

# **Application of Robust Regression M-Estimation in Handling Outliers in Open Unemployment Rate in North Sumatra**

**Penerapan Regresi Robust dengan Estimasi-M untuk Mengetahui Faktor-Faktor yang Berpengaruh terhadap Tingkat Pengangguran Terbuka di Sumatera Utara**

**Muhammad Aqil Siradj<sup>1</sup>, Irsyad Parama Dwiputra<sup>1</sup>, Avrel Chesia Berbina<sup>1</sup>, Muhammad Tsaqif Najmuddin<sup>1</sup>, Sachnaz Desta Oktarina<sup>1</sup>, Akbar Rizki<sup>1</sup>, Akmal Riza Wibisono<sup>1</sup>**

<sup>1</sup>Department of Statistics, IPB University, Indonesia

\*corresponding author: 0087siradj@apps.ipb.ac.id

## **Abstrak**

Provinsi Sumatera Utara menjadi provinsi yang memiliki tingkat pengangguran terbuka tertinggi ketiga di Indonesia per tanggal 31 desember 2021. Penelitian ini bertujuan untuk melihat faktor apa yang memengaruhi tingkat pengangguran terbuka di Sumatera Utara Tahun 2022. Pemodelan terbaik dalam menangani kasus ini terutama dengan adanya anomali pencilan dilakukan dengan regresi kekar dengan estimasi-M menggunakan pembobot Huber. Data penelitian melibatkan 25 kabupaten dan delapan kota yang bersumber dari Badan Pusat Statistik (BPS). Peubah bebas yang digunakan antara lain harapan lama sekolah (tahun), tingkat partisipasi angkatan kerja (persen), produk domestik regional bruto (rupiah/kapita), rata-rata lama sekolah (tahun), jumlah penduduk (persen), jumlah angkatan kerja (persen), kepadatan penduduk (jiwa/km<sup>2</sup>), dan pertumbuhan ekonomi (persen). Hasil penelitian menunjukkan tingkat partisipasi angkatan kerja, jumlah penduduk, jumlah angkatan kerja, dan kepadatan penduduk berpengaruh signifikan terhadap tingkat pengangguran terbuka di Sumatera Utara tahun 2022. Setiap peningkatan jumlah angkatan kerja maka tingkat pengangguran terbuka juga meningkat, sebaliknya setiap peningkatan tingkat partisipasi angkatan kerja dan kepadatan penduduk maka tingkat pengangguran terbuka akan menurun.

**Keywords:** tingkat pengangguran terbuka, pencilan, robust regression

## 1. Pendahuluan

Salah satu masalah utama yang dihadapi oleh negara berkembang seperti Indonesia adalah pengangguran. Terbatasnya lapangan kerja bagi angkatan kerja menyiratkan kegagalan kebijakan dengan konsekuensi sosial-politik dan ekonomi. Masyarakat sering kali merasa tidak puas ketika para pencari kerja tidak dapat memperoleh pekerjaan yang produktif. Pengangguran adalah perbedaan antara jumlah orang yang bersedia bekerja dan jumlah orang yang benar-benar bekerja (Indayani dan Hartono 2020). Dalam hal indikator ketenagakerjaan, pengangguran terbuka didefinisikan sebagai mereka yang tidak memiliki pekerjaan dan sedang mencari pekerjaan, sedang memulai usaha, tidak mencari pekerjaan karena merasa sulit mendapatkan pekerjaan, atau memiliki pekerjaan namun belum mulai bekerja.

Tingkat Pengangguran Terbuka (TPT) adalah persentase jumlah pengangguran terhadap jumlah angkatan kerja. Berdasarkan data Badan Pusat Statistik (BPS) pada tahun 2023 Sumatera Utara mengalami peningkatan TPT dari Februari yaitu 5,24% meningkat ke 5,89% pada bulan Agustus. Peningkatan tersebut adalah salah satu peningkatan paling tinggi pada provinsi di Indonesia pada tahun 2023. Hal tersebut dapat disebabkan oleh beberapa faktor, yaitu Harapan Lama Sekolah, Tingkat Partisipasi Angkatan Kerja (TPAK), Produk Domestik Regional Bruto, Rata-Rata Lama Sekolah, Jumlah Penduduk, Jumlah Angkatan Kerja, Kepadatan Penduduk dan Pertumbuhan ekonomi.

Memahami faktor-faktor yang memengaruhi tingkat pengangguran terbuka di Sumatera Utara menjadi penting untuk merumuskan kebijakan yang tepat dalam mengatasi permasalahan ini. Dari faktor-faktor yang memengaruhi tingkat pengangguran, akan dilakukan analisis untuk mengidentifikasi faktor yang berpengaruh signifikan pada tingkat pengangguran terbuka. Analisis akan dilakukan dengan menggunakan metode regresi linear berganda. Regresi linier berganda adalah alat statistik untuk menentukan pola hubungan antara variabel dependen dan dua atau lebih variabel independen (Maharadja *et al.* 2021). Data ekonomi seringkali mengandung *outlier* atau pengamatan yang ekstrem, sehingga dapat memengaruhi hasil analisis regresi. Oleh karena itu, digunakan metode lain untuk mempertahankan *outlier*. Pendekatan ini diperlukan untuk memastikan bahwa analisis data dengan adanya *outlier* dapat bertahan terhadap asumsi-asumsi yang mendasari analisis data. Metode ini disebut dengan metode *robust*. Regresi robust diusulkan untuk pertama kalinya pada tahun 1972 oleh Andrews ketika ada pencilan tertentu yang berdampak pada model atau ketika distribusi kesalahan tidak normal.

