REGRESI KOMPONEN UTAMA *ROBUST S-ESTIMATOR* PADA INDEKS PENDUDUK MISKIN DI INDONESIA TAHUN 2022

(Skripsi)

Oleh

ATINA FARIZA AULIA
1757031017

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2024

ABSTRACT

ROBUST S-ESTIMATOR PRINCIPAL COMPONENT REGRESSION ON THE POOR POPULATION INDEX IN INDONESIA 2022

By

ATINA FARIZA AULIA

Principal Component Regression (RKU) is one method that can be used to overcome multicollinearity problems by reducing the dimensions of correlated data into new variables that are mutually independent and are linear combinations of the original variables. If the data is contaminated by outliers, the Robust method of RKU is used using the S-Estimator which in this study uses data on the percentage of poor people in Indonesia which is influenced by eight independent variables. The purpose of this study is to determine the effectiveness of the application of the S-Estimator method in Robust RKU for modelling the index of poor people in Indonesia. The result of this study is that the Robust S-Estimator method in considered the most effective in overcoming multicollinearity and outlier problems compared to RKU Classical-OLS, RKU Robust-OLS, and RKU Classical S-Estimator with an Adjusted R^2 value of 0.6451 and an RMSE value of 0.3085.

Keywords : Principal Component Regression, Multicollinearity, Outliers, Robust, Percentage of Poor Population.

ABSTRAK

REGRESI KOMPONEN UTAMA *ROBUST S-ESTIMATOR* PADA INDEKS PENDUDUK MISKIN DI INDONESIA TAHUN 2022

Oleh

ATINA FARIZA AULIA

Regresi Komponen Utama (RKU) adalah salah satu metode yang dapat digunakan untuk mengatasi masalah multikolinearitas dengan mereduksi dimensi data yang saling berkorelasi menjadi variabel-variabel baru yang saling bebas dan merupakan kombinasi linear dari variabel aslinya. Apabila data terkontaminasi oleh pencilan, maka digunakan metode *Robust* pada RKU menggunakan estimator S yang dalam penelitian ini menggunakan data persentase penduduk miskin di Indonesia yang dipengaruhi oleh delapan variabel bebas. Tujuan dari penelitian ini adalah mengetahui efektivitas penerapan metode Estimator S pada RKU *Robust* untuk pemodelan indeks penduduk miskin di Indonesia. Hasil dari penelitian ini adalah bahwa metode *Robust* Estimator S dinilai paling efektif dalam mengatasi masalah multikolinearitas dan pencilan dibandingkan RKU Klasik-OLS, RKU *Robust*-OLS, dan RKU Klasik Estimator S dengan nilai *Adjusted* R^2 sebesar 0.6451 dan nilai RMSE nya sebesar 0.3085.

Kata kunci : Regresi Komponen Utama, Multikolinearitas, Pencilan, *Robust*, Persentase Penduduk Miskin.

REGRESI KOMPONEN UTAMA *ROBUST S-ESTIMATOR* PADA INDEKS PENDUDUK MISKIN DI INDONESIA TAHUN 2022

Oleh

Atina Fariza Aulia

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2024

Judul Skripsi : **REGRESI KOMPONEN UTAMA *ROBUST S-ESTIMATOR* PADA INDEKS PENDUDUK MISKIN DI INDONESIA TAHUN 2022**

Nama Mahasiswa : **Atina Fariza Aulia**

Nomor Pokok Mahasiswa : **1757031017**

Jurusan : **Matematika**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1.Komisi Pembimbing

Prof. Drs. Mustofa Usman, M.A., Ph.D.	**Dra. Dorrah Aziz, M.Si.**
NIP. 19570101 198403 1 020	NIP. 19610128 198811 2 001

2.Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

Ketua	: **Prof. Drs. Mustofa Usman, M.A., Ph.D.**	....
Sekretaris	: **Dra. Dorrah Aziz, M.Si.**	....
Pembahas	: **Drs. Nusyirwan, M.Si.**	....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 19711001 200501 1 002

Tanggal Lulus Ujian Skripsi : **12 Juni 2024**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama	: **Atina Fariza Aulia**
Nomor Pokok Mahasiswa	: **1757031017**
Jurusan	: **Matematika**
Judul Skripsi	: **Regresi Komponen Utama *Robust S-Estimator* Pada Indeks Penduduk Miskin Di Indonesia Tahun 2022**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 12 Juni 2024
Penulis,

METERAI TEMPEL
6568DALX160024892

Atina Fariza Aulia
NPM. 1757031017

RIWAYAT HIDUP

Penulis bernama lengkap Atina Fariza Aulia, lahir di Purbolinggo pada 19 Desember 1999. Penulis merupakan anak pertama dari tiga bersaudara, pasangan Bapak Muhammad Nurdin dan Ibu Isri'ah.

Penulis menempuh pendidikan kanak-kanak di TK Dharma Wanita pada tahun 2003-2005. Kemudian penulis melanjutkan pendidikan di SD Negeri 1 Tanjung Kesuma pada tahun 2005-2011. Kemudian melanjutkan pendidikan di SMP Negeri 1 Purbolinggo pada tahun 2011-2014 dan melanjutkan pendidikan di SMA Negeri 1 Kota Metro pada tahun 2014-2017.

Penulis melanjutkan pendidikan di perguruan tinggi dan terdaftar sebagai mahasiswa Strata Satu (S1) di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung. Selama menjadi mahasiswa penulis aktif di beberapa organisasi seperti Generasi Muda Himatika (GEMATIKA) 2017, Garuda Muda BEM F Unila 2017, Korps Muda BEM U KBM Unila 2017, anggota Keilmuan di Himpunan Mahasiswa Jurusan Matematika (HIMATIKA) periode 2018, dan anggota staff Kementerian Dalam Negeri BEM U KBM Unila periode 2018.

Pada bulan Januari tahun 2020 penulis melaksanakan Kuliah Kerja Nyata (KKN) di Pekon Sekincau, Kecamatan Sekincau, Kabupaten Lampung Barat selama 40 hari sebagai bentuk pengabdian mahasiswa terhadap Tri Dharma Perguruan Tinggi. Kemudian pada bulan Agustus tahun 2020 penulis melaksanakan Kerja Praktik (KP) di Badan Pusat Statistik Kota Metro sebagai bentuk pengembangan diri serta menerapkan ilmu yang telah didapat selama perkuliahan.

