

Analisis Regresi Robust Estimasi M Terhadap Kecelakaan Lalu Lintas Di Provinsi Sumatera Utara

Anazta Widya Ananda Br. Siagian¹, T.J. Marpaung²

¹Mahasiswa Program Studi Statistika Fakultas Vokasi Universitas Sumatera Utara,

²Program Studi Statistika Fakultas Vokasi Universitas Sumatera Utara

e-mail : anaztasiagian@gmail.com, tj.marpaung@gmail.com

Abstrak

Analisis regresi merupakan analisis statistik yang bertujuan untuk memodelkan hubungan antara variabel terikat dengan variabel bebas. Model regresi yang baik memerlukan data yang baik pula. Suatu data dikatakan baik apabila data tersebut berada di sekitar garis regresi. Kenyataannya, terkadang terdapat data yang terletak jauh dari garis regresi atau pola data keseluruhan. Data tersebut dikenal dengan istilah pencilan atau outlier. Adanya outlier dapat mengakibatkan estimasi koefisien regresi yang diperoleh tidak tepat. Diperlukan metode lain yang dapat mengestimasi masalah outlier, yaitu regresi robust. Salah satu metode estimasi dari regresi robust adalah estimasi-M. Penelitian ini bertujuan untuk mampu menerapkan regresi Robust menggunakan metode Estimasi-M pada data kecelakaan lalu lintas di Provinsi Sumatera Utara pada tahun 2018-2022. Dengan menggunakan regresi Robust metode Estimasi-M, diperoleh model persamaan regresi Robust Estimasi-M adalah $\hat{Y} = -12.2760 - 0.0472X_1 + 0.1121X_2 + 0.3342X_3 + 0.7703X_4$.

Kata kunci : Estimasi-M, Metode Kuadrat Terkecil, Pencilan, Regresi Robust

Abstract

Regression analysis is a statistical analysis that aims to model the relationship between the dependent variable and the independent variable. A good regression model requires good data too. A data is said to be good if the data is around the regression line. In fact, sometimes there are data that lie far from the regression line or the overall data pattern. These data are known as outliers or outliers. The presence of outliers can result in an inappropriate estimate of the regression coefficient obtained. We need another method that can estimate the outlier problem, namely robust regression. One of the estimation methods of robust regression is M-estimation. This study aims to be able to apply Robust regression using the M-Estimation method to traffic accident data in North Sumatra Province in 2018-2022. By using the M-Estimated Robust regression method, the M-Estimated Robust regression equation model is $\hat{Y} = -12.2760 - 0.0472X_1 + 0.1121X_2 + 0.3342X_3 + 0.7703X_4$.

Keywords : M-estimation, Least Squares Method, Outlier, Strong Regression.

1. PENDAHULUAN

Istilah regresi pertama kali diperkenalkan oleh Sir Francis Galton pada tahun 1877 dalam penelitian biogenisisnya. Regresi berguna dalam menelaah hubungan sepasang variabel atau lebih. Salah satu cara untuk mencari estimasi koefisien parameter dengan menggunakan Metode Kuadrat Terkecil atau Ordinary Least Square (OLS).

Analisis regresi merupakan analisis statistik yang bertujuan untuk memodelkan hubungan antara variabel terikat dengan variabel bebas. Model regresi yang baik memerlukan data yang baik pula. Suatu data dikatakan baik apabila data tersebut berada di sekitar garis regresi. Kenyataannya, terkadang terdapat data yang terletak jauh dari garis regresi atau pola data keseluruhan. Data tersebut dikenal dengan istilah pencilan atau outlier. Pencilan merupakan pengamatan yang

jauh dari kelompok data yang mungkin berpengaruh besar terhadap koefisien regresi (Soemartini, 2007).

Data yang mengandung pencilan dapat dianalisis dengan menggunakan regresi Robust. Menurut Olive menyatakan bahwa, "Regresi Robust adalah suatu metode regresi yang digunakan pada saat sisaan data tidak berdistribusi normal karena disebabkan oleh adanya beberapa pencilan yang mempengaruhi model". Salah satu metode estimasi dalam regresi Robust yaitu Estimasi-M. Metode ini diterapkan pada data Kecelakaan Lalu Lintas Di Provinsi Sumatera Utara pada tahun 2018-2022.

