

---

## **ANALISIS REGRESI *ROBUST M ESTIMATOR* UNTUK MENGETAHUI FAKTOR YANG MEMPENGARUHI LAMA STUDI MAHASISWA S1 STATISTIKA FMIPA UNIVERSITAS LAMBUNG MANGKURAT**

**Widawati Annisa Putri <sup>1\*</sup>, Selvi Annisa <sup>2</sup>**

<sup>1,2</sup> Program Studi Statistika Fakultas MIPA Universitas Lambung Mangkurat

\*e-mail corresponding: [widawati.nett@gmail.com](mailto:widawati.nett@gmail.com)

---

### **Abstract**

*Robust regression is a statistical technique commonly used to model relationships between variables by minimizing the impact of outlier data. The use of robust regression M Estimator works well when there are outliers in the data. In this study, robust regression M estimator analysis will be applied to student study period data. The aim of this research is to determine the significant factors influencing the study period of Statistics undergraduate students at the Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University. The results of the research show that the residual data characteristics are not normal and there are outliers in the data. Using the Robust Regression M Estimator, the F test results show that F calculated  $6.2492 > F$  table  $2.173112$ , which means rejecting  $H_0$ , indicating that the independent variables collectively have a significant effect on the dependent variable. From the t-test, it is known that the Guidance Process for students while working on their final project, the Employment Status of students, and the GPA of students significantly affect the Study Period of students.*

**Keywords:** *Robust Regression M Estimator, Study Period of Students, ULM*

---

### **1. PENDAHULUAN**

Analisis regresi linier merupakan sebuah metode yang umum digunakan dalam statistika untuk memodelkan hubungan antara satu atau lebih variabel independen dengan variabel dependen. Kelemahan dari analisis regresi linier adalah estimasi kuadrat terkecil tidak kuat terhadap *outlier* yang berimbas pada hasil koefisien estimasi regresi, yang dapat memiliki efek besar pada koefisien estimasi, *standar errors*, dapat mengakibatkan garis regresi akan menjauh atau tidak fit terhadap pengamatan dan prediksi yang dihasilkan oleh model akan kurang akurat yang dapat mengakibatkan kesalahan prediksi. *Outlier* sendiri dapat diartikan sebagai data yang secara signifikan berbeda dari nilai-nilai lain dalam data set.

Regresi robust merupakan sebuah teknik statistika yang dirancang untuk mengatasi masalah ini dengan meminimalkan dampak dari adanya *outlier* pada data, sehingga memberikan estimasi parameter regresi yang lebih stabil dan konsisten. Salah satu metode regresi robust yang umum digunakan adalah M Estimator, yang bekerja dengan mengiterasi proses pencarian estimasi parameter yang paling sesuai dengan data, dengan meminimumkan suatu fungsi obyektif yang telah ditentukan. Dalam konteks regresi robust M Estimator, Iteratively Reweighted Least Squares (IRLS) sering digunakan untuk mencapai konvergensi dalam proses iterasi tersebut.

Pada penelitian ini akan menerapkan analisis regresi robust M Estimator pada data masa studi mahasiswa S1 Statistika Fakultas MIPA di Universitas Lambung Mangkurat

---

(ULM) untuk mengetahui faktor apa saja yang dapat mempengaruhi lama studi mahasiswa S1 Statistika FMIPA ULM.

## 2. TINJAUAN PUSTAKA

### 2.1. Lama Studi

Lama Studi merupakan waktu yang ditempuh mahasiswa untuk menyelesaikan Pendidikan terhitung dari awal masuk kuliah hingga dinyatakan lulus atau telah menyelesaikan masa studinya. Untuk jenjang S1 lama studi yang dirancang 8 semester (4 tahun). Berdasarkan permendikbud No. 49 (2014) tentang Standar Nasional Pendidikan Tinggi (SNPT) mengatakan, beban belajar minimal mahasiswa S-1/D-4 adalah 144 SKS (Satuan Kredit Semester) dan untuk menuntaskan seluruh beban SKS tersebut, mahasiswa diberi batas waktu 4-5 Tahun atau setara dengan 8-10 semester.

### 2.2. Instrumen Penelitian

Instrumen penelitian adalah pedoman tertulis tentang wawancara, Pengamatan, dan pertanyaan yang dipersiapkan untuk mendapatkan informasi atau data. Untuk melihat kebaikan instrumen dapat menggunakan uji validitas dan reabilitas.

