PERBANDINGAN METODE ESTIMASI M DAN ESTIMASI MM (*METHODE OF MOMENT*) PADA REGRESI *ROBUST*

Arlinda Amalia Dewayanti[1], Edy Widodo[2]
[1]Mahasiswa Statistika Universitas Islam Indonesia, [2]Dosen Statistika Universitas Islam Indonesia
Dewayanti30@gmail.com, Edykafifa@gmail.com

Abstrak

Analisis regresi merupakan suatu metode dalam statistika untuk mengetahui hubungan antara variabel dependen dengan variabel independen. Metode yang digunakan dalam estimasi parameter pada model tersebut adalah Metode Kuadrat Terkecil (MKT).Metode ini sangat peka terhadap penyimpangan-penyimpangan asumsi pada data.Asumsi yang sering tidak terpenuhi adalah asumsi normalitas.Salah satu penyebab tidak terpenuhinya asumsi ini karena terdapat *outlier* pada data. Oleh sebab itu, digunakan metode lain untuk menangani data *outlier*. Salah satunya adalah metode regresi *robust* dengan menggunakan estimasi M dan MM (*Methode of Moment*). Metode yang digunakani adalah estimasi MKT, estimasi M, dan estimasi MM. Pada penelitian ini akan dibahas mengenai perbandingan antara M dan estimasi MM dengan metode MKT dilihat dari nilai *residual standard error*, *standard error*dan nilai koefisien regresi. Hasil yang diperoleh dengan simulasi data menunjukkan bahwa untuk data yang mengandung *outlier* estimasi parameter yang diperoleh pada metode regresi *robust* dengan metode M lebih baik digunakan dibandingkan dengan metode MKT. Sedangkan untuk data tanpa *outlier* estimasi parameter yang diperoleh dengan metode MKT lebih baik dibandingkan dengan metode estimasi M dan estimasi MM

Kata Kunci: *Estimasi M; Estimasi MM; MKT; Outlier; Regresi Robust*

1. PENDAHULUAN

Pada berbagai kasus, tidak jarang ditemukan kondisi dimana asumsi-asumsi tersebut tidak terpenuhi. Jika asumsi tidak terpenuhi akan mengakibatkan hasil estimasi parameter pada MKT kurang baik Salah satu asumsi yang tidak terpenuhi adalah asumsi normalitas. Hal ini disebabkan adanya pencilan atau *outlier* pada data pengamatan.*Outlier*adalahkasus atau data yang memiliki karakteristik unik yang terlihat sangat berbeda jauh dari observasi-observasi lainnya dan muncul dalam bentuk nilai ekstrim, baik untuk sebuah variabel (Ghozali,2005). Oleh karena itu diperlukan metode lain untuk menangani outlier yaitu Metode Regresi *Robust* (MRR).

MRR adalah metode yang digunakan dalam mengatasi pencilan tanpa menghapus data pencilan tersebut(Mashitah, dkk, 2013).Terdapat berbagai macam MRR diantaranya estimasi M, metode S, estimasi MM (*Method of Moment*), LTS (*Least Terimmed Square*) dan LMS (*Least Median Square*).