Dari data yang didapat, diketahui terdapat beberapa pencilan dalam data. Menurut Sembiring (2003), adanya pencilan dalam data dapat mengakibatkan estimasi koefisien regresi yang diperoleh tidak tepat. Penghapusan begitu saja terhadap data pencilan bukanlah langkah yang bijaksana. Terkadang pencilan memberikan informasi yang tidak bisa diberikan oleh titik data lainnya.

Dari latar belakang diatas, penulis tertarik membahas analisis regresi Robust menggunakan data kecelakaan lalu lintas dan pengambilan data di Badan Pusat Statistik Provinsi Sumatera Utara. Maka penulis membuat judul tugas akhir dengan judul "**Analisis Regresi Robust Estimasi M Terhadap Kecelakaan Lalu Lintas Di Provinsi Sumatera Utara**".

2. METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan adalah data sekunder yang diperoleh dengan cara mengambil data di Badan Pusat Statistik Provinsi Sumatera Utara. Pengambilan data sekunder merupakan data yang diperoleh peneliti secara tidak langsung melalui perantara. Penulis memperoleh data dengan cara studi kepustakaan yang bersifat teoritis yang mendukung serta relevan.

2.2 Lokasi Penelitian

Dalam melakukan penulisan tugas akhir ini, penulis mengambil data sekunder yaitu data yang diperoleh dari Badan Pusat Statistik Provinsi Sumatera Utara (BPS Sumut), Jl. Asrama No. 179, Dwi Kora, Medan Helvetia, Kota Medan. Sampel yang diambil

dalam penelitian ini adalah yang berada di Provinsi Sumatera Utara pada tahun 2018-2022.

2.3 Metode Pengumpulan Data

Untuk memperoleh data yang dibutuhkan, penulis melakukan riset dengan mengambil data di kantor Badan Pusat Statistik Provinsi Sumatera Utara. Data yang diperoleh kemudian disajikan dalam bentuk angka-angka agar gambaran yang jelas dari sekumpulan data yang diperoleh dapat diambil yang kemudian dapat ditarik kesimpulannya.

2.4 Pengolahan Data

Mengolah data yang didapat dari tahap pengumpulan untuk menyelesaikan permasalahan yang terjadi dianalisis dengan menggunakan metode Analisis Regresi Robust Estimasi M.

2.5 Langkah-Langkah Pengolahan Data Dengan Menggunakan Metode Analisis Regresi Robust Estimasi M

Langkah-langkah untuk pengolahan data dengan menggunakan metode analisis Regresi Robust adalah sebagai berikut :

1. Menyiapkan data sekunder untuk penelitian
2. Melakukan uji asumsi klasik
3. Mencari koefisien regresi dengan metode OLS
4. Menguji signifikansi parameter OLS
5. Menguji adanya data pencilan dengan metode DFFITS
6. Menganalisis model regresi robust dengan estimasi M.

Algoritma perhitungan Estimasi M :

- a. Menghitung parameter awal $\beta^0, \beta^1, \beta^2, \beta^3, \beta^4, \beta^5$
- b. Menghitung nilai sisaan $e_i = y_i - \hat{y}_i$
- c. Menghitung nilai skala regresi robust estimasi M :

$$\hat{\sigma}_M = \frac{MAD}{0,6745} = \frac{\text{median} |e_i - \text{median}(e_i)|}{0,6745}$$

- d. Menghitung nilai $u_i = \frac{e_i}{\hat{\sigma}_M}$
- e. Menghitung nilai pembobotan

$$w_i = \begin{cases} \left[1 - \left(\frac{|u_i|}{4,685} \right)^2 \right]^2 & , |u_i| \leq 4,685 \\ 0 & \text{lainnya} \end{cases}$$

- $|u_i|$
- $> 4,685$
- Menghitung parameter $\hat{\beta}_M$ dengan metode WLS dengan pembobot W_M .
 - Mengulangi Langkah b – f sampai diperoleh nilai $\hat{\beta}_M$ yang konvergen.

