IMPLEMENTASI REGRESI
ROBUST PRINCIPAL COMPONENT ANALYSIS
PADA DATA KEMISKINAN JAWA TIMUR

SKRIPSI

OLEH
ELMA AL HUSNA
NIM. 15610080

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022

IMPLEMENTASI REGRESI
ROBUST PRINCIPAL COMPONENT ANALYSIS
PADA DATA KEMISKINAN JAWA TIMUR

SKRIPSI

Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Matematika (S.Mat)

Oeh
ELMA AL HUSNA
NIM. 15610080

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022

IMPLEMENTASI REGRESI
ROBUST PRINCIPAL COMPONENT ANALYSIS
PADA DATA KEMISKINAN JAWA TIMUR

SKRIPSI

Oleh
Elma Al Husna
NIM. 15610080

Telah Disetujui untuk Diuji

Malang, 13 Juni 2022

Dosen Pembimbing I	Dosen Pembimbing II
Dr. Sri Harini, M.Si	Mohammad Nafie Jauhari, M.Si
NIP. 19731014 200112 2 002	NIDT. 19870218 20160801 1 056

Mengetahui,
Ketua Program Studi Matematika

Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005

IMPLEMENTASI REGRESI
ROBUST PRINCIPAL COMPONENT ANALYSIS
PADA DATA KEMISKINAN JAWA TIMUR

SKRIPSI

Oleh
Elma Al Husna
NIM. 15610080

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima sebagai salah satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)

Tanggal 17 Juni 2022

Ketua Penguji	: Abdul Aziz, M.Si	....
Anggota Penguji 1	: Ria Dhea Layla Nur Karisma, M.Si	....
Anggota Penguji 2	: Dr. Sri Harini, M.Si	....
Anggota Penguji 3	: Mohammad Nafie Jauhari, M.Si	....

Mengetahui,
Ketua Program Studi Matematika

Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005

PERNYATAAN KEASLIAN TULISAN

Saya yang bertandatangan di bawah ini:

Nama : Elma Al Husna

NIM. : 15610080

Judul Skripsi : Implementasi Regresi *Robust Principal Component Analysis* (ROBPCA) Pada Data Kemiskinan Jawa Timur.

menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar rujukan. Apabila di kemudian hari terbukti atau dapat dibuktikan sekripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 23 Juni 2022
Yang membuat pernyataan,

METERAI TEMPEL
CCBC6AJX741391731

Elma Al Husna
NIM. 15610080

MOTO DAN PERSEMBAHAN

"Kenali dirimu, maka engkau akan mengenal Tuhanmu"

Dengan rasa syukur kepada Allah SWT penulis persembahkan skripsi ini kepada:

Abah Nuryasin dan Ibu Muzdalipah tercinta, yang tak putus-putus memberikan doa, dukungan baik secara fisik maupun psikis, semangat, motivasi serta bimbingannya agar penulis bisa menempuh jalan yang selalu di ridhoi Allah SWT dan menjadi seseorang yang tidak lupa akan sholawat kepada Nabi Muhammad Saw sebagai pengikutnya.Tak lupa kedua adik terkasih penulis, Akmal Abid Nabawi dan Engga Habiburrahman yang dengan candaanya bisa menghibur serta membangkitkan semangat penulis untuk menyelesaikan skripsi ini.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji dan syukur bagi Allah SWT atas rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini sebagai salah satu syarat untuk memperoleh gelar Sarjana Matematika (S.Mat) di Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Shalawat serta salam juga tetap tercurah kepada junjungan kita Nabi Besar Muhammad Saw yang telah menunjukkan manusia kepada jalan yang terang (Islam).

Dalam proses penyusunan skripsi ini tidak lepas dari bantuan berbagai pihak, karena itu penulis mengucapkan terima kasih kepada:

1. Prof. Dr. M. Zainuddin, M.A, selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Sri Harini, M.Si, selaku Dekan Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Elly Susanti, M.Sc, selaku ketua Program Studi Matematika, Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Sri Harini, M.Si, selaku dosen pembimbing I yang telah banyak meluangkan waktunya untuk memberikan bimbingan, nasihat, arahan dan motivasi kepada penulis.
5. Mohammad Nafie Jauhari, M.Si, selaku dosen pembimbing II yang telah banyak meluangkan waktunya untuk memberikan bimbingan, nasihat, arahan dan motivasi kepada penulis.
6. Ria Dhea Layla Nur Karisma, M.Si., selaku sebagai Penguji Utama dalam Ujian Skripsi.
7. Abdul Aziz, M.Si., selaku sebagai Ketua Penguji dalam Ujian Skripsi.
8. Segenap civitas akademika Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang terutama seluruh dosen yang telah sabar dan ikhlas dalam mendidik dan memberikan ilmu serta bimbinganya kepada penulis.

9. Abah dan Ibu serta kedua adik tercinta yang selalu memberikan doa, dukungan, semangat dan motivasi kepada penulis saat ini.
10. Pengasuh hamba Abah Kyai Haji M. Chusaini Al-Hafidz beserta Ibu Nyai Hj. Dewi Wardah W., yang senantiasa membirak ridho dan do'anya kepada penulis sebagai santri beliau selama di PPTQ Nurul Furqon Malang.
11. Seluruh teman-teman penulis baik di ranah pesantren maupun kampus yang dengan serta merta memberikan dukungan, bantuan, serta motivasi kepada penulis sehingga skripsi ini bisa terselesaikan dengan baik.

Akhir kata, semoga skripsi ini dapat bermanfaat bagi para pembaca khususnya mahasiswa Program Studi Matematika.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Malang, 23 Juni 2022

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR SIMBOL

H : Matriks pemisalan dengan ukuran $n \times n$, dimana n berisi 7 pengamatan

S : Matriks varian kovarian

$\bar{h}$: Vektor rata-rata

σ^2 : Varian dari matriks H

Q_s : Matriks yang sudah dibakukan

$W_{(n\times n)}^{\frac{1}{2}}$: Simpangan matriks pada matriks baku

ρ : Matriks korelasi

λ : Nilai *Eigen*

l : Jarak vektor pada vektor *eigen*

H_d : Variabel baru atau *principal component* terpilih

V_d : Variabel dari matriks H_d yang sudah dibakukan

m_d : Vektor *eigen* dari matriks korelasi

ω : Pengasumsian *outlier* di dalam model regresi *principal component*

τ : Fungsi pembobot *Hubert*

ABSTRAK

Al Husna, Elma. 2022. **Implementasi Regresi *Robust Principal Component Analysis* Pada Data Kemiskinan Jawa Timur.** Skripsi. Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Sri Harini, M.Si., (II) Muhammad Nafie Jauhari, M.Si.

Kata Kunci: Multikolinieritas, *Outlier*, Regresi *Robust Principal Component Analysis*, *M-estimator*, Kemiskinan.

Regresi *Robust Component Analysis* (ROBPCA) merupakan perkembangan dari metode *Principal Component Analysis* (PCA) dengan menerapkan regresi *robust* pada metode *Principal Component Analysis* (PCA) dan *Principal Componet Regression* (PCR). Regresi ROBPCA bertujuan menangani multikolinieritas dan *outlier* pada regresi linier berganda dengan cara memperoleh variabel baru yang sederhana tanpa menghilangkan variabel asal. Metode estimasi parameter yang digunakan dalam regresi ROBPCA adalah metode *M-estimator*. Data yang digunakan pada penelitian ini adalah data kemiskinan di Provinsi Jawa Timur tahun 2019 dengan jumlah kemiskinan sebagai variabel terikat (Y), variabel bebas yang terdiri dari kepadatan penduduk (X_1), tingkat pengangguran terbuka (X_2), produk domestik regional bruto (X_3), indeks keparahan kemiskinan (X_4), rata-rata lama sekolah (X_5), indeks pembangunan manusia (X_6) dan pengeluaran per kapita riil (X_7). Dari hasil pembentukan *principal component* untuk metode regresi ROBPCA, diperoleh model kombinasi sebagai berikut:

$$H_d = m_{1d}C_1 + m_{2d}C_2 + m_{3d}C_3 + m_{4d}C_4 + m_{5d}C_5 + m_{6d}C_6 + m_{7d}C_7,$$

Implementasi regresi ROBPCA *M-estimator* pada data kemiskinan di Jawa Timur menghasilkan nilai kebaikan model sebesar 70,87% dan nilai RSE sebesar 33,817, ini menunjukkan bahwa diperoleh model yang efektif dalam menangani kasus multikolinieritas dan outlier. Model tersebut adalah:

$$\begin{aligned} Y_{ROBPCA} = {} & 73{,}025 + 0{,}0003X_1 + 8{,}892X_2 + 0{,}119X_3 + 38{,}983X_4 - 0{,}878X_5 \\ & - 0{,}142X_6 - 0{,}068X_7 \end{aligned}$$

ABSTRACT

Al Husna, Elma. 2022. **Implementation of *Robust* Principal Component Analysis Regression on East Java Poverty Data**. Thesis. Mathematics Study Program, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University of Malang. Advisors: (I) Dr. Sri Harini, M.Si., (II) Muhammad Nafie Jauhari, M.Si.

Keyword: Multicollinearity, Outliers, Regression *Robust* Principal Component Analysis, M-estimator, Poverty.

Regression *Robust* Component Analysis (ROBPCA) is a development of the Principal Component Analysis (PCA) method by applying *robust* regression to the PCA and Principal Componet Regression (PCR) methods. The ROBPCA regression aims to deal with multicollinearity and outliers in multiple linear regression by obtaining a simple new variable without eliminating the original variable. The parameter estimation method used in ROBPCA regression is the M-estimator method. The data used in this study is poverty data in East Java Province in 2019 with the amount of poverty as the dependent variable (Y), the independent variable consisting of population density (X_1), open unemployment rate (X_2), gross regional domestic product (X_3), poverty severity index (X_4), average length of schooling (X_5), human development index (X_6) and real per capita expenditure (X_7). From the results of the principal component formation for the ROBPCA regression method, the following combination model is obtained:

$$H_d = m_{1d}C_1 + m_{2d}C_2 + m_{3d}C_3 + m_{4d}C_4 + m_{5d}C_5 + m_{6d}C_6 + m_{7d}C_7.$$

The implementation of the ROBPCA M-estimator regression on poverty data in East Java resulted in a model goodness value of 70.87% and an RSE value of 33,817, this indicates that an effective model is obtained in handling cases of multicollinearity and outliers. The models are:

$$\begin{aligned} Y_{ROBPCA} = {} & 73{,}025 + 0{,}0003X_1 + 8{,}892X_2 + 0{,}119X_3 + 38{,}983X_4 - 0{,}878X_5 \\ & - 0{,}142X_6 - 0{,}068X_7 \end{aligned}$$

مستخلص البحلث

الحسنى، علما. ٢٠٢٢. **تنفيذ *Robust Principal Component Analysis* Regresi على بيانات الفقر في جاوى الشرقية**. البحث الجمعي، قسم الر ياضيات. كلية العلوم والتكنولوجية. جامعة مولانامالك إبراهييم الإسلامية الحكومية مالانج. المشرفة: (١) الدكتورة سري هاريني، الماجستير، (٢) المشرف محمد نافع جوهاري، الماجستير.

الكلمات الرئيسية: Multikolinieritas، *Outlier*، Regresi *Robust Principal Component Analysis*، *M-estimator*، فقر.

تحليل المكونات القوية للإ ن حدار (*Robust Principal Component Analysis*) هو تطوير لطريقة تحليل PCA من خلال تطبيق الإن حدار *Robust* على طرق PCA و PCR. يهدف تراجع ROBPCA إلى التعامل مع التعددية و القيم المتطرفة فى **الإنحداد** الخطي المتعدد من خلال على متغير جديد بسيط دون القضاءعلى المتغير الأصلي. طريقة تقدير المعلمات المستخدمة في إنحدار ROBPCA هي الطريقة *M-estimator*. البيانات المستخدمة في هذه الدراسة هي بيانات الفقر في مقاطعة جاوى الشرقية السنة ٢٠١٩ مع مقدار الفقر كمتغير تابع (Y)، المتغيرات المستقلة تتكون من كثافة السكان (X_1)، معدل البطالة المفتوحة (X_2)، الناتج المحلي الإجمالي الإقليمي (X_3)، مؤشر شدة الفقر (X_4)، متوسط مدة الدراسة (X_5)، مؤشر التنمية البشرية (X_6) و نصيب الفرد من الإنفاق الحقيقي (X_7). من نتائج تكوين المكونات الرئيسية لطريقة الإنحدار ROBPCA، يتم الحصول على نموذج المركب التالي:

$$H_d = m_{1d}C_1 + m_{2d}C_2 + m_{3d}C_3 + m_{4d}C_4 + m_{5d}C_5 + m_{6d}C_6 + m_{7d}C_7$$

أدى تنفيذ تراجع ROBPCA *M-estimator* على بيانات الفقر في جاوى الشرقية الى قيمة طيبة نمو جية تبلغ 70,87% و قيمة RSE تبلغ 33,817، و هذا يشير إلى أنه يتم الحصول على نموذج فعال في التعامل مع حالات التعدد و القيم المتطرفة. الحصول عليها هو:

$$Y_{ROBPCA} = 73,025 + 0,0003X_1 + 8,892X_2 + 0,119X_3 + 38,983X_4 - 0,878X_5 - 0,142X_6 - 0,068X_7$$

BAB I
PENDAHULUAN

1.1 Latar Belakang

Regresi adalah salah satu metode statistika yang digunakan dalam menentukan tingkat pengaruh suatu variabel terhadap variabel yang lain, lebih tepatnya yaitu untuk mengetahui hubungan antara variabel terikat Y (*dependent*) dengan variabel bebas X (*independent*). Model regresi yang terdiri dari satu variabel terikat dan satu variabel bebas disebut dengan regresi linier sederhana, sedangkan yang terdiri dari satu variabel terikat dan memiliki lebih dari satu variabel bebas disebut dengan regresi linier berganda. Di dalam regresi linier berganda terdapat beberapa asumsi yang harus terpenuhi untuk mendapakan model yang baik, di antaranya adalah asumsi kenormalan, homokedastisitas, tidak adanya autokorelasi (*nonautocorrelation*) dan tidak adanya multikolinieritas (*nonmulticollinierity*) (Supranto, 1994).

Multikolinieritas merupakan salah satu pelanggaran asumsi yang terjadi pada regresi linier berganda di mana adanya hubungan linier atau korelasi sangat kuat pada variabel-variabel bebas. Pada umumnya, masalah multikolinieritas ini terjadi pada data dengan jumlah observasi sedikit akan tapi memiliki banyak variabel bebas (Sumodiningrat, 2002). Menurut Draper dan Smith (1992), adanya multikolinieritas ini mengakibatkan nilai determinasi dari matriks $X'X$ mempunyai *ill condition* (kondisi tidak baik) sehingga nilai estimasi varian bagi parameter menjadi lebih besar. Dikarenakan nilai estimasi varian membesar, maka hasil estimasi menjadi tidak stabil serta keputusan yang didapat menjadi tidak lagi signifikan, sehingga diperlukan sebuah penyempurnaan model dalam menemukan model terbaik bagi regresi linier berganda. Beberapa cara yang dapat digunakan untuk membebaskan suatu model regresi dari pengaruh multikolinieritas, seperti mengumpulkan data tambahan, perbandingan dan evaluasi estimator, spesifikasi model dengan menghapus suatu variabel yang berkorelasi, metode regresi *ridge* dan metode regresi *Principal Component Analysis* (PCA) (Montgomery dkk, 2012). Penelitian ini akan melibatkan metode *Principal Component Analysis* (PCA) dalam membebaskan model regresi dari pengaruh multikolinieritas dan mendapatkan model yang baik bagi regresi linier berganda.

Metode PCA merupakan salah satu metode yang digunakan untuk menyelesaikan pelanggaran asumsi multikolinieritas yang terjadi pada regresi linier berganda. Metode PCA mengatasi pelanggaran asumsi multikolinieritas dengan cara mereduksi data berdimensi besar, yakni mentransformasikan variabel yang saling berkorelasi dari satu variabel dengan variabel lain dimana akan menghasilkan variabel baru yang bebas dari korelasi. Variabel baru yang terbentuk dinamakan dengan variabel *principal component*. Hubert, dkk (2005) dalam Larasati, dkk (2020) menjelaskan bahwa proses terbentuknya metode PCA ini memiliki dua langkah yaitu, pertama *principal component* terbentuk berdasarkan matriks varian kovarian atau matriks korelasi variabel bebas dengan memanfaatkan vektor *eigen* dan yang kedua *principal component* terpilih tersebut akan diregresikan dengan variabel terikat menggunakan metode *Ordinary Least Square* (OLS). Akan tetapi matriks varian kovarian memiliki kesensitifan dengan adanya *outlier* pada data penelitian, sehingga metode PCA memerlukan suatu metode yang kuat terhadap *outlier*. Kebutuhan metode yang kuat dalam menangani keberadaan *outlier* pada metode PCA ini maka diterapkan metode regresi *robust*.

Regresi *robust* sendiri merupakan metode regresi yang digunakan pada saat distribusi dari galat tidak normal dalam model. Metode ini merupakan alat yang penting dalam menganalisis data pengamatan dengan keberadaan *outlier* di dalamnya dan menghasilkan model yang kuat (Drapper dan Smith, 1992). Menurut Chen (2002), regresi *robust* terdiri dari lima metode estimasi di antaranya adalah *Maximum Likelihood Type* (*M-estimator*), *Least Median Square* (LMS), *Least Trimmed Square* (LTS), *Scale* (*S-estimator*) dan *Method of Moment* (*MM-estimator*). Kelima metode estimasi tersebut memiliki kelebihan dan kelemahan masing-masing, oleh karena itu penyelesaian masalah multikolinieritas dan *outlier* akan dilakukan dengan menggunakan salah satu dari metode estimasi tersebut, yaitu metode *M-estimator*. Metode *M-estimator* ini akan diterapkan pada metode PCA yang *robust* dan metode gabungan ini dikenal dengan metode regresi *Robust Principal Component Analysis* (ROBPCA).

Metode *Robust Principal Component Analysis* (ROBPCA) dihasilkan dari penggabungan konsep *Project Persuit* (PP) dan *Minimum Covariance Determinant* (MCD). Metode ini diterapkan pada data penelitian yang memiliki jumlah variabel p lebih kecil dari jumlah observasi n atau pada data dengan jumlah variabel p lebih besar dari jumlah observasi n. Pada saat jumlah variabel p lebih kecil dari jumlah observasi n, maka akan diterapkan proses dekomposisi nilai singular (Hubert dkk, 2005). Penggunaan metode ROBPCA dalam menyelesaikan multikolinieritas dan *outlier* diharapkan mendapatkan model persamaan terbaik dengan mengimplementasikannya pada data kemiskinan Jawa Timur tahun 2019.

Kemiskinan merupakan salah satu permasalahan yang masih dipersoalkan oleh setiap negara di dunia baik negara maju maupun negara berkembang, salah satunya adalah negara Indonesia. Sebagai salah satu negara berkembang, Indonesia memiliki daerah kantong kemiskinan yang tersebar merata di seluruh wilayahnya. Kemiskinan itu sendiri merupakan ketidakmampuan manusia dalam memenuhi kebutuhan hidupnya baik untuk diri sendiri maupun keluarganya (BPS, 2018). Terdapat banyak faktor yang mempengaruhi peningkatan jumlah kemiskinan di setiap tahunnya, seperti pada faktor kepadatan penduduk, tingkat pengangguran terbuka, produk domestik regional bruto, indeks keparahan kemiskinan, rata-rata lama sekolah, indeks pembangunan manusia dan pengeluaran per kapita riil.

