

STUDI KOMPARATIF METODE KUADRAT TERKECIL DENGAN METODE REGRESI ROBUST PEMBOBOT WELSCH PADA DATA YANG MENGANDUNG PENCILAN

NOVERIKA ANGGRAINI PUTRI

*Program Studi Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia,
noverikaputri@rocketmail.com*

Abstrak. Metode Kuadrat Terkecil (MKT) atau *Ordinary Least Square* (OLS) merupakan salah satu metode yang sering digunakan untuk mendapatkan nilai-nilai penduga parameter model regresi. Namun demikian, metode ini juga memiliki kelemahan, dimana penduga yang dihasilkan sangat dipengaruhi oleh adanya data yang polanya menyimpang dari pola umum data yang disebut pencilan (*outlier*). Namun demikian, tindakan membuang begitu saja suatu pencilan bukanlah tindakan yang bijaksana, karena adakalanya pencilan memberikan informasi yang cukup berarti. Oleh karena itu dibutuhkan suatu metode regresi yang kekar, tidak dipengaruhi oleh adanya pencilan. Metode regresi semacam ini dinamakan sebagai Metode Regresi *Robust*. Salah satu pembobot pada metode ini adalah pembobot *Welsch*.

Kata Kunci: Metode Kuadrat Terkecil, Metode Regresi Robust Pembobot *Welsch*, *Root Means Error* (RMSE), Pencilan (*outlier*)

1. Pendahuluan

Analisis regresi adalah analisis statistik yang digunakan untuk memodelkan hubungan antara beberapa variabel dalam suatu sistem. Hubungan antara beberapa variabel tersebut biasanya dinyatakan dalam suatu model matematika yang dinamakan model regresi. Dalam model regresi terlibat dua jenis variabel, yaitu variabel bebas dan variabel tak bebas.

Didalam menduga model regresi, Metode Kuadrat Terkecil (MKT) atau *Ordinary least square* (OLS) merupakan salah satu metode yang sering digunakan untuk mendapatkan nilai-nilai penduga parameter model regresi. Namun demikian, metode ini juga memiliki kelemahan, dimana penduga yang dihasilkan sangat dipengaruhi oleh adanya data yang polanya menyimpang dari pola umum data yang disebut pencilan (*outlier*). Suatu pencilan dalam data dapat mengakibatkan estimasi koefisien regresi yang diperoleh tidak tepat. Namun demikian, tindakan membuang begitu saja suatu pencilan bukanlah tindakan yang bijaksana, karena adakalanya pencilan memberikan informasi yang cukup berarti [2]. Oleh karena itu diperlukan suatu metode regresi yang kekar, tidak dipengaruhi oleh adanya pencilan. Metode regresi semacam ini dinamakan sebagai Metode Regresi *Robust*.

Metode regresi robust diperkenalkan oleh Andrews [2] dan merupakan metode regresi yang digunakan ketika distribusi dari galat tidak normal atau adanya beberapa pencilan yang berpengaruh pada model. Beberapa pembobot yang dimiliki oleh Regresi Robust salah satunya adalah pembobot *Welsch*.

Makalah ini bertujuan untuk membentuk model dengan Metode Regresi Robust pembobot *Welsch* dan membandingkan kedua metode tersebut berdasarkan nilai *Root Means Square Error* (RMSE) pada analisis regresi linier sederhana.

2. Analisis Regresi Linier

Analisis regresi merupakan teknik statistika yang digunakan untuk memodelkan hubungan antara variabel tak bebas dengan satu atau lebih variabel bebas. Terdapat dua jenis analisis regresi linear, yaitu regresi linear sederhana dan regresi linear berganda.

Analisis regresi linier sederhana merupakan analisis regresi yang melibatkan satu variabel bebas dan satu variabel tak bebas. Model regresi linier sederhana dapat dinyatakan sebagai berikut [2].

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

dengan Y_i merupakan nilai variabel tak bebas untuk pengamatan ke- i , X_i merupakan nilai variabel bebas untuk pengamatan ke- i , β_0, β_1 merupakan koefisien regresi dan ε_i merupakan nilai galat untuk pengamatan ke- i .

Analisis regresi linear berganda merupakan metode regresi yang melibatkan satu variabel tak bebas dan beberapa variabel bebas. Model regresi linier berganda dapat dinyatakan sebagai berikut.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

dengan Y_i merupakan nilai variabel tak bebas untuk pengamatan ke- i , X_{ik} merupakan nilai variabel bebas ke- k untuk pengamatan ke- i , β_k merupakan koefisien regresi ke- k dan ε_i merupakan nilai galat untuk pengamatan ke- i .