KATA INSPIRASI

"Maka ingatlah kepada-Ku, aku pun akan ingat kepadamu. Bersyukurlah kepada-Ku dan janganlah kamu ingkar kepada-Ku"

(QS. Al-Baqarah : 152)

"Tuhanmu lebih mengetahui apa yang ada dalam dirimu. Jika kamu adalah orang-orang yang shaleh, sesungguhnya Dia adalah Maha Pengampun bagi orang orang yang bertaubat"

(QS. Al-Isra : 25)

"Ketetapan Allah pasti datang, maka janganlah kamu meminta agar dipercepat (datang)nya"

(QS. An-Nahl :1)

"Siapa yang bertakwa kepada Allah, niscaya Dia akan membukakan jalan keluar baginya dan memberikan rezeki dari arah yang tidak dia duga. Siapa yang bertawakal kepada Allah, niscaya Allah akan mencukupinya. Sesungguhnya Allah lah yang menuntaskan urusanNya, sesungguhnya Allah telah mengadakan bagi tiap-tiap suatu ketentuan"

(Ayat 1000 Dinar)

"Tidak ada yang dapat mengubah takdir kecuali do'a"

(Anonim)

PERSEMBAHAN

Puji dan syukur kepada Allah SWT atas segala nikmat dan karunia-Nya. Shalawat beserta salam senantiasa tercurahkan kepada baginda besar Nabi Muhammad SAW, keluarga, serta sahabat-sahabatnya yang mulia.

Kupersembahkan karya sederhana ini kepada:

Kedua Orang Tua Tercinta

Yang selalu memberikan kasih sayang, do'a, dan dukungan di setiap langkahku. Terima kasih untuk selalu menjadi rumah untukku dan atas ridho Abi Umi yang telah memudahkan setiap perjalanan hidupku.

Adik Tersayang

Semoga apa yang telahku lakukan selalu bisa menjadi contoh baik dan motivasi kelak di hidup kalian.

Dosen Pembimbing dan Pembahas

Terima kasih kepada dosen pembimbing dan pembahas yang sangat berjasa dalam membimbing, memberikan arahan, motivasi, dan ilmu yang bermanfaat.

Sahabat-sahabat Terbaik

Almamater Tercinta, Universitas Lampung

SANWACANA

Puji syukur penulis ucapkan kepada Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul "Regresi Komponen Utama *Robust S-Estimator* pada Indeks Penduduk Miskin di Indonesia Tahun 2022".

Penulis menyadari bahwa skripsi ini dapat terselesaikan dengan baik karena bimbingan, dukungan, saran, serta do'a dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis menyampaikan terima kasih kepada:

1. Bapak Prof. Drs. Mustofa Usman, M.A., Ph.D., selaku Dosen Pembimbing I yang telah bersedia memberikan arahan, bimbingan, kritik, dan saran bagi penulis selama penyusunan skripsi ini.
2. Ibu Dra. Dorrah Aziz, M.Si., selaku Dosen Pembimbing II yang telah memberikan saran serta arahan bagi penulis selama penyusunan skripsi ini.
3. Bapak Drs. Nusyirwan, M.Si., selaku Dosen Penguji yang telah memberikan kritik dan saran serta evaluasi sehingga terselesaikannya skripsi ini.
4. Bapak Subian Saidi, S.Si., M.Si., selaku Dosen Pembimbing Akademik yang telah memberikan bimbingan dan arahan selama masa perkuliahan.
5. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Bapak Dr. Eng. Heri Satria, S.Si., M.Si., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Seluruh Dosen, Staff, dan Karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
8. Orang tua tercinta dan adik-adik Rima dan Alya, serta seluruh keluarga besar yang senantiasa memberikan kasih sayang tiada terkira, serta memberikan do'a untuk kesuksesan penulis.

9. dr. Farhana Fitri Amalia atas dukungan, motivasi, dan semangat selama masa perkuliahan.
10. Sahabat-sahabat penulis Indah, Dewi, Dhea, Epmi, Yustika, Nita, Inas, Eka, Indah Ai, dan Vira yang siap mendengarkan keluh kesah, memberikan dukungan, motivasi, serta bantuan dalam hal apapun.
11. Inas, Mayda, Iqbal, Nupus, Bella, dan Diah yang telah saling memberikan bantuan, dukungan, serta motivasi di akhir-akhir masa perkuliahan ini.
12. Teman-teman “Lambetika C” atas kebersamaan dan kenangan indah selama masa perkuliahan.
13. Teman-teman Mahasiswa Jurusan Matematika Angkatan 2017.
14. Semua pihak yang telah membantu penulis dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa dalam penulisan skripsi ini masih jauh dari kata sempurna, sehingga informasi tambahan, saran, dan kritik untuk pengembangan lebih lanjut sangat penulis harapkan. Semoga skripsi ini dapat berguna dan bermanfaat bagi kita semua.

Bandar Lampung, 12 Juni 2024
Penulis,

Atina Fariza Aulia

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

Gambar Halaman

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis regresi merupakan salah satu metode statistik yang sering digunakan dalam kehidupan sehari-hari. Tujuan analisis regresi adalah untuk mengetahui rentang hubungan antara suatu variabel dependen dengan salah satu atau lebih variabel independen. Apabila analisisnya hanya mencakup satu variabel independen maka disebut analisis regresi linear sederhana. Sebaliknya, jika analisisnya mencakup beberapa variabel independen maka disebut analisis regresi linear berganda.

Masalah umum dalam analisis regresi linear berganda adalah adanya multikolinearitas. Multikolinearitas meningkatkan estimasi variansi, sehingga sulit memperoleh estimasi yang akurat, hal tersebut menyebabkan interval estimasi cenderung lebih besar dan statistik uji t menjadi kecil akibatnya variabel independen secara statistik tidak signifikan terhadap variabel dependen, serta koefisien determinasi (R^2) masih tergolong tinggi (Widarjono, 2005).

Menurut Montgomery & Peck (1992), masalah multikolinearitas dapat diatasi dengan menggunakan regresi komponen utama. Dalam mengatasi masalah ini, regresi komponen utama dapat dilakukan dalam dua tahap. Tahap pertama melakukan analisis komponen utama menggunakan vektor eigen dari matriks kovarian atau korelasi variabel independen. Tahap kedua menggunakan metode *Ordinary Least Squares* (OLS) untuk melakukan regresi komponen utama terpilih dengan variabel dependen.