3. HASIL DAN PEMBAHASAN

3.1 Penyajian Data

Data yang digunakan dalam tugas akhir ini adalah data sekunder dari laporan Badan Pusat Statistik (BPS) yang terdiri dari 140 observasi yaitu data Jumlah Kecelakaan Lalu Lintas di Provinsi Sumatera Utara Menurut Kabupaten/Kota pada tahun 2018-2022.

3.2 Analisis Data

Berikut merupakan Langkah – Langkah analisis Regresi Robust menggunakan Rstudio :

- Import Data csv ke Rstudio
 - Untuk melihat data, masukkan function : View(Data_TA)
 - Statistika Deskriptif
- Pada hasil summary data pada program Rstudio, terdapat analisis deskriptif di setiap variable sebagai berikut :

```
#Ringkasan Data#
> summary(Data_TA)
  X1      X2      X3      X4      Y
Min. : 0.00  Min. : 0.00  Min. : 8.0  Min. : 9.0  Min. : 12.00
1st Qu.: 48.75 1st Qu.: 72.25 1st Qu.: 67.0 1st Qu.: 54.0 1st Qu.: 56.25
Median : 180.50 Median : 194.00 Median : 131.5 Median : 135.5 Median : 162.00
Mean : 213.93  Mean : 232.86  Mean : 217.2  Mean : 200.6  Mean : 230.89
3rd Qu.: 291.00 3rd Qu.: 293.75 3rd Qu.: 282.2 3rd Qu.: 257.0 3rd Qu.: 282.50
Max. : 1212.00  Max. : 1334.00  Max. : 1585.0  Max. : 1345.0  Max. : 1665.00
```

Gambar 3.1 Statistika Deskriptif

- Estimasi model menggunakan Regresi OLS, menggunakan function :


```
>fit.ols <- lm(Y ~ X1+X2+X3+X4, data = Data_TA)
>summary(fit.ols)
```

```
> #Estimasi model menggunakan regresi OLS#
> fit.ols <- lm(Y ~ X1+X2+X3+X4, data = Data_TA)
> summary(fit.ols)

Call:
lm(formula = Y ~ X1 + X2 + X3 + X4, data = Data_TA)

Residuals:
    Min       1Q   Median       3Q      Max
-39.319 -14.368   1.766  10.822  54.791

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -12.7908    7.0229  -1.821  0.0816 .
X1           -0.1290    0.1437  -0.898  0.3785
X2            0.1849    0.1358   1.362  0.1865
X3            0.2449    0.1539   1.591  0.1253
X4            0.8727    0.1677   5.203 2.82e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.36 on 23 degrees of freedom
Multiple R-squared:  0.9947, Adjusted R-squared:  0.9938
F-statistic: 1084 on 4 and 23 DF, p-value: < 2.2e-16
```

Gambar 3.2 Hasil Summary Regresi OLS

$$\hat{Y} = -12.7908 - 0.1290X1 + 0.1849X2 + 0.2449X3 + 0.8727X4$$

Dengan nilai Multiple R-squared sebesar 0,9947 maka ini memiliki makna bahwa variabel-variabel bebas (X1, X2, X3, X4) mampu menjelaskan variasinya terhadap variabel terikat (Y) sebesar 99,47% dan sisanya sebesar 0,53% dijelaskan atau dipengaruhi oleh variabel-variabel lainnya yang tidak dimasukkan ke dalam model persamaan.

Kemudian dengan nilai p-value sebesar 2,2e-16 yang lebih kecil dari alpa 0,05 untuk F-statistic sehingga kita harus tolak H0, kondisi ini menunjukkan bahwa seluruh variabel bebas (X1, X2, X3, X4) secara bersama-sama mampu mempengaruhi variabel terikat (Y) pada tingkat alpha 5%.