2.1. Metode Kuadrat Terkecil

Metode Kuadrat Terkecil (MKT) adalah salah satu metode penduga parameter dalam model regresi. Metode ini lebih sering digunakan daripada metode penduga parameter yang lain [1].

Dengan metode kuadrat terkecil diperoleh nilai dugaan dari b_0 dan b_1 dengan meminimumkan jumlah kuadrat sisaan (JKS), sehingga diperoleh persamaan berikut

$$b_0 = Y - b_1 X,$$

$$b_1 = \frac{\sum_{i=1}^n x_i y_i - [(\sum_{i=1}^n x - i)(\sum_{i=1}^n y_i)] \setminus n}{\sum_{i=1}^n x_i^2 - ((\sum_{i=1}^n x_i)^2) \setminus n}$$

2.2. External Studentization

Pencilan adalah data pengamatan yang tidak mengikuti pola umum model regresi yang dihasilkan atau tidak mengikuti pola secara keseluruhan. Pencilan dapat diidentifikasi melalui beberapa metode, salah satunya dengan melakukan uji *External Studentization (R-student)*.

Hipotesis dari pengujian ini adalah

$$H_0 : \Delta_i = 0 \text{ (tidak terdapat pencilan),}$$

$$H_1 : \Delta_i \neq 0 \text{ (terdapat nilai pencilan).}$$

Statistik uji yang digunakan adalah

$$t_i = \frac{Y_i - \hat{Y}_i}{s_{-i} \sqrt{1 - h_{ii}}},$$

dengan kriteria uji pada pengujian ini adalah $|t_i| > t_{\alpha \setminus n, n-p-1}$ maka tolak H_0 , sebaliknya jika $|t_i| \leq t_{\alpha \setminus n, n-p-1}$ maka terima H_0 .

3. Metode Regresi Robust

Regresi Robust diperkenalkan oleh Andrews (1972) dan merupakan metode regresi yang digunakan ketika distribusi dari galat tidak normal atau adanya beberapa pencilan yang berpengaruh pada model. Beberapa metode penduga dalam regresi robust salah satunya adalah Estimasi-M.

3.1. Estimasi-M

Salah satu metode estimasi regresi robust paling luas digunakan adalah *Estimasi-M*, yang dikenalkan oleh Huber. *Estimasi-M* meminimalisasi fungsi ρ yang merupakan fungsi dari sisaan dengan fungsi obyektif sebagai berikut

$$\sum_{i=1}^n \rho(\varepsilon_i) = \sum_{i=1}^n \rho\left(Y_i - \sum_{j=1}^k X_{ij} \beta_j\right).$$

Kemudian dicari turunan parsial pertama fungsi obyektif pada persamaan di atas terhadap $\beta_j, j = 0, 1, \dots, k$ dan disamakan dengan nol dan sisaan dibakukan, maka diperoleh

$$\sum_{i=1}^n x_{ij} \psi \left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{s} \right) = 0. \quad (3.1)$$

Kemudian persamaan (3.1) ditulis menjadi

$$\sum_{i=1}^n x_{ij} w_i \left(y_i - \sum_{j=0}^k x_{ij} \beta_j \right),$$

dan dapat ditulis dalam bentuk matriks sebagai

$$b = (X'WX)^{-1} X'WY.$$

3.1.1. Pembobot Welsch

Pembobot Welsch adalah salah satu pembobot dalam regresi robust metode estimasi-M. Fungsi Welsch dinyatakan sebagai berikut;

$$\rho(u) = \frac{c}{2} [1 - \exp(-(\frac{c}{2})^2)].$$

Fungsi pembobot Welsch dinyatakan sebagai berikut

$$W(u) = \frac{\psi(u)}{u} = \frac{u[\exp(-(\frac{c}{2})^2)]}{u} = [\exp(-(\frac{c}{2})^2)],$$

dimana $\psi(u)$ merupakan turunan parsial dari fungsi welsch terhadap u serta $c = 2,9846$.