Kajian ini bertujuan untuk mendapatkan model terbaik yang mampu menjelaskan dan memprediksi tingkat pengangguran di Sumatera Utara secara akurat. Model ini diharapkan dapat memberikan pemahaman yang menyeluruh mengenai variabel-variabel yang memengaruhi tingkat pengangguran di kawasan ini yang

dapat menjadi dasar untuk mengembangkan langkah-langkah yang efektif untuk menanggulangnya.

## 2. Metodologi

### 2.1. Bahan dan Data

Data yang digunakan dalam penelitian ini bersumber dari Badan Pusat Statistik (BPS) dan literatur ilmiah daring yang relevan dengan topik penelitian pada tahun 2022. Sampel yang dipilih adalah kabupaten-kabupaten di Sumatera Utara. Variabel yang diamati meliputi Harapan Lama Sekolah, Tingkat Partisipasi Angkatan Kerja (TPAK), Produk Domestik Regional Bruto (PDRB), Rata-Rata Masa Sekolah, Jumlah Penduduk, Jumlah Angkatan Kerja, dan Kepadatan Penduduk. Melalui analisis statistik, kami berusaha memahami korelasi antara variabel-variabel tersebut dengan tingkat pengangguran terbuka di setiap kabupaten.

Tabel 1: *Data Variabel Dependen dan Independen*

Variabel	Nama Variabel	Satuan	Sumber Pustaka
Y	Tingkat Pengangguran Terbuka (TPT)	Persen	(Febrina, 2023)
X1	Harapan Lama Sekolah	Tahun	(Sisnita dan Prawoto, 2017)
X2	Tingkat Partisipasi Angkatan Kerja	Persen	(Chandra <i>et al.</i> , 2020)
X3	Produk Domestik Regional Bruto	Ribu Rupiah/Kapita	(Viryanto <i>et al.</i> , 2024)
X4	Rata-Rata Lama Sekolah	Tahun	(Viryanto <i>et al.</i> , 2024)
X5	Kepadatan Penduduk	Jiwa/Km <sup>2</sup>	(Muljaningsih, 2022)
X6	Pertumbuhan ekonomi	Persen	(Arianto <i>et al.</i> , 2015)
X7	Jumlah Angkatan Kerja	Jiwa	(Adriantoro <i>et al.</i> , 2020)

Penjelasan mengenai peubah-peubah penjelas di atas sebagai berikut :

1. Harapan Lama Sekolah (X1)

Menurut penelitian yang dilakukan oleh Sukirno (2008, sebagaimana dikutip dalam Sisnita dan Prawoto 2017), terdapat bukti yang menunjukkan bahwa kenaikan tingkat pengangguran terbuka tidak semata disebabkan oleh rendahnya tingkat pendidikan. Peneliti memperhitungkan faktor harapan lama sekolah sebagai salah satu variabel yang berpotensi memengaruhi tingkat pengangguran terbuka.

2. Tingkat Partisipasi Angkatan Kerja (X2)

Menurut Chandra *et al.* (2020), terdapat korelasi antara besarnya dan juga luasnya lapangan kerja yang tersedia dengan penurunan tingkat pengangguran. Semakin banyak lapangan kerja yang tersedia, semakin tinggi pula potensi penurunan tingkat pengangguran terbuka. Tingkat partisipasi angkatan kerja dianggap sebagai salah satu variabel yang penting dalam mengukur tingkat pengangguran terbuka.

3. Produk Domestik Regional Bruto (X3)

Menurut penelitian yang dilakukan oleh Arzial dan Marwan (2019,

sebagaimana dikutip dalam Viryanto *et al.* 2024), ada bukti yang mengindikasikan adanya pengaruh negatif dan signifikan antara produk domestik regional bruto (PDRB) dengan tingkat pengangguran terbuka. Hal ini menandakan bahwa peningkatan PDRB berpotensi mengurangi tingkat pengangguran terbuka. Ketika PDRB suatu wilayah meningkat, kemungkinan besar tingkat pengangguran terbuka akan mengalami penurunan.

#### 4. Rata-Rata Lama Sekolah (X4)

Menurut penelitian yang dilakukan oleh Ramiayu (2016, sebagaimana dikutip oleh Viryanto *et al.*, 2024), disebutkan bahwa rata-rata lama sekolah memiliki dampak positif terhadap tingkat pengangguran terbuka, sehingga jika rata-rata lama sekolah meningkat, tingkat pengangguran terbuka juga akan meningkat.

#### 5. Kepadatan Penduduk (X5)

Menurut penelitian yang dilakukan oleh Muljaningsih (2022), peningkatan pengangguran disebabkan oleh migrasi penduduk desa ke kota karena kurangnya lapangan kerja di pedesaan sehingga penduduk lebih memilih untuk mencari pekerjaan di kota.

#### 6. Pertumbuhan Ekonomi (X6)

Menurut penelitian yang dilakukan oleh Arianto *et al.* (2015), peningkatan pertumbuhan ekonomi dapat diikuti oleh peningkatan tingkat pengangguran terbuka. Salah satu alasan utama adalah berdirinya banyak perusahaan yang membutuhkan tenaga kerja.

#### 7. Jumlah Angkatan Kerja (X7)

Berdasarkan penelitian yang dilakukan oleh N.P.N. Yanti dan I.K. Sudibia (2019), pertumbuhan penduduk yang meningkat akan menyebabkan berkurangnya lapangan pekerjaan yang tersedia, yang pada gilirannya meningkatkan jumlah pengangguran.