Peran dalam upaya pengentasan kemiskinan yang terjadi di tengah-tengah masyarakat tidak hanya dilakukan oleh pemerintah saja, masyarakat yang terkategorikan mampu juga termasuk yang ikut andil di dalamnya. Peran yang dilakukan antar masyarakat ini adalah gambaran dari sikap tolong-menolong antar sesama makhluk ciptaanNya. Menghindari sikap tersebut sangat dilarang di dalam Islam seperti halnya tidak mau memberikan santunan atau membiarkan mereka yang tidak mampu mengalami kelaparan, sebagaimana firman Allah SWT di dalam Al-Qur'an surat Al-Mudatsir ayat 40-44 yang artinya sebagai berikut: (Shihab, 2003)

"Mereka di dalam surga, saling tanya menanya, tentang para pendurhaka: "Apakah yang memasukkan kamu ke dalam Saqar?" Mereka menjawab, "Kami dahulu tidak termasuk orang-orang yang shalat, dan kami tidak (pula) memberi makan orang miskin."

Menurut Shihab (2003) di dalam tafsir Al-Misbah, adanya kemiskinan terjadi akibat kekurangan harta benda atau oleh sebab lain seperti keteraniayaan, kerendahan hati dan sebagainya. Dan yang dimaksudkan pada ayat 44, tidak menunaikan zakat atau tidak bersedekah merupakan suatu pengembangan dari keburukan hubungan terhadap sesama manusia.

Penelitian ini merupakan perkembangan dari penelitian Pratiwi (2016) yang menyatakan bahwa metode *ridge robust* estimasi-M menghasilkan nilai *standard error* yang sangat besar sehingga perlu dilakukan analisis dengan metode berbeda untuk mendapatkan nilai *standard error* yang lebih kecil. Salah satu metode yang dapat digunakan untuk menangani data dengan pelanggaran asumsi multikolinieritas dan *outlier* antara lain adalah metode regresi ROBPCA.

Penelitian ini merujuk pada Jurnal Ilmu Komputer dengan judul “Metode ROBPCA dan Metode Clara (*Clustering Large Area*) pada data dengan *Outlier*” oleh Susilowati, dkk (2020), yang menyatakan bahwa metode ROBPCA merupakan suatu model PCA yang *robust* terhadap *outlier* dan lebih efisien (mampu menghasilkan lebih sedikit jumlah *principal component*) daripada model PCA klasik. Selanjutnya penelitian oleh Larasati, dkk (2020), yang meneliti tentang analisis regresi komponen utama *robust* menggunakan metode *Minimum Covariance Determinant* (MCD) dan metode *Least Trimmed Square* (LTS) dimana menghasilkan model terbaik untuk metode regresi komponen utama *robust* dengan diperolehnya nilai rata-rata bias yang lebih kecil dibandingkan regresi komponen utama klasik, dalam hal ini menunjukkan bahwa metode regresi komponen utama yang *robust* merupakan metode yang kekar dan efektif terhadap keberadaan *outlier*. Berdasarkan pemaparan di atas, maka penulis akan melanjutkan penelitian tentang metode regresi ROBPCA dengan judul penelitian “Implementasi Regresi *Robust Principal Component Analysis* pada Data Kemiskinan Jawa Timur”.

1.2 Rumusan Masalah

Berdasarkan latar belakang pada penelitian ini, maka rumusan masalah yang diambil adalah sebagai berikut:

1. Bagaimana bentuk persamaan *principal component* yang dihasilkan untuk metode regresi *Robust Principal Component Analysis* (ROBPCA)?
2. Bagimana model regresi *Robust Principal Component Analysis* (ROBPCA) dengan metode *M-estimator* yang diimplementasikan pada data Kemiskinan Jawa Timur tahun 2019 yang mengandung multikolinieritas dan *outlier*?

1.3 Tujuan Masalah

Berdasarkan latar belakang tersebut maka tujuan penelitian yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Mengetahui bentuk persamaan *principal component* pada metode regresi *Robust Principal Component Analaysis* (ROBPCA).
2. Mengetahui model regresi *Robust Principal Component Analysis* (ROBPCA) dengan metode *M-estimator* yang dimplementasikan pada data Kemiskinan Jawa Timur tahun 2019 yang mengandung multikolinieritas dan *outlier*.

1.4 Manfaat Penelitian

Adapun manfaat yang di peroleh dari penelitian ini adalah sebagai berikut:

1. Bagi Peneliti

   Penelitian ini merupakan kesempatan bagi peneliti untuk mengimplementasikan metode ROBPCA dalam menyelesaikan masalah pelanggaran asumsi berupa multikolinieritas dan adanya *outlier* pada data kemiskinan Jawa Timur tahun 2019.

2. Bagi Pembaca
   a. Penelitian ini dapat dijadikan sebagai bahan rujukan dan pengembangan dalam pembelajaran bidang statistika.
   b. Sebagai contoh studi kasus mata kuliah pilihan bidang statistika yang dipelajari di bangku kuliah.
   c. Penelitian ini dapat memberikan metode alternatif prihal metode-metode lain yang digunakan untuk menyelesaikan masalah pelanggaran asumsi berupa mulikolinieritas dan *outlier* pada data penelitian.
3. Bagi Lembaga
   a. Penelitian ini dapat menambah wawasan dan pengetahuan dalam keilmuan matematika, terkhususnya bidang statistika.
   b. Membandingkan penelitian yang sudah ada dengan metode lain.
   c. Mengimplementasikan ilmu matematika terutama dalam bidang statistika.

1.5 Batasan Masalah

Batasan masalah yang diambil berdasarkan rumusan masalah pada penelitian di atas adalah:

1. Metode estimasi parameter yang digunakan adalah metode *M-estimator*.
2. Memusatkan penelitian pada masalah multikolinieritas dan *outlier*.
3. Data yang digunakan data Kemiskinan Jawa Timur tahun 2019.

4. Variabel yang digunakan dalam penelitian ini adalah jumlah kemiskinan Jawa Timur tahun 2019 yang meliputi kepadatan penduduk, tingkat pengangguran terbuka, produk domestik regional bruto, indeks keparahan kemiskinan, rata-rata lama sekolah, indeks pembangunan manusia dan pengeluaran per kapita riil.

1.6 Definisi Istilah

Untuk memperjelas maksud dan tujuan dalam penenlitian ini agar lebih terfokuskan pada judul penelitian maka peneliti mengambil definisi istilah sebagai berikut:

1. Implementasi

   Implementasi yang dimaksud dalam penelitian ini adalah proses penerapan suatu model dengan metode yang sudah ditentukan peneliti pada data kemiskinan Jawa Timur tahun 2019 dengan beberapa indikator atau variabel parameter yang sudah ditentukan, yaitu:

   - Jumlah Kemiskinan
   - Kepadatan Penduduk
   - Tingkat Pengangguran Terbuka (TPT)
   - Produk Domestik Regional Bruto (PDRB)
   - Indeks Keparahan Kemiskinan (IKK)
   - Rata-rata Lama Sekolah
   - Indeks Pembangunan Manusia (IPM)
   - Pengeluaran per Kapita Riil

2. Metode Regresi *Robust Principal Component Analysis* (ROBPCA)

   Metode ROBPCA yang digunakan peneliti dalam penelitian ini merupakan gabungan dari dua metode yaitu metode PCA dan metode *robust*. Metode ini bertujuan untuk menyelesaikan kasus multikolinieritas dan *outlier* yang terjadi dalam regresi linier berganda. Cara kerja metode ini yaitu dengan mereduksi variabel dalam model dan membentuk variabel baru yang dinamakan dengan *principal component*, dimana proses reduksi ini tidak akan menghilangkan variabel asal yang digunakan dalam penelitian ini dan proses ini disebut dengan transformasi variabel.

BAB II
KAJIAN TEORI

2.1 Regresi Linier Berganda

Regresi linier berganda merupakan metode pada regresi linier yang digunakan untuk menentukan pola hubungan dari variabel terikat Y dengan lebih dari satu variabel bebas X dan digunakan untuk mengestimasi nilai dari variabel terikat Y dengan nilai dari variabel bebas $(X_1, X_2, X_3, ..., X_k)$ yang sudah diketahui. Model persamaan regresi linier berganda dengan k variabel bebas secara umum adalah sebagai berikut: (Wohon dkk, 2017)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + \varepsilon_i, \tag{2.1}$$

dengan

Y_i : Variabel terikat (nilai yang diprediksi),

X_i : Variabel bebas ke-i, $i = 1,2,3, \dots, k$,

β_0 : Konstanta (nilai Y apabila $X_1, X_2, X_3 \dots, X_k = 0$),

β_k : Koefisien regresi variabel bebas X_{ki} (nilai peningkatan ataupun penurunan), dan

ε_i : Galat acak berdistribusi $N(0, \sigma^2)$.

2.2 Uji Asumsi Klasik

Uji asumsi klasik merupakan uji asumsi yang harus terpenuhi dalam melakukan uji regresi pada regresi linier berganda. Tujuan dari uji asumsi klasik adalah menentukan kevalidan suatu penelitian dengan data yang dihasilakan bersifat tidak bias, konsisten dan penaksiran koefisien regresinya efisian (Gujarati, 2004). Berikut adalah beberapa uji asumsi yang dilakukan pada penelitian ini.

2.2.1 Uji Normalitas

Pengujian asumsi pertama pada regresi linier adalah terpenuhinya uji normalitas, yaitu pengujian model regresi yang bertujuan untuk memeriksa adanya distribusi normal pada variabel bebas dan terikat (Supranto, 1994). Menurut Nasrum (2018), uji normalitas memiliki beberapa metode uji salah satunya adalah Kolmogorov-Smirnov yang merupakan uji pencocokan kurva (*Goodness of Fit Test*) untuk distribusi data secara umum. Berikut hipotesis uji yang digunakan dalam uji normalitas:

H_0 : Residual berdistribusi normal,

H_1 : Residual tidak berdistribusi normal,

dengan statistik uji yang didefinisikan sebagai berikut:

$$D_{hitung} = \max_{l \leq i \leq k}\left(\left|F(z_i) - F_{k_{i-l}}(X_i)\right|, \left|F(z_i) - F_{k_i}(X_i)\right|\right), \quad i = 1,2,3, \dots, k \quad (2.2)$$

dimana:

$F(z_i)$: Fungsi distribusi kumulatif teoritis ke-i (Normal Baku Z),

$F_k(X_i)$: Fungsi distribusi kumulatif data observasi ke-i.

Prinsip uji normalitas menggunakan *Kolmogrov-Smirnov* yaitu mencari simpangan terbesar (D) dari fungsi distribusi kumulatif data observasi (empiris) terhadap fungsi distribusi kumulatif teoritisnya. Kriteria pengambilan kesimpulan pada uji ini adalah apabila $D \leq D_{tabel}$ atau $P_{value} > \alpha$ maka H_0 diterima artinya sampel berdistribusi normal. Sebaliknya apabila $D > D_{tabel}$ atau $P_{value} < \alpha$ maka H_0 ditolak, artinya sampel tidak berdistribusi normal (Nasrum, 2018).

2.2.2 Heterokedastisitas

Menurut Ghozali (2013), uji heterokedastisitas merupakan uji untuk mendeteksi adanya ketidaksamaan varian dari residual satu observasi ke observasi lain. Terjadi heterokedastisitas apabila varian dari residual berbeda dan apabila varian dari residual tetap maka disebut dengan homokedastisitas. Deteksi ada tidaknya heterokedastisitas dalam model dapat dilakukan dengan melakukan uji glejser. Uji glejser dapat dilakukan dengan cara meregresikan semua variabel bebas terhadap nilai absolut residual ($|e_i|$), apabila semua variabel bebas signifikan secara statistik, maka dalam model terdapat heterokedastisitas. Berikut bentuk fungsional dari uji glejser: (Alghifari,1997)

$$|e_i| = \beta_i X_i + \varepsilon_i. \quad (2.3)$$

Hipotesis uji:

H_0 : Tidak terdapat heterokedastisitas,

H_1 : Terdapat heterokedastisitas

keputusan yang diambil untuk uji glejser adalah apabila nilai P_{value} lebih besar dari nilai tingkat kepercayaan (α) sebesar 5% atau $P_{value} > \alpha = 0{,}05$ maka terima H_0, yaitu tidak terdapat heteroedastisita pada model regresi linier berganda dan berlaku sebaliknya (Gujarati, 2004).

2.3 Metode Estimasi Parameter *Ordinary Least Square* (OLS)

Metode *Ordinary Least Square* (OLS) merupakan metode estimasi yang banyak digunakan dalam menentukan nilai-nilai estimasi pada model regresi. Tujuan utama dari metode OLS adalah untuk mendapatkan estimasi dari β dengan cara meminimumkan jumlah kuadrat residual (*error*) (Montgomery dkk, 2012). Estimasi parameter pada metode ini bisa dilakukan dengan pendekatan matriks. Bentuk matriks yang diberikan pada model persamaan (2.1) adalah:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{k1} \\ 1 & X_{12} & X_{22} & \dots & X_{k2} \\ 1 & X_{13} & X_{23} & \dots & X_{k3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \dots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_k \end{bmatrix}, \tag{2.4}$$

sehingga model ini dapat disederhanakan menjadi persamaan berikut:

$$Y = X\beta + \varepsilon, \tag{2.5}$$

dengan:

Y : Vektor kolom $n \times 1$ dari variabel terikat Y,

X : Matriks $n \times (k+1)$ dari variabel bebas X,

β : Vektor kolom $(k+1) \times 1$ dari parameter yang tak diketahui $\beta_0\beta_1, \beta_2, \beta_3, \dots, \beta_k$, dan

ε : Vektor kolom $n \times 1$ dari galat ε_i.

Misalkan sampel pada Y diberikan, maka aturan yang memungkinkan dalam pemakaian sampel untuk memperoleh estimasi dari β adalah dengan membuat $\varepsilon = Y - X\beta$ seminimum mungkin. Hasil yang diharapkan dalam aturan ini adalah memperoleh komponen sistematik yang lebih berperan dari pada komponen stokastiknya, apabila komponen stokastiknya yang lebih berperan artinya hanya diperoleh sedikit mungkin informasi tentang Y variabel terikat. Dengan kata lain, X tidak mampu menjelaskan Y (Aziz, 2010).Untuk tujuan ini diperlukan pemilihan parameter β seminimum mungkin, sehingga

$$
\begin{aligned}
S &= \varepsilon'\varepsilon \\
&= (Y - X\beta)'(Y - X\beta) \\
&= (Y' - \beta'X')(Y - X\beta) \\
&= Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta \\
&= Y'Y - (Y'X\beta)' - \beta'X'Y + \beta'X'X\beta \\
&= Y'Y - \beta'X'Y - \beta'X'Y + \beta'X'X \\
&= Y'Y - 2\beta'X'Y + \beta'X'X\beta.
\end{aligned}
\tag{2.6}
$$

Selanjutnya mencari turunan parsial S terhadap β untuk meminimumkan persamaan (2.6):

$$
\begin{aligned}
\frac{dS}{d\beta'} &= 0 - 2X'Y + X'X\beta + (\beta'X'X)' \\
&= -2X'Y + X'X\beta + X'X\beta \\
&= -2YX' + 2X'X\beta.
\end{aligned}
\tag{2.7}
$$

dan menyamadengankan persamaan (2.6) dengan nol, maka diperoleh:

$$
\begin{aligned}
2YX' - 2X'X\beta &= 0 \\
2YX' &= 2X'X\beta \\
X'Y &= X'X\beta.
\end{aligned}
\tag{2.8}
$$

Sehingga dari persamaan (2.7), estimasi param eter β secara OLS diperoleh dengan persamaan matriks sebagai berikut:

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y, \tag{2.9}$$

dengan estimator kuadrat terkecil untuk varian σ^2:

$$\hat{\sigma}_{OLS} = \frac{\hat{e}'\hat{e}}{n-k} = \frac{\left(Y - X\hat{\beta}_{OLS}\right)'\left(Y - X\hat{\beta}_{OLS}\right)}{n-k}, \tag{2.10}$$

(Aziz, 2010).

2.4 Pengujian Hipotesis dalam Regresi Linier Berganda

Menurut Waluyo (2001), pengujian hipotesis yang dilakukan pada suatu pernyataan atau suatu tanggapan berkaitan dengan parameter populasi disebut dengan hipotesis statistik. Berikut beberapa uji hipotesis yang dilakukan pada penelitian ini:

2.4.1 Uji Kesesuaian Model Regresi (Uji F)

Uji kesesuaian model merupakan uji yang digunakan dalam memastikan adanya hubungan atau tidak adanya hubungan linier antara variabel terikat Y dengan variabel bebas $X_1, X_2, \dots, X_k$ secara serentak. Hipotesis yang sesuai adalah (Montgomery dkk, 2012)

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

H_1 : $\beta_i \neq 0$, untuk setidaknya satu i ($i = 1,2,$ ...,k),

Penolakan hipotesis nol menjelaskan bahwa setidaknya satu dari regresi $X_1, X_2, \dots, X_k$ memberikan kontribusi yang signifikan terhadap model. Berikut statistik uji untuk uji F:

$$F_{hitung} = \frac{SS_R/k}{SS_{Res}/(n-k-1)} = \frac{MS_R}{MS_{Res}}, \tag{2.11}$$

keterangan:

SS_R : Jumlah kuadrat regresi,

SS_{Res} : Jumlah kuadrat residu,

n : Banyaknya observasi, dan

k : Banyaknya variabel bebas yang memengaruhi variabel terikat.

Kriteria pengambilan kesimpulan untuk uji F adalah tolak H_0 jika $F_{hitung} > F_{(\alpha;k,n-k-1)}(F_{tabel})$, artinya ada pengaruh antara variabel bebas terhadap variabel terikat dengan tingkat kesalahan tertentu. Sebaliknya terima H_0 jika $F_{hitung} < F_{tabel(\alpha;k,n-k-1)}$, artinya tidak ada pengaruh antara variabel bebas terhadap variabel terikat.

2.4.2 Uji Parsial Parameter (Uji t)

Uji parsial parameter bertujuan untuk mengetahui seberapa jauh pengaruh dari variabel bebas X_i terhadap variabel terikat Y secara individual. Berikut hipotesis untuk uji t: (Montgomery dkk, 2012)

H_0 : $\beta_i = 0$,

H_1 : $\beta_i \neq 0$,

Statistik uji untuk uji t yaitu:

$$t_{hitung} = \frac{\hat{\beta}_i}{se_{\hat{\beta}_i}}, (i = 1,2, \dots, k) \tag{2.12}$$

keterangan:

$\hat{\beta}_i$: Nilai parameter untuk β_i,

$se_{\hat{\beta}_i}$: Simpangan baku dari $\hat{\beta}_i$.

Kriteria uji untuk uji t adalah jika $|t_{hitung}| > t_{(\frac{\alpha}{2};n-k-1)}$ atau $p - value < \alpha$ maka tolak H_0, sebaliknya jika $|t_{hitung}| < t_{(\frac{\alpha}{2};n-k-1)}$ atau $p - value > \alpha$ maka terima H_0.