5. Uji Asumsi Klasik

```
> ##Uji Asumsi Klasik##
> #Uji Normalitas Residual
> shapiro.test(fit.ols$residuals)
```

Shapiro-Wilk normality test

```
data: fit.ols$residuals
W = 0.97665, p-value = 0.764
```

Gambar 3.3 Uji Normalitas Residual

Hipotesis :

H0 : Residual berdistribusi normal

H1 : Residual tidak berdistribusi normal

Tingkat signifikansi : $\alpha = 0,05$

Berdasarkan output diatas dari Rstudio dapat dilihat bahwa p-value = 0,764. Maka p-value $> \alpha$, H0 diterima. Artinya residual berdistribusi normal.

a. Uji Homoskedastisitas

```
> bptest(fit.ols, studentize = FALSE, data = Dat)
```

```
Breusch-Pagan test
```

```
data: fit.ols
BP = 1.898, df = 4, p-value = 0.7545
```

Gambar 3.4 Uji Homoskedastisitas

Hipotesis :

H0 : Asumsi Kehomogenan variansi residual terpenuhi

H1 : Asumsi Kehomogenan variansi residual tidak terpenuhi

Tingkat signifikansi : $\alpha = 0,05$

Berdasarkan output diatas dari Rstudio dapat dilihat bahwa p-value = 0,7545. Maka p-value > α , H0 diterima. Artinya asumsi Kehomogenan variansi residual terpenuhi

b. Uji Auto Korelasi

```
> #Uji Autokorelasi Residual
> library(lmtest)
> dwtest(fit.ols)
```

```
Durbin-watson test
```

```
data: fit.ols
DW = 2.3454, p-value = 0.8351
alternative hypothesis: true autocorrelation is greater
```

Gambar 3.5 Uji Autokorelasi

Hipotesis :

H0 : Tidak terdapat autokorelasi pada residual

H1 : Terdapat autokorelasi pada residual

Tingkat signifikansi : $\alpha = 0,05$

Berdasarkan output diatas dari Rstudio dapat dilihat bahwa p-value = 0,8351. Maka p-value > α , H0 diterima. Artinya tidak terdapat autokorelasi pada residual

c. Uji Multikolinieritas

```
> vif(fit.ols)
```

```
      x1      x2      x3
52.98173 55.42460 93.05473 82.138
```

Gambar 3.6 Uji Multikolinieritas

Hipotesis :

H0 : Asumsi multikolinieritas terpenuhi

H1 : Asumsi multikolinieritas tidak terpenuhi

Tingkat signifikansi : $\alpha = 0,05$

Daerah Kritis : H0 ditolak jika = VIF > 10

Berdasarkan output diatas dari Rstudio dapat dilihat nilai VIF pada masing – masing variable > 10.

Maka H0 ditolak, Artinya asumsi multikolinieritas tidak terpenuhi.

Dari hasil uji asumsi Klasik :

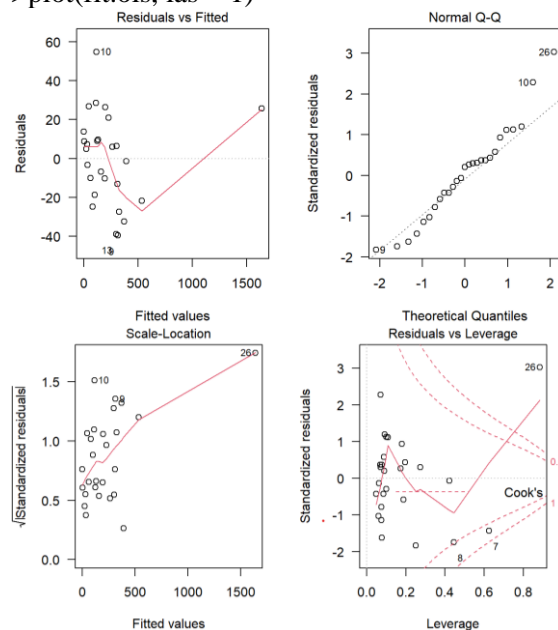
Asumsi normalitas residual = Terpenuhi
 Uji Homoskedastisitas = Terpenuhi
 Asumsi autokorelasi = Terpenuhi
 Uji Multikolinieritas = Tidak Terpenuhi

Tidak terpenuhinya suatu asumsi, banyak faktor yang mempengaruhinya, diantaranya adalah adanya kemungkinan bahwa jumlah observasi belum mewakili populasi, variable belum mewakili model, atau adanya outlier.