## 2.2. Metode Penelitian

### 2.2.1. Regresi Linear Berganda

Model regresi linear berganda adalah model analisis regresi yang menjelaskan hubungan antara peubah penjelas (variabel independen) yang berjumlah lebih dari satu dengan peubah respon (variabel dependen) (Ngumar 2008 dalam Triyanto 2019). Model dari regresi linear berganda adalah:

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \epsilon_i$$

$y_i$  = Peubah respon (Variabel terikat)

$\beta_0$  = Intersep

$\beta_j$  = Parameter regresi ( $j=1,2,3\dots k$ )

$x_{ij}$  = Peubah penjelas ( $i=1,2,3\dots k$ )

$\epsilon_i$  = Galat regresi

Nilai dari parameter regresi ditemukan dengan metode *Ordinary Least Square* (OLS). OLS adalah metode regresi yang digunakan dengan cara meminimumkan jumlah kesalahan (error) kuadrat untuk mendapatkan dugaan parameter regresi ( $\beta$ ) (Makarti dan Karim, 2017). Rumus yang digunakan untuk menemukan nilai pendugaan parameternya adalah:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (1)$$

## 2.2.2. Uji Signifikansi Parameter Regresi Linear Berganda

### 2.2.2.1. Uji Simultan (Uji F)

Uji simultan (Uji F) adalah uji yang digunakan untuk mengetahui signifikansi pengaruh dari peubah penjelas secara bersamaan atau simultan terhadap peubah respon. Ketika nilai F-hitung lebih besar dibandingkan dengan F-tabel dan signifikansi lebih kecil dibandingkan  $\alpha(0,05)$ , maka dapat disimpulkan bahwa peubah penjelas berpengaruh signifikan terhadap peubah respon secara bersamaan. Hipotesis dari uji f adalah sebagai berikut:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0 \text{ (Tidak ada peubah penjelas yang berpengaruh signifikan)}$$

$$H_1: \text{minimal terdapat satu } \beta_i \neq 0$$

(Peubah penjelas  $X_i$  berpengaruh signifikan terhadap model)

### 2.2.2.2. Uji Parsial ( Uji t)

Uji parsial (Uji t) merupakan uji yang digunakan untuk mengetahui signifikansi pengaruh secara parsial antara peubah penjelas dan peubah respon. Ketika nilai t-hitung lebih besar dibandingkan dengan t-tabel dan signifikansi lebih kecil dibandingkan  $\alpha(0,05)$ , maka dapat disimpulkan bahwa peubah penjelas berpengaruh signifikan terhadap peubah respon secara parsial. Jika nilai Hipotesis dari uji t adalah sebagai berikut:

$$H_0: \beta_i = 0 \text{ (Peubah penjelas } X_i \text{ tidak berpengaruh signifikan terhadap model)}$$

$$H_1: \beta_i \neq 0 \text{ (Peubah penjelas } X_i \text{ berpengaruh signifikan terhadap model)}$$

## 2.2.3. Metode Pendugaan Parameter Regresi Linear Berganda Stepwise

Metode stepwise adalah paduan dari metode *Backward elimination* dan *Forward selection*. Model pertama yang dibentuk adalah model dengan seluruh peubah penjelas yang akan dieliminasi satu per satu kemudian peubah yang telah dieliminasi akan diimputasi kembali ke dalam model. Proses pemilihan model dilakukan berdasarkan nilai dari *Akaike Information Criterion* (AIC).

## 2.2.4. Uji Asumsi Klasik

Analisis regresi berganda terdapat asumsi yang harus terpenuhi agar model sesuai dengan data yang disajikan atau dapat diketahui juga sebagai *Best Linear Unbiased Estimator* (BLUE). Pengujian asumsi-asumsi tersebut, dapat dilakukan dengan melakukan uji asumsi klasik yang terdiri atas uji multikolinearitas, uji autokorelasi, dan uji heteroskedastisitas.

#### **2.2.4.1. Uji Normalitas**

Normalitas adalah ketika data memiliki sebaran yang bersifat normal. Uji normalitas adalah uji yang berfungsi untuk menguji sebaran dari populasi data bersifat normal atau tidak. Uji dilakukan untuk menganalisis kondisi normalitas data adalah uji Kolmogorov-Smirnov.

#### **2.2.4.2. Uji Multikolinearitas**

Multikolinearitas adalah kondisi ketika terjadi korelasi antara variabel penjelas atau hubungan antar variabel penjelas yang bersifat tidak bebas. Mendeteksi adanya kondisi multikolinearitas dapat dilakukan dengan mencari nilai dari *Variance Inflation Factor* (VIF). Untuk mendapatkan nilai VIF dapat menggunakan rumus di bawah.

$$VIF_i = \frac{1}{1-R_i^2} \quad (2)$$

Nilai VIF yang melebihi 10 memiliki arti bahwa terdapat multikolinearitas yang serius (Ryan 1997 dalam Sriningsih et al. 2018).

#### **2.2.4.3. Uji Autokorelasi**

Autokorelasi adalah salah satu pelanggaran asumsi OLS yang menyatakan bahwa dalam pengamatan yang berbeda tidak terdapat autokorelasi antar error (Tinungki, 2016). Uji yang dilakukan untuk menganalisis kondisi autokorelasi dari data adalah uji Durbin-Watson.

#### **2.2.2.4. Uji Heteroskedastisitas**

Heteroskedastisitas merupakan kondisi dalam model regresi berganda ketika model dari persamaan memiliki varians yang tidak tetap atau tidak konstan (Astuti dan Indarto 2020). Metode yang digunakan untuk mendeteksi heteroskedastisitas dalam penelitian ini adalah uji Breusch-Pagan.