Namun analisis ini sangat sensitif terhadap pencilan (*outlier*) yang membuat estimasi parameter regresi menjadi bias.

Menurut Chen (2002), regresi *robust* merupakan alat yang tepat untuk menganalisis data yang terkontaminasi oleh pencilan dan akan memberikan hasil yang resisten. Regresi *robust* terdiri dari 5 metode estimasi, antara lain *M-estimator*, *S-estimator*, *MM-estimator*, *Least Median Square* (LMS)-*estimator*, dan *Least Trimmed Square* (LTS)-*estimator*.

Di Indonesia, kemiskinan merupakan masalah yang sangat penting yang sampai saat ini penanggulangannya masih terus diupayakan. Kemiskinan tidak cukup apabila hanya diukur melalui kondisi ekonomi saja namun juga dilihat permasalahan sosial dan ketidakstabilan politik dalam negeri. Berdasarkan hal tersebut, penelitian ini akan mengkaji analisis Regresi Komponen Utama *Robust* menggunakan metode *S-Estimator* untuk menangani pencilan pada data penduduk miskin di Indonesia pada tahun 2022.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengkaji analisis regresi komponen utama dengan metode *S-Estimator* serta mengetahui ketegaran metode *S-Estimator* pada analisis regresi komponen utama dengan melihat kepekaannya terhadap pencilan.

1.3 Manfaat Penelitian

Dengan adanya penelitian ini, peneliti bermaksud untuk menambah pengetahuan bagi penulis dan memberikan referensi bagi akademisi untuk mengembangkan penelitian serta menjadi perbandingan untuk peneliti selanjutnya.

II. TINJAUAN PUSTAKA

2.1 Analisis Regresi

Regresi linear adalah metode statistik yang digunakan untuk mengetahui hubungan antara variabel dependen (Y) dengan satu atau lebih variabel independen (X). Regresi linear sederhana merupakan model regresi linear yang terdiri dari satu variabel dependen (Y) dan satu variabel independen (X) sedangkan regresi linear berganda merupakan model regresi yang terdiri atas satu variabel dependen dan memiliki lebih dari satu variabel independen. Model regresi berganda adalah model yang digunakan untuk menguji ketergantungan suatu variabel respon terhadap dua atau lebih variabel penjelas (Gujarati, 2006).

Menurut Sembiring (2003) secara umum model regresi adalah:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \tag{2.1}$$

dengan:

Y_i	: variabel tidak bebas untuk i = 1, 2, …, n
$\beta_0, \beta_1, \beta_2, \dots, \beta_k$	: parameter / koefisien regresi
$X_{1i}, X_{2i}, \dots, X_{ki}$	: variabel bebas untuk i = 1, 2, …, n
ε_i	: galat

Persamaan (2.1) jika ditulis dalam bentuk matriks adalah sebagai berikut:

$$Y = X\beta + \varepsilon \tag{2.2}$$

atau

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}; X = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix}; \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}; \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

persamaan (2.2) adalah bentuk umum persamaan regresi dalam lambang matriks, bentuk umum Y merupakan vektor dependen berukuran n × 1, X merupakan matriks peubah bebas berukuran n × (k+1), β adalah vektor parameter berukuran (k+1) × 1, dan ε vektor galat berukuran n × 1. Ada sebanyak k + 1 parameter yang harus ditaksir dari data, termasuk β_0. Seperti terdahulu taksirannya dapat ditulis dalam bentuk umum persamaan $\hat{Y} = Xb$, dengan $b = \hat{\beta}$.

2.2 Metode Kuadrat Terkecil

Menurut Sembiring (1995) metode kuadrat terkecil merupakan suatu metode yang digunakan untuk menaksir nilai parameter regresi dengan meminimumkan jumlah kuadrat galat dari model regresi yang terbentuk. Metode ini memerlukan beberapa asumsi klasik yang harus dipenuhi antara lain kenormalan, ragam yang homogen, dan non autokorelasi, dengan meminimumkan fungsi:

$$J = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \qquad (2.3)$$

untuk mengestimasi β_0 dan β_1, maka dapat dipilih b_0 dan b_1 sebagai nilai estimator yang dapat meminimumkan *J*. Nilai b_0 dan b_1 dapat ditentukan dengan menurunkan persamaan (2.3) terhadap β_0 dan β_1 maka akan diperoleh:

$$\frac{\partial J}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0$$

atau

$$\sum_{i=1}^{n} y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} x_i = 0 \qquad (2.4)$$

Kemudian J diturunkan terhadap β_1 dan samakan dengan 0

$$\frac{\partial J}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i = 0$$

atau

$$\sum_{i=1}^{n} y_i x_i - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 = 0 \tag{2.5}$$

pada persamaan (2.4) dan (2.5) nilai β_0 dan β_1 diganti dengan taksirannya masing-masing yaitu b_0 dan b_1. Persamaan tersebut kemudian menjadi suatu sistem persamaan linear yang disebut dengan persamaan normal.

$$\sum_{i=1}^{n} b_0 + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{2.6}$$

$$b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i \tag{2.7}$$

dengan $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ dan $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ maka persamaan (2.6) menjadi:

$$\begin{aligned} b_0 &= \frac{\sum_{i=1}^{n} y_i}{n} - b_1 \frac{\sum_{i=1}^{n} x_i}{n} \\ &= \bar{y} - b_1 \bar{x} \end{aligned}$$

Persamaan (2.7) menjadi:

$$\sum_{i=1}^{n} y_i x_i - \left\{\frac{\sum_{i=1}^{n} y_i}{n} - b_1\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)\right\}\left(\sum_{i=1}^{n} x_i\right) - b_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} y_i x_i - \frac{(\sum_{i=1}^{n} y_i) - (\sum_{i=1}^{n} x_i)}{n} - b_1 \left\{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right\} = 0$$

Jadi,

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

selanjutnya dapat disederhanakan menjadi:

$$b_1 = \frac{\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{2.8}$$

Seperti pada regresi linear sederhana, nilai $\beta_0, \beta_1, \dots, \beta_k$ diganti dengan $b_0, b_1, \dots, b_k$. Penaksir tersebut dapat diperoleh dengan cara meminimumkan jumlah kuadrat *error* (galat).

$$J = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 \beta_{i1} - \beta_2 \beta_{i2} - \cdots - \beta_k \beta_{ik})^2 \tag{2.9}$$