2.5 Uji Ukuran Kebaikan Model (*Goodness of Fit*)

Kebaikan suatu model dapat ditentukan dengan beberapa metode uji pengukuran, berikut adalah metode uji pengukuran kebaikan model pada penelitian ini:

2.5.1 *Adjusted* R^2

Untuk mengetahui ketepatan atau kecocokan garis regresi yang terbentuk dalam mewakili kelompok data hasil observasi, perlu adanya pengecekan sejauh mana model yang terbentuk mampu menerangkan kondisi yang sebenarnya yang dikenal dengan nama koefisien determinasi R^2 yang disesuaikan (*Adjusted*

R^2). Nilai dari koefisien determinasi adalah suatu ukuran yang menunjukkan besarnya pengaruh dari variabel penjelas terhadap respon. Berikut rumus dari koefisien determinasi: (Siagian & Sugiarto, 2006)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}, \tag{2.13}$$

di mana $0 \leq R^2 \leq 1$, apabila $R^2 = 0$ artinya tidak ada hubungan di antara X dan Y atau secara model regresi yang terbentuk tidak tepat untuk medefinisikan Y dan apabila nilai $R^2 = 1$ artinya garis regresi yang terbentuk dapat mendefinisikan Y secara sempurna.

Menurut Aziz (2007), di dalam koefisien determinasi R^2 terdapat koefisien determinasi tersesuaikan (*Adjusted R-Square*) dengan rumus sebagai berikut:

$$R_{adj}^2 = 1 - \frac{\sum_{i=1}^{n}\frac{(\hat{Y}_i - \bar{Y})^2}{n-k}}{\sum_{i=1}^{n}\frac{(Y_i - \bar{Y})^2}{n-1}}. \tag{2.14}$$

2.5.2 *Residual Standard Error* (RSE)

Residual Standard Error (RSE) merupakan metode estimasi simpangan baku (*standard deviation*) dari ε, di mana ε adalah selisih antara nilai yang diprediksi dengan nilai yang diamati. RSE secara umum dapat didefinisikan sebagai berikut: (Faizia dkk, 2019)

$$RSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n-k-1}}, \tag{2.15}$$

di mana MSE (*Mean Squared Error*) adalah rata-rata residual kuadrat.

2.6 Multikolinieritas

Menurut Gujarati (2004), istilah multikolinieritas pertama kali dikemukakan oleh Ragnar Frisch pada tahun 1934 yang artinya terdapatnya hubungan linier kuat di antara variabel bebas atau terdapatnya korelasi pada model regresi, terutama pada model regresi linier berganda. Adanya mutikolinieritas pada model regresi linier berganda ini menimbulkan kondisi kurang baik (*ill condition*) atau hampir singular pada matriks $X^T X$ dan mengakibatkan nilai estimasi varian bagi parameter regresi menjadi lebih tinggi, serta ada sedikit variabel bebas signifikan atau bahkan tidak signifikan.

2.6.1 Deteksi Multikolinieritas

Menurut Alghifari (1997), cara sederhana yang dilakukan dalam mendiagnosis terdapatnya multkolinieritas pada model regresi adalah sebagai berikut:

1. Melalui nilai t_{hitung}, R^2 dan F Rasio. Jika R^2 tinggi maka nilai F rasio akan tinggi, sedangkan sebagian besar koefisien regresi tidak signifikan yang artinya nilai t_{hitung} sangat rendah, maka kemungkinan terjadi kasus multikolinieritas pada model regresi.
2. Menentukan koefisien korelasi antara variabel bebas yang satu dengan variabel bebas yang lain. Jika antara dua variabel bebas memiliki korelasi spesifik (misalnya, koefisien korelasi variabel yang tinggi antara variabel bebas atau tanda koefisien korelasi variabel bebas berbeda dengan tanda koefisien regresinya), maka terjadi multikolinieritas pada model regresi.
3. Membuat persamaan regresi antar variabel bebas. Jika koefisien regresinya signifikan, maka terjadi multikolinieritas pada model regresi.

Selain dari beberapa cara di atas ada beberapa cara lain yang digunakan dalam mendeteksi adanya multikolinieritas sebagai berikut: (Montgomery dkk, 2012)

1. Deteksi adanya multikolinieritas pada model regresi dapat dilakukan dengan menggunakan nilai *Variance Inflation Factory* (VIF). *Variance Inflation Factory* (VIF) yaitu faktor perubahan variansi dalam variabel bebas ke-i. Pada VIF besarnya nilai bergantung pada nila koefisian determinasi (R^2) yang dihasilkan. Berikut persamaan yang digunakan dalam menghitung VIF:

$$VIF = \frac{1}{(1 - R_i^2)} = \frac{1}{Tolarance}, \qquad (2.16)$$

dengan R_i^2 adalah koefisien determinasi ke-i, di mana $i = 1,2,3, \dots, k$. *Tolerance* (TOL) dalam perhitungan VIF memiliki ketentuan semakin rendah nilai yang diperolehnya maka semangkin besar kemungkinan terjadinya multikolinieritas antar variabel. Batas nilai yang digunakan untuk VIF adalah antara 5 sampai 10 dan batas nilai dari *tolerance* adalah 0.1. Apabila nilai yang diperoleh $VIF \geq 10$ dan $tolerance < 0.1$ maka terjadi multikolinieritas yang kuat diatara variabel bebas dan sebaliknya.

2. Deteksi multikolinieritas berikutnya yaitu melalui koefisien korelasi. Nilai koefisien korelasi bisa diperoleh dengan menggunakan rumus perhitungan dari korelasi *pearson*, yaitu:

$$r_{x_j x_i} = \frac{n \sum x_j x_i - \sum x_j \sum x_i}{\sqrt{n \sum x_j^2 - \left(\sum x_j\right)^2} \sqrt{n \sum x_i^2 - (\sum x_i)^2}}, \qquad (2.17)$$

keterangan:

$r_{x_j x_i}$: Koefisien korelasi antara x_j dan x_i,

n : Jumlah observasi,

x_j : Variabel pertama pada observasi ke-j ($j = 1,2,3, \ldots, k$),

x_i : Variabel kedua pada observasi ke-i ($i = 1,2,3, \ldots, k$).

Apabila nilai dari koefisien korelasi yang diperoleh mendekati 0.8 maka ada multkolinieritas dalam model.

3. Pendeteksian dengan melihat *eigenvalue* (λ). Karena jumlah *eigenvalue* harusa sama dengan jumlah variabel bebas, sehingga apabila diperoleh *eigenvalue* sangat kecil atau mendekati 0.05 maka multikolinieritas terindikasi pada model. Fungsi dari *eigenvalue* yang digunakan untuk mengukur adanya multikolinieritas variabel secara keseluruhan adalah fungsi dari *conditional index* (CI), yaitu:

$$CI = \sqrt{\frac{\lambda(p-1)}{\lambda(1)}}, \tag{2.18}$$

di mana $p-1$ adalah jumlah prediktor, $\lambda(p-1)$ dan $\lambda(1)$ adalah niali minimum dan maksimum dari *eigenvalue*. Nilai CI yang diperoleh selalu lebih besar dari 1, sehingga semakin tinggi nilai CI maka multikolinieritas akan terdeteksi lebih kuat. Ketentuan untuk nilai CI adalah apabila nilai $CI < 15$ menunjukkan multikolinieritas lemah, nilai $15 < CI < 30$ menunjukkan multikolinieritas sedang dan nilai $CI > 30$ menunjukkan multikolinieritas kuat.

4. Menggunakan perbandingan pada nilai koefisian determinasi individual (r^2) dengan nilai koefisien determinasi serentak (R^2), dengan kriteria pengujiannya adalah apabila $r^2 > R^2$ maka terjadi multikolinieritas dan

sebaliknya apabila $r^2 < R^2$ maka tidak terjadi multikolinieritas pada model regresi.

2.6.2 Dampak Multikolinierita

Menurut Widarjono (2005), dampak dari terdapatnya multikolinieritas pada model regresi adalah:

1. Estimator masih tetap bersifat *Best Linier Unbiased Estimator* (BLUE), akan tetapi memiliki varian yang besar sehingga akan sulit mendapatkan estimasi yang tepat.
2. Dari dampak yang dihasilkan oleh nomer satu, maka interval estimasi akan cenderung lebih lebar dan nilai hitung statistik uji t akan kecil, sehingga membuat variabel terikat secara statistik tidak signifikan memengaruhi variabel bebas.
3. Meskipun secara individual variabel terikat tidak memengaruhi variabel bebas dengan uji t, akan tetapi nilai dari koefisien determinasi (R^2) masih tetap tinggi.

2.6.3 Pemecahan Masalah Multikolinieritas

Pemecahan masalah multikolinieritas pada model regresi bisa dilakukan dengan beberapa cara, yaitu sebagai berikut: (Alghifari, 1997)

1. Menghilangkan salah satu atau beberapa variabel yang memiliki korelasi tinggi dari model regresi.
2. Melakukan penambahan data. Cara ini akan berguna apabila dapat memastikan adanya multikolinieritas pada model regresi.
3. Melakukan transformasi variabel.

Menurut Montgomery dkk, (2012), dari pemecahan di atas terdapat beberapa metode lain yang dapat digunakan untuk menyelesaikan masalah multikolinieritas, yakni metode regresi *ridge* dan metode regresi *principal component*.

2.7 Outlier

Outlier atau pencilan dapat diartikan sebagai data yang tidak mengikuti pola pada model atau data yang keluar dari model dan tidak berada dalam daerah selang kepercayaan (Sembiring, 1995). Adanya *outlier* dapat memberikan beberapa pengaruh terhadap kegagalan asumsi regresi yaitu tidak terpenuhinya asumsi normalitas *error*, adanya masalah heteroskidastisitas, dan juga berpengaruh terhadap hasil taksiran model regresi di mana taksiran interval koefisien regresi dapat menjadi lebar. Analisis *outlier* dikelompokkan dalam suatu analisis residual dengan melihat nilai residual (e_i) atau nilai sisa pengamatan yang secara matematis didefinisikan sebagai

$$e_i = Y - \hat{Y}, \tag{2.19}$$

dengan (Y) adalah variabel dependen dan $(\hat{Y})$ adalah nilai dependen hasil taksiran model (Sembiring, 1995).

Mendeteksi *outlier* pada variabel bebas dapat dilakukan dengan melihat *error* pada model regresi. Ada beberapa langkah yang dapat digunakan untuk menyelesaikan masalah *outlier* salah satunya adalah dengan *Difference in Fit Standardized* (DFFITS). Berikut rumus perhitungan dari DFFITS:

$$(DFFITS_i) = t_i \left(\frac{h_i}{1 - h_i}\right)^{\frac{1}{2}}, \tag{2.20}$$

dengan

$$t_i = e_i \sqrt{\frac{n-p-1}{SSE(1-h_i)-e_i^2}}, \tag{2.21}$$

dimana

t_i : *Studentized deleted residual* untuk kasus ke-i,

e_i : Residual ke-i,

h_i : Elemen baris ke-i kolom ke-i dari matriks $\hat{\beta} = X(X'X)^{-1}X'$.

Dari sini sebuah data bisa dikatakan mengandung *outlier* apabila $|DFFITS| > 2\sqrt{\frac{p}{n}}$ untuk data dengan ukuran besar, di mana $p = k + 1$ adalah banyaknya ukuran data dan n adalah banyaknya jumlah observasi (Montgomery dkk, 2012).

2.8 Regresi *Robust*

Regresi *robust* merupakan salah satu metode regresi yang digunakan untuk menganalisis data yang terkontaminasi dengan *outlier*. Tujuan utama dari metode ini adalah mendeteksi *outlier* dan memberikan hasil estimasi yang resisten atau stabil dengan adanya *outlier*. Metode regresi *robust* menangani kasus *outlier* dengan cara mengabaikan identifikasi adanya *outlier* dan tidak menghilangkan data yang mengandung *outlier* didalam proses penanganannya (Chen, 2002). Estimasi yang resisten adalah estimasi yang relatif tidak terpengaruh dengan adanya perubahan yang terjadi, baik itu perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data.

Kelebihan dari metode regresi *robust* adalah tidak terlalu responsif terhadap keberadaan penyimpangan-penyimpangan yang sering terjadi pada asumsi ideal dibandingkan dengan metode *least square*, sehingga penggunaannya merupakan pengganti dari metode *least square* (Draper & Smith, 1992). Berikut bentuk persamaan estimasi parameter dari regresi *robust*:

$$\hat{\beta}_{robust} = (X'WX)^{-1}X'WY, \tag{2.22}$$

di mana

Y : Variabel tak bebas,

X : Variabel bebas,

$\hat{\beta}_{robust}$: Koefisien regresi *robust* ($i = 1,2,3, \dots, k$), dan

W : Matriks pembobot.

2.9 *Principal Component Analysis* (PCA)

Principal Component Analysis (PCA) pertama kali diperkenalkan oleh Karl Pearson pada tahun 1901, dikembangkan oleh Harold Hotelling pada tahun 1933 dan Rao pada tahun 1964 yang hingga saat ini pengembangan PCA masih terus dilakukan (Draper & Smith, 1992). *Principal Component Analysis* adalah suatu teknik analisis statistik untuk mentransformasi variabel-variabel asal yang saling berkorelasi satu dengan yang lainnya menjadi satu set variabel baru agar terbebas dari korelasi. Secara umum tujuan dari PCA adalah mereduksi dimensi data. *Principal Component Analysis* muncul sebagai solusi bagi proses pengumpulan data yang terdiri dari variabel dengan jumlah banyak dan akan diperoleh variabel baru dengan jumlah lebih sedikit namun tetap mampu menjelaskan keragaman data (Richard & Wichern, 2007).

Menurut Larasati, dkk (2020), proses penyelesaian masalah multikolinieritas pada model regresi dengan metode PCA terdiri dari dua langkah, yaitu *principal component* dibentuk dengan menggunakan vektor *eigen* dari matriks varian kovarian dan diregresikan terhadap variabel terikat Y dengan metode *least square*, langkah kedua adalah dengan menggunakan matriks korelasi. *Principal component* merupakan suatu kombinasi linier berdasarkan pada skala pengukuran variabel acak $X_1, X_2, \dots, X_k$ yang sama dan memiliki struktur matriks varian kovarian Σ yang kemudian akan dihasilkan nilai *eigen* λ sebanyak q ($\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_k \geq 0$). Dari variabel acak tersebut dibentuk variabel baru H_d sebagai *principal component* dengan kombinasi linier dari X_j.

Sebelum persamaan kombinasi linier untuk H_d diperoleh, maka terlebih dahulu dibentuk varian kovarian untuk *principal component* H_d dari variabel asal X_j, yaitu:

$$\begin{aligned} Var(H_d) &= E\left[\left(H_d - E(H_d)\right)\left(H_d - E(H_d)\right)'\right] \\ &= E\left[\left(m_d'X - E(m_d'X)\right)\left(m_d'X - E(m_d'X)\right)'\right] \\ &= E\left[\left(m_d'X - m_d'E(X)\right)\left(m_d'X - m_d'E(X)\right)'\right] \\ &= E[(m_d'X - m_d'\mu)(m_d'X - m_d'\mu)'] \\ &= E[(m_d'X - m_d'\mu)(X'm_d - \mu'm_d)] \\ &= m_d'E[(X-\mu)(X-\mu)']m_d \\ &= m_d'\Sigma m_d = \lambda_d \end{aligned} \quad . \qquad (2.23)$$

dan

$$Cov(H_d, H_j) = m_d'\Sigma \mathrm{m}_j, \qquad d \neq j \qquad (2.24)$$

dengan $d = 1,2,3, \dots, k$. Dimana λ_d merupakan nilai *eigen* ke-d dari matriks varian kovarian Σ dan m_d merupakan vektor *eigen* ke-d dari matriks varian kovarian Σ yang berpadanan dengan nilai *eigen* λ. *Principal component* ke-d yang merupakan kombinasi linier dari variabel asal bertujuan memaksimumkan varian dari variabel H_d, membebaskan dari korelasi antara *principal component* yang satu dengan yang lain serta bersifat orthogonal, sehingga H_d harus memenuhi syarat batasan $m_d'\mathrm{m}_j =$

0 dan $m'_d m_d = 1$. Dari varian kovarian *principal component* yang sudah terbentuk maka kombinasi linier untuk H_d yaitu:

$$H_d = m'_d X = m_{1d} X_1 + m_{2d} X_2 + \cdots + m_{kd} X_k, \qquad (2.25)$$

dengan bentuk penjabarannya sebagai berikut:

$$\begin{aligned} H_1 &= m'_1 X = m_{11} X_1 + m_{12} X_2 + \cdots + m_{1k} X_k \\ H_2 &= m'_2 X = m_{21} X_1 + m_{22} X_2 + \cdots + m_{2k} X_k \\ &\vdots \\ H_k &= m'_k X = m_{k1} X_1 + m_{k2} X_2 + \cdots + m_{kk} X_k \end{aligned} .$$

Menurut Sanusi dan Dewi (2019), λ ke-d merupakan akar ciri yang diperoleh dari persamaan $|\Sigma - \lambda I| = 0$ dengan $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_k \geq 0$ dan vektor ciri m_d sebagai pembobot dari transformasi linier variabel asal yang diperoleh dari $(\Sigma - \lambda_d I) m_d = 0$. Proporsi dari total varian populasi yang disumbangkan oleh *principal component* ke-d adalah

$$\frac{\lambda_d}{\lambda_1 + \lambda_2 + \lambda_3 + \cdots + \lambda_k} \times 100\%, \qquad d = 1,2,3, \ldots, k. \qquad (2.26)$$

Gasperz (1995) dalam Safidah (2014), menyatakan bahwa *principal component* yang terbentuk dengan matriks varian kovarian akan memiliki susunan yang bergantung pada ukuran satuan, di mana ukuran satuan ini digunakna untuk mengukur variabel-variabel bebas dan pengukuran untuk variabel-variabel bebas dengan skala yang belum sama. Sehingga permasalahan ini diselesaikan dengan pembentukan *principal component* menggunakan matriks korelasi, yaitu mentransformasikan variabel asal X ke variabel baku C. Berikut adalah rumus yang digunakan untuk pambakuan variabel:

$$\begin{aligned} C_1 &= \frac{(X_1 - \bar{X}_1)}{\sqrt{\sigma_{11}}} \\ C_2 &= \frac{(X_2 - \bar{X}_2)}{\sqrt{\sigma_{22}}}, \\ &\vdots \\ C_k &= \frac{(X_k - \bar{X}_k)}{\sqrt{\sigma_{kk}}} \end{aligned} \tag{2.27}$$

dengan matriks kovarian C berikut:

$$\begin{aligned} Cov\,(C) &= \left(W^{\frac{1}{2}}\right)^{-1} \Sigma \left(W^{\frac{1}{2}}\right)^{-1}. \\ &= \rho \end{aligned} \tag{2.28}$$

Principal component yang terbentuk sebagai kombinasi linier dari variabel yang dibekukan (C_d) adalah:

$$H_d = m'_d C = m_{1d} C_1 + m_{2d} C_2 + \cdots + m_{kd} C_q, \tag{2.29}$$

Tulak, dkk (2017), menyatakan bahwa secara singkat langkah-langkah yang diterapkan dalam menggunakan metode PCA sebagai berikut:

1. Menghitung matriks varian kovarian S.
2. Dari matriks varian kovarian sampel akan diperoleh nilai *eigen*, yaitu $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_k$ di mana $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0$.
3. Menghitung nilai *eigen* dari matriks varian kovarian Σ dengan rumus
$$|\Sigma - \lambda I| = 0. \tag{2.30}$$
4. Menghitung vaktor *eigen* dari matriks varian kovarian Σ dengan rumus
$$(\Sigma - \lambda I) H_d = 0. \tag{2.31}$$
5. Menormalisasikan vektor *eigen* dengan rumus
$$H_d = \frac{m}{\sqrt{m'm}}. \tag{2.32}$$
6. Membentuk *principal component* ke-$1,2,3, \dots, d$ dengan memanfaatkan persamaan (2.29).