#### **2.2.5. Kondisi Amatan**

Kondisi amatan adalah kondisi dimana suatu amatan berbeda dengan amatan lainnya. Kondisi amatan dapat berbentuk pencilan, leverage atau amatan berpengaruh. Titik pencilan adalah titik pada variabel tidak penjelas yang berbeda jauh dengan titik lainnya. Ketika terdapat titik pencilan, ada kemungkinan untuk model regresi yang terbentuk menjadi lebih besar atau kecil dibandingkan model yang sesuai. Titik leverage adalah titik pada variabel penjelas yang berbeda jauh dengan amatan lainnya, tetapi masih mengikuti linearitas variabel tidak penjelas amatan lainnya.

Amatan berpengaruh adalah gabungan antara titik pencilan dan titik leverage,

yang berarti amatan tersebut tidak mengikuti pola dari data lainnya. Meskipun amatan tersebut berbeda dengan lainnya, amatan tersebut masih berpengaruh terhadap keseluruhan data. Pendeteksian amatan berpengaruh pada penelitian menggunakan nilai *Different of Fits* (DFFITS). DFFITS digunakan untuk mengetahui pengaruh suatu pengamatan ke- $i$  terhadap model regresi yang ditinjau dari nilai fitnya (Indra *et al.*, 2013). Besar dari nilai DFFITS dapat dihitung dengan rumus di bawah.

$$DFFITS_i = t_i \sqrt{\frac{h_{ii}}{1-h_{ii}}} \quad (3)$$

Suatu amatan akan dikatakan berpengaruh apabila nilai DFFITS lebih besar dibandingkan nilai dari  $2\sqrt{\frac{p}{n}}$

### 2.2.6. Robust Regression

*Robust Regression* merupakan metode regresi yang berfungsi untuk menangani kondisi distribusi dari error tidak normal atau ketika terdapat *outlier* yang berpengaruh terhadap hasil akhir model (Olive 2005 dalam Setiarini dan Listyani 2017). Metode regresi robust yang akan digunakan untuk membuat model adalah metode estimasi M (Maximum likelihood type) yang dikenalkan oleh Huber. Estimasi yang dikenalkan oleh Huber ini menganalisis data dengan mengasumsikan bahwa sebagian besar yang terdeteksi pencilan pada peubah penjelas. Metode estimasi M digunakan bersamaan dengan pembobot huber dengan rumus pembobot.

$$w(u_i) = \begin{cases} 1, & |u_i| \leq c \\ \frac{c}{|u_i|}, & |u_i| > c \end{cases} \quad (4)$$

## 2.3 Metode Analisis

Metode analisis data yang dilakukan oleh peneliti adalah menggunakan *software R studio*. Prosedur analisis data mencakup tahapan-tahapan sebagai berikut :

1. Melakukan eksplorasi data untuk melihat informasi mengenai tingkat pengangguran terbuka di Provinsi Sumatera Utara dan variabel independen yang memengaruhinya.
2. Melakukan pengujian multikolinearitas antar variabel penjelas. Suatu model regresi dianggap bebas dari indikasi multikolinearitas jika semua variabel penjelas memiliki nilai VIF di bawah 10.
  - Jika semua variabel penjelas memiliki nilai VIF di bawah 10, analisis data dapat dilanjutkan ke tahap berikutnya.
  - Jika terdapat setidaknya satu variabel penjelas dengan nilai VIF di atas 10, perlu dilakukan seleksi dan eliminasi variabel-variabel penjelas tertentu hingga semua nilai VIF berada di bawah 10.

3. Melakukan pendugaan parameter model regresi menggunakan metode stepwise dengan tahapan sebagai berikut:
  - Pada setiap iterasi, akan ditambahkan variabel yang signifikan dan menghilangkan variabel yang tidak signifikan
  - Proses tersebut akan terus berulang hingga tidak ada lagi perubahan di dalam model, dan sudah menemukan hasil terbaik.
4. Mendeteksi keberadaan pencilan, titik leverage, dan amatan berpengaruh dengan menggunakan metode DFFITS:
  - Jika tidak ada amatan yang terdeteksi sebagai amatan berpengaruh, prosedur analisis data dapat dilanjutkan ke tahap berikutnya.
  - Jika terdapat sejumlah amatan yang terdeteksi sebagai amatan berpengaruh, amatan tersebut perlu dihapus dari kumpulan data.
5. Melakukan pengujian asumsi regresi linear berganda, termasuk asumsi Gauss-Markov (rataan sisaan sama dengan nol, ragam sisaan homogen, dan sisaan yang bebas satu sama lain):
  - Jika semua asumsi regresi linear berganda terpenuhi, prosedur analisis data dapat dilanjutkan ke tahap berikutnya.
  - Jika ada beberapa asumsi regresi linear berganda yang tidak terpenuhi, perlu dilakukan penanganan terhadap data, seperti transformasi data dan pembobotan amatan.
6. Melakukan regresi kekar estimasi-M dengan untuk mengatasi anomali yang muncul.
7. Mengidentifikasi model regresi linear terbaik serta variabel-variabel penjelas yang terbukti berpengaruh signifikan terhadap tingkat pengangguran terbuka.
8. Melakukan pengujian kembali terhadap asumsi diagnostik model Gauss-Markov yaitu rataan sisaan sama dengan nol, homoskedastisitas ragam sisaan, sisaan saling bebas, dan sisaan menyebar normal.