Jika J diturunkan secara parsial terhadap $\beta_0, \beta_1, \dots, \beta_k$ dan samakan dengan nol maka akan didapatkan:

$$\begin{aligned}
\frac{\partial J}{\partial \beta_0} &= -2 \sum (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik}) = 0 \\
\frac{\partial J}{\partial \beta_1} &= -2 \sum (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik}) x_{i1} = 0 \\
\frac{\partial J}{\partial \beta_2} &= -2 \sum (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik}) x_{i2} = 0 \\
&\vdots \\
\frac{\partial J}{\partial \beta_k} &= -2 \sum (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik}) x_{ik} = 0
\end{aligned}$$

Selanjutnya mengganti parameter dengan penaksirnya, setelah disusun kembali maka dapat dituliskan:

$$\begin{aligned}
\sum y_i &= n b_0 + b_1 \sum x_{i1} + b_2 \sum x_{i2} + \cdots + b_k \sum x_{ik} \\
\sum y_i x_{i1} &= b_0 \sum x_{i1} + b_1 \sum x_{i1}^2 + b_2 \sum x_{i2} x_{i1} + \cdots + b_k \sum x_{ik} x_{i1} \\
\sum y_i x_{i2} &= b_0 \sum x_{i2} + b_1 \sum x_{i1} x_{i2} + b_2 \sum x_{i2}^2 + \cdots + b_k \sum x_{ik} x_{i2} \\
&\vdots \\
\sum y_i x_{ik} &= b_0 \sum x_{ik} + b_1 \sum x_{i1} x_{ik} + b_2 \sum x_{i2} x_{ik} + \cdots + b_k \sum x_{ik}^2
\end{aligned} \tag{2.10}$$

Persamaan (2.10) merupakan persamaan normal. Berikut jika ditulis dalam lambang matriks:

$$(X^t X) b = X^t Y \tag{2.11}$$

atau

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i2} & & \sum_{i=1}^{n} X_{ik} \\ \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i1}^2 & \sum_{i=1}^{n} X_{i2}\, X_{i1} & \cdots & \sum_{i=1}^{n} X_{ik}\, X_{i1} \\ \sum_{i=1}^{n} X_{i2} & \sum_{i=1}^{n} X_{i1}\, X_{i2} & \sum_{i=1}^{n} X_{i2}^2 & & \sum_{i=1}^{n} X_{ik}\, X_{i2} \\ & \vdots & & \ddots & \vdots \\ \sum_{i=1}^{n} X_{ik} & \sum_{i=1}^{n} X_{i1}\, X_{ik} & \sum_{i=1}^{n} X_{i2}\, X_{ik} & \cdots & \sum_{i=1}^{n} X_{ik}^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & & 1 \\ X_{11} & X_{21} & \cdots & X_{n1} \\ X_{12} & X_{22} & & X_{n2} \\ & \vdots & \ddots & \vdots \\ X_{ik} & X_{2k} & \cdots & X_{nk} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}$$

Bila X^tX non singular sehingga estimasi kuadrat terkecil dari b pada persamaan (2.11) kemudian dikalikan dengan invers X^tX, sehingga menjadi:

$$(X^TX)^{-1}X^tXb = (X^TX)^{-1}X^tY$$

$$b = (X^tX)^{-1}X^tY \qquad (2.12)$$

2.3 Uji Signifikansi

2.3.1 Uji F (Simultan)

Menurut Montgomery & Peck (1992) uji ini digunakan untuk menguji keberartian regresi apabila terdapat hubungan linear antara variabel dependen y dengan variabel independen $x_1, x_2, \ldots, x_k$ secara bersama-sama. Uji-uji yang dilakukan pada model adalah sebagai berikut:

a. Hipotesis

$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$

$H_1: \text{terdapat } \beta_j \neq 0, \text{dengan } j = 1,2, \ldots, k$

b. Statistik uji

$$F_{hitung} = \frac{JKR/k}{JKS/(n-k-1)} = \frac{MSR}{MSE}$$

c. Kriteria uji

$$H_0 \text{ditolak jika } F_{hitung} > F_{(\alpha,k,n-k-1)} \text{ atau } p - value < \alpha$$

2.3.2 Uji t (Parsial)

Uji t digunakan untuk menguji apakah terdapat pengaruh yang signifikan antara masing-masing variabel independen terhadap model regresi linear. Berikut adalah langkah-langkahnya:

a. Hipotesis

$H_0 : \beta_j = 0$ (koefisien parameter variabel x_j tidak signifikan terhadap y)

$H_1 : \beta_j \neq 0$ (koefisien parameter variabel x_j signifikan terhadap y)

b. Statistik uji

$$t_{hitung} = \frac{\hat{\beta}_j}{Se(\hat{\beta}_j)}$$

c. Kriteria uji

$$H_0 \text{ditolak jika } |t_{hitung}| > t_{(\alpha/_2,n-k-1)} \text{atau } p - value < \alpha$$

2.4 Ukuran Kecocokan Model

2.4.1 Koefisien Determinasi (*Adjusted* R^2)

Menurut Montgomery & Peck (2012) koefisien determinasi (R^2) digunakan untuk mengukur seberapa baik suatu model dalam menjelaskan variasi variabel dependen. Koefisien determinasi memiliki nilai antara nol dan satu. Nilai R^2 yang kecil berarti kemampuan variabel independen dalam menjelaskan variasi variabel dependen sangat terbatas.

Nilai yang mendekati satu berarti variabel independen memberikan hampir semua informasi yang diperlukan untuk memprediksi variasi variabel dependen.

$$R^2_{Adj,k} = 1 - \frac{JKS/(n-k-1)}{JKT/(n-1)} \tag{2.13}$$

2.4.2 *Mean Square Error* (MSE)

Mean Square Error (MSE) adalah metrik yang umum digunakan dalam mengukur rata-rata dari kuadrat selisih antara nilai yang diestimasi dan nilai sebenarnya. Rumus MSE dinyatakan sebagai berikut:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2.14}$$

dengan:

n : jumlah titik data

y_i : nilai sebenarnya dari titik data ke-*i*

$\hat{y}_i$: nilai yang diprediksi (diestimasi) dari titik data ke-*i*

Nilai MSE yang lebih rendah menunjukkan kesepakatan yang lebih baik antara nilai yang diprediksi dan nilai sebenarnya, sementara nilai MSE yang lebih tinggi menunjukkan kesepakatan yang lebih buruk.