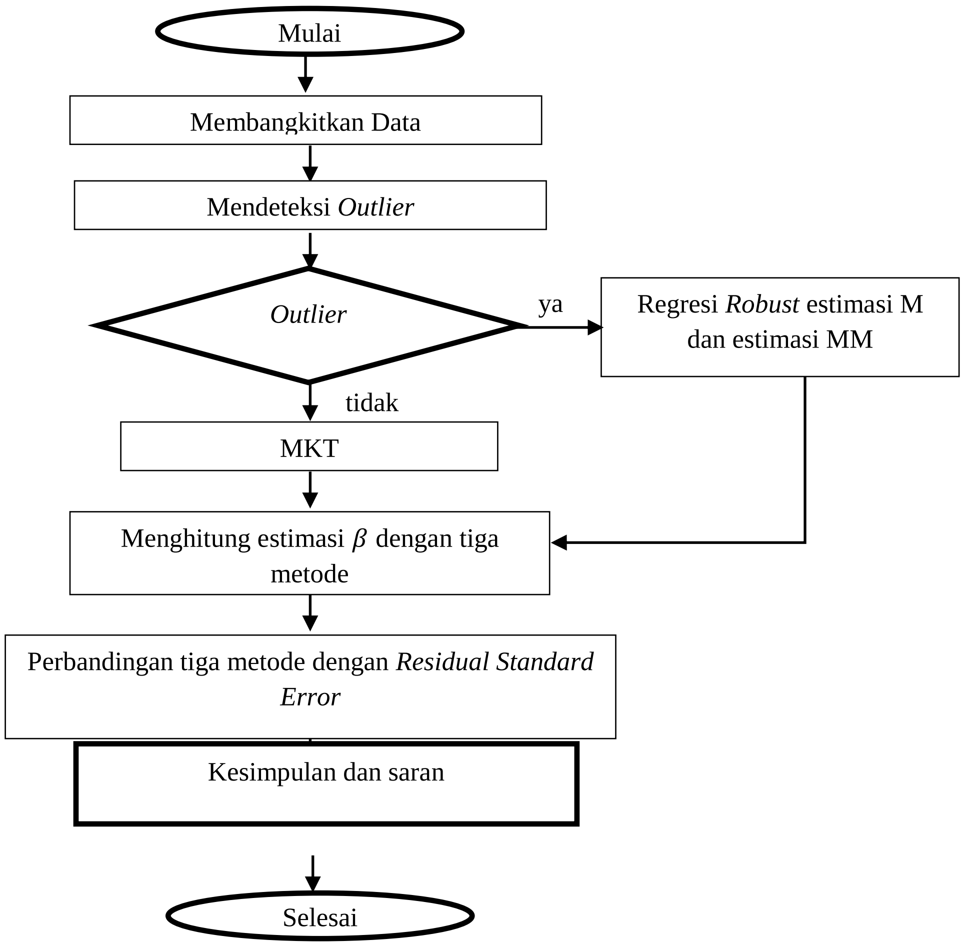
Estimasi M merupakan MRR yang baik untuk menduga parameter yang disebabkan oleh *youtlier* dan memiliki *breakdownpoint* 1/n(Bekti, 2011). Estimasi M dilakukan dengan cara memberi bobot pada e_i kemudian perhitungan nilai parameter dilakukan dengan WLS(*Weight Least Square*). Selain itu estimasi MM berusaha untuk mempertahankan sifat *robust*dan *resistant* dari estimasi S. Selain itu, metode ini juga mempertahankan sifat efisien dari estimasi M (Susanti, dkk, 2013).

Penelitian-penelitian yang pernah dilakukan mengenai metode estimasi M dan estimasi MM pada regresi *robust*, antara lain Safitri (2015) mengenai "Perbandingan Metode Estimasi M Dan Estimasi MM (*Method Of Moment*) Pada Regresi *Robust*". Selain itu ada pula penelitian mengenai "Analisis Regresi Pada Data *Outlier* dengan Menggunakan *Least Trimmed Square* (LTS) dan MM-Estimasi" yang dilakukan oleh Nurcahyadi (2010).

Oleh karena itu, dilakukan penelitian untuk membandingkan metode estimasi M dan estimasi MM pada Regresi *Robust*.Tujuan lain dari penilitian ini adalah untuk mengetahui tingkat keakuratan metode-metode tersebut dalam mengestimasi data yang mengandung *outlier*.Selain itu, dilakukan simulasi dengan menggunakan *software* R 2.14.2.

2. METODE PENELITIAN

Data yang digunakan adalah simulasi, metode penelitian diuraikan dengan diagram alur sebagai berikut ini:



Gambar 2.1Alur Tahapan Analisis Data

3. HASIL PENELITIAN DAN PEMBAHASAN

Regresi *robust* merupakan metode yang digunakan untuk mengatasi *outlier* tanpa menghapus data *outlier* tersebut. Pada regresi *robust*, menurut Rousseeuw(1987) metode yang sering digunakan dalam mengestimasi parameter diantaranya adalah *Huber* estimasi M (*Maximum Likelihood* Type) dan estimasi MM (*Method of Moment*)

a. Estimasi M

Metode ini mengasumsikan bahwa sebagian besar yang terdeteksi *outlier* berada pada variabel independen. Estimasi M meminimumkan fungsi objektif (ρ) dari fungsi residualnya (e_i). Fungsi tersebut dapat dilihat dalam persamaan

$$\min_{\hat{\beta}} \sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = \min_{\hat{\beta}} \sum_{i=1}^{n} \rho\left(\frac{y_i - x_i'\hat{\beta}}{s}\right) \tag{1}$$