### 2.3. Validitas

Validitas adalah karakteristik yang digunakan untuk memastikan bahwa suatu instrumen memberikan pengukuran yang akurat dan dapat diandalkan terhadap variabel yang diukur. Pada uji validitas setiap butir pertanyaan diukur dengan menjumlah keseluruhan respon dari pertanyaan yang digunakan pada setiap variabel. Instrumen dapat dikatakan baik apabila koefisien validitas  $\geq 0.50$ . Dengan pemahaman tersebut, untuk melakukan diagnosis terhadap besaran koefisien validitas, dapat menggunakan pedoman Interpretasi *Uncorrected Correlation Coefficients* dalam studi validitas prediktif, berikut :

Tabel 1. Pedoman Interpretasi *Uncorrected Correlation Coefficients*

Koefisien Validitas	Interpretasi
$> 0.50$	Sangat Berguna
$0.21 \text{ sd } 0.50$	Dapat Berguna
$0.11 \text{ sd } 0.20$	Tergantung Keadaan
$< 0.11$	Tidak Berguna

Uji validitas dapat menggunakan korelasi *product moment* dengan rumus :

$$r_{xy} = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}} \quad (1)$$

Dimana :

$n$  Jumlah responden

$X$  Skor total yang diperoleh dari seluruh item variabel  $x$

$Y$  Skor total yang diperoleh dari seluruh item variabel  $y$

## 2.4. Reliabilitas

Reliabilitas adalah istilah yang digunakan untuk mengukur sejauh mana kekonsistenan suatu hasil pengukuran apabila pengukuran diulang dua kali atau lebih. Uji reliabilitas umumnya digunakan untuk mengetahui apakah data yang akan digunakan dapat diandalkan atau akurat. Uji reliabilitas mengukur variabel dengan menggunakan data dari pertanyaan yang sama dalam uji validitas dan telah dinyatakan valid. Pada umumnya uji reliabilitas untuk data penelitian dan kuesioner menggunakan rumus *Cronbach Alpha* dengan statistik uji:

$$r_{11} = \left[ \frac{k}{(k-1)} \right] \left[ 1 - \frac{\sum \sigma_b^2}{(\sigma_t^2)} \right] \quad (2)$$

Dimana :

$r_{11}$  koefisien reliabilitas instrumen       $\sum \sigma_b^2$  jumlah varian butir

$k$  jumlah butir pertanyaan       $\sigma_t^2$  varian skor total

Adapun kategori koefisien reliabilitas guilford adalah sebagai berikut :

Tabel 2. Koefisien Reliabilitas Guilford

Koefisien Reliabilitas ( $r_{11}$ )	Tingkat Reliabilitas
$0.80 < r_{11} \leq 1.00$	Sangat tinggi
$0.60 < r_{11} \leq 0.80$	Tinggi
$0.40 < r_{11} \leq 0.60$	Sedang
$0.20 < r_{11} \leq 0.40$	Rendah
$-1.00 < r_{11} \leq 0.20$	Sangat Rendah

## 2.5 Pendeteksian *Outlier*

*Outlier* adalah data yang berbeda dari data yang ada di populasi. Kehadiran *outlier* memiliki dampak penting dalam analisis data. Adanya *outlier* dapat menyebabkan bentuk distribusi data menjadi tidak normal, serta dapat berpengaruh terhadap hasil pengujian parameter. Data *outlier* dapat dikenali dengan pemeriksaan secara visual dari data mentah atau dari *boxplot*, diagram pencar/*scatterplot*. Selain itu pendeteksian *outlier* pada data juga dapat dilakukan dengan Metode *DfFITS* (*Difference fitted value FITS*) atau *Standarized DfFITS*. *DfFITS* adalah ukuran pengaruh yang disebabkan oleh pengamatan ke- $i$  terhadap  $\hat{Y}_i$ . Nilai *DfFITS* <sub>$i$</sub>  diperoleh dari :

$$|DfFITS|_i = \frac{\hat{Y}_i - \hat{Y}_{i-1}}{s_{i-1}^2 - \sqrt{h_{ii}}} \quad (3)$$

Dengan  $h_{ii}$  adalah nilai pengaruh

Dimana amatan ke- $i$  dapat dicurigai sebagai *outlier* jika nilai :

$|DfFITS|_i > 1$ , untuk  $n \leq 30$ , dan

$|DfFITS|_i > 2 \left( \frac{p}{n} \right)^{1/2}$ , untuk  $n > 30$

Dengan  $p$  adalah banyak parameter dan  $n$  adalah banyaknya pengamatan. Pada penelitian ini akan menggunakan *Boxplot*. *Boxplot* adalah salah satu cara dalam statistika deskriptif untuk mendeteksi *outlier* pada data dengan visualisasi. Dalam *boxplot* dapat diketahui informasi terkait kuartil pertama (Q1) atau mengurangi 25% dari data terendah dengan persamaan  $Q1 = 1(n + 1)/4$ , median (Q2) atau nilai tengah dapat menggunakan persamaan  $Q2 = 2(n + 1)/4$ , kuartil tertinggi (Q3) atau mengurangi 25% dari data tertinggi dengan persamaan  $Q3 = 3(n + 1)/4$ , *Interquartil Range (IQR)* atau selisih nilai Q1 dan Q3 dimana  $IQR = Q3 - Q1$ , *outlier*.