2.5 Multikolinearitas

Menurut Ryan (1997) multikolinearitas merupakan kondisi dimana terdapat korelasi antara variabel independen. Besaran (*quality*) yang dapat digunakan untuk mendeteksi adanya multikolinearitas adalah faktor inflasi ragam (*Variance Inflation Factor* / VIF). VIF digunakan sebagai kriteria untuk mendeteksi multikolinearitas pada regresi linear yang melibatkan lebih dari dua variabel independen. Nilai VIF yang lebih besar dari 10 menunjukkan adanya masalah multikolinearitas yang signifikan.

VIF untuk koefisien regresi-j dapat dijelaskan sebagai berikut:

$$VIF_{(j)} = \frac{1}{1 - R_j^2} \quad ; j = 1,2, \dots , p \tag{2.15}$$

dengan R_j^2 adalah koefisien determinasi antara X_j dengan variabel independen lainnya pada persamaan atau model dugaan.

2.6 Pencilan

Pencilan (*outlier*) adalah pengamatan yang letaknya jauh dari pengamatan-pengamatan lainnya, hal ini dapat diartikan sebagai data yang tidak mengikuti pola umum dalam model atau sebagian data yang berada diluar jangkauan dan tidak berada dalam interval kepercayaan (Sembiring, 1995).

Data pencilan bisa dikenali dengan pemeriksaan visual dari data aslinya sedangkan untuk dua atau lebih variabel independen dibutukan bantuan menggunakan uji statistik tertentu yaitu regresi diagnostik yang dapat membantu pendeteksian pencilan. Regresi diagnostik statistik digunakan untuk memeriksa tiga karakteristik yang berpotensi merupakan data pencilan antara lain *leverage*, *discrepancy*, dan *influence*.

Ukuran dari *influence* adalah kombinasi dari ukuran lainnya yaitu *discrepancy* dan *leverage* yang menjelaskan mengenai bagaimana perubahan dari persamaan regresi berubah ketika kasus ke-*i* dihapus dari himpunan data. Salah satu metode yang bisa digunakan adalah DFFITS (*Difference in Fit Standardized*).
Rumus DFFITS dapat ditulis sebagai berikut:

$$DFFITS_i = t_i \sqrt{\frac{h_{ii}}{1 - h_{ii}}} \tag{2.16}$$

dengan $t_i$merupakan *externally studentized residuals* untuk kasus ke-*i* didefinisikan sebagai berikut:

$$t_i = \frac{e_i}{\sqrt{MS_{residual(i)}1 - h_{ii}}} \qquad (2.17)$$

dan h_{ii} merupakan *leverage* didefinisikan sebagai berikut:

$$h_{ii} = \frac{1}{n} + \frac{(X_i - M_X)^2}{\sum x^2} \qquad (2.18)$$

jika $t_i$dan h_{ii} meningkat maka nilai DFFITS juga akan meningkat hal ini menunjukan bahwa kasus tersebut mempunyai pengaruh yang signifikan terhadap hasil analisis regresi. Jika kasus ke-*i* terletak tepat pada garis regresi sehingga $\widehat{Y_i}$ tidak mengalami perubahan ketika *i* dihilangkan, maka DFFITS = 0.

Himpunan data dikatakan pencilan jika nilai $|DFFITS| > 2\sqrt{p/n}$ untuk kumpulan data yang berukuran besar, dengan p = k + 1 dan n adalah banyaknya observasi (Montgomery & Peck, 1992).

Selain itu pendeteksian pencilan juga dapat dilakukan dengan bantuan jarak Mahalanobis. Jarak Mahalanobis pada dasarnya dihitung dengan mengukur setiap objek dengan mengatur besar kecilnya titik pusat dari himpunan objek tersebut, jarak Mahalanobis tidak hanya dapat mengatasi permasalahan perbedaan skala data namun juga memperhitungkan pengaruh korelasi antar variabel (Seber,2005).

Beberapa kelebihan pada jarak ini adalah dapat diterapkan pada data awal tanpa standarisasi dan juga dapat memberikan bobot yang sama pada setiap variabel untuk memperhitungkan korelasi antar variabel yang mungkin ada.

2.7 Analisis Komponen Utama

Analisis komponen utama adalah teknik statistik yang dapat digunakan untuk menjelaskan struktur variasi sekumpulan variabel melalui sejumlah variabel baru, variabel-variabel baru ini tidak bergantung satu sama lain dan merupakan kombinasi linear dari variabel aslinya. Variabel baru tersebut disebut komponen utama (*principal component*). Secara umum, tujuan analisis komponen utama adalah mereduksi dimensi data yang besar dan saling berkorelasi menjadi dimensi data yang kecil dan tidak berkorelasi (Jollife, 2002).

Menurut Johnson & Wichern (2007) secara aljabar linear komponen utama adalah kombinasi linear khusus dari p peubah acak $X_1, X_2, X_3, \dots, X_P$. Secara geometris kombinasi linear ini merupakan sistem koordinat baru hasil perputaran sistem koordinat asli di sekitar $X_1, X_2, X_3, \dots, X_P$. Sumbu baru tersebut adalah arah dengan variabilitas maksimum dan memberikan kovariansi yang lebih sederhana.

Komponen utama sepenuhnya bergantung pada matriks kovarian yang disimbolkan dengan $\sum$ (atau matriks korelasi ρ) dari komponen utama peubah-peubah $X_1, X_2, X_3, \dots, X_P$. Pengembangan komponen utama tidak memerlukan asumsi multivariat normal.