Dengan $s = \frac{median|e_i - median(e_i)|}{0{,}6745}$ merupakan skala dari suatu estimasi *robust* dan $\rho\left(\frac{e_i}{s}\right)$ merupakan fungsi yang memberikan kontribusi pada masing-masing residual pada fungsi objektif.

Berdasarkan persamaan (1), dapat diperoleh estimasi parameter dari persamaan regresi.Sehingga diperoleh hasil sebagai berikut.

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{y_i - x_i'\hat{\beta}}{s}\right) = 0; j = 0,1,\dots,m \tag{2}$$

Dengan ψmerupakan fungsi *influence* yang digunakan dalam memperoleh bobot, x_{ij}adalah observasi ke-*i* pada respon ke-*j* dan x_{i0}=1.

Estimasiparameter dengan metode M disebut juga dengan *Iteratively Reweighted Least* Squares *(IRLS)*. Penyelesaian metode *IRLS* pada persamaan (2) menghasilkan persamaan (3) berikut ini.

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{y_i - x_i'\hat{\beta}}{s}\right) = \sum_{i=1}^{n} x_{ij}\frac{\psi\left(\frac{y_i - x_i'\hat{\beta}}{s}\right)}{\left(\frac{y_i - x_i'\hat{\beta}}{s}\right)}\left(\frac{y_i - x_i'\hat{\beta}}{s}\right) = 0, j = 0,1,\dots,k \tag{3}$$

Jika persamaan tersebut dinotasikan dengan matrik dimana W_0merupakan diagonal matriks "*weight*" yang berukuran n x n. Regresiterboboti tersebut dapat digunakan sebagai alat untuk mendapatkan estimasi M.Sehingga estimasi parameter menjadi:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X'W_0X})^{-1}\boldsymbol{X'W_0y} \tag{4}$$

Prosedur estimasi dengan menggunakan estimasi M atau *IRLS*diuraikan sebagai berikut :

- Dihitung estimasi dariβ,dinotasikan $\hat{\beta}$ menggunakan MKT, sehingga didapatkan $\hat{y}_{i0}$dan $\varepsilon_{i0} = y_i - \hat{y}_{i0}$, *(i = 1, 2, ... n)* yang diperlakukan sebagai nilai awal ($y_i$adalah hasil observasi).

- Dari nilai-nilai residual ini dihitung $\hat{\sigma}_0$ dan pembobot awal $w_{i0} = \frac{\psi(\varepsilon_{i0}^*)}{(\varepsilon_{i0}^*)}$. Nilai $\psi(\varepsilon_{i0}^*)$ dihitung sesuai fungsi Huber, dan $\varepsilon_{i0}^* = \frac{\varepsilon_{i0}}{\hat{\sigma}_0}$.
- Disusun matrik pembobot berupa matrik diagonal dengan elemen $w_{10}, w_{20}, \dots, w_{n0}$.
- Dihitungestimasi koefisien regresi,$\hat{\beta}_{Robust\ ke\ 1} = (\boldsymbol{X}'\boldsymbol{W_0}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W_0}\boldsymbol{y}$
- Dengan menggunakan nilai $\hat{\beta}_{Robust\ ke\ 1}$ dihitung pula $\sum_{i=1}^{n}|y_i - \hat{y}_{i1}|$atau $\sum_{i=1}^{n}|\varepsilon_{i1}|$.
- Selanjutnya langkah 2 sampai dengan 5 diulang sampai didapatkan $\sum_{i=1}^{n}|\varepsilon_{im}|$konvergen. Dengan kata lain, $\sum_{i=1}^{n}|\varepsilon_{im}|$cukup kecil untuk *j=0,1,...,k*.
- Jika diambil nilai terstandarisasi dari *e*, maka berdasarkan simulasi yang dilakukan oleh Huber , dipilih nilai *k=1.345*, sehingga diperoleh persamaan $w_i$sebagai berikut.