4. Metode Penelitian

Data yang akan dianalisis pada penelitian ini dibangkitkan dengan menggunakan *software MINITAB versi 16*. Data terdiri dari satu variabel bebas dan satu variabel tak bebas. Pembangkitan data dilakukan untuk sampel berukuran $n = 20, 40, 60$ dan 100 dengan proporsi pencilan (p) sekitar $10 - 20$ dari ukuran sampel. Penelitian ini dilakukan dengan tahapan sebagai berikut.

- (1) Tahap pembangkitan data.
 - (a) Tetapkan nilai β_0 dan β_1 .
 - (b) Bangkitkan nilai x_i sebanyak n dari sebaran Uniform(2,5).
 - (c) Bangkitkan $(n - np)$ nilai ε_1 dari sebaran normal $N \sim (0, 1)$.
 - (d) Bangkitkan pn nilai ε_2 dari sebaran normal $N \sim (5, 1)$. Nilai ε_2 ini akan digunakan untuk mendeteksi pencilan.
 - (e) Gabungkan data pada ε_1 dengan ε_2 yang menghasilkan data baru ε .
 - (f) Tentukan nilai untuk y_i melalui persamaan

$$y_i = \beta_0 + \beta_1 x_1 + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

- (2) Tahap Metode Kuadrat Terkecil
 - (a) Duga β_0 dan β_1 dengan metode kuadrat terkecil.
 - (b) Hitung nilai RMSE (*Root Means Square Error*) dari metode kuadrat terkecil.
- (3) Tahap Metode Regresi Robust Pembobot Welsch
 - (a) Melakukan estimasi regresi dengan metode kuadrat terkecil (MKT).
 - (b) Mendeteksi pencilan dengan menggunakan *External Studentization* (*R-student*).
 - (c) Menentukan penduga parameter β dengan metode kuadrat terkecil.
 - (d) Menentukan vektor sisaan, $\varrho_i = y_i - \hat{y}_i$.
 - (e) Menentukan fungsi bobot $u_{i(i)} = \frac{\varrho_i}{s}$.
 - (f) Menetuan matriks pembobot $W = [\exp(-(\frac{u_{i(i)}}{2,9846})^2)]$.
 - (g) Menghitung parameter b dengan $b = (X'WX)^{-1}X'WY$.
 - (h) Mengulangi langkah 4 sampai 7 agar diperoleh nilai b yang sama dengan sebelumnya.

- (i) Hitung nilai RMSE.
- (4) Tahap Perbandingan Metode Kuadrat Terkecil dengan Metode Regresi *Robust* dengan Pembobot *Welsch*.
- (a) Lakukan langkah 1 sampai 3 sebanyak 100 kali.
- (b) Hitung rata-rata nilai RMSE untuk Metode Kuadrat Terkecil.
- (c) Hitung rata-rata nilai RMSE untuk Metode Regresi *Robust* dengan Pembobot *Welsch*.
- (d) Bandingkan nilai rata-rata RMSE yang didapatkan dari kedua metode tersebut.

Lakukan langkah 1 sampai langkah 4 untuk semua kombinasi $n = 20, 40, 60$ dan 100 dengan proporsi pencilan sekitar 10 – 20.

5. Hasil dan Pembahasan

Data amatan dibangkitkan dan diolah dengan menggunakan *software MINITAB versi 16*. Data yang dibangkitkan adalah data dengan ukuran sampel 20,40,60, dan 100. Simulasi dilakukan sebanyak 100 kali ulangan. Data sampel yang telah dibangkitkan dengan ukuran sampel 20 dapat dilihat pada Tabel 1.

Tabel 1. Data bangkitan dengan $n = 20$

No	Y	X
1	14,5992	4,82947
2	13,7816	4,64891
3	13,5992	4,60341
4	12,3222	4,51905
5	13,9224	4,45953
6	12,3062	4,24072
7	12,2514	4,20615
8	10,4504	4,06565
9	11,5477	3,84134
10	11,6540	3,65625
11	10,7092	3,58502
12	12,0693	3,48318
13	9,8206	3,43769
14	10,3149	3,28754
15	9,3590	3,09958
16	7,6088	2,78484
17	8,0603	2,36869
18	8,2366	2,29456
19	12,1828	2,04006
20	14,5012	2,01628

Dengan menggunakan *software MINITAB versi 16*, diperoleh model prediksi dengan kuadrat terkecil $\hat{Y} = 7,04 + 1,24X_1$ dengan nilai $RMSE = 1,844$.