7. Menghitung proporsi varian dengan rumus (2.26).

Sedangkan untuk kriteria pemilihan *principel component*, sebagai berikut:

1. Pemilihan komponen-komponen utama yang memiliki proporsi kumulatif keragaman data asal dengan total minimal 80%.
2. *Principal component* terpilih adalah komponen-komponen utama yang memiliki nilai *eigen* $\lambda_d > 1$ dan apabila nilai *eigen* $\lambda_d < 1$ maka komponen-kompenen tersebut tidak signifikan atau bisa diabaikan, di mana pemilihan *principal component* diperoleh dari matriks korelasi.
3. Pemeriksaan *scree plot*. *Scree plot* merupakan plot antara nilai *eigen* λ_d dengan *principal component* yang terbentuk, di mana pemilihan *principal component* dengan *scree plot* ini memiliki sifat grafis atau visual.

2.9.1 Uji Asumsi *Principal Component Analysis* (PCA)

Uji asumsi metode PCA bertujuan untuk mengetahui kelayakan suatu data sampel dalam proses analisis dan dapat digunakan untuk analisis lebih lanjut, uji asumsi pada metode PCA ini disebut juga dengan uji asumsi kecukupan data. Berikut uji asumsi dalam metode PCA:

a. Uji *Kaiser Meyer Olkin* (KMO)

Uji KMO merupakan uji yang membandingkan besarnya koefisien korelasi yang diteliti dengan besarnya koefisien parsial. Nilai yang diperoleh dari uji KMO diharuskan lebih besar dari 0,05 agar data bias dianalisis lebih lanjut dengan menggunakan metode analisis komponen utama atau *principal component analysis* (PCA) (Verdian, 2019). Berikut adalah hipotesis untuk uji KMO: (Simamora, 2005)

H_0 : Data cukup untuk dianalisis dengan *principal component*,

H_1 : Data tidak cukup untuk dianalisis dengan *principal component*,

dengan statistik uji:

$$KMO = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} r_{ij}^2}{\sum_{i=1}^{p}\sum_{j=1}^{p} r_{ij}^2 + \sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}^2}, \tag{2.33}$$

untuk $i = 1,2,3, \dots, p$ dan $j = 1,2,3, \dots, p$.

b. Uji Bartlett

Menurut Marisson (1990), uji dalam mengetahui adanya hubungan atau korelasi antar variabel dapat dilakukan dengan menguji independensi (kebebasan) antar variabel dengan menggunakan metode *bartlett sphericity*. Variabel dikatakan memiliki sifat independensi atau saling bebas apabila matriks korelasi antar variabel membentuk matriks identitas. Berikut hipotesisi yang doambil untuk uji *bartlett sphericity*:

H_0 : $\mathrm{R} = \mathrm{I}$ (Variabel independen tidak berkorelasi),

H_1 : $\mathrm{R} \neq \mathrm{I}$ (Variabel independen berkorelasi),

dengan statistik uji:

$$\chi^2 = \left\{n - 1 - \frac{2p+5}{6}\right\} \ln|\mathrm{R}, \tag{2.34}$$

dimana:

n : Julah observasi,

p : Jumlah variabel independen,

R : Matriks korelasi dari masing-masing variabel independent.

Keputusan yang diambil untuk uji *bartlett sphericity* ini yaitu, apabila nilai signifikan kurang dari nilai α atau apabila nilai $\chi^2_{hitung} > \chi^2_{(\alpha, \frac{1}{2}p(p-1))}$ maka

H_0 ditolak, sehingga dapat disimpulkan bahwa terdapat korelasi antar variabel independen.

2.9.2 *Principal Component Regression* (PCR)

Metode *principal component analysis* (PCR) merupakan suatu metode analisis yang mengkombinasikan antara analisis regresi dengan *principal component analysis* (PCA), di mana analisis regresi sebagai penentu ada tidaknya hubungan antara variabel terikat dengan variabel bebas dan PCA sebagai penyederhana variabel pengamatan dengan cara mereduksi dimensinya (Supriyadi, 2017). Metode PCR difungsikan untuk meminimumkan masalah multikolinieritas yang terjadi di dalam model regresi linier berganda dengan cara tidak menghilangkan variabel bebas yang bersinggungan dengan kolinieritas.

Menurut Nelwin dkk, (2019), proses *principal component regression* dilakukan setelah analisis model menggunankan metode *principant component analysis* secara sempuran mendapatkan nilai *principal component*. *Principal component* terpilih diregresikan bersama variabel terikat menggunakan metode *ordinary least square* (OLS). Metode *principal component regression* menghasilkan model persamaan yang sama dengan model regresi pada umumnya, berikut persamaannya:

$$\hat{Y} = \beta_0 + \beta_1 H_{1d} + \beta_2 H_{2d} + \cdots + \beta_d H_{kd} + \varepsilon, \tag{2.35}$$

dengan notasi matriks dari persamaan (2.35) berikut:

$$Y = H\beta + \varepsilon, \tag{2.36}$$

dengan matriks

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_d \end{bmatrix} = \begin{bmatrix} 1 & H_{11} & H_{12} & \cdots & H_{1d} \\ 1 & H_{21} & H_{22} & \cdots & H_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & H_{d1} & H_{d2} & \cdots & H_{kd} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_d \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_d \end{bmatrix},$$

dan estimasi parameter dengan metode *Ordinary Least Square* (OLS) terhadap $\hat{\beta}$ sebagai berikut:

$$\hat{\beta} = (H'H)^{-1}H'Y. \tag{2.37}$$

Bentuk matriks varian kovarian untuk $\widehat{m}$ adalah

$$Var(\hat{\beta}) = \sigma^2(H'H)^{-1}. \tag{2.38}$$

2.10 *Regresi Robust Principal Component Analysis* (ROBPCA)

Menurut Larasati dkk, (2020), metode regresi *robust principal component analysis* (ROBPCA) adalah sebagai metode pengembangan dari metode PCA (*principal component analysis*) yang dikembangkan oleh Hubert dkk, (2005) dan merupakan suatu metode yang *robust* terhadap keberadaan *outlier* pada data pengamatan. Keberadaan *outlier* ini sangat sensitif untuk metode PCA dalam menyelesaikan masalah multikolinieritas pada model regresi linier berganda. Metode ROBPCA dalam mengatasi masalah multikolieriatas yaitu dengan menggunakan matriks varian kovarian *robust* pada PCA dan metode regresi *robust* pada tahap analisis regresi.

Metode ROBPCA diterapkan pada data penelitian yang memiliki jumlah variabel $d < q$ jumlah observasi atau suatu data dengan jumlah variabel $d > q$ jumlah observasi. Ketika variabel $d < q$, maka dilakukan dengan dekomposisi nilai singular (Sanusi dan S., 2019).

2.10.1 Metode *Minimum Covariance Determinant* (MCD)

Metode *Minimum Covariance Determinant* (MCD) adalah metode estimasi yang *robust* untuk rata-rata dan matriks varian kovarian dari sebagian data pengamatan yang memiliki kovarian minimum. Metode MCD bersifat resisten terhadap keberadaan *outlier* di dalam data pengamatan, sehingga sangat berguna untuk mendeteksi *outlier*. Tujuan dari metode MCD adalah menemukan g pengamatan dari q data pengamatan dengan matriks kovariannya memiliki determinan terkecil. Metode MCD dapat dihitung apabila $g > q$, jika kebalikannya maka matriks kovarian dari setiap subset g memiliki determinan 0 dan untuk menyelesaikannya setidaknya dibutuhkan $q > 2q$ (Hubert dkk, 2005).

Menurut Larasati, dkk (2021), nilai standar dari g adalah

$$g = \left[\frac{(q + k + 1)}{2}\right]. \tag{2.39}$$

Metode estimasi MCD mudah dihitung dan ditemukan apabila jumlah q pengamatan kecil. Apabila jumlah q pengamatan besar maka akan banyak kombinasi subsampel dari g yang akan dicari dan ditemukan. Oleh karena itu, metode MCD menggunakan langkah-langkah perhitungan dari metode FAST-MCD, di mana persamaan yang digunakan dalam metode ini adalah.

$$\bar{X}_{MCD} = \frac{1}{g}\sum_{i \in g}^{g} x_i\,, \tag{2.40}$$

dan

$$S_{MCD} = \frac{1}{g}\sum_{i \in g}^{g} [x_i - \bar{X}_{MCD}][x_i - \bar{X}_{MCD}]^T\,, \tag{2.41}$$

dengan

g : Elemen pengamatan,

$\bar{X}_{MCD}$: Vektor rata-rata MCD,

S_{MCD} : Matriks varian kovarian MCD.

Apabila $\det(S_1) = 0$ maka perhitungan berhenti dan apabila $\det(S_1) \neq 0$ maka proses perhitungan dapat dilanjutkan dengan menghitung jarak *robust*, yaitu *robust distance* (RD) untuk setiap pengamatan. Selanjutnya mengurutkan dari yang terkecil hingga terbesar menggunakan persamaan RD sebagai berikut: (Larasati dkk, 2020)

$$RD_i = \sqrt{(x_i - \bar{X}_{MCD})^T S_{MCD}^{-1} (x_i - \bar{X}_{MCD})}, \tag{2.42}$$

dan untuk subsampel-subsampel seterusnya akan diambil sebanyak h pengamatan dengan jarak terkecil hingga mencapai konvergen $(S_{i+1}) = (S_1)$. Proses perhitungan berikutnya adalah memilih himpunan yang memiliki determinan S_{MCD} terkecil, serta mencari $\bar{X}_{MCD}$ dan S_{MCD} dari himpunan H terpilih.

2.10.2 *Weighted Least Square* (WLS)

Metode *Weighted Least Square* (WLS) atau kuadrat terkecil berbobot merupakan salah satu metode yang digunakan untuk mengatasi heterokedastisitas, yaitu terdapatnya varian *error* tidak konstan pada model regresi linier berganda. Metode WLS memiliki prinsip yang sama dengan metode *Ordinary Least Square* (OLS) karena merupakan perkembangan dari metode OLS sendiri. Perbedaan ada pada metode WLS dengan adanya penambahan variable baru berupa variabel w, dimana w adalah pembobot. Fungsi untuk WLS adalah: (Montgomery dkk, 2012)

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} w_i \,(Y_i - \beta_0 - \beta_1 X_i)^2, \tag{2.43}$$

persamaan normal untuk *least square* diperoleh sebagai berikut:

- Model untuk regresi linier sederhana dengan pembobot:

$$\sum_{i=1}^{n} w_i \, Y_i = \hat{\beta}_0 \sum_{i=1}^{n} w_i + \hat{\beta}_1 \sum_{i=1}^{n} w_i \, X_i^2. \tag{2.44}$$

- Model untuk regresi linier berganda dengan pembobot:

$$\sum_{i=1}^{n} w_i \, X_i Y_i = \beta_0 \sum_{i=1}^{n} w_i \, X_i + \hat{\beta}_1 \sum_{i=1}^{n} w_i \, X_i^2 + \hat{\beta}_1 \sum_{i=1}^{n} w_i \, X_i^2 \ldots + \hat{\beta}_1 \sum_{i=1}^{n} w_i \, X_i^2. \tag{2.45}$$

Sedangkan persamaan normal untuk WLS adalah

$$X' w X \hat{\beta} = X' w Y, \tag{2.46}$$

model persamaan estimasi metode WLS adalah

$$\hat{\beta} = (X' w X)^{-1} X' w Y. \tag{2.47}$$

Dari Munawwir (2008), untuk estimasi awal metode WLS didapatkan dari parameter $\left(\hat{\beta}^{(1)}\right)$ dan galat pertama yang distandarisasikan $\left(\varepsilon_i^{(1)}\right)$ untuk metode WLS. Nilai pembobot awal untuk metode WLS yaitu $w_i^{(1)} = w(\varepsilon_i^{(0)})$, dimana $\left(\varepsilon_i^{(0)}\right)$ merupakan galat yang distandarisasi dari metode OLS. Dengan $w_i^{(1)}$ sebagai fungsi pembobot maka untuk persamaan (2.43) untuk regresi linier berganda dapat ditulis ulang menjadi

$$\sum_{i=1}^{n} w_i^{(1)} \, X_i Y_i = \beta_0 \sum_{i=1}^{n} w_i^{(i)} \, X_i + \hat{\beta}_1 \sum_{i=1}^{n} w_i^{(i)} \, X_i^2. \tag{2.48}$$

Untuk persamaan normal WLS pada persamaan (2.44) menjadi

$$X' w_i^{(1)} X \hat{\beta} = X' w_i^{(1)} Y, \tag{2.49}$$

persamaan (2.47) ini merupakan persamaan matriks, dengan elemen diagonal $w_1, w_2, \dots, w_n$ maka bentuk matriks diagonal berukuran $n \times n$ dapat ditulis sebagai berikut:

$$w^{(q)} = \begin{pmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ 0 & 0 & w_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & w_n \end{pmatrix}. \tag{2.50}$$

Untuk persamaan estimasi metode WLS pada persamaan (2.45) dapat ditulis ulang menjadi

$$\hat{\beta}^{(1)} = \left(X' w_i^{(1)} X\right)^{-1} X' w_i^{(1)} Y, \tag{2.51}$$

dengan $w_i^{(1)}$ fungsi pembobot. Sedangkan untuk iterasi berikutnya diperoleh dengan cara yang sama, yaitu dengan hasil sebagai berikut:

$$\hat{\beta}^{(2)} = \left(X' w_i^{(2)} X\right)^{-1} X' w_i^{(2)} Y,$$

$$\hat{\beta}^{(3)} = \left(X' w_i^{(3)} X\right)^{-1} X' w_i^{(3)} Y,$$

$$\vdots$$

$$\hat{\beta}^{(q)} = \left(X' w_i^{(q)} X\right)^{-1} X' w_i^{(q)} Y.$$

dimana $\hat{\beta}^{(q)}$ adalah matriks estimator berukuran $n \times 1$ untuk iterasi ke-q dan $w_i^{(q)}$ adalah matriks fungsi pembobot pada iterasi ke-q.

2.10.3 *M-estimator*

Metode estimasi *M-estimator* (*maximum-likelihood type*) merupakan metode estimasi yang dikemukakan oleh Hubert pada tahun 1973 dan merupakan metode estimasi paling sederhana di antara metode estimasi pada regresi *robust* yang lain. Pada prinsipnya *M-estimator* merupakan estimasi yang meminimumkan suatu fungsi sisaan ρ dengan persamaan berikut: (Montgomery dkk, 2012)

$$min_{\beta} \sum_{i=1}^{n} \rho(e_i) = min_{\beta} \sum_{i=1}^{n} \rho\left(y_i - \sum_{i=1}^{n} X_{iq}\beta_q\right). \tag{2.52}$$

Persamaan (2.55) merupakan persamaan dari skala invariant yang akan diperoleh dengan melakukan penstandarisasi terlebih dahulu terhadap sebuah skala estimasi *robust* $\hat{\sigma}$, sehingga persamaannya menjadi

$$min \sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}}\right) = min \sum_{i=1}^{n} \rho\left[\frac{y_i - \sum_{q=0}^{n} X_{iq}\beta_q}{\hat{\sigma}}\right], \tag{2.53}$$

di mana pemilihan skala estimasi *robust* $\hat{\sigma}$ menggunakan rumus dari *Median Absolute Deviation* (MAD) sebagai berikut:

$$\hat{\sigma} = \frac{MAD}{0{,}6745} = \frac{median|e_i - median(e_i)|}{0{,}6745}, \tag{2.54}$$

dengan konstanta $0{,}6745$ skala estimasi *robust* $\hat{\sigma}$ mendekati tak bias jika n besar dan residu berdistribusi normal.

Menurut Christophe, dkk (2000), untuk *M-estimator* secara implisit didefinisikan sebagai berikut: (Croux & Haesbroeck, 2000)

$$t_n = \sum_{i=1}^{n} w_1\{d(x_i, t_n, G_n)\}x_n\,, \tag{2.55}$$

$$G_n = \frac{1}{n}\sum_{i=1}^{n} w_2\{d^2(x_i, t_n, G_n)\}(x_i - t_n)(x_i - t_n)^T\,, \tag{2.56}$$

di mana $w_1$dan w_2 sebagai fungsi pembobot, sedangkan jarak statistik antara x_i dan t diukur dalam matriks yang diinduksi oleh matriks *positive definite* G.

Menurut Fox (2011), metode estimasi parameter dengan *M-estimator* dilakukan menggunakan estimasi iterasi yang disebut dengan IRLS (*iteratively reweighted least square*), berikut langkah-langkahnya:

1. Mengetimasi parameter regresi (β) menggunakan metode *Ordinary Least Square* (OLS) dan akan diperoleh nilai residual e_i.
2. Menghitung nilai skala estimasi *robust* $\hat{\sigma}$:

$$\hat{\sigma} = \frac{MAD}{0{,}6745} = \frac{median|e_i - median(e_i)|}{0{,}6745}, \tag{2.57}$$

3. Menghitung pembobot dengan $\psi(e_i^*)$ sebagai fungsi pembobot *Hubert*, persamaan dapat didefinisikan sebagai berikut:

$$w_{(e_i)} = \frac{\psi(e_i^*)}{e_i^*}, \tag{2.58}$$

di mana nilai $e_i^* = \frac{e_i}{\hat{\sigma}}$. Fungsi pembobot *Hubert* memiliki kriteria sebagai berikut:

$$w(e_i^*) = \begin{cases} 1, & |e_i^*| \leq c \\ \dfrac{c}{e_i^*}, & |e_i^*| > c \end{cases}, \tag{2.59}$$

dengan e_i adalah residual ke-i dan nilai c (*tuning constant*) dalam regresi *robust* yaitu menentukan kuatnya *estimator* terhadap *outlier* sebagai efisiensi *estimator* dalam ketiadaan *outlier*. Apabila diambil $\alpha = 5\%$, maka *M-estimator* dengan fungsi *Hubert* menggunakan niai $c = 1{,}345$.

4. Menyusun matriks pembobot berupa matriks diagonal berukuran $n \times n$ dengan elemen diagonal $w_1; w_2; \dots; w_n$.
5. Menghitung estimasi parameter regresi *robust* untuk $l + 1$ iterasi berikut:

$$\hat{\beta}_{robust} = (X'w_lX)^{-1}X'w_lY. \tag{2.60}$$

Mengulang langkah 2 sampai langkah 5 sampai diperoleh $\hat{\beta}_{robust}$ yang konvergen, artinya perubahan antara $\hat{\beta}_{robust}$ ke-$l+1$ dan $\hat{\beta}_{robust}$ ke-l lebih kecil dari 0,1%.

2.11 Kemiskinan dan Faktor- Faktor yang Memengaruhi

Kemiskinan merupakan ketidakmampuan manusia dalam memenuhi kebutuhan hidupnya baik untuk diri sendiri maupun keluarganya. Dalam mengukur kemiskinan BPS menggunakan konsep dari kemampuan memenuhi kebutuhan dasar, yang mana dengan pendekatan ini kemiskinan dapat dipandang sebagai ketidakmampuan dari sisi ekonomi dalam memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran (BPS, 2018). Terdapat banyak faktor yang mempengaruhi peningkatan jumlah kemiskinan di setiap tahunnya, seperti pada faktor kepadatan penduduk, tingkat pengangguran terbuka, produk domestik regional bruto, indeks keparahan kemiskinan, rata-rata lama sekolah, indeks pembangunan manusia dan pengeluaran per kapita.