## 2.6. Analisis Regresi Berganda

Regresi merupakan metode statistika yang biasa digunakan untuk mengetahui hubungan antar satu variabel dependen dengan dua atau lebih variabel independen. Regresi linier berganda digunakan untuk memodelkan hubungan antar dua atau lebih variabel independen. Model setiap pengamatan dinyatakan dengan statistik uji :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_j X_{ij} + \varepsilon_i \quad (4)$$

Dimana :

- $Y_i$  Nilai variabel respon dalam amatan ke- $i$ ,
- $X_i$  Variabel bebas yang diketahui nilainya dalam amatan ke- $i$ ,
- $\varepsilon_i$  Nilai galat
- $\beta_0$  Nilai intersep model regresi
- $\beta_j$  Parameter regresi ke- $j$
- $i$  Amatan (1,2,..., $n$ )
- $j$  Variabel (1,2,..., $k$ )

Apabila persamaan (4) dinyatakan dengan notasi matriks akan menjadi :

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & \dots & X_{1,k-1} \\ 1 & X_{21} & \dots & X_{2,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{n,k-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (5)$$

*Estimator* untuk persamaan (4) dengan menggunakan metode kuadrat terkecil (MKT) adalah :

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y} \quad (6)$$

Dengan  $\mathbf{X}'$  merupakan transformasi data dari  $\mathbf{X}$ .

Dengan mengalikan kedua ruas dari persamaan (6) dengan invers dari  $(\mathbf{X}'\mathbf{X})$ . Sehingga didapat *estimator* kuadrat terkecil dari  $\hat{\boldsymbol{\beta}}$  :

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (7)$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (8)$$

Dimana  $\hat{\boldsymbol{\beta}}$  adalah vektor dari parameter yang diestimasi berukuran  $(k + 1) \times 1$ ,  $\mathbf{X}$  adalah matriks variabel independen berukuran  $n \times (k + 1)$  dan  $\mathbf{Y}$  adalah vektor observasi dari variabel dependen berukuran  $n \times 1$ .

## 2.7. Uji Normalitas *Kolmogorov Smirnov*

Pada penelitian ini melakukan uji Normalitas menggunakan Q-Q plot dan *Kolmogorov Smirnov* karena data yang digunakan >50. Q-Q plot adalah visualisasi yang sering digunakan untuk menguji apakah residu regresi berdistribusi normal. Jika residu regresi berdistribusi normal, Q-Q plot akan menunjukkan pola sebaran data yang sejalan dengan garis linier. Uji normalitas *Kolmogorov Smirnov* ini bertujuan untuk mengetahui bahwa data berdistribusi normal atau tidak. Dalam menentukan hipotesis yang diambil dalam uji normalitas, diantaranya :

$H_0$  : Sampel data berdistribusi normal

$H_1$  : Sampel data tidak berdistribusi normal

Dengan statistik uji :

$$D = \max|F_0(x) - S_n(x)| \quad (9)$$

Dimana :

$D$  Statistik uji *Kolmogorov Smirnov*

$\max$  Nilai maksimum dari selisih antara distribusi empiris data dan distribusi kumulatif teoritis

$F_0(x)$  Distribusi kumulatif empiris dari data, perkiraan distribusi empiris berdasarkan data yang telah diurutkan

$S_n(x)$  Distribusi kumulatif teoritis yang berdistribusi normal

Dengan daerah Kritis atau biasa disebut daerah penolakan dimana tolak  $H_0$  jika  $D > D_{(\alpha,n)}$  atau  $sig < \alpha$ .

## 2.8. Analisis Regresi *Robust M Estimator*

Regresi *robust* dikembangkan oleh Rousseeuw dan Leroy pada tahun 1987. Regresi *robust* merupakan metode yang dapat digunakan untuk mengatasi masalah dari metode kuadrat terkecil yang disebabkan oleh data yang terkontaminasi oleh *outlier*. Regresi *robust* memiliki lima metode estimasi yaitu *M-Estimator*, *LMS-Estimator*, *LTS-Estimator*, *S-Estimator* dan *MM-Estimator*. Pada penelitian ini akan menggunakan *M-Estimator* dikarenakan salah satu kelebihan dari metode ini dapat digunakan untuk data yang berdistribusi tidak normal dan mengandung *outlier*, karena regresi *robust M estimator* menggunakan fungsi kerugian yang berbasis deviasi absolut. Deviasi absolut dihitung dengan mengambil nilai absolut dari selisih antara nilai observasi dan nilai rata-rata. Dengan menggunakan deviasi absolut, regresi *robust* dapat memberikan estimasi parameter yang lebih tahan terhadap *outlier* dan data yang tidak berdistribusi normal.

Dalam regresi *robust M estimator*, menggunakan fungsi objektif untuk mengevaluasi seberapa baik suatu model. Fungsi objektif adalah fungsi yang dioptimalkan dalam suatu logaritma untuk memprediksi seberapa baik suatu model dan meminimumkan

kesalahan dalam memprediksi model. Fungsi objektif yang umum digunakan dalam regresi *robust M estimator* adalah *Huber*. Dengan fungsi objektif *M estimator*, sebagai berikut :

$$\min \sum_{i=1}^n \rho(e_i) \quad (10)$$

fungsi objektif *Huber* dalam persamaan berikut :

$$\rho(e) = \begin{cases} \frac{e^2}{2}, & |e| \leq k \\ k|e| - \frac{k^2}{2}, & |e| > k \end{cases} \quad (11)$$