Data pencilan dapat dilihat terhadap posisi dan sebaran data yang lainnya sehingga akan dievaluasi apakah data pencilan tersebut perlu dihilangkan atau tidak. terdapat beberapa cara untuk mendeteksi pencilan salah satunya dengan melakukan uji *External Studentization (R-student)*. Hasil dari nilai *R-student* dapat dilihat pada Tabel 2.

Tabel 2. Hasil dari nilai *R-student*

No	Y	x_1	TRES1
1	14,5992	4,82947	0,92605
2	13,7816	4,64891	0,55986
3	13,5992	4,60341	0,48542
4	12,322	4,51905	-0,17507
5	13,9224	4,45953	0,76889
6	12,3062	4,24072	0,00828
7	12,2514	4,20615	0,00168
8	10,4504	4,06565	-0,90637
9	11,5477	3,84134	-0,13497
10	11,6540	3,65625	0,04683
11	10,7092	3,58502	-0,41843
12	12,0693	3,48318	0,38900
13	9,8206	3,43769	-0,81397
14	10,3149	3,28754	-0,43391
15	9,3590	3,09958	-0,84483
16	7,6088	2,78484	-1,72425
17	8,0603	2,36869	-1,12942
18	8,2366	2,29456	-0,096893
19	12,1828	2,04006	1,76445
20	14,5012	2,01628	4,18503

Pengamatan akan dikategorikan sebagai pencilan adalah $t_i > t_{\frac{\alpha}{2}, n-p-1} = 1,73406$, dengan taraf nyata 10%. Dari tabel diatas dapat dilihat bahwa data ke-19 dan ke-20 adalah data pencilan.

Proses perhitungan estimasi-M dengan pembobot welsch dimulai dengan menentukan dugaan awal koefisien regresi, yang diperoleh dari metode kuadrat terkecil $\hat{Y} = 7,04 + 1,24X_1$. Kemudian olah data tersebut sehingga diperoleh nilai $\hat{Y}_1, \varepsilon_1, u_1, w_1$ pada iterasi pertama. Hasilnya dapat dilihat pada Tabel 3.

Tabel 3. Nilai $\hat{Y}_1, \varepsilon_1, u_1, w_1$

No	Y	\hat{Y}_1	ε_1	u_1	w_1
1	14,5992	13,02854	1,57905	1,28314	0,83124
2	13,7816	12,80465	0,98498	0,80039	0,93061
3	13,5992	12,74823	0,85896	0,69799	0,94678
4	12,3222	12,64362	-0,31354	-0,25479	0,99274
5	13,9224	12,56982	1,36031	1,10539	0,87182
6	12,3062	12,29849	0,01507	0,01225	0,99998

No	Y	\hat{Y}_1	ε_1	u_1	w_1
7	12,2514	12,25563	0,00307	0,00249	1,00000
8	10,4504	12,08141	-1,62395	-1,31962	0,82243
9	11,5477	11,80326	-0,24892	-0,20227	0,99542
10	11,6540	11,57375	0,08660	0,07037	0,99944
11	10,7092	11,48542	-0,77005	-0,62574	0,95700
12	12,0693	11,35914	0,71620	0,58199	0,96269
13	9,8206	11,30274	-1,47626	-1,19961	0,85082
14	10,3149	11,11655	-0,79602	-0,64685	0,95412
15	9,3590	10,88348	-1,51921	-1,23451	0,84275
16	7,6088	10,49320	-2,87962	-2,33998	0,54081
17	8,0603	9,97718	-1,91282	-1,55436	0,76244
18	8,2366	9,88525	-1,64479	-1,33656	0,81829
19	12,1828	9,56967	2,61655	2,12621	0,60200
20	14,5012	9,54019	4,96439	4,03406	0,16091

Selanjutnya, data diatas diolah dengan menggunakan rumus

$$b = (X'WX)^{-1}X'WY.$$

Selanjutnya dengan menggunakan b pada iterasi pertama, kita dapat menentukan nilai $\hat{Y}_2, \varepsilon_2, u_2, w_2$ pada iterasi kedua Iterasi berlanjut hingga memperoleh b yang sama dengan iterasi sebelumnya. Hasil perhitungan b pada tiap-tiap iterasi disajikan dalam Tabel 4.