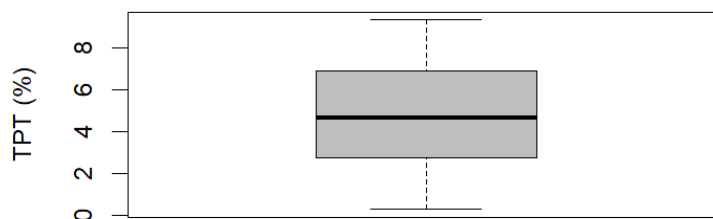
### **3. Hasil dan Pembahasan**

#### **3.1. Eksplorasi Data**

Variabel dependen yaitu tingkat pengangguran terbuka digunakan pada penelitian ini berfokus pada provinsi Sumatera Utara pada tahun 2022. Eksplorasi pada variabel dependen bertujuan untuk melihat adanya anomali pada data yang diteliti.

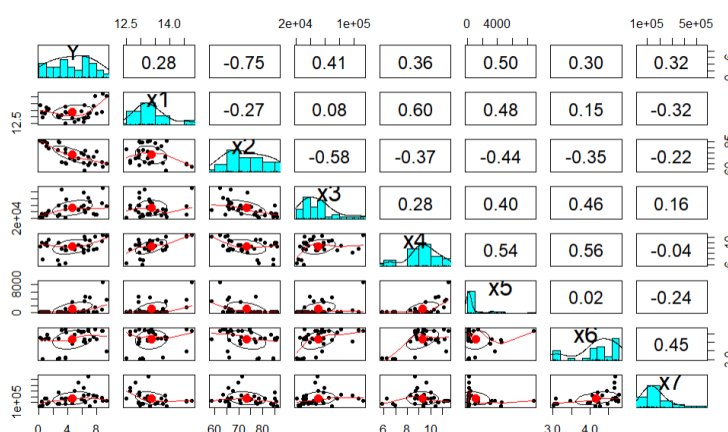
Gambar 1: Box Plot Tingkat Pengangguran Terbuka Sumatera Utara 2022





Gambar 1 memberikan sebaran yang disajikan melalui visualisasi Box Plot. Terlihat pada gambar di atas bahwa sebaran data pada Tingkat Pengangguran Terbuka di Sumatera Utara 2022 tidak memiliki pencilan yang mengindikasikan bahwa data tersebut cukup baik dalam sebaran nilainya. Eksplorasi Penyajian sebaran juga dilihat dari variabel-variabel independen yang dibantu dengan penyajian visualisasi dengan *scatter plot* dan histogram guna melihat adanya bentuk atau pola kenormalan dan arah korelasi tiap variabel independen (X) terhadap hubungan variabel dependen (Y).

Gambar 2: *Scatter Plot, Histogram, dan Korelasi antar Variabel*



Beberapa peubah memiliki sebaran yang ekstrim yaitu peubah yang tidak menyebar normal yaitu Kepadatan penduduk (X5), pertumbuhan ekonomi (X6), dan jumlah angkatan kerja (X8). Baris pertama pada gambar menjelaskan korelasi tiap variabel independen (X) yang memengaruhi variabel dependen (Y). Gambar di atas memperlihatkan ada korelasi negatif yaitu pada peubah tingkat partisipasi angkatan kerja (X2) yang menjelaskan hubungan berbanding terbalik. Terlihat bahwa korelasi yang menunjukkan sebab akibat yang sangat baik terhadap Tingkat Pengangguran Terbuka yaitu tingkat partisipasi angkatan kerja (X2) dengan nilai korelasi yaitu -0,75.

### 3.2. Pendugaan Model Regresi Awal dan Pendeteksian Multikolinearitas

Tabel 2: Pendugaan Model Regresi Awal

Peubah	Koefisien	p-value	VIF
Intersep	12.16	0.2982	

X1	0.5195	0.5149	1.9524
X2	-0,1976	0.0008 *	1.8570
X3	$-1.482 \times 10^{-5}$	0.4545	2.1511
X4	-0.203	0.6354	3.4828
X5	$5.260 \times 10^{-4}$	0.0488 *	2.5163
X6	0.2150	0.8295	3.2460
X7	$7.261 \times 10^{-6}$	0.0488 *	1.7316
<i>Adjusted R<sup>2</sup></i>		0.5868	
<i>p-value</i>		$6.902 \times 10^{-5} *$	

Keterangan : \* (signifikan di taraf nyata 5%)

Pemodelan awal dengan metode regresi linear berganda, tabel di atas menjelaskan hanya variabel tingkat partisipasi angkatan kerja (X2), kepadatan penduduk penduduk (X5), dan jumlah angkatan kerja (X7) secara signifikan memengaruhi model dengan taraf nyata 5%. *Adjusted R-Squared* menyimpulkan bahwa variabel-variabel independen yang menjelaskan model dengan persentase sebesar 58,68% terhadap variabel dependen yaitu tingkat pengangguran terbuka di Sumatera Utara tahun 2022. Pemodelan di atas belum dikatakan model terbaik karena dapat terjadi hubungan antara variabel independen yang menjadi anomali pada pemodelan regresi. Pendeteksian multikolinearitas dilakukan guna mencari pemilihan model terbaik dalam penelitian ini. Multikolinearitas pada pemodelan dapat diukur ketika nilai *Variance Inflation Factor* (VIF) > 10. Pendeteksian dalam variabel independen sesuai tabel diatas yaitu tidak terdapat hasil multikolinearitas dalam pemodelan awal regresi. Sebelumnya, tidak semua variabel independen pada pemodelan awal signifikan pada taraf nyata 5% maka dilakukan seleksi peubah untuk mencari pemodelan terbaik.