Dimana, pada *Huber* nilai  $k = 1.345$  dan  $\rho$  adalah fungsi obyektif dari residual. Turunan fungsi dari persamaan 11 sebagai berikut :

$$\varphi(y) = \begin{cases} e, & |e| \leq k \\ k \cdot \text{sgn}(e), & |e| > k \end{cases} \quad (12)$$

Dimana,  $\text{sgn}(e)$  atau *sign function* digunakan untuk menentukan suatu bilangan positif, negative atau nol (0), yang dapat didefinisikan sebagai berikut:

$$\text{sgn}(e) = \begin{cases} 1 & \text{jika } e > k \\ 0 & \text{jika } e = k \\ -1 & \text{jika } e < k \end{cases} \quad (13)$$

Pada umumnya regresi *robust M estimator* meminimumkan fungsi obyektif dengan :

$$\sum_{i=1}^n \rho(\mu_i) = \sum_{i=1}^n \rho\left(\frac{e_i}{s}\right) = \sum_{i=1}^n \rho\left(\frac{y_i - \sum_{j=0}^k X_{ij}\beta_j}{s}\right) \quad (14)$$

Dimana,  $\mu_i = \frac{e_i}{s}$  sering mengacu pada residu yang telah ditransformasi oleh fungsi pengaruh ( $\psi$ ) dan  $s$  adalah skala estimasi *robust*. Estimasi  $s$  yang digunakan adalah :

$$s = \frac{MAD}{0.6745} = \frac{\text{median}|\varepsilon_i - \text{median}(\varepsilon_i)|}{0.6745} \quad (15)$$

Dengan menggunakan turunan parsial pertama fungsi  $\rho$  terhadap  $\beta_j (j = 0, 1, \dots, k) = 0$ , maka untuk meminimumkan persamaan (2.14) diperoleh :

$$\sum_{i=1}^n X_{ij} \psi\left(\frac{y_i - X_i \beta_j}{s}\right) X_i = 0, \quad (16)$$

Dimana  $\psi$  adalah *influence function* yang digunakan untuk memperoleh bobot (*weight*). Dengan fungsi pembobot :

$$w(\mu_i) = \frac{\psi\left(\frac{y_i - \sum_{j=0}^k X_{ij}\beta_j}{s}\right)}{\left(\frac{y_i - \sum_{j=0}^k X_{ij}\beta_j}{s}\right)} \quad (17)$$

Parameter pada regresi *robust M Estimator* dilakukan dengan estimasi *Iteratively Reweighted Least Square* (IRLS). Iterasi ini membutuhkan proses iterasi dimana nilai  $w_i$  akan berubah pada setiap iterasi. Kemudian persamaan (17) dapat ditulis :

$$\sum_{i=1}^n w_i X_{ik} (y_i - \sum_{j=0}^k X_{ij} \beta_j) = 0, \quad j = 0, 1, \dots, k \quad (18)$$

Dengan  $\beta_j$  adalah perkiraan parameter regresi yang akan berubah pada setiap iterasi IRLS dan  $s$  adalah skala atau sebaran dari residual yang dihitung dalam setiap iterasi.

Persamaan (18) diselesaikan dengan IRLS. Persamaan (8) dapat ditulis :

$$\hat{\beta}_j = (X'WX)^{-1}X'Wy \quad (19)$$

Iterasi akan berhenti jika  $\hat{\beta}_j$  konvergen yaitu selisih nilai  $\hat{\beta}_j^{(m+1)}$  dan  $\hat{\beta}_j^{(m)}$  mendekati 0.

## 2.9 Uji F

Uji  $F$  dilakukan untuk mengetahui apakah variabel independen yang dimasukkan dalam model berpengaruh secara bersama-sama terhadap variabel dependen. Berikut hipotesis yang akan digunakan :

$H_0$  : variabel independen secara simultan tidak berpengaruh terhadap variabel dependen

$H_1$  : variabel independen secara simultan berpengaruh terhadap variabel dependen

Dengan statistik uji :

$$F_n = \frac{R^2/k}{\frac{1-R^2}{n-k-1}} \quad (20)$$

Dimana :

$F_n$  Statistik yang dihitung

$R$  Koefisien determinasi

$k$  Jumlah variabel independen dalam model regresi

$n$  Banyak amatan

Dengan daerah Kritis atau biasa disebut daerah penolakan dimana tolak  $H_0$  jika

$F_{hitung} > F_{(\alpha, n-k-1)}$  atau nilai  $p_{value} < \alpha$

## 2.10 Uji T

Uji  $t$  digunakan untuk menunjukkan seberapa jauh pengaruh dari variabel independen terhadap variabel dependen dengan menganggap variabel independen adalah konstan.

Berikut hipotesis yang digunakan :

$H_0$ : variabel independen ke- $k$  tidak berpengaruh signifikan terhadap variabel dependen

$H_1$ : variabel independen ke- $k$  berpengaruh signifikan terhadap variabel dependen

Dengan statistik uji :

$$t = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \quad (21)$$

Dimana :

$\hat{\beta}_k$  Estimasi koefisien regresi pada variabel independen ke- $k$

$SE(\hat{\beta}_k)$  Standar error dari koefisien regresi pada variabel independen ke- $k$

Dengan daerah Kritis atau biasa disebut daerah penolakan dimana tolak  $H_0$  jika  $t > t_{(\frac{\alpha}{2}, n-k-1)}$  atau nilai  $p_{value} < \alpha$ .