Komponen utama (W) dapat dibentuk menggunakan matriks kovarian atau matriks korelasi. Misalkan vektor acak $X^T = [X_1, X_2, \dots, X_P]$ memiliki matriks kovarian dengan *eigen value* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Sehingga kombinasi linearnya adalah sebagai berikut:

$$
\begin{aligned}
W_1 &= a_1'X = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\
W_2 &= a_2'X = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\
&\vdots \qquad\qquad\qquad\qquad\qquad \vdots \\
W_p &= a_p'X = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p
\end{aligned}
$$

untuk mencari variansi $W_i$dan kovariansi $\left(W_{i,}W_k\right)$ digunakan persamaan sebagai berikut:

$$Var(W_i) = a_i' \sum a_i \qquad i = 1,2, \dots, p \qquad (2.19)$$

$$Cov\left(W_{i,}W_k\right) = a_i' \sum a_i \qquad i, k = 1,2, \dots, p \qquad (2.20)$$

Karena komponen utama juga dapat diperoleh dari variabel yang distandarisasi, maka komponen utama yang diperoleh dibentuk sebagai kombinasi linear dari variabel yang distandarkan yaitu $Z_1, Z_2, \dots, Z_k$ sebagai berikut:

$$\begin{gathered} W_1 = a_1'Z = a_{11}Z_1 + a_{12}Z_2 + \cdots + a_{1p}Z_p \\ W_2 = a_2'Z = a_{21}Z_1 + a_{22}Z_2 + \cdots + a_{2p}Z_p \\ \vdots \qquad\qquad\qquad\qquad \vdots \\ W_p = a_p'Z = a_{p1}Z_1 + a_{p2}Z_2 + \cdots + a_{pp}Z_p \end{gathered}$$

Beberapa cara dalam menentukan seberapa banyak jumlah komponen utama yang digunakan yaitu dengan melihat persentase total varian ketika j (j < k) buah komponen yang dipilih mampu menerangkan varian sekitar 80% sampai 90%. Komponen yang diambil tersebut sudah dapat menggantikan variabel k aslinya tanpa banyak kehilangan informasi. Jumlah komponen utama juga dapat diketahui dengan menggunakan *scree plot* dimana *scree plot* tersebut merupakan plot antara λ_j dan j. Untuk menentukan jumlah komponen utama yaitu dengan melihat tikungan tajam pada *scree plot*. Jumlah komponen yang diambil adalah yang nilai eigen relatif kecil dan semua berukuran sama.

2.8 Regresi Komponen Utama

Analisis regresi komponen utama merupakan teknik analisis komponen utama yang dikombinasikan dengan analisis regresi, dimana analisis komponen utama dijadikan sebagai tahap analisis, analisis komponen utama merupakan analisis yang memperkecil dimensi variabel dengan tujuan menyederhanakan variabel yang diamati dengan cara mereduksi dimensinya (Mariana, 2013).

Variabel bebas pada regresi komponen utama berupa hasil kombinasi linear dari variabel asal Z, yang disebut sebagai komponen utama. Koefisien penduga dari metode ini diperoleh melalui penyusutan dimensi komponen utama, dimana subset komponen utama yang dipilih harus tetap mempertahankan keragaman yang sebesar-besarnya, dimana Z adalah hasil normal baku dari variabel X.

Menurut Notiragayu & Nisa (2008), regresi komponen utama merupakan regresi dari peubah tak bebas terhadap komponen-komponen utama yang tidak saling berkorelasi, dimana setiap komponen utama merupakan kombinasi linear dari semua peubah bebas yang telah dispesifikasikan sejak awal.

Bentuk persamaan regresi dalam bentuk peubah asal *X* dapat ditulis dengan:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \qquad (2.21)$$

dengan:

Y = peubah tak bebas

β_i = parameter-parameter regresi

X_i = peubah bebas ke-i $(i = 1,2, \dots, k)$

ε = galat

Peubah baru sebagai komponen utama (W) adalah hasil transformasi dari peubah asal (X) yang modelnya dalam bentuk matriks adalah W=AX dan komponen utama ke-j ditulis dengan $W_j = a_{1j}X_1 + a_{2j}X_2 + \cdots + a_{kj}X_k = a_j^T X$ dimana vektor pembobot a_j^T diperoleh dengan memaksimumkan keragaman komponen utama ke-j, yaitu $S_{W_{j_j}}^2 = a_j^T S a_j$ dengan kendala $a_j^T a_j = 1$ serta $a_h^T a_j = 0$, untuk h≠j. Vektor pembobot a_j^T diperoleh dari matriks kovarian ∑ yang diduga dengan matriks S, yaitu $S = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T$. Vektor a_j^T yang memenuhi kendala diatas adalah vektor eigen dari matriks kovarian ∑.

Model regresi komponen utama dapat ditulis sebagai berikut:

$$Y = \beta_0 + \beta_1 W_1 + \beta_2 W_2 + \cdots + \beta_j W_j + \varepsilon, \text{ dengan } j \le k \qquad (2.22)$$

2.9 Metode *S-Estimator*

Rousseeuw & Leroy (1987), menyatakan bahwa estimasi regresi yang memiliki *breakdown point* tinggi salah satunya adalah *S-estimator* yang diperkenalkan oleh Rousseeuw & Yohai (1984). *S-estimator* didefinisikan dengan:

$$\hat{\beta} = \min_{\beta} s\left(e_1(\beta), e_2(\beta), \dots, e_n(\beta)\right) \tag{2.23}$$

e_i merupakan residual ke-i dari β dan s $(e_1, e_2, \dots, e_n)$ didefinisikan solusi dari:

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{e_1}{s}\right) = \frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{s}\right) = K \tag{2.24}$$

Untuk mencapai *breakdown point* 50% maka $K = E_{\Phi}\rho(u_i) = 0.1995$ dan c $= 1.547$. Menurut Maronna, Martin, & Yohay (2006):

$$s = \sqrt{\frac{1}{nK}\sum_{i=1}^{n} w_i {e_i}^2} \tag{2.25}$$

Pembobot $w_i = w(u_i) = \frac{\rho(u_i)}{{u_i}^2}$ untuk iterasi berikutnya. Untuk iterasi pertama menggunakan:

$$s = \frac{MAD}{0{,}6745} = \frac{median\,|e_i - median\,(e_i)|}{0{,}6745} \tag{2.26}$$

Dengan penurunan parsial pertama dari ρ terhadap $\beta_j (j = 0,1, \dots, k)$ disamakan dengan 0, sehingga:

$$\sum_{i=1}^{n} x_{ij}\Psi\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{s}\right) = 0, \quad j = 0,1, \dots, k \tag{2.27}$$

Draper & Smith (1998), memberikan solusi dengan mendefinisikan fungsi pembobot:

$$w(u_i) = \frac{\Psi\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{s}\right)}{\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{s}\right)} \tag{2.28}$$

dan $w_i = w(u_i)$. Kemudian estimasi persamaan (2.18) dapat ditulis:

$$\sum_{i=1}^{n} x_{ij} w_i \left(y_i - \sum_{y=0}^{k} x_{ij} \beta_j \right) = 0, \quad j = 0,1,\dots,k \tag{2.29}$$