### 3.3. Pendugaan Parameter Regresi dengan Metode *Stepwise*

Tabel 3: Seleksi Peubah menggunakan Metode *stepwise*

Model	AIC	Interpretasi
Model dengan X1, X2, X3, X4, X5, X6, X7	43.83	Semua peubah ditambahkan
Model dengan X1, X2, X3, X4, X5, dan X7	41.89	Peubah X6 dihilangkan
Model dengan X1, X2, X3, X5, dan X7	40.16	Peubah X4 dihilangkan
Model dengan X2, X3, X5, dan X7	38.49	Peubah X1 dihilangkan
Model dengan X2, X5, dan X7	38.49	Peubah X3 dihilangkan

Seleksi peubah menggunakan metode *stepwise* menunjukkan beberapa langkah hingga mendapat pemodelan terbaik sesuai dengan AIC yang dihasilkan setiap iterasi. Kebaikan suatu model dilihat dari nilai AIC terkecil dari hasil seleksi peubah. Hasil pemodelan terbaik didapat dengan mengurangi peubah harapan lama sekolah (X1), produk domestik regional bruto (X3), rata-rata lama sekolah (X4), dan pertumbuhan ekonomi (X6) sehingga model regresi terbaik yaitu  $\hat{Y} = \beta_0 + \beta_1 X_2 + \beta_2 X_5 + \beta_3 X_7 + \varepsilon$ . Langkah selanjutnya mendeteksi titik pencilan, leverage, dan amatan berpengaruh agar model yang didapat tidak menimbulkan anomali dan ketidaklayakan asumsi.

### 3.4. Pendeteksian Titik Pencilan, Leverage, dan Amatan Berpengaruh

Terdapat beberapa anomali pemodelan dalam regresi linear berganda yaitu titik pencilan dan leverage. Metode *Different of Fits* (DFFITS) digunakan peneliti untuk menentukan apakah anomali tersebut termasuk amatan berpengaruh. Berikut tabel pendeteksian titik pencilan, leverage, dan amatan berpengaruh.

Tabel 4: *Tabel Pendeteksian Titik Pencilan, Leverage, dan Amatan Berpengaruh*

Observasi	$ r_i $	DFFITS <sub>i</sub>	$ h_{ii} $	Keterangan
4	2.7577	0.5749	0.0417	Pencilan
13	0.1828	-0.1526	0.4106	Leverage
30	1.0754	-1.2433	0.5724	Leverage dan Amatan Berpengaruh

Observasi amatan disebut sebagai pencilan jika  $|r_i| > 2$ , maka amatan ke 4 yaitu kabupaten Tapanuli Tengah merupakan pencilan dari gugus data tersebut. Titik leverage memiliki syarat yaitu jika nilai  $h_{ii} > \frac{2p}{n} = 0.18$ . Terdapat titik leverage pada amatan ke 13 dan 30. Amatan berpengaruh merupakan amatan yang tidak dapat dihilangkan dari data, menggunakan rumus  $|DFFITS_i| > \text{nilai pembandingan DFFITS}$  yaitu  $2\sqrt{\frac{3}{33}} \approx 0.6030$  maka didapat amatan berpengaruh berada di amatan ke 30. Hasil dari pendeteksian kondisi amatan menjelaskan terdapat titik pencilan yang menjadi anomali pada data menjadikan dasar peneliti untuk melanjutkan analisis model dengan regresi kekar. Analisis dengan model tersebut digunakan karena regresi kekar sensitif terhadap anomali sehingga dapat memperbaiki model dengan estimasi koefisien yang lebih akurat.

### 3.5. Uji Asumsi Klasik

Pemodelan regresi linear berganda perlu memenuhi beberapa asumsi sebelum melanjutkan tahap analisis yang disebut sebagai *Best Linear Unbiased Estimator* (BLUE). Uji Asumsi klasik akan menguji asumsi Gauss-Markov yaitu nilai harapan sisaan memiliki nilai sama dengan nol, homoskedastisitas ragam sisaan, sisaan saling bebas, dan sisaan menyebar normal. Uji tersebut memiliki kriteria jika

$p\text{-value} > 0,05$  mengindikasikan syarat dari asumsi terpenuhi secara keseluruhan. Tabel di bawah menyajikan uji-uji yang menjadi syarat asumsi pada regresi linear berganda.

Tabel 5: Uji Asumsi Klasik terhadap Data Variabel yang telah dieliminasi

Asumsi	Jenis Uji	p-value	Keputusan
$E(\varepsilon) = 0$	<i>t-test</i>	1	Tidak tolak $H_0$
Homoskedastisitas	<i>Breusch-Pagan test</i>	0.6353	Tidak tolak $H_0$
Sisaan saling bebas	<i>Durbin-Watson test</i>	0.7193	Tidak tolak $H_0$
Normalitas sisaan	<i>Kolmogorov-Smirnov test</i>	0.3939	Tidak tolak $H_0$

Terlihat dari tabel di atas semua uji asumsi klasik terpenuhi karena  $p\text{-value} >$  taraf nyata yaitu 5% yang mengindikasikan bahwa pemodelan yang telah dibuat valid dan dapat dilanjutkan dengan metode regresi kekar.