$$w_i = \begin{cases} 1, & jika\ |e_i| \le 1.345 \\ \frac{1.345}{|e_i|}, & jika\ |e_i| > 1.345 \end{cases} \tag{5}$$

b. Estimasi MM

Pada umumnya digunakan fungsi *Tukey Bisquareβ* baik pada estimasi S maupun estimasi M. Persamaan dari estimasi MM adalah sebagai berikut:

$$\tilde{\beta}_{MM} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) \tag{6}$$

Langkah pertama dalam estimasi ini adalah mencari nilai estimasi S. Bentuk persamaan estimasi S dapat dilihat seperti berikut ini.

$$\tilde{\beta}_S = \underset{\beta}{\operatorname{argmin}}\, s(e_1, e_2, \dots, e_k) \tag{7}$$

Dengan *s* adalah estimasi skala *robust* yang memenuhi $\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = K$. *K*merupakan konstan yang didefinisikan sebagai $K = E[\Phi(\rho)]$. Φ adalah distribusi normal standar. Nilai *breakdown* dari estimasi S dapat ditulis

$$\frac{K}{max\,\rho(e)} = 0{,}5. \tag{8}$$

Setelah diketahui nilai estimasi S, langkah selanjutnya adalah menetapkan parameter-parameter regresi menggunakan estimasi M. Berikut ini merupakan langkah-langkah estimasi parameter pada model linier berganda dengan regresi *robust* estimasi MM:

- Menghitung nilai estimasi awal koefisien $\hat{\beta}_j$ dan residual e_i dari regresi *robust* dengan *high breakdown point* (estimasi S) dengan pembobotan *tukey bisquare*.
- Residual e_i pada langkah pertama dilakukan untuk menghitung skala estimasi s, dan dihitung pula pembobot awal w_i.

- Residual e_i dengan skala estimasi s pada langkah kedua digunakan dalam iterasi awal sebagai estimasi *WLS* (*Weight Least Square*) untuk menghitung koefisien regresi.
- Menghitung bobot baru w_i dengan skala estimasi dari iterasi awal *WLS*. Perhitungan estimasi koefisien regresi menggunakan metode ini menggunakan persamaan $\widehat{\boldsymbol{\beta}}^{(\boldsymbol{m+1})} = (\boldsymbol{X'W^mX})^{-\boldsymbol{1}}\boldsymbol{X'W^my}$.
- Mengulang langkah 2 sampai 4 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).Jika diambil nilai terstandarisasi dari *e*, maka berdasarkan simulasi yang dilakukan oleh Tukey, dipilih nilai *k=4.685*, sehingga diperoleh persamaan $w_i$sebagai berikut.

$$w_i = \begin{cases} \left[1 - \left(\frac{e_i}{4.685}\right)^2\right]^2, jika\ |e_i| \le 4.685 \\ 0, \qquad\qquad\qquad jika\ |e_i| > 4.685 \end{cases} \tag{3.9}$$

c. Studi Kasus

Berdasarkan data hasil pembangkitan menggunakan *software*R 2.14.2 sebanyak 50 data, akan dilihat model yang terbentuk dari beberapa metode serta keakuratan masing-masing metode tersebut. Ada dua analisis yang dilakukan pada data yang dibangkitkan, yaitu analisis data yang mengandung *outlier* dan analisis data tanpa *outlier* (menghilangkan data *outlier*). Perhitungan juga dilakukan dengan *software* R. Berikut ini pembahasannya:

1) Data *Outlier*

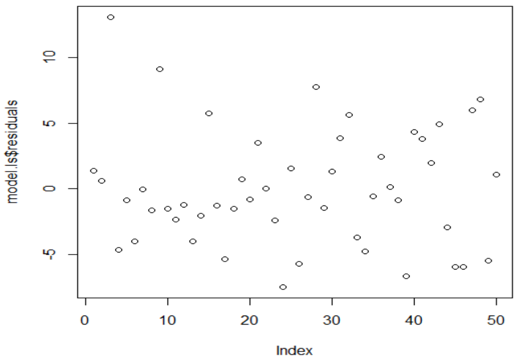
Pada kasus ini, nilai *cutoff* untuk masing-masing metode ditentukan berdasarkan jumlah sampel (*n*) yaitu 50 dan banyaknya parameter (*p*).Disini akan dilihat outlier pada arah *y* dimana dengan nilai dari$|R - student| > 2$ maka teridentifikasi adanya outlier.