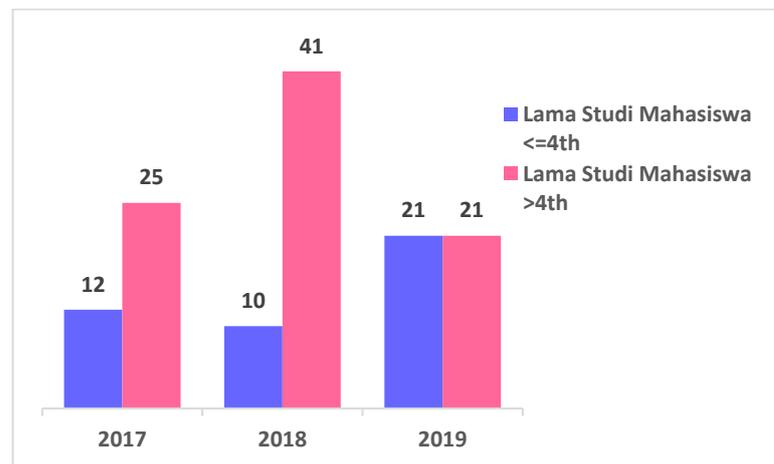
### 3. METODE PENELITIAN

Sumber data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari sub bagian akademik FMIPA ULM dan data primer yang diperoleh dari instrument penelitian dengan target responden ialah Mahasiswa Statistika FMIPA ULM Angkatan 2017 hingga 2019. Data yang digunakan merupakan data mahasiswa/i S1 Statistika FMIPA ULM Angkatan 2017 hingga 2019 dimana terdapat 130 data mahasiswa/i. Variabel Independen pada penelitian ini adalah Status Tempat Tinggal, Motivasi Mahasiswa, Proses Bimbingan, Status Berorganisasi, Status Bekerja, dan IPK. Variabel dependen pada penelitian ini adalah Lama Studi Mahasiswa.

Prosedur penelitian dimulai dari melakukan uji validitas dan reliabilitas terhadap instrumen penelitian yang akan digunakan, melakukan analisis statistika deskriptif, mengestimasi koefisien regresi dengan Metode Kuadrat Terkecil (MKT), melakukan uji asumsi normalitas *Kolmogorov Smirnov*, melakukan pemeriksaan *outlier*, mengestimasi koefisien regresi *robust M estimator*, menganalisis faktor yang mempengaruhi variabel dependen.

### 4. HASIL DAN PEMBAHASAN

#### 4.1 Analisis Deskriptif



**Gambar 1.** Masa Studi Mahasiswa S1 Statistika FMIPA ULM

Berdasarkan Gambar 1, ada 66,92% mahasiswa Statistika FMIPA ULM yang masa studinya melebihi 4 tahun. Pada angkatan 2017 terdapat sebanyak 67,57% dari 37 mahasiswa angkatan 2017 yang memiliki masa studi lebih dari 4 tahun dan terdapat 32,43% dari 37 mahasiswa Angkatan 2017 yang memiliki masa studi kurang dari sama dengan 4 tahun, pada angkatan 2018 terdapat sebanyak 80,39% dari 51 mahasiswa angkatan 2018 yang memiliki masa studi lebih dari 4 tahun dan terdapat 19,61% dari

51 mahasiswa Angkatan 2018 yang memiliki masa studi kurang dari sama dengan 4 tahun, pada angkatan 2019 terdapat sebanyak 50% dari 42 mahasiswa angkatan 2019 yang memiliki masa studi lebih dari 4 tahun dan 50% dari 42 mahasiswa angkatan 2019 yang memiliki masa studi kurang dari sama dengan 4 tahun.

#### 4.2 Analisis Regresi Berganda

Regresi linier adalah suatu metode yang digunakan untuk menyatakan pola hubungan antar variabel Y dan variabel X. Jika variabel X bernilai lebih dari satu dapat menggunakan regresi linier berganda.

**Tabel 3** Hasil Estimasi Parameter Regresi Berganda

	Estimasi Koefisien Regresi	t-tabel	t-hitung	p-value	Keputusan
<i>(Intercept)</i>	7.187747	1.979439	10.909	<2e-16***	Tolak $H_0$
STT2 ( $X_{1(2)}$ )	0.036402	1.979439	0.395	0.693810	Terima $H_0$
Motivasi ( $X_2$ )	-0.008434	1.979439	-0.496	0.620653	Terima $H_0$
PB ( $X_3$ )	-0.010879	1.979439	-1.571	0.606138	Terima $H_0$
SB2( $X_{4(2)}$ )	0.291918	1.979439	2.820	0.005602**	Tolak $H_0$
SO2( $X_{5(2)}$ )	0.052293	1.979439	0.540	0.590450	Terima $H_0$
IPK ( $X_6$ )	-0.691504	1.979439	-3.836	0.000198***	Tolak $H_0$
<i>Adj. R-Square</i>			0.2125		
<i>F-hitung</i>			6.802 ( <i>p-value</i> 3.017e-06)		
<i>Standard error</i>			0.5216		
<i>F-tabel</i>			2.173112		