### 3.6. Mengatasi Anomali dengan Regresi Kekar Estimasi-M

Pencarian model terbaik pada kasus titik pencilan dan leverage yang menjadi amatan berpengaruh dapat diatasi menggunakan regresi kekar. Peneliti menggunakan pendekatan metode regresi kekar estimasi-M dengan pembobot huber yang membutuhkan pembobot ( $w_i$ ). Nilai residu ( $e_i$ ) didapat dari summary model dengan metode kuadrat terkecil. Nilai sigma duga ( $\hat{\sigma}$ ) didapat dengan rumus  $\hat{\sigma} = 1.668$ . Nilai  $u_i$  didapat dengan rumus  $u_i = \frac{e_i}{1.668}$ . Penyajian nilai pembobot dibagi menjadi dua kondisi, kondisi pertama ketika  $|u_i| \leq c$ , maka nilai pembobot diberikan nilai 1. Kondisi kedua jika  $|u_i| \geq c$ , maka nilai pembobot diberikan dengan nilai  $\frac{c}{u_i}$ . Nilai koefisien ( $c$ ) untuk pembobot huber adalah 1.345. Pendugaan estimasi *Iteratively Reweighted Least Square* (IRLS) digunakan pada regresi kekar estimasi-M yang menjalankan proses iterasi yang merubah besarnya nilai pembobot ( $w_i$ ) terhadap parameter. Menggunakan software *R Studio* didapat hasil regresi kekar pada iterasi ke-7.

Tabel 6: Penyajian Regresi Kekar dengan Pembobot Huber

Peubah	Koefisien	p-value
Intersep	16.23	$2.77 \times 10^{-5} *$
X2	-0.1812	0.0001 *
X5	-0,000485	0.0077 *
X7	$6.377 \times 10^{-6}$	0.0201 *
<i>Adjusted R<sup>2</sup></i>		0.6645
<i>p-value</i>		$1.184 \times 10^{-7} *$

Keterangan : \* (signifikan di taraf nyata 5%)

Tabel di atas menjelaskan penyajian iterasi dalam regresi kekar, didapat bahwa pada iterasi ke-8 model regresi sudah memenuhi kriteria untuk menjadi model regresi kekar yaitu diperoleh nilai koefisien duga konvergen dengan hasil iterasi sebelumnya. Pemodelan regresi terbaik yang dilakukan dengan metode regresi kekar dijelaskan sebagai berikut.

$$\hat{Y} = 16.23 - 0.1812 X_2 - 0.000485 X_5 + 6.377 \times 10^{-6} X_7$$

Uji pengujian model regresi menggunakan uji simultan dan uji parsial. Uji simultan dilihat dari *p-value* tabel karena *p-value* kurang dari 0.05 maka peubah bebas secara simultan berpengaruh terhadap peubah respons yaitu tingkat pengangguran terbuka di Sumatera Utara. Uji parsial dilihat dari *p-value* setiap peubah bebas, tabel di atas menjelaskan bahwa setiap peubah bebas yaitu tingkat partisipasi angkatan kerja ( $X_2$ ), kepadatan penduduk ( $X_5$ ), dan jumlah angkatan kerja ( $X_7$ ) lulus uji parsial dimana setiap peubah bebas memiliki pengaruh terhadap tingkat pengangguran terbuka di Sumatera Utara.

Setiap kenaikan tingkat partisipasi angkatan kerja ( $X_2$ ) sebesar satu satuan maka menurunkan tingkat pengangguran terbuka sebesar 0.1812 persen. Kenaikan kepadatan penduduk dalam satuan satuan akan menurunkan tingkat pengangguran terbuka sebesar 0.000485 persen. Sebaliknya, setiap kenaikan jumlah angkatan kerja sebesar satu satuan maka menaikkan tingkat pengangguran terbuka sebesar  $6.377 \times 10^{-6}$ . Setelah itu, uji asumsi klasik dilakukan guna melihat model regresi yang telah dibuat dapat diinterpretasikan dengan valid.

Tabel 7: Uji Asumsi Klasik terhadap Model Regresi Kekar dengan Pembobot Huber

Asumsi	Jenis Uji	p-value	Keputusan	Uji
$E(\varepsilon) = 0$	<i>t-test</i>	0.6961	Tidak tolak $H_0$	
Homoskedastisitas	<i>Breusch-Pagan test</i>	0.7309	Tidak tolak $H_0$	
Sisaan saling bebas	<i>Durbin-Watson test</i>	0.7193	Tidak tolak $H_0$	
Normalitas sisaan	<i>Kolmogorov-Smirnov test</i>	0.3838	Tidak tolak $H_0$	

asumsi klasik pada pemodelan terbaik menggunakan metode regresi kekar estimasi-M terpenuhi karena *p-value* > lebih dari taraf nyata yaitu 5% yang menjelaskan pemodelan yang dibuat sudah valid dan dapat diinterpretasikan.

#### 4. Kesimpulan

Penelitian ini melihat faktor-faktor yang menyebabkan tingkat pengangguran terbuka di Sumatera Utara tahun 2022. Tidak hanya melihat faktor tetapi peneliti melihat anomali dalam kasus model regresi tersebut. Setelah menggunakan regresi kekar estimasi-M dengan taraf nyata 5% terdapat variabel indepen yang signifikan

yaitu tingkat partisipasi angkatan kerja (X2), kepadatan penduduk (X5), dan jumlah angkatan kerja (X7). Nilai *Adj-R squared* pada model regresi menunjukkan nilai 66.45% menyimpulkan bahwa peubah bebas di atas menjelaskan kontribusi terhadap persentase tingkat pengangguran terbuka di Sumatera Utara dengan 33.55% dipengaruhi oleh peubah bebas lainnya yang tidak ada dalam model ini. Pemodelan yang diatas menggunakan regresi kekar dengan estimasi-M memberikan model yang lebih baik dengan peubah signifikan dan nilai *Adj-R squared* yang bertambah dibandingkan dengan pendugaan model awal yang memiliki anomali titik pencilan yang menunjukkan bahwa regresi kekar dapat mengatasi anomali tersebut pada amatan data penelitian.