Berdasarkan kriteria tersebut, dapat diketahui observasi-observasi yang merupakan *outlier*.Pada hasil yang diperoleh dapat dilihat bahwa terdapat data observasi yang merupakan data *outlier* pada arah *y* dengan melihat hasil perhitungan dari nilai *r-student*.Data yang *outlier* pada arah *y* adalah data observasi ke-3 dengan nilai 3.27 dan data observasi ke-8 dengan nilai 2.12.

Hasil model persamaan regresi yang terbentuk dari estimasi nilai parameter metode MKT adalah:

$$\hat{y} = 13.7722 + 4.211.9x_1 + 2.0432x_2 + \varepsilon \tag{10}$$

Estimasi pada metode ini menghasilkan *residualstandarderror*sebesar 4.514.Dilihat pada plot residual yang dihasilkan mayoritas residual terletak disekitar 0. Namun, ada beberapa residual yang terletak jauh dari 0.Hal ini mengindaksikan bahwa data tidak memenuhi asumsi kenormalan karena terdapat *outlier* dalam data. Berikut ini merupakan plot residual metode MKT.



Gambar1. Plot Residual Metode MKT (Data *Outlier*)

Pada gambarteridentifikasi adanya *outlier* sehinggaterdapat metode lain yang dapat digunakan menangani *outlier* yaitu regresi *robust*. Regresi *robust* pertama didapatkan hasil model persamaan regresi yang terbentuk dari estimasi nilai parameter metode estimasi Madalah:

$$\hat{y} = 14.1208 + 4.1049x_1 + 1.9016x_2 + \varepsilon \quad (11)$$

Estimasi pada metode ini menghasilkan *residualstandarderror*sebesar 4.101.Regresi *robust* kedua didapatkan hasilmodel persamaan regresi yang terbentuk dari estimasi nilai parameter metode estimasi MMadalah:

$$\hat{y} = 12.9063 + 4.2523x_1 + 1.8939x_2 + \varepsilon \quad (12)$$

Estimasi pada metode ini menghasilkan *residualstandarderror*sebesar 4.345.Berdasarkan hasil estimasi parameter menggunakan tiga metode tersebut diperoleh *residualstandard error* secara keseluruhan seperti berikut:

Tabel 1.*Residual Standard Error* Pada Tiga Metode Pada Data *Outlier*

Metode	*Residual Standard Error*
MKT	4.515
Estimasi M	4.101
Estimasi MM	4.345

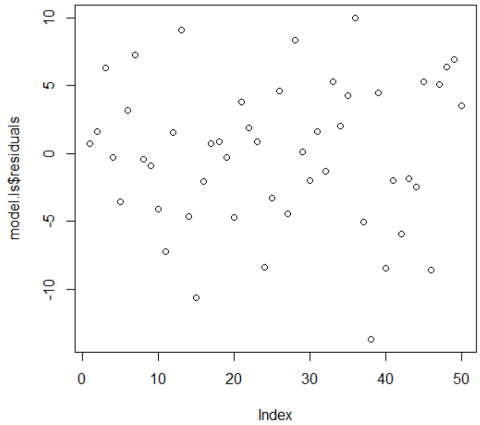
Pada tabel 1.dapat dilihat bahwa metode estimasi M memiliki *residualstandard error* paling kecil diantara metode yang lain yaitu sebesar 4.101. Oleh sebab itu, metode estimasi M lebih baik digunakan untuk mengatasi *outlier* dari pada dua metode yang lain.

2) Data Tanpa *Outlier*

Pada data tanpa outlier dilihat pada arah *y* tidak teridentifikasi adanya *outlier*. Kemudian dalam hasil perhitungan yang dilakukan dengan MKT didapatkan model persamaan regresi yang terbentuk dari estimasi nilai parameter metode MKT adalah:

$$\hat{y} = 15.15432 + 14.02770x_1 + 1.98165x_2 + \varepsilon \quad (13)$$

Estimasi pada metode ini menghasilkan *residualstandarderror*sebesar 5.431.Plot residual pada metode MKT dapat dilihat bahwa semua residual terletak disekitar 0.Residual menyebar dan terletak disekitar 0.Hal ini mengartikan bahwa data memenuhi asumsi. Berikut ini merupakan plot residual metode MKT.