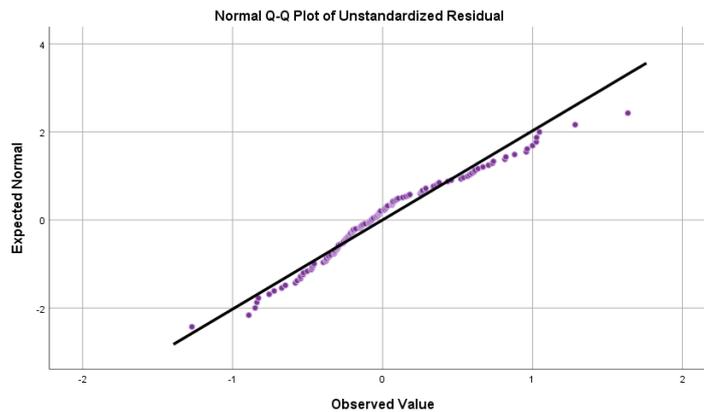
Nilai *adj. R-square* sebesar 0.2125 berarti bahwa 21.25% keragaman dari Lama Studi Mahasiswa dapat dijelaskan oleh variabel-variabel yang ada di dalam model. Dari Tabel 8 didapat model regresi sebagai berikut :

$$Y = 7.187747 + 0.036402X_{1(2)} - 0.008434X_2 - 0.010879X_3 + 0.291918X_{4(2)} + 0.052293X_{5(2)} - 0.691504X_6$$

- Pada nilai *intercept* menunjukkan bahwa jika semua variabel independen sama dengan nol (0) maka nilai variabel LMS sebesar. Koefisien *intercept* adalah nilai prediksi dari variabel Y Ketika Mahasiswa tinggal di Rumah Orang Tua, tidak memiliki Motivasi untuk menyelesaikan tugas akhir, memiliki proses bimbingan selama pengerjaan tugas akhir yang buruk, tidak berorganisasi, tidak bekerja dan memiliki IPK nol (0), maka Lama Studi Mahasiswa menjadi 7.187747 tahun.
- Nilai koefisien regresi Status Bekerja Mahasiswa sebesar 0.291918 yang berarti, jika mahasiswa yang berkerja meningkat satu satuan maka Lama Studi Mahasiswa akan bertambah 0.291918 dengan asumsi variabel lain konstan.
- Nilai koefisien regresi IPK sebesar 0.691504 yang berarti, jika peningkatan IPK mahasiswa sebesar satu satuan maka Lama Studi Mahasiswa akan berkurang 0.691504 dengan asumsi variabel lain konstan.

### 4.3 Uji Normalitas

Asumsi normalitas dapat dilakukan melalui 2 cara, yaitu secara grafik dapat menggunakan Q-Q plot dan secara pengujian hipotesis dapat menggunakan *Kolmogorov-Smirnov Test*. Hasil asumsi normalitas dapat dilihat dari Q-Q Plot pada Gambar 9.



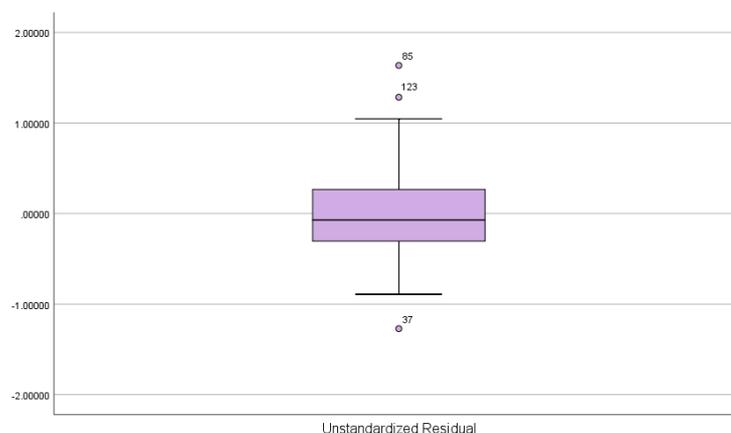
**Gambar 2** Q-Q Plot Data Residual

Dari Gambar , dapat diduga bahwa data residual tidak berdistribusi normal karena terlihat sebaran data tidak mengikuti garis linier yang ditandai dengan warna merah, untuk membuktikan benar atau tidaknya dugaan tersebut, maka dilakukan pengujian statistik *Kolmogorov-Smirnov Test*.

Berdasarkan hasil uji ternyata didapatkan statistik uji sebesar 0.25 dengan nilai  $p$  – *value* 0.001434 dimana nilai  $p$  – *value* lebih kecil dari 0.05 sehingga dapat disimpulkan menolak  $H_0$ , maka data tidak berdistribusi normal dan hal ini sejalan dengan kesimpulan yang didapat dari Gambar 2.

### 4.4 Pendeteksian *Outlier*

Diketahui dari hasil uji normalitas bahwa data residual tidak berdistribusi normal karena itu muncul dugaan bahwa adanya *outlier* pada data, sehingga dilakukan pendeteksian *outlier* pada data menggunakan *boxplot*.