Persamaan (2.22) diselesaikan dengan *Iteratively Reweighted Least Square* (IRLS). Pada notasi matriks persamaan (2.22) dapat ditulis:

$$\widehat{\beta_J} = (X'WX)^{-1} X' W_y \tag{2.30}$$

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

III. METODE PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun akademik 2023/2024, bertempat di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

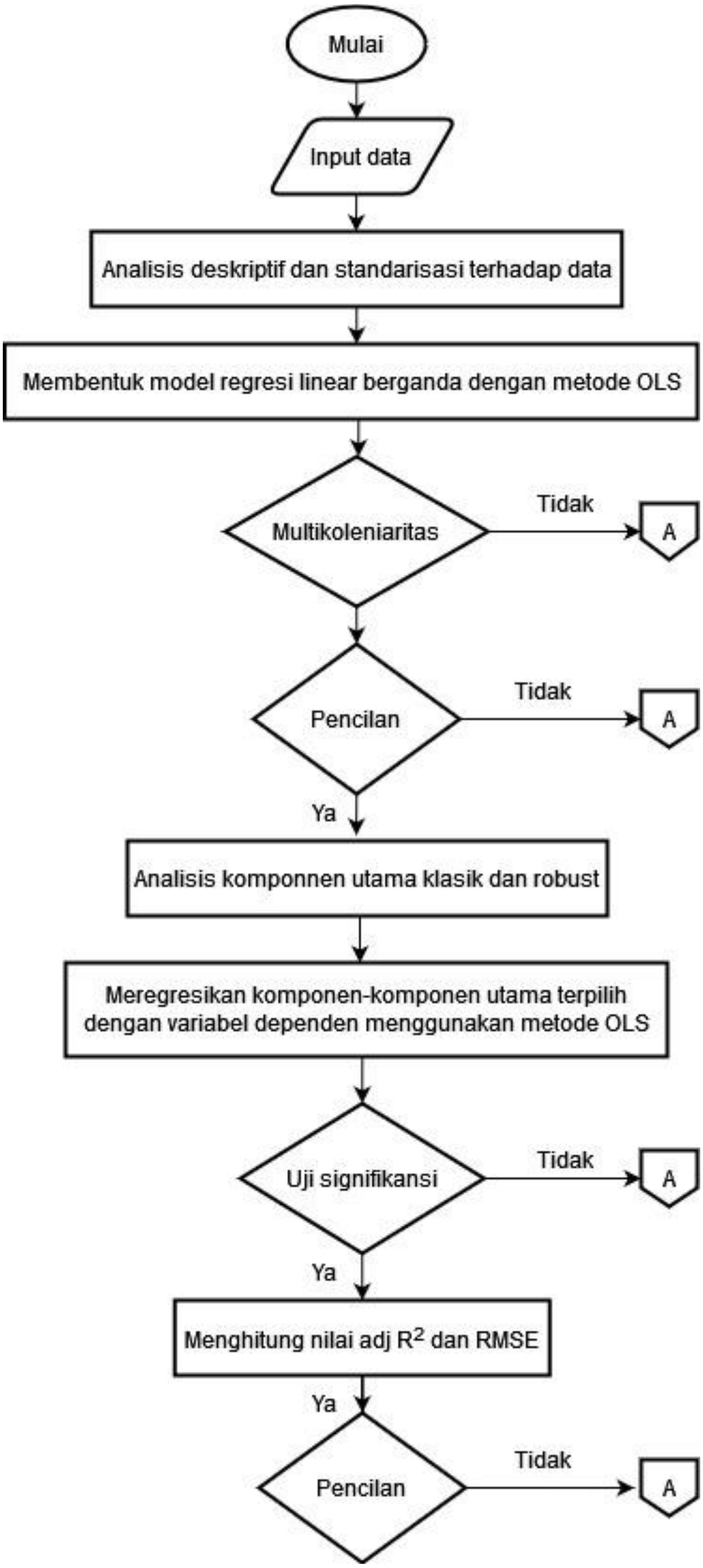
Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari *website* Badan Pusat Statistik (BPS) Indonesia *bps.go.id* dan Kementerian Ketenagakerjaan Republik Indonesia *kemnaker.go.id*. Data yang digunakan terdiri dari dua himpunan variabel yaitu variabel dependen dan variabel independen, dimana variabel dependen Y yaitu data persentase penduduk miskin di Indonesia tahun 2022 sedangkan variabel independen X_1 yaitu angka harapan hidup, X_2 yaitu harapan lama sekolah, X_3 yaitu rata-rata lama sekolah, X_4 yaitu pengeluaran rill kapita per tahun, X_5 yaitu penduduk usia kerja, X_6 yaitu jumlah pekerja anak, X_7 yaitu tingkat pengangguran terbuka, dan X_8 yaitu angka melek huruf di Indonesia tahun 2022.