Berikut beberapa faktor yang dapat digunakan untuk mengukur tingkat kemiskinan yang terjadi di Provinsi Jawa Timur pada tahun 2019:

1. Kepadatan Penduduk

   Data kependudukan merupakan informasi penting dalam proses pembangunan dalam kesejahteraan suatu daerah, karena penduduk berperan sebagai subyek sekaligus obyek dalam pembangunan. Bertambahnya jumlah penduduk akan mengakibatkan konsekuensi dalam penyediaan beberapa kebutuhan menjadi bertambah, seperti halnya pada kebutuhan pangan, perumahan, fasilitas kesehatan dan seluruh kebutuhan juga harus disediakan lebih banyak lagi atau bisa dikatakan akan meningkat dua kali lipat dari sebelum-sebelumnya. Dari sini bisa disimpulkan bahwa kepadatan penduduk memiliki pengaruh yang kuat dalam proses pengamatan tingkat persentase kemiskinan di setiap periodenya (BPS, 2018).

2. Tingkat Pengangguran Terbuka

   Tingkat pengangguran terbuka merupakan persentase jumlah pengangguran terhadap jumlah angkatan kerja. Besaran tingkat pengangguran terbuka ini sebagai cerminan untuk tingkat kemakmuran dan kesejahteraan masyarakat dengan alasan pengangguran merupakan salah satu penyebab adanya hambatan dalam mencapai kemakmuran masyarakat. (BPS, 2018)

3. Produk Domestik Regional Bruto (PDRB)

   Produk domestik regional bruto (PDRB) merupakan nilai tambahan yang diterima oleh faktor-faktor produksi yang ikut serta dalam proses produksi di suatu daerah dalam jangka waktu tertentu (BPS, 2021)

4. Indeks Keparahan Kemiskinan (IKK)

   Indeks keparahan kemiskinan (IKK) merupakan gambaran mengenai penyebaran pengeluaran antar penduduk miskin, dimana semakin tinggi angka dari IKK maka akan semakin tinggi pula angka dari ketimpangan pengeluaran antar penduduk miskin (BPS, 2021).

5. Rata-rata Lama Sekolah

   Rata-rata lama sekolah merupakan rata-rata jumlah tahun yang dihabiskan oleh penduduk dengan usia 15 tahun ke atas dalam menempuh semua jenis pendidikan yang pernah dijalani. Bagi mereka yang tamat SD diperhitungkan lama sekolah selama enam tahun, tamat SMP diperhitungkan selama Sembilan tahun dan tamat SMA diperhitungkan selama 12 tahun tanpa memperhitungkan apakah pernah tinggal kelas atau tidak (BPS, 2018).

6. Indeks Pembangunan Manusia (IPM)

   Indeks pembangunan manusia (IPM) pertama kali diperkenalkan oleh *United Nation Development Programme* (UNDP) pada tahun 1990 dan merupakan salah satu indikator penting yang digunakan untuk mengukur keberhasilan suatu upaya membagun kualitas kehidupan manusia baik berupa masyarakat atau penduduk. Selain itu IPM dapat menentukan peringkat pembangunan suatu wilayah atau negara baik itu tingkat provinsi, kota atau kabupaten. IPM dapat menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan dan yang lainnya (BPS, 2021).

7. Pengeluaran Per Kapita

   Pengeluaran per kapita merupakan biaya yang dikeluarkan sebagai konsumsi semua anggota rumah tangga selama sebulan baik itu berasal dari proses pembelian maupun produksi sendiri dan dibagi dengan banyaknya anggota rumah tangga dalam rumah tangga tersebut (BPS, 2021).

2.12 Kajian Teori dengan Teori Pendukung

Susilowati, dkk., (2020) menjelaskan tentang metode *Robust Principal Component Analysis* (ROBPCA) dan metode *Clustering Large Area* (Clara) dalam menyelesaikan *outlier* pada sebuah data tentang indeks kebahagiaan dunia tahun 2018. Dengan menggunakan metode ROBPCA diperoleh *principal component* sebanyak 3 komponen yang dapat menjelaskan sebesar 92,6% dari total variansi data, sedangkan untuk metode PCA didapat sebanyak *principal component* yang dapat menjelaskan sebesar 85,89% dari total varians. Dengan hasil demikian maka dapat ditentukan bahwasanya metode ROBPCA lebih efisien dalam menyelesaikan kasus outlier pada data dari pada metode PCA biasa.

Larasati, dkk., (2020) telah melakukan penelitian dengan tema analisis regresi komponen utama *robust* dengan menggunakan dua metode estimasi berbeda, yaitu metode *Minimum Covariance Determinant* (MCD) dan *Least Trimmed Square* (LTS). Proses analisis pada metode analisis komponen utama *robust* dapat dilakukan dengan menggunakan persamaan matriks dari matriks kovarian *robust* MCD dan metode regresi *robust* LTS, dimana estimasi nilai koefisien RKU *robust* yang digunakan adalah

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{h} y_i - \beta_1 \sum_{i=1}^{h} Q_i}{h}, \tag{2.61}$$

$$\hat{\beta}_1 = \frac{h \sum_{i=1}^{h} y_i Q_i - \beta_1 \sum_{i=1}^{h} y_i \sum_{i=1}^{h} Q_i}{h}. \tag{2.62}$$

Hasil simulasi yang diperoleh yaitu pada sampel dengan ukuran persentase pencilan sebesar 5%-25%, niali rata-rata bias dan nilai MSE pada metode RKU klasik selalu naik secara konstan dalam setiap penambahan persentase pencilan dan pada metode RKU *robust* dihasilkan nilai rata-rata bias yang lebih kecil dari RKU klasik. Hal ini menunjukkan bahwa metode RKU *robust* menggunakan MCD-LTS merupakan metode yang kekar dan efektif dalam menyelesaikan multikolinieritas juga outlier dibandingkan dengan RKU klasik yang sangat sensitif terhadap keberadaan outlier.

Nelwin, dkk., (2019) telah melakukan penelitian dengan tema yang diambil berupa regresi komponen utama *robust* dengan menggunakan metode estimasi S dalam menganalisis pengaruh jumlah pengangguran di Jawa Timur. Dalam melihat efektifitas keberhasilan metode RKU *robust* menyelesaikan kasus outlier pada data, peneliti membandingakan empat model yang diperoleh dari kombinas dua metode berbeda. Hasil perbandingan tersebut yaitu model RKU klasik dengan OLS menghasilakan R^2 disesuaikan sebesar 0,5143 dan 0,6869 untuk nilai RSE, model RKU *robust* dengan OLS menghasilkan R^2 disesuaikan sebesar 0,5720 dan 0,6440 untuk nilai RSE, model RKU klasik dengan estimasi S menghasilkan R^2 disesuaikan sebesar 0,9135 dan 0,4141 untuk nilai RSE dan model RKU *robust* dengan estimasi S menghasilkan R^2 disesuaikan sebesar 0,9165 dan 0,4073 untuk nilai RSE. Jadi dari penjabaran hasil perbandingan tersebut diperoleh model terbaik yaitu model RKU *robust* dengan estimasi S yang memiliki nilai R^2 disesuaikan tertinggi dan nilai RSE terendah.

Safidah (2014) telah menentukan proses pembentukan *principal component* dalam penelitiannya dengan tema analisis dekomposisi spektral dengan metode *Principal Component Analysis*. Pembentukan *principal component* merupakan proses dasar di dalam metode regresi *robust principal component analysis* untuk menyelesaikan masalah multikolinieritas dan *outlier*, di mana *principal component* merupakan kombinasi linier yang didasarkan pada skala pengukuran dari variabel acak X_k. Terdapat dua langkah dalam menentukan *principal component*, yaitu:

1. Menggunakan matriks varian kovarian, merupakan matriks dengan varian antar variabel di dalamnya. Matriks varian kovarian dapat ditentukan berdasarkan model dari persamaan vektor *eigen* pada persamaan (2.28), dalam persamaan tersebut terdapat matriks varian kovarian, nilai *eigen* dan vektor *eigen* yang harus ditentukan terlebih dahulu. Pertama-tama yaitu

memisalakan sebuah matriks berukuran $n \times n$ di dalam matriks pemisalan X yang diberikan untuk menentukan *principal component*:

$$X = \begin{bmatrix} X_{11} & X_{12} & X_{13} & \cdots & X_{1n} \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2n} \\ X_{31} & X_{32} & X_{33} & \cdots & X_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & X_{n3} & \cdots & X_{nn} \end{bmatrix}. \tag{2.63}$$

Sebelum mencari nilai dari vektor rata-rata, nilai rata-rata dari setiap kolom berupa $\bar{x}_d$ harus dicari terlebih dahulu, yaitu

$\bar{x}_1 = \frac{X_{11}+X_{21}+X_{31}+X_{41}+\cdots+X_{n1}}{n}$,

$\bar{x}_2 = \frac{X_{12}+X_{22}+X_{32}+X_{42}+\cdots+X_{n2}}{n}$,

$\vdots$

$\bar{x}_n = \frac{X_{1n}+X_{2n}+X_{3n}+X_{4n}+\cdots+X_{nn}}{n}$,

dan diperoleh persamaan sebagai berikut:

$$\bar{x} = [\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \dots, \bar{x}_n] = \frac{1}{n} X^T 1, \tag{2.64}$$

sehingga:

$$\begin{aligned} \bar{x} &= \frac{1}{n} \begin{bmatrix} X_{11} & X_{12} & X_{13} & \cdots & X_{1n} \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2n} \\ X_{31} & X_{32} & X_{33} & \cdots & X_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & X_{n3} & \cdots & X_{nn} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \frac{1}{n} \begin{bmatrix} X_{11} + X_{21} + X_{31} + \cdots + X_{n1} \\ X_{12} + X_{22} + X_{32} + \cdots + X_{n2} \\ X_{13} + X_{23} + X_{33} + \cdots + X_{n3} \\ \vdots \\ X_{1n} + X_{2n} + X_{3n} + \cdots + X_{nn} \end{bmatrix} . \\ &= \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \\ \vdots \\ \bar{x}_n \end{bmatrix} \end{aligned} \tag{2.65}$$

$\bar{x}$ merupakan matriks rata-rata yang berukuran $n \times 1$. Selanjutnya menentukan varian dari matriks X dengan persamaan berikut:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(X-\bar{x})(X-\bar{x})^T}{n}, \qquad i = 1,2,3,\dots,n. \tag{2.66}$$

Selanjutnya merubah matriks pada (2.64) yang berukuran $n \times 1$ ke dalam matriks berukuran $n \times n$ dengan cara mengalikan $\bar{x}$ dengan transpose vektor 1, yaitu:

$$\begin{aligned}\bar{x} &= \begin{bmatrix}\bar{x}_1\\ \bar{x}_2\\ \bar{x}_3\\ \vdots\\ \bar{x}_n\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & \cdots & 1\end{bmatrix}\\ &= \begin{bmatrix}\bar{x}_1 & \bar{x}_1 & \bar{x}_1 & \bar{x}_1 & \cdots & \bar{x}_1\\ \bar{x}_2 & \bar{x}_2 & \bar{x}_2 & \bar{x}_2 & \cdots & \bar{x}_2\\ \bar{x}_3 & \bar{x}_3 & \bar{x}_3 & \bar{x}_3 & \cdots & \bar{x}_3\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\ \bar{x}_n & \bar{x}_n & \bar{x}_n & \bar{x}_n & \cdots & \bar{x}_n\end{bmatrix}.\end{aligned} \tag{2.67}$$

Selanjutnya menentukan matriks diagonal untuk simpangan rata-rata Q_s, yaitu dengan cara mengurangi matriks X dengan matriks $\bar{x}$ pada persamaan (2.67) sebagai berikut:

$$\begin{aligned}Q_s &= X - \bar{x}\\ &= \begin{bmatrix}X_{11} & X_{12} & X_{13} & \cdots & X_{1n}\\ X_{21} & X_{22} & X_{23} & \cdots & X_{2n}\\ X_{31} & X_{32} & X_{33} & \cdots & X_{3n}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ X_{n1} & X_{n2} & X_{n3} & \cdots & X_{nn}\end{bmatrix} - \begin{bmatrix}\bar{x}_1 & \bar{x}_1 & \bar{x}_1 & \bar{x}_1 & \cdots & \bar{x}_1\\ \bar{x}_2 & \bar{x}_2 & \bar{x}_2 & \bar{x}_2 & \cdots & \bar{x}_2\\ \bar{x}_3 & \bar{x}_3 & \bar{x}_3 & \bar{x}_3 & \cdots & \bar{x}_3\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\ \bar{x}_n & \bar{x}_n & \bar{x}_n & \bar{x}_n & \cdots & \bar{x}_n\end{bmatrix}.\\ &= \begin{bmatrix}X_{11}-\bar{x}_1 & X_{12}-\bar{x}_1 & X_{13}-\bar{x}_1 & \cdots & X_{1n}-\bar{x}_1\\ X_{21}-\bar{x}_2 & X_{22}-\bar{x}_2 & X_{23}-\bar{x}_2 & \cdots & X_{2n}-\bar{x}_2\\ X_{31}-\bar{x}_3 & X_{32}-\bar{x}_3 & X_{33}-\bar{x}_3 & \cdots & X_{3n}-\bar{x}_3\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ X_{n1}-\bar{x}_n & X_{n2}-\bar{x}_n & X_{n3}-\bar{x}_n & \cdots & X_{nn}-\bar{x}_n\end{bmatrix}\end{aligned} \tag{2.68}$$

Untuk mendapatkan varian atau nilai kuadrat bisa dilakukan dengan cara memperoleh jumlah kuadrat simpangan (JK) terlebih dahulu, yaitu:

$$
\begin{aligned}
JK &= Q_s Q_s^T \\
&= (X-\bar{x})(X-\bar{x})^T \\
&= \begin{bmatrix} X_{11}-\bar{x}_1 & X_{12}-\bar{x}_1 & X_{13}-\bar{x}_1 & \cdots & X_{1n}-\bar{x}_1 \\ X_{21}-\bar{x}_2 & X_{22}-\bar{x}_2 & X_{23}-\bar{x}_2 & \cdots & X_{2n}-\bar{x}_2 \\ X_{31}-\bar{x}_3 & X_{32}-\bar{x}_3 & X_{33}-\bar{x}_3 & \cdots & X_{3n}-\bar{x}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{n1}-\bar{x}_n & X_{n2}-\bar{x}_n & X_{n3}-\bar{x}_n & \cdots & X_{nn}-\bar{x}_n \end{bmatrix}. \\
&\begin{bmatrix} X_{11}-\bar{x}_1 & X_{21}-\bar{x}_2 & X_{31}-\bar{x}_1 & \cdots & X_{n1}-\bar{x}_n \\ X_{12}-\bar{x}_1 & X_{22}-\bar{x}_2 & X_{32}-\bar{x}_1 & \cdots & X_{n2}-\bar{x}_n \\ X_{13}-\bar{x}_1 & X_{23}-\bar{x}_2 & X_{33}-\bar{x}_1 & \cdots & X_{n3}-\bar{x}_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{1n}-\bar{x}_1 & X_{2n}-\bar{x}_2 & X_{3n}-\bar{x}_1 & \cdots & X_{nn}-\bar{x}_n \end{bmatrix}
\end{aligned} \tag{2.69}
$$

Dari nilai JK yang sudah diketahui, varian bisa ditentukan dengan persamaan

$$\sigma_{ij}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_{ij}-\bar{x}_{ij}\right)\left(X_{ij}-\bar{x}_{ij}\right)^T, \qquad i,j = 1,2,3,\dots,n \tag{2.70}$$

sehingga;

$$\sigma_1^2 = \frac{(X_{11}-\bar{x}_1)(X_{11}-\bar{x}_1)+(X_{12}-\bar{x}_1)(X_{12}-\bar{x}_1)+\cdots+(X_{1n}-\bar{x}_1)(X_{1n}-\bar{x}_1)}{n},$$

$$\sigma_{12}^2 = \frac{(X_{11}-\bar{x}_1)(X_{21}-\bar{x}_2)+(X_{12}-\bar{x}_1)(X_{22}-\bar{x}_2)+\cdots+(X_{1n}-\bar{x}_1)(X_{2n}-\bar{x}_2)}{n},$$

$$\vdots$$

$$\sigma_n^2 = \frac{(X_{n1}-\bar{x}_n)(X_{n1}-\bar{x}_n)+(X_{n2}-\bar{x}_n)(X_{n2}-\bar{x}_n)+\cdots+(X_{nn}-\bar{x}_n)(X_{nn}-\bar{x}_n)}{n},$$

dari varian matriks X yang sudah diperoleh, maka matriks varian kovarian Σ dapat dibentuk sebagai berikut:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \sigma_{13}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma_2^2 & \sigma_{23}^2 & \cdots & \sigma_{2n}^2 \\ \sigma_{31}^2 & \sigma_{32}^2 & \sigma_3^2 & \cdots & \sigma_{3n}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \sigma_{n3}^2 & \cdots & \sigma_n^2 \end{bmatrix}. \tag{2.71}$$

2. Menentukan matriks korelasi (ρ). Sebelum menentukan matriks korelasi, kovariannya harus dicari terlebih dahulu. Langkah untuk mencari kovarian yang pertama yaitu dengan menghitung matriks baku yang didalamnya

terdapat simpangan baku, dimana terdapat asumsi yang harus terpenuhinya yaitu:

$$Var(X_i, X_j) = \begin{cases} \sigma_{ij}^2, & i = j \\ 0, & i \neq j \end{cases}. \tag{2.72}$$

Simpangan baku merupakan akar dari varian, dengan persamaannya berikut untuk matriks baku berukuran $n \times n$;

$$\begin{aligned} W_{n\times n} &= \sqrt{\mathrm{S}} \\ &= \sqrt{\sigma_{ij}^2}, \end{aligned} \tag{2.73}$$

sehingga diperoleh:

$$\begin{aligned} W_{n\times n} &= \begin{bmatrix} \sqrt{\sigma_1^2} & \sqrt{\sigma_{12}^2} & \sqrt{\sigma_{13}^2} & \cdots & \sqrt{\sigma_{1n}^2} \\ \sqrt{\sigma_{21}^2} & \sqrt{\sigma_2^2} & \sqrt{\sigma_{23}^2} & \cdots & \sqrt{\sigma_{2n}^2} \\ \sqrt{\sigma_{31}^2} & \sqrt{\sigma_{32}^2} & \sqrt{\sigma_3^2} & \cdots & \sqrt{\sigma_{3n}^2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sqrt{\sigma_{n1}^2} & \sqrt{\sigma_{n2}^2} & \sqrt{\sigma_{n3}^2} & \cdots & \sqrt{\sigma_n^2} \end{bmatrix}. \\ &= \begin{bmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_n \end{bmatrix} \end{aligned} \tag{2.74}$$

Selanjutnya dalam menentukan korelasi dari matriks X maka variabel baku ditentukan terlebih dahulu dengan dengan cara menghitung invers dari matriks baku, yaitu:

$$(W_{n\times n})^{-1} = \frac{1}{Det(W_{n\times n})} Adj(W_{n\times n}), \tag{2.75}$$

dan didapatkan nilai determinan;

$$Det(W_{n\times n}) = [\sigma_1 \quad \sigma_2 \quad \sigma_3 \quad \cdots \quad \sigma_n], \tag{2.76}$$

$$J = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \cdots & c_{2n} \\ c_{31} & c_{32} & c_{33} & \cdots & c_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & c_{n3} & \cdots & c_{nn} \end{bmatrix}, \tag{2.77}$$

dimana:

$$c_{11} = (-1)^{1+1}M_{11} = \begin{vmatrix} \sigma_2 & 0 & \cdots & 0 \\ 0 & \sigma_3 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{vmatrix} = \sigma_2\sigma_3 \ldots \sigma_n,$$

$$c_{12} = (-1)^{1+2}M_{12} = \begin{vmatrix} 0 & 0 & \cdots & 0 \\ 0 & \sigma_3 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{vmatrix} = 0,$$

$$c_{22} = (-1)^{2+2}M_{22} = \begin{vmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_3 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{vmatrix} = \sigma_1\sigma_3 \ldots \sigma_n,$$

$$\vdots$$

$$c_{nn} = (-1)^{n+n}M_{nn} = \begin{vmatrix} \sigma_n & 0 & \cdots & 0 \\ 0 & \sigma_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{vmatrix} = \sigma_n\sigma_n \ldots \sigma_n.$$