**Gambar 3** Pendeteksi *Outlier*

Berdasarkan Gambar 3, menyatakan bahwa terdapat *outlier* dengan menggunakan *boxplot* berdasarkan distribusi data pada data ke- 37, 85, dan 123. Selain itu untuk mengidentifikasi *outlier* dapat menggunakan metode *DFFITS*, Dimana observasi dikatakan *outlier* jika nilai  $|DfFITS_i| > 2 \left(\frac{p}{i}\right)^{1/2} = 0.464$ .

**Tabel 4** *Outlier DffITS*

Observasi	$ DfFITS_i $	$2 \left(\frac{p}{i}\right)^{1/2}$
3	5.337	
6	5.059	
26	6.498	
37	2.333	
55	4.773	0.464
85	8.419	
123	5.999	
127	4.774	
129	6.56	

Dari Tabel 5, diketahui bahwa dengan menggunakan *DffITS* didapat *outlier* yang dapat berpengaruh terhadap hasil model regresi pada observasi 3,6,26,37,55,85,123,125,129. Pada observasi 37, 85, 123 merupakan *outlier* yang terdapat pada distribusi data dan dapat mempengaruhi hasil dari model regresi.

Berdasarkan hasil dari uji normalitas yang didapat menyatakan bahwa data tidak berdistribusi normal dan terdapat *outlier*, sehingga hasil dari analisis regresi linier berganda di atas tidak akurat. Jika asumsi normalitas tidak terpenuhi maka hal tersebut juga dapat mempengaruhi keakuratan dari hasil uji F dan uji t. Oleh karena itu penulis memutuskan untuk menggunakan analisis Regresi *Robust M Estimator* untuk mengatasi masalah yang terdapat pada data.

#### 4.5 Regresi *Robust M Estimator*

Berdasarkan hasil dari uji normalitas yang didapat menyatakan bahwa data tidak berdistribusi normal dan terdapat *outlier* pada data, sehingga regresi linier berganda tidak dapat digunakan karena adanya ketidak normalan distribusi data dan adanya *outlier* pada data dapat membuat hasil regresi tidak akurat. Jika asumsi normalitas tidak terpenuhi maka hal tersebut juga dapat mempengaruhi keakuratan hasil. Oleh karena itu penulis memutuskan untuk menggunakan analisis Regresi *Robust M Estimator* untuk mengatasi masalah *outlier* yang terdapat pada data.

Residu yang diperoleh dari model MKT (Metode Kuadrat Terkecil) digunakan untuk pendugaan *M estimator* menggunakan persamaan 10 dengan fungsi objektif *Huber*

menggunakan persamaan 11 kemudian dilakukan proses iterasi dengan persamaan 19 hingga mendapatkan hasil koefisien regresi yang cenderung konstan pada iterasi ke-10. Iterasi dilakukan hingga mendapatkan hasil koefisien regresi konvergen. Pada penelitian ini koefisien regresi konvergen pada iterasi ke-10. Sehingga didapat nilai estimasi parameter Regresi *Robust M Estimator* sebagai berikut :

$$\hat{\beta}_j = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \\ \hat{\beta}_6 \end{bmatrix} = \begin{bmatrix} 7.543249 \\ -0.132835 \\ -0.002343 \\ -0.053435 \\ 0.212522 \\ 0.057193 \\ -0.624936 \end{bmatrix}$$

Berdasarkan hasil estimasi parameter menggunakan *M Estimator*, diperoleh model regresi yang terbentuk dari nilai parameter regresi *robust M Estimator* sebagai berikut :

$$Y = 7.543249 - 0.132835 X_{1(2)} - 0.002343 X_2 - 0.053435 X_3 + 0.212522 X_{4(2)} + 0.057193 X_{5(2)} - 0.624936 X_6$$

Dari persamaan di atas dapat diketahui koefisien *intercept* adalah nilai prediksi dari variabel Y Ketika Mahasiswa tinggal di Rumah Orang Tua, tidak memiliki Motivasi untuk menyelesaikan tugas akhir, memiliki proses bimbingan selama pengerjaan tugas akhir yang buruk, tidak berorganisasi, tidak bekerja dan memiliki IPK nol (0), maka Lama Studi Mahasiswa menjadi 7.543249 tahun. Selain itu didapat nilai *adj. R-square* sebesar 0.6157 berarti bahwa 61.57% keragaman dari Lama Studi Mahasiswa dapat dijelaskan oleh variabel-variabel yang ada di dalam model, sementara sisanya 38.43% dijelaskan oleh variabel di luar model. Jika dibandingkan dengan Regresi Linier Berganda memiliki nilai *adj. R-square* (0.2125) yang lebih kecil dari nilai *adj. R-Square* Regresi *Robust M Estimator* (0.6157), yang berarti pada data penelitian ini Regresi *Robust M Estimator* lebih baik dalam menangani data yang terkontaminasi *outlier*.

