menggunakan Pembobotan Bisquare Tukey dan Welsch dalam mengatasi Data Outlier*, Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya, **8**(4) .(2019).