Tabel 4. Hasil estimasi parameter tiap-tiap iterasi

Iterasi	b_0	b_1
MKT	7,04	1,24
Iterasi 1	4,81	1,79
Iterasi 2	2,53	2,36
Iterasi 3	2,02	2,49
Iterasi 4	1,91	2,52
Iterasi 5	1,88	2,53
Iterasi 6	1,86	2,53
Iterasi 7	1,86	2,53

Proses berhenti pada iterasi ke-7, karena nilai b yang baru sama dengan nilai b sebelumnya. Jadi model regresi linearnya adalah

$$\hat{Y} = 1,86 + 2,53X_1$$

dengan nilai $RMSE = 0,490$.

Dengan demikian untuk $n = 20$ yang mengandung pencilan sebesar 10 %, penggunaan parameter dengan Metode Regresi Roust Pembobot Welsch melalui tujuh iterasi menghasilkan model yang lebih baik dibandingkan dengan MKT, karena nilai RMSE metode regresi pembobot welsch lebih kecil dari nilai RMSE MKT. Nilai simpangan baku galat (RMSE) untuk berbagai ukuran sampel dan beberapa ukuran pencilan melalui kedua metode ditampilkan pada Tabel 5.

Tabel 5 Nilai simpangan baku galat (RMSE) untuk berbagai ukuran sampel dan banyak pencilan

No	N	Banyaknya Pencilan	Banyaknya Iterasi*	RMSE* MKT	RMSE* Welsch
1	20	2	8	1,638	0,699
2	20	3	8	1,779	0,902
3	20	4	7	1,753	1,363
4	40	4	6	1,828	0,758
5	40	6	6	2,057	0,783
6	40	8	7	2,224	0,900
7	60	6	6	1,840	0,790
8	60	9	6	2,076	0,778
9	60	12	6	2,260	0,890
10	100	10	6	1,806	0,788
11	100	15	6	2,049	1,990
12	100	20	6	2,221	0,895

Keterangan. RMSE*(nilai rata-rata RMSE), Iterasi*(rata-rata iterasi)

Berdasarkan Tabel 5 dapat dilihat bahwa untuk berbagai ukuran sampel dan proporsi pencilan sekitar 10 %- 0 %, pendugaan parameter Metode regresi robust pembobot welsch menghasilkan model yang lebih baik daripada model regresi yang dihasilkan oleh MKT. Ini dapat dilihat berdasarkan nilai RMSE yang dihasilkan oleh metode regresi robust pembobot welsch lebih kecil dibandingkan nilai RMSE yang dihasilkan oleh metode kuadrat terkecil.

6. Kesimpulan

Model regresi yang diperoleh dengan Metode Regresi Robust Pembobot Welsch lebih baik dari model regresi yang diperoleh dengan menggunakan Metode Kuadrat Terkecil (MKT), untuk berbagai ukuran sampel dan banyak pencilan. Hal ini ditunjukkan oleh nilai simpangan baku galat (RMSE) yang dihasilkan kedua metode tersebut. Dalam hal ini nilai RMSE hasil dari MRR-W selalu lebih kecil dibandingkan nilai RMSE pada MKT.

7. Ucapan Terima kasih

Penulis mengucapkan terima kasih kepada ibu Izzati Rahmi H.G, M.Si dan ibu Hazmira Yozza, M.Si atas bimbingannya dalam penelitian ini. Terima kasih juga kepada ibu Dr. Maiyastri, Bapak Yudiantri Asdi, M.Sc, Bapak Dr.Dodi Devianto dan Bapak Narwen, M.Si yang telah memberikan masukan dan saran sehingga makalah ini dapat diselesaikan dengan baik.

Daftar Pustaka

- [1] Myers, R.H. 1990. **Classical and Modern Regression with Applications**, PWS-KENT, Boston

- [2] Montgomery, D. C., dan Peck, E. A. 1992. **Introduction to Linear Regression Analysis**. New York, Wiley-Interscience Publication
- [3] Ryan, T.P., 1997, **Modern Regression Models**, John Wiley dan Sons, New York
- [4] Cahyawati, D. dkk, 2007, Efektivitas Metode Regresi Robust Penduga Welsch dalam Mengatasi Pencilan Pada Pemodelan Regresi Linear Berganda, *Jurnal Penelitian Sains* **12**(1A), UP2M FMIPA Univ. Sriwijaya
- [5] Fox, J. 2002. **Robust Regression**. New York
- [6] Pratiwi, Kishartya, 2012, *Model Regresi Robust Dengan Pembobot Welsch Dan Pembobot Ramsay*, Skripsi Sarjana Sains Matematika, tidak dipublikasikan, FMIPA UNS