3.3 Metode Penelitian

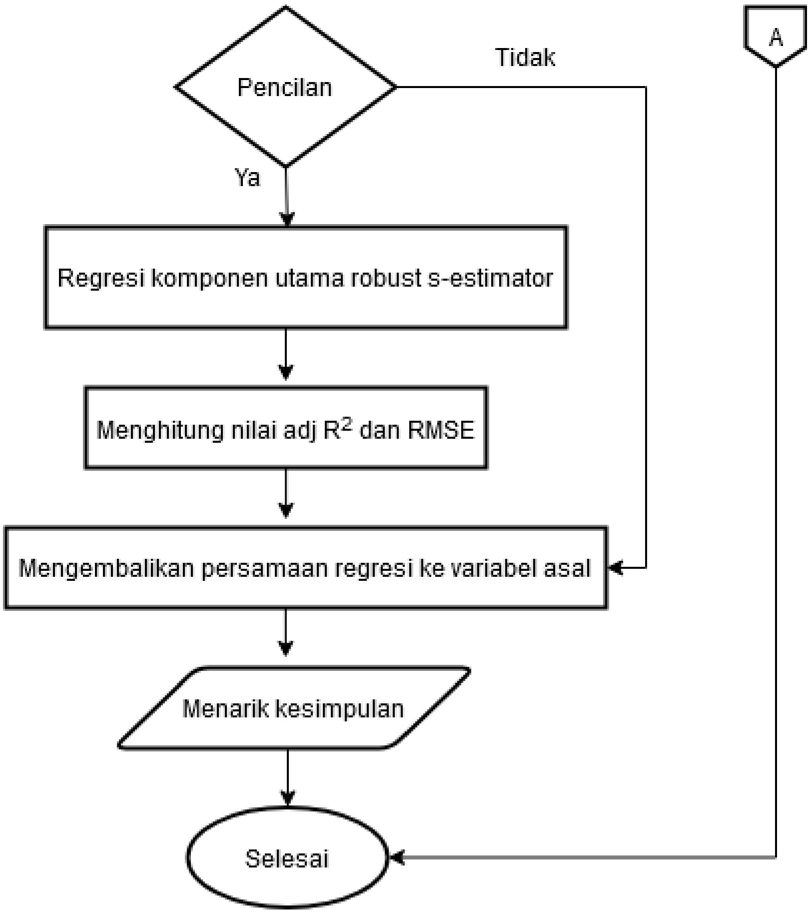
Penelitian ini dilakukan dengan bantuan perangkat lunak (*software*) *R Studio*. Adapun tahapan yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Melakukan analisis deskriptif dan standarisasi terhadap data.
2. Melakukan analisis regresi linear berganda dengan metode *Ordinary Least Square* (OLS).
3. Melakukan pemeriksaan asumsi non multikolinearitas dengan melihat nilai *Variance Inflation Factor* (VIF) dan nilai koefisien korelasi *pearson* antar variabel independen.
4. Mendeteksi adanya pencilan pada variabel independen menggunakan jarak mahalanobis.
5. Melakukan analisis komponen utama klasik dan *robust* terhadap variabel independen.
6. Meregresikan komponen-komponen utama terpilih dari analisis komponen utama klasik dan *robust* dengan variabel dependen menggunakan metode OLS.
7. Melakukan uji signifikansi model regresi melalui uji F dan uji koefisien regresi secara individu melalui uji t.
8. Melakukan deteksi multikolinearitas yang akan memperoleh hasil bahwa masalah multikolinearitas sudah teratasi.
9. Menghitung nilai *adjusted* R^2 dan nilai RSE.
10. Mendeteksi adanya pencilan pada dua buah model menggunakan metode DFFITS.
11. Terdapatnya pencilan maka dilakukan analisis komponen utama *robust*, yaitu dengan meregresikan komponen-komponen utama terpilih dengan variabel dependen menggunakan metode *S-Estimator*.
12. Menghitung nilai *adjusted* R^2 dan nilai RSE.
13. Mengembalikan persamaan regresi ke bentuk semula kemudian menarik kesimpulan.

3.4 Diagram Alir (F*lowchart*)



Gambar 1a. Diagram Alir.



Gambar 1b. Diagram Alir.

V. KESIMPULAN

Berdasarkan hasil dan pembahasan, maka diperoleh kesimpulan bahwa model Regresi Komponen Utama *Robust S-Estimator* dinilai paling efektif dan efisien dalam mengatasi masalah multikoleniaritas dan *outlier*. Didapat persamaan yaitu:

$$\hat{Y} = -0.00294X_1 + 0.3206X_2 - 0.1048X_3 + 0.00000738X_4 + 0.00000716X_5 + 0.00005X_6 - 0.0194X_7 - 0.0776X_8$$

Menunjukan bahwa persentase penduduk miskin (Y) berdasarkan 34 provinsi di Indonesia tahun 2022 dipengaruhi oleh angka harapan hidup (X_1), harapan lama sekolah (X_2), rata-rata lama sekolah (X_3), pengeluaran rill kapita per tahun (X_4), penduduk usia kerja (X_5), jumlah pekerja anak (X_6), tingkat pengangguran terbuka (X_7), dan angka melek huruf (X_8) sebesar 64,51% dan nilai RMSE nya sebesar 0,3085.

DAFTAR PUSTAKA

Chen, C. 2002. *Robust Regression and Outlier Detection with The ROBUSTREG Procedure*. SAS Institue : Cary, NC. 265-27.

Draper, N.R. & Smith, H. 1998. *Applied Regression Analysis*. 3rdEdition. John Wiley & Sons, New York.

Gujarati, D. 2006. *Basic Econometrics*. Diterjemahkan oleh Sumarto Zain. Erlangga, Jakarta.

Johnson, R.A. & Wichern, D.W. 2007. *Applied Multivariate Statistical Analysis*. 6thEdition. Pearson Prentice Hall, New Jersey.

Jollife, I.T. 2002. *Principal Component Analysis*. 2ndEdition. SpringerVerlag. Inc, New York.

Mariana. 2013. Analisis Komponen Utama. *Jurnal Matematika dan Pembelajarannya*. **2**(2):112-113.

Maronna, R.A., Martin, R.D. & Yohai, V.J. 2006. *Robust Statistics Theory and Methods*. John Wiley & Sons, New York.

Montgomery, D.C. & Peck, E.A. 1992. *Introduction to Linear Regression Analysis*. John Wiley & Sons, New York.

Montgomery, D.C. & Peck, E.A. 2012. *Introduction to Linear Regression Analysis*. *5th Edition*. John Wiley & Sons, New York.

Notiragayu & Nisa, K. 2008. Analisis Regresi Komponen Utama Robust untuk Data Mengandung Pencilan. *Jurnal Sains MIPA*. **14**(1):45-50.

Rousseeuw, P.J. & Leroy, A.M. 1987. Robust Regression and Outlier Detection. *Wiley Series in Probability and Mathematical Statistics*. John Wiley & Sons, New York.

Rousseeuw, P.J. & Yohai, V. 1984. Robust Regression by Means of S Estimators. *Robust and Nonlinear Time* Series *Analysis*. Springer Verlag, New York, pp. 256-272.

Ryan, T. P. 1997. *Modern Regression Methods*. John Wiley & Sons, New York.

Seber, A.K. 1983. *Multivariate Observations*. Aucklands, New York.

Sembiring, R.K. 1995. *Analisis Regresi Terapan*. Penerbit ITB, Bandung.

Sembiring, R.K. 2003. *Analisis Regresi*. Ed. Ke-2. Penerbit ITB, Bandung.

Widarjono, A. 2005. *Ekonometrika: Teori dan Aplikasi untuk Ekonomi dan Bisnis*. Ed. Ke-1. Ekonisia, Yogyakarta.