#### 4.6 Uji F

Uji F yaitu pengujian yang digunakan untuk mengetahui apakah semua variabel *X* yang ada pada model memiliki dampak secara bersama-sama terhadap variabel *Y*. Dari Tabel 9 Diketahui bahwa  $F_{hitung} = 6.2492 > F_{tabel} = 2.173112$  dengan  $p - value 9.405 \times 10^{-6} < 0.05$  yang berarti tolak  $H_0$ . Artinya variabel Status Tempat Tinggal, Motivasi, Proses Belajar, Status Berkerja, Status Organisasi, IPK secara bersama-sama berpengaruh signifikan terhadap Lama Studi Mahasiswa pada  $\alpha = 5\%$ . Dengan kata lain terdapat cukup bukti bahwa minimal terdapat satu variabel independen yang berpengaruh secara signifikan terhadap variabel dependen.

#### 4.7 Uji T

Uji t yaitu pengujian yang digunakan untuk menguji bagaimana pengaruh masing-masing variabel  $X$  secara individu terhadap variabel  $Y$ .

**Tabel 5 Uji T**

Variabel	t-tabel	t-hitung	Keputusan
STT2 ( $X_{1(2)}$ )	1.979439	-1.771	Tidak Signifikan
Motivasi ( $X_2$ )	1.979439	-0.168	Tidak Signifikan
PB ( $X_3$ )	1.979439	-3.037	Signifikan
SB2 ( $X_{4(2)}$ )	1.979439	2.531	Signifikan
SO2 ( $X_{5(2)}$ )	1.979439	0.731	Tidak Signifikan
IPK ( $X_6$ )	1.979439	-3.048	Signifikan

Berdasarkan Tabel 11 diketahui bahwa variabel Proses Bimbingan, Status Bekerja, IPK berpengaruh signifikan terhadap Lama Studi Mahasiswa, dengan:

- Dugaan koefisien regresi pada variabel Proses Bimbingan sebesar 0.053435 yang berarti, jika Proses Bimbingan mahasiswa berjalan dengan baik maka Lama Studi Mahasiswa akan berkurang sebesar 0.053435 tahun dengan asumsi variabel lain konstan.
- Dugaan koefisien regresi pada variabel Status Bekerja ( $Y_a$ ) sebesar 0.212522. Artinya mahasiswa yang bekerja memiliki Lama Studi 0.212522 tahun lebih lama dibandingkan dengan mahasiswa yang tidak bekerja, dengan asumsi variabel lain konstan.
- Dugaan koefisien regresi pada variabel IPK sebesar 0.624936 yang berarti, jika adanya peningkatan IPK mahasiswa sebesar satu satuan maka Lama Studi Mahasiswa akan berkurang sebesar 0.624936 tahun dengan asumsi variabel lain konstan.

#### 5. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah diperoleh dari analisis di atas, maka dapat disimpulkan bahwa :

- Karakteristik data pada penelitian ini, didapat bahwa data residual tidak berdistribusi normal dan terkontaminasi *outlier* pada observasi 3,6,26,37,55,85,123,125,129, serta diketahui bahwa 66,92% mahasiswa S1 Statistika FMIPA ULM memiliki Lama Studi  $> 4$  tahun.
- Berdasarkan regresi *robust M estimator* diperoleh  $R^2 = 0.6335$ , faktor-faktor yang mempengaruhi Lama Studi Mahasiswa S1 Statistika FMIPA ULM adalah Proses Bimbingan mahasiswa selama mengerjakan tugas akhir, Status Bekerja mahasiswa, dan IPK mahasiswa merupakan faktor yang berpengaruh terhadap Lama Studi Mahasiswa.

## DAFTAR PUSTAKA

- [1] Y. Utami, "Uji Validitas dan Uji Reabilitas Instrument Penilaian Kinerja Dosen.," *Sains dan Teknologi*, 2023.
- [3] I. G. A. M. Srinadi, "Pengaruh Outlier Terhadap Estimator Parameter Regresi dan Model Regresi Robust," *Konferensi Nasional Matematika XVII 2014*, 2014.
- [4] P. J. & L. A. M. Rousseeuw, *Robust Regression and Outlier Detection*, USA: John Wiley & sons, 2005.
- [5] M. Bakrie, "Perihal Masa Studi Jenjang Strata 1 Maksimal 5 Tahun," 14 Januari 2016. [Online]. Available: <https://lldikti13.kemdikbud.go.id/2016/01/14/perihal-masa-studi-jenjang-strata-1-maksimal-5-tahun>
- [6] S. Azwar, *Reliabilitas dan Validitas*, Yogyakarta: Pustaka Belajar, 2012.
- [10] F. Barus, "Mendeteksi Outlier pada Data Multivariat dengan Metode Jarak Mahalanobis-Minimum Covariance Determinant (MMCD)," *IJM: Indonesia Journal of Multidisciplinary*, 1(3), 1164-1172, 2023.
- [7] M. D. & H. D. E. Lauro, "The simple boxplot method for an effective prediction," *IOP Conference Series: Materials Science and Engineering*, 2020.
- [8] D. d. R. Montgomery, *Applied Statistics and Probability for Engineers.*, United States of America: John Wiley and Sons, 2003.