Adjoin matriks diperoleh dari matriks (2.77), yaitu;

$$Adj(W_{n\times n}) = J^T$$
$$= \begin{bmatrix} \sigma_2\sigma_3\sigma_4 \cdots \sigma_n & 0 & \cdots & 0 \\ 0 & \sigma_1\sigma_3\sigma_4 \cdots \sigma_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_1\sigma_2\sigma_3 \cdots \sigma_n \end{bmatrix}, \qquad (2.78)$$

dari determinan matriks pada persamaan (2.76) dan adjoin matrik pada persamaan (2.78), maka diperoleh invers matriks dengan persamaan (2.75) sebagai berikut:

$$(W_{n\times n})^{-1} = \frac{1}{Det(W_{n\times n})} Adj(W_{n\times n})$$
$$= \frac{1}{\sigma_1\sigma_2\sigma_3\cdots\sigma_n}\begin{bmatrix} \sigma_2\sigma_3\sigma_4\cdots\sigma_n & 0 & \cdots & 0 \\ 0 & \sigma_1\sigma_3\sigma_4\cdots\sigma_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_1\sigma_2\sigma_3\cdots\sigma_n \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{\sigma_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sigma_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\sigma_n} \end{bmatrix} \quad . \quad (2.79)$$

Setelah variabel baku didapatkan, maka matriks korelasi bisa tentukan dengan rumus

$$\rho = (W_{n\times n})^{-1}\Sigma(W_{n\times n})^{-1}$$
$$= \begin{bmatrix} \frac{1}{\sigma_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sigma_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\sigma_n} \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \sigma_{13}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma_2^2 & \sigma_{23}^2 & \cdots & \sigma_{2n}^2 \\ \sigma_{31}^2 & \sigma_{32}^2 & \sigma_3^2 & \cdots & \sigma_{3n}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \sigma_{n3}^2 & \cdots & \sigma_n^2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sigma_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\sigma_n} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{\sigma_1^2}{\sigma_1} & \frac{\sigma_{12}^2}{\sigma_1} & \frac{\sigma_{13}^2}{\sigma_1} & \cdots & \frac{\sigma_{1n}^2}{\sigma_1} \\ \frac{\sigma_{21}^2}{\sigma_2} & \frac{\sigma_2^2}{\sigma_2} & \frac{\sigma_{23}^2}{\sigma_2} & \cdots & \frac{\sigma_{2n}^2}{\sigma_2} \\ \frac{\sigma_{31}^2}{\sigma_3} & \frac{\sigma_{32}^2}{\sigma_3} & \frac{\sigma_3^2}{\sigma_3} & \cdots & \frac{\sigma_{3n}^2}{\sigma_3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{n1}^2}{\sigma_n} & \frac{\sigma_{n2}^2}{\sigma_n} & \frac{\sigma_{n3}^2}{\sigma_n} & \cdots & \frac{\sigma_{nn}^2}{\sigma_n} \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sigma_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\sigma_n} \end{bmatrix}$$
$$= \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2n} \\ \rho_{31} & \rho_{32} & 1 & \cdots & \rho_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \rho_{n3} & \cdots & 1 \end{bmatrix} \quad (2.80)$$

$$= \begin{bmatrix} \frac{\sigma_1^2}{\sigma_1\sigma_1} & \frac{\sigma_{12}^2}{\sigma_1\sigma_2} & \frac{\sigma_{13}^2}{\sigma_1\sigma_3} & \cdots & \frac{\sigma_{1n}^2}{\sigma_1\sigma_n} \\ \frac{\sigma_{21}^2}{\sigma_2\sigma_1} & \frac{\sigma_2^2}{\sigma_2\sigma_2} & \frac{\sigma_{23}^2}{\sigma_2\sigma_3} & \cdots & \frac{\sigma_{2n}^2}{\sigma_2\sigma_n} \\ \frac{\sigma_{31}^2}{\sigma_3\sigma_1} & \frac{\sigma_{32}^2}{\sigma_3\sigma_2} & \frac{\sigma_3^2}{\sigma_3\sigma_3} & \cdots & \frac{\sigma_{3n}^2}{\sigma_3\sigma_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{n1}^2}{\sigma_n\sigma_1} & \frac{\sigma_{n2}^2}{\sigma_n\sigma_2} & \frac{\sigma_{n3}^2}{\sigma_n\sigma_3} & \cdots & \frac{\sigma_{nn}^2}{\sigma_n\sigma_n} \end{bmatrix}$$

di mana persamaan yang diperoleh untuk korelasinya adalah

$$\rho_{ij} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_{ij}-\bar{x}_i}{\sigma_{ii}}\right)\left(\frac{X_{ji}-\bar{x}_j}{\sigma_{jj}}\right), \tag{2.81}$$

dengan melihat asumsi pada persamaan (2.72), maka untuk $i = j$ akan diperoleh $\rho = 1$ sehingga apabila diberikan ρ_{33}

$$\rho_{33} = \sum_{i=1}^{4}\left(\frac{X_{33}-\bar{x}_3}{\sigma_3}\right)\left(\frac{X_{33}-\bar{x}_3}{\sigma_3}\right) = \frac{(X_{33}-\bar{x}_3)(X_{33}-\bar{x}_3)}{\sigma_3\sigma_3} = 1,$$

untuk $i \neq j$ akan diperoleh $\rho = 0$, apabila diberikan ρ_{31} maka

$$\rho_{31} = \sum_{i=1}^{4}\left(\frac{X_{31}-\bar{x}_3}{\sigma_3}\right)\left(\frac{X_{13}-\bar{x}_1}{\sigma_1}\right) = \frac{\left(X_{31}-\bar{h}_3\right)(X_{13}-\bar{x}_1)}{\sigma_3\sigma_1} = 0.$$

Selanjutnya menentukan nilai *eigen* (λ) dengan merujuk pada persamaan (2.77), maka persamaan untuk mencari nilai *eigen* pada matriks korelasi adalah

$$|\rho - \lambda I| = 0, \tag{2.82}$$

di mana

$$|\rho - \lambda I| = 0$$

$$\left|\begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2n} \\ \rho_{31} & \rho_{32} & 1 & \cdots & \rho_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \rho_{n3} & \cdots & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}\right| = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (2.83)$$

$$\left|\begin{bmatrix} 1-\lambda & \rho_{12} & \rho_{13} & \cdots & \rho_{1n} \\ \rho_{21} & 1-\lambda & \rho_{23} & \cdots & \rho_{2n} \\ \rho_{31} & \rho_{32} & 1-\lambda & \cdots & \rho_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \rho_{n3} & \cdots & 1-\lambda \end{bmatrix}\right| = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

sehingga didapatkan nilai *eigen* sebagai berikut:

$$\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \vdots \\ \lambda_n \end{bmatrix}. \quad (2.84)$$

Selanjutnya adalah menentukan vektor *eigen* dengan persamaan (2.28), maka persamaan vektor *eigen* untuk matriks korelasi adalah

$$(\rho - \lambda I)a = 0, \quad (2.85)$$

dimana

$$(\rho - \lambda)a = 0$$

$$\left(\begin{bmatrix} 1-\lambda & \rho_{12} & \rho_{13} & \cdots & \rho_{1n} \\ \rho_{21} & 1-\lambda & \rho_{23} & \cdots & \rho_{2n} \\ \rho_{31} & \rho_{32} & 1-\lambda & \cdots & \rho_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \rho_{n3} & \cdots & 1-\lambda \end{bmatrix}\right) \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix} = 0.$$

Kasus di atas dapat diatasi dengan cara mencari eselon baris, yaitu:

$$(1-\lambda_1)a_1 + \rho_{12}a_2 + \rho_{13}a_3 + \cdots + \rho_{1n}a_n = 0,$$

$$\rho_{21}a_1 + (1-\lambda_1)a_2 + \rho_{23}a_3 + \cdots + \rho_{2n}a_n = 0,$$

$$\vdots$$

$$\rho_{n1}a_1 + \rho_{n2}a_2 + \rho_{n3}a_3 + \cdots + (1-\lambda_1)a_n = 0. \quad (2.86)$$

Untuk memenuhi $a_n^T a_n = 1$, maka iterasi tersebut dinormalkan melalui jarak *Euclid* sebagai berikut:

Jarak vektor:

$$l = \sqrt{((a_1)^2 + (a_2)^2 + (a_3)^2 + \cdots + (a_n)^2)}, \tag{2.87}$$

Vektor *eigen*:

$$\left[\frac{a_1}{l}, \frac{a_2}{l}, \frac{a_3}{l}, \cdots, \frac{a_n}{l}\right], \tag{2.88}$$

sehingga diperoleh vektor normal a_n^T. Untuk mendapatkan vektor normal dilakuka perhitungan sebagai berikut:

$$\begin{aligned} a_{11} &= \frac{a_{11}}{\sqrt{((a_1)^2 + (a_2)^2 + (a_3)^2 + \cdots + (a_n)^2)}} \\ a_{21} &= \frac{a_{21}}{\sqrt{((a_1)^2 + (a_2)^2 + (a_3)^2 + \cdots + (a_n)^2)}}, \\ &\vdots \\ a_{n1} &= \frac{a_{n1}}{\sqrt{((a_1)^2 + (a_2)^2 + (a_3)^2 + \cdots + (a_n)^2)}} \end{aligned} \tag{2.89}$$

dihasilkan vektor normal a_n',

$$\begin{bmatrix} a_{11} & a_{21} & a_{31} & \cdots & a_{n1} \\ a_{12} & a_{22} & a_{32} & \cdots & a_{n2} \\ a_{13} & a_{23} & a_{33} & \cdots & a_{n3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \cdots & a_{nn} \end{bmatrix}. \tag{2.90}$$

Dari langkah-langkah pembentukan *principal component* diatas maka dapat dibentuk persamaan kombinasi linier sebagai berikut:

$$\begin{aligned} C_1 &= a_1'X = a_{11}X_1 + a_{21}X_2 + a_{31}X_3 + \cdots + a_{n1}X_n \\ C_2 &= a_2'X = a_{12}X_1 + a_{22}X_2 + a_{32}X_3 + \cdots + a_{n2}X_n \\ &\vdots \\ C_n &= a_n'X = a_{1n}X_1 + a_{2n}X_2 + a_{3n}X_3 + \cdots + a_{nn}X_n \end{aligned}, \tag{2.91}$$

atau $C = a_n' = a'\Sigma a$.

2.13 Kajian Agama Mengenai Metode Regresi *Robust Principal Component Analysis*

Metode regresi *Robust Principal Component Analysis* (ROBPCA) merupakan gabungan dari dua metode, yakni metode PCA dan metode *robust*. Tujuan dari metode ROBPCA ini adalah untuk menangani masalah multikolinieritas dan outlier, dengan prinsip dasarnya yaitu menyederhanakan variabel yang diamati dengan cara mereduksi dimensinya (Tulak, dkk., 2019). Dari proses mereduksi data ini akan terbentuk sebuah variabel baru yang kemudian tidak menghilangkan variabel asalnya dan model yang dihasilkan akan terbebas dari korelasi. Di mana proses penyelesaian ini dapat menggambarkan bagaimana seseorang dalam memperbaiki suatu masalah di dalam hidupnya tanpa harus melupakan masalah tersebut, karena adanya masalah dalam hidup ini merupakan sebuah liku yang ada di dalam alur kehidupan dan harus dihadapi untuk mendapatkan tujuan hidup yang lebih bermanfaat lagi. Sesuai yang terkandung di dalam Al Qur’an surat Al Baqarah ayat 216 yang artinya: (Shihab, 2005)

“Diwajibkan atas kamu berperang, padahal itu tidak menyenangkan bagimu. Tetapi boleh jadi kamu tidak menyenangi sesuatu, padahal itu baik bagimu, dan boleh jadi kamu menyukai sesuatu, padahal itu tidak baik bagimu. Allah mengetahui, sedang kamu tidak mengetahui”.

Shihab (2005) dalam tafsir Al Misbah, menafsirkan bahwa sesuatu yang tidak disenangi belum pasti tidak baik, terutama ketetapan Allah SWT. Oleh karenanya, ketika ada perintah atau larangan dari Allah SWT yang terkesan tidak mengenakkan, sudah seharusnya menanamkan rasa optimisme di dalam jiwa dan berkata bisa jadi di balik ketetapan tersebut terdapat sesuatu yang baik maupun bernilai. Demikian pula sebaliknya, apabila seseorang mendapatkan nikmat kebahagiaan hidup di dunia berupa kelimpahan rezeki atau sebagainya, semestinya tidak berbahagia secara berlebihan hingga lupa diri. Karena bisa jadi di balik kenikmatan hidup yang disukai tersebut terdapat mudharat yang tidak disangka-sangka atau bisa dikatakan terjadinya istidraj di dalamnya. Ayat ini mengajarkan bagaimana manusia harus berpasrah diri atau berserah diri hanya kepada Allah SWT dengan cara menjalankan hidup secara seimbang, yaitu menjalani kehidupan dengan tidak menjauhkan rasa optimisme ketika dilanda suatu kesedihan dan tidak pula larut dalam kegembiraan sampai melupakan daratan.

Allah SWT berfirman dalam Al Qur’an surat Shad ayat 17, yang artinya: (Basyir, 2011)

“Bersabarlah atas segala apa yang mereka katakan, dan ingatlah hamba Kami Daud yang mempunyai kekuatan, sesungguhnya dia amat taat (kepada Tuhan)”.

Dari ayat tersebut Allah SWT menceritakan prihal hamba dan rasul-Nya, yakni Dawud AS., bahwa dia dianugerahi kekuatan. Dalam ayat ini diungkapkan bahwa dengan kata al-aid yang artinya kekuatan, yakni kekuatan dalam ketaatan. Qatadah mengatakan bahwa “Dawud AS., dianugerahi kekuatan dalam mengerjakan ibadah dan memberinya pengetahuan tentang Islam”. Di dalam Shahihain dinyatakan, bahwa Rasulullah Saw bersabda, “Shalat yang paling dicintai Allah SWT adalah shalat Dawud, puasa yang paling dicintai Allah SWT adalah puasa Dawud. Beliau tidur setengah malam, bangun pada sepertiganya dan tidur seperempatnya. Beliau puasa satu hari dan berbuka satu hari. Beliau tidak lari jika berjumpa musuh dan sesungguhnya beliau adalah orang yang awwab, yaitu golongan orang yang segera kembali kepada Allah SWT dalam seluruh perkara dan keadaan”. Dalam tafsir Muyassar oleh Basyir (2011), Allah SWT memerintahkan kepada Rasulullah Saw agar selalu mengingatkan kaumnya akan kisah Nabi Dawud AS., sang pemilik kekuatan atas musuh-musuh Allah SWT dan selalu sabar menaati-Nya. Kekuatan yang dimaksud dalam ayat ini adalah kekuatan dalam menaati Allah SWT dan kekuatan dalam memahami agama.

Berdasarkan penafsiran Al Qur’an surat Shaad di atas, penulis mengiterpretasikan bahwasanya kekuatan dalam ketaatan kepada Allah SWT yang dimiliki nabi Dawud AS., adalah sebuah gambaran bagi kita sebagai makhluk ciptaan-Nya agar selalu istiqamah dalam hal beribadah dan dalam melakukan hal-hal kebaikan. Dimana dalam metode ROBPCA yang merupakan gabungan dari dua metode dengan metode satu sebagai penguat dari metode satunya, yaitu metode *robust* yang kekar terhadap suatu masalah yang biasa terjadi dalam metode PCA, dengan menguatkan setiap komponen pilihan dalam metode PCA untuk menyelesaikan masalah *outlier* pada data penelitian. Kekekaran yang dimiliki regresi *robust* ini membantu metode PCA dengan tidak merubah model yang ada di metode tersebut, sesuai perintah Allah SWT agar kita senantiasa mensyukuri nikmat yang Allah SWT berikan kepada kita dengan tidak mengubah segala sesuatu yang diberikan-Nya kepada kita.

BAB III
METODOLOGI PENELITIAN

3.1 Jenis Penelitian

Jenis penelitian tentang implementasi metode regresi *Robust Principal Component Analysis* (ROBPCA) pada data kemiskinan di Jawa Timur ini menggunakan pendekatan deskriptif kuantitatif dengan studi literatur. Deskriptif kuantitatif dilakukan dengan menyusun data dan menganalisis sesuai dengan bahan penelitian, sedangkan studi literatur adalah pengumpulan bahan-bahan pustaka yang dibutuhkan oleh peneliti sebagai acuan dalam menyelesaikan penelitian mengenai metode ROBPCA.

3.2 Data dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder tentang kemiskinan Jawa Timur pada tahun 2019. Data diperoleh dari *website* Badan Pusat Statistika (BPS) Jawa Timur yang dipublikasikan berbentuk dokumen dengan judul "Provinsi Jawa Timur dalam angka" pada tahun 2020. Unit observasi dalam penelitian ini menyertakan 38 Kabupaten/Kota di Provinsi Jawa Timur dan terlampir pada (Lampiran 3).

3.3 Lokasi Penelitian

Data yang diperoleh untuk penelitian ini diambil secara online dari *website* Badan Pusan Statitika (BPS) Jawa Timur yang dipublish di internet pada tahun 2020.

3.4 Teknik Pengumpulan Data

Dari pemaparan sumber data diatas maka teknik pengumpulan data yang diambil oleh peniliti adalah studi kepustakaan. Studi kepustakaan merupakan proses pengambilan bahan-bahan penelitian dengan cara mengambil sumber dari beberapa jurnal, artikel, karya ilmiah atau literatur-literatur yang dipublikasikan di internet dan layak digunakan sebagai sumber penelitian.

Dalam penelitian ini data yang akan diimplementasikan dengan metode regresi ROBPCA adalah data kemiskinan Jawa Timur tahun 2019 yang bersumber dari *website* BPS Jawa Timur. Proses pengambilan data yaitu dengan menentukan indikator utama berupa kemiskinan. Dari indikator utama tersebut peneliti memilih variabel-variabel pendukung yang sesuai dengan syarat-syarat yang sudah peneliti tentukan.

3.5 Instrumen Penelitian

Dalam penelitian ini yang penulis menggunakan alat bantu aplikasi berupa *software* SPSS 26, Minitab 18 dan NCSS 22. Sedangkan variabel-variabel dari data yang digunakan untuk diimplemantasikan dengan metode regresi ROBPCA adalah data jumlah kemiskinan Jawa Timur tahun 2019 sebagai variabel terikat dan variabel bebas X sebagai berikut:

X_1 : Kepadatan penduduk,

X_2 : Tingkat pengangguran terbuka,

X_3 : Produk domestik regional bruto,

X_4 : Indeks keparahan kemiskinan,

X_5 : Rata-rata lama sekolah,

X_6 : Indeks pembangunan manusia,

X_7 : Pengeluaran per kapita riil.

3.6 Teknik Analisis Data

Langkah-langkah analisis data dalam penelitian ini adalah sebagai berikut:

3.6.1 Membentuk *Principal Component*

Berikut adalah langkah-langkah yang digunakan untuk membentuk *principal component* pada metode ROBPCA:

1. Membentuk sebuah matrik, yakni matriks varian kovarian antar variabel bebas dengan ukuran $n \times n$.
2. Mencari persamaan untuk matriks varian kovarian dengan menggunakan langkah-langkah perhitungan dari metode FAST-MCD.
3. Membentuk matriks korelasi.
4. Mencari nilai *eigen* dan vektor *eigen*.
5. Menentukan *principal component* dengan membentuk sebuah persamaan kombinasi linier.