6. Pendeteksi Outlier pada data menggunakan Plot, dengan function :

```
> par(mfrow = c(1,2))
```

```
> plot(fit.ols, las = 1)
```



Gambar 3.7 Pendeteksi Outlier OLS Menggunakan Plot

Pada grafik Residual vs Fitted di pojok kiri atas menunjukkan bahwa data tersebar membentuk suatu pola tertentu. Atau bisa dibilang tidak terdapat perbedaan varian residual. Sehingga disimpulkan bahwa model memenuhi asumsi homoskedastisitas.

Pada grafik normal Q-Q yang berada di pojok kanan atas, menunjukkan bahwa titik-titik berada di sekitar garis lurus, sehingga dapat disimpulkan bahwa

residual tersebar normal. Model regresi memenuhi asumsi normalitas.

Pada grafik Scale-Location dilihat bahwa variabilitas (varians) dari titik residual meningkat dengan nilai variable hasil yang dipasang, menunjukkan varians yang tidak konstan dalam kesalahan residual (Heteroskedastisitas).

Pada grafik Residuals vs Leverage terdapat tiga titik data berada diluar garis jarak Cook's. Plot mengidentifikasi pengamatan yang berpengaruh sebagai #7, #8 #9 dan #26.

7. Estimasi M, menggunakan function :

```
>library(MASS)
>library(sfsmisc)
>fit.M = rlm(Y ~ X1+X2+X3+X4, data
= Data_TA, method = "M")
>summary(fit.M)
> library(sfsmisc)
> fit.M = rlm(Y ~ X1+X2+X3+X4, data = Data_TA, method = "M")
> summary(fit.M)

Call: rlm(formula = Y ~ X1 + X2 + X3 + X4, data = Data_TA, metho
"M")
Residuals:
    Min       1Q   Median       3Q      Max
-47.3797 -12.2683  0.6114  11.0232  58.3035

Coefficients:
            Value      Std. Error t value
(Intercept) -12.7173    7.8119   -1.6279
X1           -0.0445    0.1598   -0.2785
X2            0.1135    0.1510    0.7514
X3            0.3350    0.1712    1.9563
X4            0.7665    0.1866    4.1083

Residual standard error: 18.37 on 23 degrees of freedom
> #Melihat Iterasi
> print(fit.M)
Call:
rlm(formula = Y ~ X1 + X2 + X3 + X4, data = Data_TA, method = "
converged in 11 iterations

Coefficients:
            x1            x2            x3            x4
-12.27604770  -0.04724885   0.11213671   0.33418969   0.7702954

Degrees of freedom: 27 total; 22 residual
(1 observation deleted due to missingness)
Scale estimate: 19.3
```

Gambar 3.8 Estimasi M

$$\hat{Y} = -12,7173 - 0,0445X1 + 0,1135X2 + 0,3350X3 + 0,7665X4$$

Dengan residual Standart Error = 18,37 dan jumlah iterasi sebanyak 11 iterasi.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan , maka diperoleh kesimpulan sebagai berikut :

1. Pada data jumlah kecelakaan lalu lintas di Provinsi Sumatera Utara pada Tahun 2018-2022 terdapat empat pencilan yaitu, data pada pengamatan ke-7, 8, 9 dan 26.
2. Model analisis regresi Robust dengan Estimasi-M pada data jumlah kecelakaan lalu lintas di Provinsi

Sumatera Utara pada Tahun 2018-2022 adalah :

$$\hat{Y} = -12,7173 - 0,0445X1 + 0,1135X2 + 0,3350X3 + 0,7665X4$$

Untuk konstan a = -12,7173. Artinya apabila X1, X2, X3, dan X4 (2018, 2019, 2020, 2021) sebesar 0, maka Y (2022) sebesar -12,7173.