Gambar 2. Plot Residual Metode MKT (Data tanpa *Outlier*)

Namun, untuk mengetahui secara pasti dilakukan perbandingan dengan metode estimasi M dan metode estimasi MM. Berikut ini merupakan hasil estimasi menggunakan metode estimasi M.

Model persamaan regresi yang terbentuk dari estimasi nilai parameter metode estimasi M adalah:

$$\hat{y} = 15.1184 + 14.0355x_1 + 1.9987x_2 + \varepsilon \qquad (14)$$

Estimasi pada metode ini menghasilkan *residualstandarderror*sebesar 5.839.Pada estimasi MM, model model persamaan regresi yang terbentuk dari estimasi nilai parameter adalah:

$$\hat{y} = 15.1563 + 14.0355x_1 + 1.9923x_2 + \varepsilon \qquad (15)$$

Estimasi pada metode ini menghasilkan *residualstandarderror*sebesar 5.732.Berdasarkan hasil estimasi parameter menggunakan tiga metode tersebut diperoleh *residualstandard error* secara keseluruhan seperti berikut.

Tabel 2.*Residual Standard Error* Pada Tiga Metode Pada Data Tanpa *Outlier*

Metode	*Residual Standard Error*
MKT	5.431
Estimasi M	5.839
Estimasi MM	5.732

Pada tabel 2.dapat dilihat bahwa MKT memiliki *residualstandard error* paling kecil diantara metode yang lain yaitu sebesar 5.431. Oleh sebab itu, MKT lebih baik digunakan untuk data tanpa mengandung *outlier* dari pada dua metode yang lain.

4. SIMPULAN

Pada kasus data mengandung *outlier* pada arah *y*, metode yang paling baik digunakan adalah metode estimasi M pada regresi *robust* dibandingkan dengan MKT dengan model regresi yag didapatkan adalah

$$\hat{y} = 14.1208 + 4.1049\mathrm{x}_1 + 1.9016\mathrm{x}_2 + \varepsilon$$

Metode estimasi MM. Pada kasus data tanpa *outlier*¸ metode yang paling baik digunakan adalah MKT dengan model yang didapatkan yaitu:

$$\hat{y} = 15.15432 + 14.02770x_1 + 1.98165x_2 + \varepsilon$$

Perbandingan dilakukan dengan melihat nilai *residual standard error*. Jika nilai *residual standard error*paling kecil maka model tersebutlah yang layak digunakan.

5. DAFTAR PUSTAKA

Bekti. (2011). *Materi Statistik.* Diakses dari https://statisticsanalyst.files.wordpress.com/2011/10/11.doc.

Ghozali, I. (2005). *Analisis Multivariate dengan Program SPSS Ed 3.* Semarang, S: Badan Penerbit Universitas Diponegoro.

Mashitah., Arif Wibowo., & Diah Indriani. (2013). Metode Robust Regression on Ordered Statistics (ROS) pada Data Tersensor Kiri dengan Outlier. *Jurnal Biometrika dan Kependudukan,* **II**(2), 148–157.

Nurcahyadi, H. (2010). *Analisis Regresi pada Data Outlier dengan Menggunakan Least Trimmed Square (LTS) dan MM-Estimasi.* Jakarta, J: Universitas Islam Negeri Syarif Hidayatullah.

Safitri, A.D. (2015). *Perbandingan Metode Estimasi M Dan Estimasi MM (Method Of Moment) Pada Regresi Robust*. Yogyakarta, Y: Universitas Islam Indonesia.

Susanti, Y., Pratiwi, H., &Sulistiowati, S. (2013). M Estimation, S Estimation, And Mm Estimation In Robust Regression. *International Journal of Pure and Applied Mathematics*, **91**(3), 349-360, doi: http://dx.doi.org/10.12732/ijpam.v91i3.7.

Rousseeuw, P. J. (1987). *Robust Regression and Outlier Detection.* New York, NY: Wiley and Sons.