## Daftar Pustaka

- Arianto CE, Sumarsono S, Adenan M. 2015. Pengaruh jumlah penduduk dan angka pengangguran terhadap pertumbuhan ekonomi Kabupaten Jember [artikel]. Jember: Fakultas Ekonomi, Universitas Jember.
- Arizal M, Marwan. 2019. Pengaruh produk domestik regional bruto dan indeks pembangunan manusia terhadap tingkat pengangguran terbuka di Provinsi Sumatera Barat. *EcoGen*. 2(3):433–442. doi: <http://dx.doi.org/10.24036/jmpe.v2i3.7414>.
- Astuti F, Indarto MR. 2020. Pengaruh struktur modal dan ukuran perusahaan terhadap profitabilitas pada perusahaan food & beverage yang terdaftar di bursa efek indonesia periode 2010-2015. *Cakrawangsa Bisnis*. 1(2):235-250. doi: <http://dx.doi.org/10.35917/cb.v1i2.194>.
- Cahyani FN, Muljaningsih S. 2022. Analisis pengaruh jumlah penduduk, tingkat pengangguran dan tingkat pendidikan terhadap kemiskinan di Kabupaten Gresik. *Jurnal Lampung*. 8(1):1-10. doi: <http://dx.doi.org/10.35906/jep.v8i1.977>.
- Chandra AS, Yulmardi Y, Erfit E. 2020. Pengaruh pertumbuhan penduduk, inflasi, investasi, upah minimum dan kesempatan kerja terhadap pengangguran di Kota Jambi. *Jurnal Paradigma Ekonomika*, 15(2):197–212. doi: <https://doi.org/10.22437/paradigma.v15i2.10321>.
- Chen, Colin .2002. Robust Regression and Outlier Detection with the Robustreg Procedure. SUGI paper 265-27. SAS Institute : Cary, NC
- Freund RJ, Wilson WJ, Sa P. 2006. Regression analysis statistical modeling of a response variable second edition. Boston: Academic Press.
- Hastie T, Tibshirani R, Friedman J. 2008. The Elements of Statistical Learning Data Mining, Inference, and Prediction. Ed ke-2. New York (US):Springer.
- Indayani S, Hartono B. 2020. Analisis pengangguran dan pertumbuhan ekonomi sebagai akibat pandemi Covid-19. *JEM*. 18(2):201-208. doi: <https://doi.org/10.31294/jp.v17i2>
- Indra S, Vionandra D, Sriningsih R. 2013. Pendeteksian data pencilan dan pengamatan berpengaruh pada beberapa kasus data menggunakan metode diagnostik. *Journal of Mathematics UNP*. 1(2): 67-74.
- Maharaja AN, Maulana I, Dermawan BA. 2021. Penerapan metode regresi linear berganda untuk prediksi kerugian negara berdasarkan kasus tindak pidana korupsi. *JAIC*. 5(1):95-102.

- Ramiayu DD. 2015. Analisis pengaruh rata-rata lama sekolah, upah minimum, dan pertumbuhan ekonomi terhadap tingkat pengangguran terbuka kabupaten/kota Jawa Timur. *Jurnal Ilmiah Mahasiswa FEB*. 4(2):1-16.
- Setiarini Z, Listyani E. 2017. Analisis regresi robust estimasi-s menggunakan pembobot Welsch dan Tukey Bisquare. *Jurnal Matematika*. 6(1):48-55.
- Sisnita A, Prawoto N. 2017. Analisis faktor-faktor yang memengaruhi tingkat pengangguran terbuka di Provinsi Lampung (Periode 2009-2015). *Journal of Economics Research and Social Sciences*. 1(1):1-7.
- Sriningsih M, Hatidja D, Prang JD. 2018. Penanganan multikolinearitas dengan menggunakan analisis regresi komponen utama pada kasus impor beras di Provinsi Sulut. *Jurnal Ilmiah Sains*. 18(1):18-24. doi: <https://doi.org/10.35799/jis.18.1.2018.19396>.
- Susanti Y, Pratiwi H, Sulistijowati SH, Liana T. 2014. M Estimation, S estimation, and MM estimation. *International Journal of Pure and Applied Mathematics*, 91(3):349-360. doi: <http://dx.doi.org/10.12732/ijpam.v91i3.7>.
- Tinungki GM. 2016. Metode pendeteksian autokorelasi murni dan autokorelasi tidak murni. *Jurnal Matematika, Statistika dan Komputasi*. 13(1):46-54.
- Triyanto E, Sismoro H, Laksito AD. 2019. Implementasi algoritma regresi linear berganda untuk memprediksi produksi padi di kabupaten bantul. *Jurnal Teknologi dan Sistem Informasi Univrab*. 4(2):73-86. doi: <https://doi.org/10.36341/rabit.v4i2.666>.
- Viryanto FA, Yusri TAS, Mahmudin R. 2024. Analisis regresi data panel pada dampak pendidikan terhadap tingkat pengangguran terbuka di Provinsi Kepulauan Bangka Belitung. *Jurnal Sains Matematika dan Statistika*. 10(1):31-36. doi: <http://dx.doi.org/10.24014/jsms.v10i1.26228>.
- Yanti NPN, Sudibia IK. 2019. Analisis faktor-faktor yang memengaruhi partisipasi kerja penduduk lanjut usia di Kelurahan Sesetan Kota Denpasar. *Jur. Ekon. Pembang. Fak. Ekon. dan Bisnis Univ. Udayana*. 8(1):118-147.