Untuk $\beta_1 = -0,0445$. Nilai tersebut menunjukkan pengaruh negative (berlawanan arah) antara variable X1 dan Y. Artinya setiap peningkatan X1 sebesar 1 satuan, maka Y mengalami penurunan sebesar -0,0445.

Untuk $\beta_2 = 0,1135$. Nilai tersebut menunjukkan pengaruh positif (searah) antara variable X2 dan Y Artinya setiap peningkatan X2 sebesar 1 satuan akan meningkatkan Y sebesar 0,1135.

Untuk $\beta_3 = 0,3350$. Nilai tersebut menunjukkan pengaruh positif (searah) antara variable X3 dan Y Artinya setiap peningkatan X3 sebesar 1 satuan akan meningkatkan Y sebesar 0,3350.

Untuk $\beta_4 = 0,7665$. Nilai tersebut menunjukkan pengaruh positif (searah) antara variable X4 dan Y Artinya setiap peningkatan X4 sebesar 1 satuan akan meningkatkan Y sebesar 0,7665.

3. Banyaknya iterasi pada pengolahan data jumlah kecelakaan lalu lintas di Provinsi Sumatera Utara pada Tahun 2018-2022 adalah sebanyak 11 iterasi.

5. SARAN

Adapun saran dari penulis yang mungkin berguna bagi semua pihak adalah:

1. Penulis berharap untuk penelitian selanjutnya dapat menggunakan estimasi Robust yang lainnya seperti Estimasi S (*Robust Scale*), MM (*Method of Moment*), atau LMS (*Least Median Square*).
2. Penulis berharap untuk penelitian selanjutnya dapat melakukan perbandingan metode-metode yang bisa mengatasi outlier selain regresi robust seperti menghilangkan data outlier ataupun mentransformasi data.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan untuk semua pihak yang telah membantu sehingga penulis dapat menyelesaikan penelitian ini.

DAFTAR PUSTAKA

- [1] Annisa, Lucia. 2016. *Analisis Resiko Kecelakaan Lalu Lintas Berdasarkan Pengetahuan, Penggunaan Jalur, Dan Kecepatan Berkendara*. Jawa Timur : Universitas Airlangga
 - [2] Badan Pusat Statistik Provinsi Sumatera Utara. 2023. *Jumlah Kecelakaan Lalu Lintas dan Kerugian 2020-2022*. Medan : Badan Pusat Statistik.
 - [3] Chen, Colin. 2002. *Robust Regression and Outlier Detection with the Robustreg Procedure*. SAS Institute : Cary, NC
 - [4] Ghozali, I. 2018. *Aplikasi Analisis Multivariate IMB SPSS*. Jakarta : Universitas Diponegoro
 - [5] Olive, D.J. (2005). *Applied Robust Statistics*. Carbondale: Southern Illinois University.
 - [6] Peraturan Pemerintah Republik Indonesia No. 43 Tahun 1992 Tentang Lalu lintas dan Angkutan Jalan.
 - [7] Sembiring, R.K. 2003. *Analisis Regresi, Edisi Kedua*. Bandung : ITB
 - [8] Soermatini. 2007. *Pencilan (Outlier)*. Jatinangor: Universitas Padjajaran
 - [9] Undang-Undang Nomor 22 Tahun 2009 tentang Lalu Lintas dan Angkutan Jalan.
 - [10] Widarjono, Agus. 2007. *Ekonometrika Teori dan Aplikasi untuk Ekonomi dan Bisnis. Edisi Kedua*. Yogyakarta. Fakultas Ekonomi UII
 - [11] Yuliana Susanti. 2008. *Estimasi M dan Sifat-Sifatnya pada Regresi Linier Robust*, Jurnal Math-Info, Vol. 1, Surakarta : UNS
-