3.6.2 Implementasi Data Kemikinan dengan Metode ROBPCA

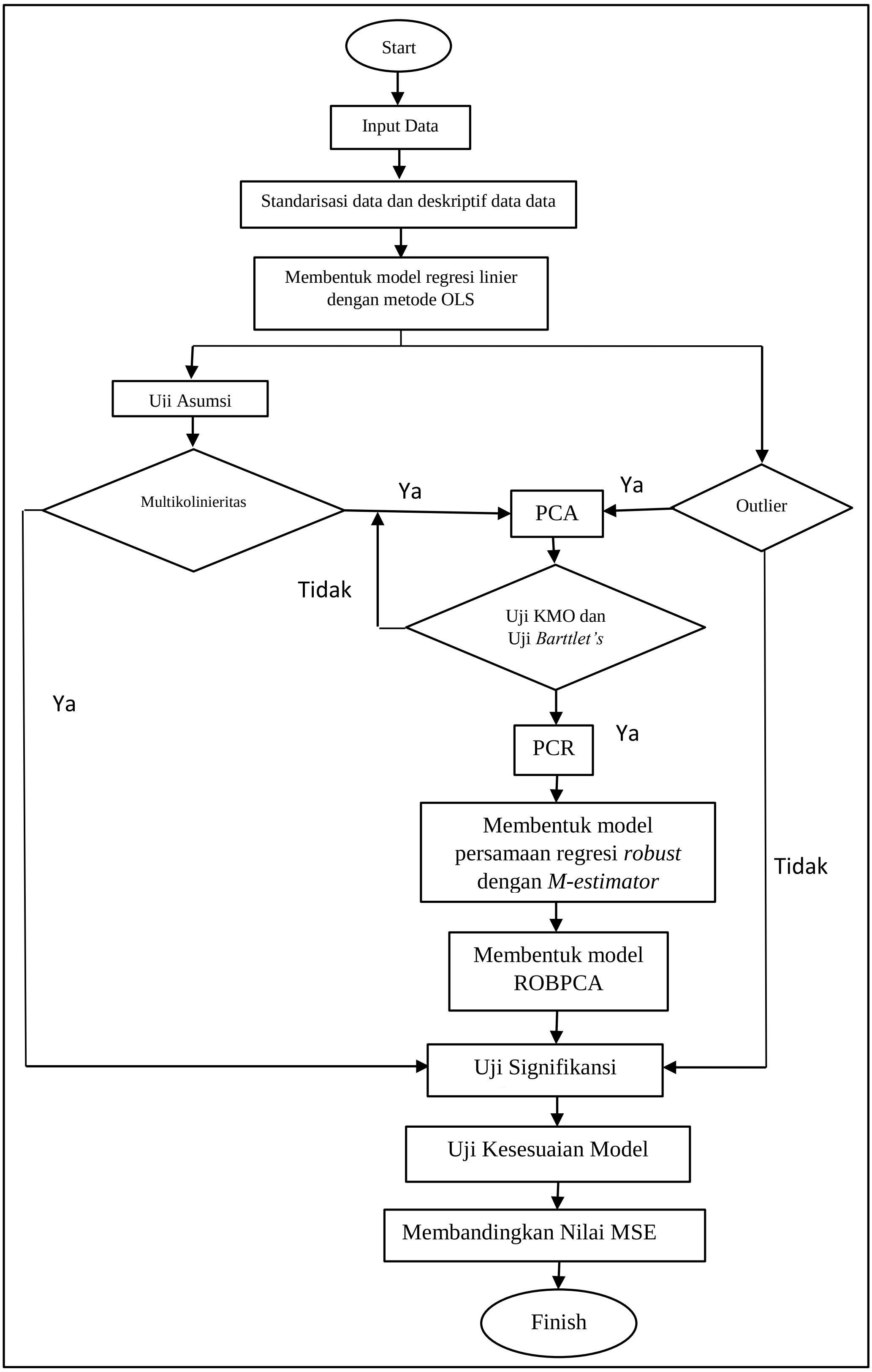
Berikut langkah-langkah yang digunakan untuk menyelesaikan multikolinieritas dengan metode ROBPCA dan implementasi metode ROBPCA pada data kemiskinan Jawa Timur tahun 2019:

1. Mengumpulkan data kemiskinan Jawa Timur tahun 2019.
2. Mendefinisikan variabel terikat Y dan variabel bebas X dari data yang diperoleh.
3. Melakukan analisis deskripif pada data penelitian.
4. Melakukan standarisasi data penelitian.
5. Memodelkan regresi linier berganda dengan metode *ordinary least square* (OLS).

6. Melakukan uji asumsi klasik untuk mendeteksi adanya multikolinieritas pada model regresi linier berganda:
    a. Uji Normalitas
    b. Uji Multikolinieritas
    c. Uji Heteroskidastisitas
7. Penyelesaian multikolinieritas dengan PCA
    a. Membentuk *principal component*.
    b. Melakukan uji asumsi untuk metode PCA dengan uji KMO dan uji *Barttlet's*.
    c. Pemilihan *principal componen* dengan kriteria yang ditentukan, yaitu yang memiliki nilai *eigen* $\lambda \geq 1$.
    d. Menghitung proporsi kumulatif varian yang bisa dijelaskan dengan *principal component* terpilih.
    e. Memodelkan *principal component* yang terpilih dengan model kombinasi linier.
8. Meregresikan *principal component* yang sudah dimodelkan dengan variabel terikat menggunakan metode OLS dan dinamakan dengan model *Principal Component Regression* (PCR).
9. Melakukan uji signifikansi parameter model PCR dengan uji serentak (uji F) dan uji koefisien regresi dengan uji individual (uji t).
10. Menghitung nilai $Adjusted$ (R^2) dan nilai RSE dari model PCR.
11. Melakukan pengecekan adanya multikolinieritas dan *outlier* pada model yang dihasilkan oleh nomer 4 dengan menggunakan metode $DFFIT's$ dan $Levarage$.

12. Apabila multikolinieritas terselesaikan, namun *outlier* ditemukan maka dilakukan analisis dengan metode ROBPCA, yaitu dengan meregresikan *principal component* terpilih dengan variabel terikat menggunakan metode estimasi berupa *M-estimator*.
13. Melakukan uji signifikansi parameter model regresi ROBPCA dengan uji serentak (Uji F) dan uji koefisien regresi dengan uji individual (Uji t).
14. Menghitung nilai $Adjusted\ (R^2)$ dan nilai RSE dari model regresi ROBPCA.
15. Mengembalikan persamaan regresi ROBPCA ke bentuk variabel asal.
16. Menarik kesimpulan.

3.7 *Flowchart*



Gambar 3. 1 *Flowchart*

BAB IV
PEMBAHASAN

4.1 Metode Regresi *Robust Principal Component Analysis* (ROBPCA*) M-Estimator*

Metode regresi *Robust Principal Component Analysis* (ROBPCA) merupakan gabungan dari dua metode, yakni terdiri dari metode PCA dan metode *robust*. Metode PCA yang memiliki kesensitifan terhadap keberadaan *outlier* maka metode *robust* diperlukan untuk menyelesaikannya, sehingga tujuan dari metode ROBPCA ini adalah selain untuk menyelesaikan multikolinieritas yang sering terjadi pada model regresi linier berganda juga untuk menyelesaikan *outlier* yang terkandung pada data penelitian. Metode estimasi yang digunakan dalam hal ini adalah metode *M-estimastor*. Berikut proses-proses dalam menyelesaikan permasalahan pada penelitian ini:

4.1.1 Pembentukan *Principal Component*

Berikut adalah proses pembentukan *principal component*:

1. Membentuk matriks varian kovarian dengan matriks yang diberikan untuk membentuk *principal component*. Pada penelitian ini menggunkan sampel penelitian dengan $n = 7$, maka diberikan matriks H berukuran 7×7, berdasarkan persamaan (2.62) yaitu:

$$H = \begin{bmatrix} H_{11} & H_{12} & H_{13} & \cdots & H_{17} \\ H_{21} & H_{22} & H_{23} & \cdots & H_{27} \\ H_{31} & H_{32} & H_{33} & \cdots & H_{37} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ H_{71} & H_{72} & H_{73} & \cdots & H_{77} \end{bmatrix}, \qquad (4.1)$$

Dengan menggunakan langkah-langkah perhitungan dari metode FAST-MCD pada persamaan (2.39) untuk menentukan *principal component*, maka persamaan untuk matriks varian kovarian adalah

$$S = \frac{1}{7}\sum_{d=1}^{7}\left(H_d - \bar{h}\right)\left(H_d - \bar{h}\right)^T, \qquad d = 1,2,3,\dots,7. \tag{4.2}$$

dimana $\bar{h}$ adalah vektor rata-rata yang dapat ditentukan dengan persamaan (2.38). Berikut nilai rata-rata dari setiap kolom untuk vektor rata-rata;

$\bar{h}_1 = \frac{H_{11}+H_{21}+H_{31}+H_{41}+\cdots+H_{71}}{7},$

$\bar{h}_2 = \frac{H_{12}+H_{22}+H_{32}+H_{42}+\cdots+H_{72}}{7},$

$\bar{h}_3 = \frac{H_{13}+H_{23}+H_{33}+H_{43}+\cdots+H_{73}}{7},$

$\vdots$

$\bar{h}_7 = \frac{H_{17}+H_{27}+H_{37}+H_{47}+\cdots+H_{77}}{7},$

dan diperoleh berdasarkan persamaan (2.63);

$$\begin{aligned}\bar{h} &= \left[\bar{h}_1, \bar{h}_2, \bar{h}_3, \bar{h}_4, \dots, \bar{h}_7\right] \\ &= \frac{1}{7}H^T 1 \\ &= \frac{1}{7}\begin{bmatrix} H_{11} & H_{21} & H_{31} & \cdots & H_{71} \\ H_{12} & H_{22} & H_{32} & \cdots & H_{72} \\ H_{13} & H_{23} & H_{33} & \cdots & H_{73} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ H_{17} & H_{27} & H_{37} & \cdots & H_{77} \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \bar{h}_1 \\ \bar{h}_2 \\ \bar{h}_3 \\ \vdots \\ \bar{h}_7 \end{bmatrix}\end{aligned} \tag{4.3}$$

Selanjutnya merubah matriks pada (4.4) yang berukuran 7×1 ke dalam matriks berukuran 7×7 sesuai dengan persamaan (2.64), yaitu:

$$\bar{h} = \begin{bmatrix} \bar{h}_1 \\ \bar{h}_2 \\ \bar{h}_3 \\ \vdots \\ \bar{h}_7 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \\ = \begin{bmatrix} \bar{h}_1 & \bar{h}_1 & \bar{h}_1 & \bar{h}_1 & \cdots & \bar{h}_1 \\ \bar{h}_2 & \bar{h}_2 & \bar{h}_2 & \bar{h}_2 & \cdots & \bar{h}_2 \\ \bar{h}_3 & \bar{h}_3 & \bar{h}_3 & \bar{h}_3 & \cdots & \bar{h}_3 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{h}_7 & \bar{h}_7 & \bar{h}_7 & \bar{h}_7 & \cdots & \bar{h}_7 \end{bmatrix}. \tag{4.5}$$

Dari persamaan (2.64) diperoleh matriks diagonal untuk simpangan rata-rata G_s, yaitu:

$$G_s = H - \bar{h} \\ = \begin{bmatrix} H_{11} - \bar{h}_1 & H_{12} - \bar{h}_1 & H_{13} - \bar{h}_1 & \cdots & H_{17} - \bar{h}_1 \\ H_{21} - \bar{h}_2 & H_{22} - \bar{h}_2 & H_{23} - \bar{h}_2 & \cdots & H_{27} - \bar{h}_2 \\ H_{31} - \bar{h}_3 & H_{32} - \bar{h}_3 & H_{33} - \bar{h}_3 & \cdots & H_{37} - \bar{h}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ H_{71} - \bar{h}_7 & H_{72} - \bar{h}_7 & H_{73} - \bar{h}_7 & \cdots & H_{77} - \bar{h}_7 \end{bmatrix}, \tag{4.6}$$

Untuk mendapatkan varian terlebih dahulu menghitung jumlah kuadrat (JK), yaitu dengan cara mengalikan G_s pada persamaan (4.6) dengan transposenya sesuai pada persamaan (2.68). Sehingga diperoleh nilai untuk JK sebagai berikut:

$$JK = G_s G_s^T \\ = \begin{bmatrix} H_{11} - \bar{h}_1 & H_{12} - \bar{h}_1 & H_{13} - \bar{h}_1 & \cdots & H_{17} - \bar{h}_1 \\ H_{21} - \bar{h}_2 & H_{22} - \bar{h}_2 & H_{23} - \bar{h}_2 & \cdots & H_{27} - \bar{h}_2 \\ H_{31} - \bar{h}_3 & H_{32} - \bar{h}_3 & H_{33} - \bar{h}_3 & \cdots & H_{37} - \bar{h}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ H_{71} - \bar{h}_7 & H_{72} - \bar{h}_7 & H_{73} - \bar{h}_7 & \cdots & H_{77} - \bar{h}_7 \end{bmatrix}, \\ \begin{bmatrix} H_{11} - \bar{h}_1 & H_{12} - \bar{h}_1 & H_{13} - \bar{h}_1 & \cdots & H_{17} - \bar{h}_1 \\ H_{21} - \bar{h}_2 & H_{22} - \bar{h}_2 & H_{23} - \bar{h}_2 & \cdots & H_{27} - \bar{h}_2 \\ H_{31} - \bar{h}_3 & H_{32} - \bar{h}_3 & H_{33} - \bar{h}_3 & \cdots & H_{37} - \bar{h}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ H_{71} - \bar{h}_7 & H_{72} - \bar{h}_7 & H_{73} - \bar{h}_7 & \cdots & H_{77} - \bar{h}_7 \end{bmatrix}$$

dari persamaan (4.2), maka matriks varian kovarian S dapat diperoleh sebagai berikut:

$$S = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \sigma_{13}^2 & \sigma_{14}^2 & \sigma_{15}^2 & \sigma_{16}^2 & \sigma_{17}^2 \\ \sigma_{21}^2 & \sigma_2^2 & \sigma_{23}^2 & \sigma_{24}^2 & \sigma_{25}^2 & \sigma_{26}^2 & \sigma_{27}^2 \\ \sigma_{31}^2 & \sigma_{32}^2 & \sigma_3^2 & \sigma_{34}^2 & \sigma_{35}^2 & \sigma_{36}^2 & \sigma_{37}^2 \\ \sigma_{41}^2 & \sigma_{42}^2 & \sigma_{43}^2 & \sigma_4^2 & \sigma_{45}^2 & \sigma_{46}^2 & \sigma_{47}^2 \\ \sigma_{51}^2 & \sigma_{52}^2 & \sigma_{53}^2 & \sigma_{54}^2 & \sigma_5^2 & \sigma_{56}^2 & \sigma_{57}^2 \\ \sigma_{61}^2 & \sigma_{62}^2 & \sigma_{63}^2 & \sigma_{64}^2 & \sigma_{65}^2 & \sigma_6^2 & \sigma_{67}^2 \\ \sigma_{71}^2 & \sigma_{72}^2 & \sigma_{73}^2 & \sigma_{74}^2 & \sigma_{75}^2 & \sigma_{76}^2 & \sigma_7^2 \end{bmatrix}. \quad (4.7)$$

2. Membentuk matriks korelasi (ρ). Proses pembentukan matrik korelasi yang pertama yaitu menentukan kovarin matriks dengan menghitung matriks baku terlebih dahulu. Berikut simpangan baku yang diperoleh dengan merujuk pada persamaan (2.73);

$$W_{(7\times7)} = \begin{bmatrix} \sqrt{\sigma_1^2} & \sqrt{\sigma_{12}^2} & \sqrt{\sigma_{13}^2} & \cdots & \sqrt{\sigma_{17}^2} \\ \sqrt{\sigma_{21}^2} & \sqrt{\sigma_2^2} & \sqrt{\sigma_{23}^2} & \cdots & \sqrt{\sigma_{27}^2} \\ \sqrt{\sigma_{31}^2} & \sqrt{\sigma_{32}^2} & \sqrt{\sigma_3^2} & \cdots & \sqrt{\sigma_{37}^2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sqrt{\sigma_{71}^2} & \sqrt{\sigma_{72}^2} & \sqrt{\sigma_{73}^2} & \cdots & \sqrt{\sigma_7^2} \end{bmatrix}, \quad (4.8)$$

$$= \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_1^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_6^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_7^2 \end{bmatrix}$$

dan untuk variabel baku dapat ditentukan dengan cara menghitung invers dari matriks baku sesuai dengan persamaan (2.74). Sebelumnya diberikan matriks C untuk menentukan adjoin pada matriks baku sebagai berikut:

$$C = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} & m_{15} & m_{16} & m_{17} \\ m_{21} & m_{22} & m_{23} & m_{24} & m_{25} & m_{26} & m_{27} \\ m_{31} & m_{32} & m_{33} & m_{34} & m_{35} & m_{36} & m_{37} \\ m_{41} & m_{42} & m_{43} & m_{44} & m_{45} & m_{46} & m_{47} \\ m_{51} & m_{52} & m_{53} & m_{54} & m_{55} & m_{56} & m_{57} \\ m_{61} & m_{62} & m_{63} & m_{64} & m_{65} & m_{66} & m_{67} \\ m_{71} & m_{72} & m_{73} & m_{74} & m_{75} & m_{76} & m_{77} \end{bmatrix},$$

dimana

$$m_{11} = (-1)^{1+1}M_{11} = \begin{vmatrix} \sigma_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_7 \end{vmatrix} = \sigma_2\sigma_3\sigma_4\sigma_5\sigma_6\sigma_7,$$

$$m_{22} = (-1)^{2+2}M_{22} = \begin{vmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_7 \end{vmatrix} = \sigma_1\sigma_3\sigma_4\sigma_5\sigma_6\sigma_7,$$

$$m_{33} = (-1)^{3+3}M_{33} = \begin{vmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_7 \end{vmatrix} = \sigma_1\sigma_2\sigma_4\sigma_5\sigma_6\sigma_7,$$

$$m_{44} = (-1)^{4+4}M_{44} = \begin{vmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_7 \end{vmatrix} = \sigma_1\sigma_2\sigma_3\sigma_5\sigma_6\sigma_7,$$

$$m_{55} = (-1)^{5+5}M_{55} = \begin{vmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_7 \end{vmatrix} = \sigma_1\sigma_2\sigma_3\sigma_4\sigma_6\sigma_7,$$

$$m_{66} = (-1)^{6+6}M_6 = \begin{vmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_7 \end{vmatrix} = \sigma_1\sigma_2\sigma_3\sigma_4\sigma_5\sigma_7,$$

$$m_{77} = (-1)^{7+7} M_7 = \begin{vmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_6 \end{vmatrix} = \sigma_1\sigma_2\sigma_3\sigma_4\sigma_5\sigma_6,$$

dan berdasarkan persamaan (2.77) diperoleh:

$$Adj(W_{7\times7}) = C^T$$
$$= \begin{bmatrix} \sigma_2\sigma_3\sigma_4\sigma_5\sigma_6\sigma_7 & 0 & \cdots & 0 \\ 0 & \sigma_1\sigma_3\sigma_4\sigma_5\sigma_6\sigma_7 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_1\sigma_2\sigma_3\sigma_4\sigma_5\sigma_6 \end{bmatrix},$$

maka diperoleh nilai invers dengan merujuk pada persamaan (2.74), yaitu:

$$(W_{7\times7})^{-1} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sigma_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\sigma_7} \end{bmatrix}.$$

Dari variabel baku yang sudah diperoleh, maka matriks korelasi dapat ditentukan dengan merujuk persamaan (2.78) dan dihasilkan:

$$\rho = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} & \rho_{15} & \rho_{16} & \rho_{17} \\ \rho_{21} & 1 & \rho_{23} & \rho_{24} & \rho_{25} & \rho_{26} & \rho_{27} \\ \rho_{31} & \rho_{32} & 1 & \rho_{34} & \rho_{35} & \rho_{36} & \rho_{37} \\ \rho_{41} & \rho_{42} & \rho_{43} & 1 & \rho_{45} & \rho_{46} & \rho_{47} \\ \rho_{51} & \rho_{52} & \rho_{53} & \rho_{54} & 1 & \rho_{56} & \rho_{57} \\ \rho_{61} & \rho_{62} & \rho_{63} & \rho_{64} & \rho_{65} & 1 & \rho_{67} \\ \rho_{71} & \rho_{72} & \rho_{73} & \rho_{74} & \rho_{75} & \rho_{76} & 1 \end{bmatrix}. \quad (4.9)$$

3. Menentukan nilai *eigen* (λ). Berikut persamaan untuk mendapatkan nilai *eigen* yang merujuk pada persamaan (2.27);

$$|\rho - \lambda I| = 0, \quad (4.10)$$

maka dengan merujuk pada langkah persamaan (2.82) diperoleh nilai *eigen* sebagai berikut:

$$|\rho - \lambda I| = 0$$

$$\left|\begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{17} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{27} \\ \rho_{31} & \rho_{32} & 1 & \cdots & \rho_{37} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{71} & \rho_{72} & \rho_{73} & \vdots & 1 \end{bmatrix} - \begin{bmatrix} \lambda \\ \lambda \\ \lambda \\ \vdots \\ \lambda \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}\right| = 0$$

$$\left|\begin{bmatrix} 1-\lambda & \rho_{12} & \rho_{13} & \rho_{14} & \rho_{15} & \rho_{16} & \rho_{17} \\ \rho_{21} & 1-\lambda & \rho_{23} & \rho_{24} & \rho_{25} & \rho_{26} & \rho_{27} \\ \rho_{31} & \rho_{32} & 1-\lambda & \rho_{34} & \rho_{35} & \rho_{36} & \rho_{37} \\ \rho_{41} & \rho_{42} & \rho_{43} & 1-\lambda & \rho_{45} & \rho_{46} & \rho_{47} \\ \rho_{51} & \rho_{52} & \rho_{53} & \rho_{54} & 1-\lambda & \rho_{56} & \rho_{57} \\ \rho_{61} & \rho_{62} & \rho_{63} & \rho_{64} & \rho_{65} & 1-\lambda & \rho_{67} \\ \rho_{71} & \rho_{72} & \rho_{73} & \rho_{74} & \rho_{75} & \rho_{76} & 1-\lambda \end{bmatrix}\right| = 0 \quad , \qquad (4.11)$$

dimana dari proses perhitungan (4.11) tersebut diperoleh nilai *eigen* sebagai berikut:

$$\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \end{bmatrix}. \qquad (4.12)$$

Untuk vektor *eigen* diperoleh dengan persamaan (2.28), yaitu:

$$(\rho - \lambda)H = 0$$

$$\left(\begin{bmatrix} 1-\lambda & \rho_{12} & \rho_{13} & \cdots & \rho_{17} \\ \rho_{21} & 1-\lambda & \rho_{23} & \cdots & \rho_{27} \\ \rho_{31} & \rho_{32} & 1-\lambda & \cdots & \rho_{37} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{71} & \rho_{72} & \rho_{73} & \cdots & 1-\lambda \end{bmatrix}\right)\begin{bmatrix} H_1 \\ H_2 \\ H_3 \\ \vdots \\ H_7 \end{bmatrix} = 0'$$

dengan mengikuti persamaan (2.85), perhitungan tersebut diselesaikan dengan eselon baris dan diperoleh hasil sebagai berikut:

$$(1-\lambda_1)m_1 + \rho_{12}m_2 + \rho_{13}m_3 + \cdots + \rho_{17}m_7 = 0,$$

$$\rho_{21}m_1 + (1-\lambda_1)m_2 + \rho_{23}m_3 + \cdots + \rho_{27}m_7 = 0,$$

$$\vdots$$

$$\rho_{71}m_1 + \rho_{72}m_2 + \rho_{73}m_3 + \cdots + (1-\lambda_1)m_7 = 0.$$

Dengan mengikuti aturan pada persamaan (2.86) dan (2.87) maka diperoleh vektor normal m_d' dan perhitungan untuk vektor normal tersebut adalah:

$$m_{11} = \frac{m_{11}}{\sqrt{((m_1)^2 + (m_2)^2 + (m_3)^2 + (m_4)^2 + (m_5)^2 + (m_6)^2 + (m_7)^2)}}$$
$$m_{21} = \frac{m_{21}}{\sqrt{((m_1)^2 + (m_2)^2 + (m_3)^2 + (m_4)^2 + (m_5)^2 + (m_6)^2 + (m_7)^2)}}$$
$$m_{31} = \frac{m_{31}}{\sqrt{((m_1)^2 + (m_2)^2 + (m_3)^2 + (m_4)^2 + (m_5)^2 + (m_6)^2 + (m_7)^2)}}$$
$$m_{41} = \frac{m_{41}}{\sqrt{((m_1)^2 + (m_2)^2 + (m_3)^2 + (m_4)^2 + (m_5)^2 + (m_6)^2 + (m_7)^2)}},$$
$$m_{51} = \frac{m_{51}}{\sqrt{((m_1)^2 + (m_2)^2 + (m_3)^2 + (m_4)^2 + (m_5)^2 + (m_6)^2 + (m_7)^2)}}$$
$$m_{61} = \frac{m_{61}}{\sqrt{((m_1)^2 + (m_2)^2 + (m_3)^2 + (m_4)^2 + (m_5)^2 + (m_6)^2 + (m_7)^2)}}$$
$$m_{71} = \frac{m_{71}}{\sqrt{((m_1)^2 + (m_2)^2 + (m_3)^2 + (m_4)^2 + (m_5)^2 + (m_6)^2 + (m_7)^2)}}$$

dan dapat ditulis dalam bentuk matriks sebagai berikut:

$$M = \begin{bmatrix} m_{11} & m_{21} & m_{31} & m_{41} & m_{51} & m_{61} & m_{71} \\ m_{12} & m_{22} & m_{32} & m_{42} & m_{52} & m_{62} & m_{72} \\ m_{13} & m_{23} & m_{33} & m_{43} & m_{53} & m_{63} & m_{73} \\ m_{14} & m_{24} & m_{34} & m_{44} & m_{54} & m_{64} & m_{74} \\ m_{15} & m_{25} & m_{35} & m_{45} & m_{55} & m_{65} & m_{75} \\ m_{16} & m_{26} & m_{36} & m_{46} & m_{56} & m_{66} & m_{76} \\ m_{17} & m_{27} & m_{37} & m_{47} & m_{57} & m_{67} & m_{77} \end{bmatrix}$$

Dari langkah-langkah pembentukan *principal component* diatas maka dapat dibentuk persamaan sebagai berikut:

$$\begin{aligned} H_1 &= m_1'C = m_{11}C_1 + m_{21}C_2 + m_{31}C_3 + m_{41}C_4 + m_{51}C_5 + m_{61}C_6 + m_{71}C_7 \\ H_2 &= m_2'C = m_{12}C_1 + m_{22}C_2 + m_{32}C_3 + m_{42}C_4 + m_{52}C_5 + m_{62}C_6 + m_{72}C_7 \\ H_3 &= m_3'C = m_{13}C_1 + m_{23}C_2 + m_{33}C_3 + m_{43}C_4 + m_{53}C_5 + m_{63}C_6 + m_{73}C_7 \\ H_4 &= m_4'C = m_{14}C_1 + m_{24}C_2 + m_{34}C_3 + m_{44}C_4 + m_{54}C_5 + m_{64}C_6 + m_{74}C_7, \quad (4.13) \\ H_5 &= m_5'C = m_{15}C_1 + m_{25}C_2 + m_{35}C_3 + m_{45}C_4 + m_{55}C_5 + m_{65}C_6 + m_{75}C_7 \\ H_6 &= m_6'C = m_{16}C_1 + m_{26}C_2 + m_{36}C_3 + m_{46}C_4 + m_{56}C_5 + m_{66}C_6 + m_{76}C_7 \\ H_7 &= m_7'C = m_{17}C_1 + m_{27}C_2 + m_{37}C_3 + m_{47}C_4 + m_{57}C_5 + m_{67}C_6 + m_{77}C_7 \end{aligned}$$

persamaan (4.13) merupakan persamaan kombinasi linier dari *principal component* H_d, di mana C_d adalah variabel yang sudah dibakukan dan m_d adalah vektor *eigen* dari matriks korelasi sampel. Model yang diperoleh untuk regresi *principal component* yaitu:

$$Y = \beta_0 + \beta_1 H_1 + \beta_2 H_2 + \beta_3 H_3 + \beta_4 H_4 + \beta_5 H_5 + \beta_6 H_6 + \beta_7 H_7 + \varepsilon. \quad (4.14)$$

4.1.2 Estimasi Parameter Model Regresi *ROBPCA* dengan Metode *M-estimator*

Proses estimasi parameter linier metode regresi *robust principal component analysis* menggunakan metode *robust M-estimator*. Dari model regresi *principal component* pada persamana (4.13) diperoleh bentuk matriks sebagia berikut:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_7 \end{bmatrix} = \begin{bmatrix} 1 & H_{11} & H_{12} & H_{13} & \dots & H_{17} \\ 1 & H_{21} & H_{22} & H_{23} & \dots & H_{27} \\ 1 & H_{31} & H_{32} & H_{33} & \dots & H_{37} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & H_{71} & H_{72} & H_{73} & \dots & H_{77} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_7 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_7 \end{bmatrix}, \quad (4.15)$$

dengan model persamaan

$$Y = H\beta + \varepsilon, \quad (4.16)$$

dari persamaan (4.16) diperoleh bentuk *error* untuk model regresi *principal component* sebagai berikut:

$$\varepsilon = Y - H\beta. \quad (4.17)$$

Proses selanjutnya adalah mengestimasi model regresi *principal component* menggunakan metode *Weighted Least Square* (WLS) yaitu dengan cara menambahkan fungsi pembobot pada modelnya. Misalkan diberikan fungsi pembobot berupa w_i dengan $i = 1,2,3, \dots, 7$ dengan menambahkannya ke dalam persamaan (4.17) maka diperoleh *error* dengan pembobot sebagai berikut:

$$\varepsilon = Yw_i - Hw_i\beta. \tag{4.18}$$

Dari persamaan (4.18), untuk memperoleh parameter β seminimum mungkin sesuai cara pada persamaan (2.6) yaitu dengan meminimumkan jumlah kuadrat *error*, sehingga diperoleh

$$\begin{aligned} S &= \varepsilon'\varepsilon \\ &= (Yw_i - Hw_i\beta)'(Yw_i - Hw_i\beta) \\ &= (Y'w_i' - H'w_i'\beta')(Yw_i - Hw_i\beta) \\ &= Y'w_iY - Y'w_iH\beta - H'\beta'Yw_i + H'\beta w_iH\beta' \\ &= Y'w_iY - (Y'w_iH\beta)' - H'\beta'Yw_i + H'\beta w_iH\beta' \\ &= Y'w_iY - Yw_iH'\beta' - H'\beta'Yw_i + H'\beta w_iH\beta' \\ &= Y'w_iY - 2Yw_iH'\beta' + H'\beta w_iH\beta' \end{aligned} \quad , \tag{4.19}$$

dan melakukan turunan parsial pertama S terhadap β' sesuai dengan persamaan (2.7) maka diperoleh

$$\begin{aligned} \frac{dS}{d\beta'} &= \frac{d(Y'w_iY)}{d\beta'} - \frac{d(2Yw_iH'\beta')}{d\beta'} + \frac{d(H'\beta w_iH\beta')}{d\beta'} \\ &= 0 - 2Yw_i'H' + H'\beta w_iH + (H'\beta'w_iH)' \\ &= 0 - 2Yw_i'H' + H'\beta w_iH + H\beta w_iH' \\ &= -2Yw_i'H' + 2H'w_iH\beta \end{aligned} \quad . \tag{4.20}$$

Pada persamaan (4.20) disamadenganka nol sesuai pada persamaan (2.8), yaitu:

$$\begin{aligned} -2Yw_i'H' + 2H'w_iH\beta &= 0 \\ 2H'w_iH\beta &= 2Yw_i'H' \\ H'w_iH\beta &= Yw_i'H' \\ \beta &= (H'w_iH)^{-1}Yw_i'H' \end{aligned} \quad , \tag{4.21}$$

maka diperoleh estimator untuk persamaan regresi *principal component* terboboti sebagai berikut:

$$\hat{\beta} = (H'w_iH)^{-1}Yw_i'H'. \tag{4.22}$$

dengan w_i merupakan matriks diagonal pembobot yang berukuran 7×7 dan berisi elemen-elemen diagonal berupa $w_1, w_2, w_3, \ldots, w_7$, sehingga dapat ditulis matriks untuk w_i sebagai berikut:

$$w_i = \begin{bmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & w_7 \end{bmatrix}. \tag{4.24}$$

dimana $i = 1,2,3,\dots,7$.

Proses selanjutnya yaitu mencari nilai untuk pembobot w_i dengan menggunakan rumus fungsi pembobot, di mana fungsi pembobot yang digunakan adalah fungsi *Hubert*. Berikut bentuk dari fungsi *Hubert* dari persamaan (2.57):

$$\begin{aligned} w_i &= w(\varepsilon_i^*) \\ &= \frac{\psi(\varepsilon_i^*)}{\varepsilon_i} \\ &= \begin{cases} 1, & |\varepsilon_i| < 0' \\ \dfrac{c}{|\varepsilon_i|}, & |\varepsilon_i| \geq 0 \end{cases} \end{aligned}$$

dimana $c = 1{,}345$. Berikunya dilakukan estimasi dengan *M-estimator*, di mana proses estimasi dilakukan dengan cara IRLS (*iteratively reweighted least square*). Pada estimasi nilai yang diperoleh yaitu w_i yang akan berubah dalam setiap iterasinya, sehingga didapatkan $\hat{\beta}^{(0)}, \hat{\beta}^{(1)}, \hat{\beta}^{(2)}, \dots, \hat{\beta}^{(l)}, \hat{\beta}^{(l+1)}$. Misal diberikan estimasi awal berupa $\hat{\beta}^{(0)}$, maka:

$$\hat{\beta}^{(0)} = \left(H' w_i^{(0)}\, H\right)^{-1} H' w_i^{(0)} Y, \tag{4.25}$$

dengan menghitung pembobot dari $w_i^{(0)}$ dari $\hat{\beta}^{(0)}$, yaitu:

$$w_i^{(0)} = \frac{\psi(|\varepsilon_i|)}{|\varepsilon_i|}$$
$$= \frac{\psi\left(\frac{\varepsilon_i}{\hat{\sigma}}\right)}{\frac{\varepsilon_i}{\hat{\sigma}}}$$
$$= \frac{\psi\left(\frac{Y - H\hat{\beta}^{(0)}}{\frac{MAD(H)}{0{,}6745}}\right)}{\frac{Y - H\hat{\beta}^{(0)}}{\frac{MAD(H)}{0{,}6745}}}.$$

Estimasi satu langkah selanjutnya diperoleh dari $w_i^{(0)}$ dengan $\hat{\beta}^{(1)}$ sebagai pengggganti dari $\hat{\beta}^{(0)}$, sehingga diperoleh:

$$\hat{\beta}^{(1)} = \left(H'w_i^{(0)}H\right)^{-1}H'w_i{}^{(0)}Y.$$

Parameter seterusnya sampai iterasi ke-l dapat dinyatakan dengan $w^{(l-1)}$, yaitu:

$$w_i^{(l-1)} = \frac{\psi\left(\frac{Y - H\hat{\beta}^{(l-1)}}{\frac{MAD(H)}{0{,}6745}}\right)}{\frac{Y - H\hat{\beta}^{(l-1)}}{\frac{MAD(H)}{0{,}6745}}},$$

dan diperoleh

$$\hat{\beta}^{(l)} = \left(H'w_i^{(l-1)}H\right)^{-1}H'w_i^{(l-1)}Y, \qquad (4.26)$$

dan untuk $w_i^{(l)}$ pembobot akan diperoleh estimator bagi β sebagai berikut:

$$\hat{\beta}^{(l+1)} = \left(H'w_i^{(l)}H\right)^{-1}H'w_i^{(l)}Y, \qquad (4.27)$$

iterasi akan selesai apabila sudah diperoleh estimasi yang bersifat konvergen, yaitu pada saat selisih antara nilai $\hat{\beta}^{(l)}$ dan $\hat{\beta}^{(l+1)}$ mendekati nol atau kurang dari 0,001 atau 0,1%.

4.2 Implementasi Regresi ROBPCA pada Data Kemiskinan Jawa Timur tahun 2019

4.2.1 Deskripsi Data

Penelitian ini menggunakan data sekunder tentang indeks keparahan kemiskinan Jawa Timur pada tahun 2019 dan bersumber dari Badan Pusat Statistika (BPS) dengan 38 kota/kabupaten. Variabel yang digunakan didalam penelitian ini adalah variabel terikat jumlah penduduk miskin (Y) dengan variabel bebas kepadatan penduduk (X_1), tingkat pengangguran terbuka (X_2), produk domestik regional bruto (X_3), indeks keparahan kemiskinan (X_4), rata-rata lama sekolah (X_5), indeks pembangunan manusia (X_6) dan pengeluaran per kapita riil (X_7).

Statistik deskriptif dalam penelitian ini bertujuan untuk mengetahui faktor-faktor yang mempengaruhi jumlah penduduk miskin di Jawa Timur tahun 2019. Statistik deskriptif dapat dilihat dari ukuran pemusatan dan penyebaran data yang diperoleh melalui *software* SPSS 26 dengan tabel yang diperoleh sebagai berikut:

Tabel 4.1 Deskriptif Data

	N	Minimum	Maksimum	Rata-rata	Standart Deviasi	Variansi
Y	38	6,63	246,60	108,218	66,511	4423,725
X1	38	279	8262	1922,34	2179,87	4751825,745
X2	38	0,95	6,04	3,7534	1,112	1,237
X3	38	5	411	44,05	69,028	4764,862
X4	38	0,05	1,080	0,383	0,236	0,056
X5	38	4,55	11,13	7,798	1,616	2,611
X6	38	62	82	71,69	5,142	26,444
X7	38	8,72	178,54	96,479	49,075	2408,383

Berdasarkan tabel (4.1) dapat diketahui bahwa rata-rata untuk untuk jumlah penduduk miskin di Jawa Timur tahun 2019 adalah 108.2176 jiwa. Besarnya angka tersebut disebabkan beberapa faktor penduga yang menjadi pengaruh terhadap jumlah penduduk miskin di Jawa Timur. Faktor-faktor tersebut adalah kepadatan penduduk, tingkat pengangguran terbuka, produk domestik regional bruto, indeks keparahan kemiskinan, rata-rata lama sekolah, indeks pembangunan manusia dan pengeluaran per kapita riil.

4.2.2 Hasil Estimasi Parameter Menggunakan Metode OLS

Berdasarkan data indeks keparahan kemiskinan di Jawa Timur tahun 2019 terdapat delapan faktor yang diduga menjadi pengaruh indeks keparahan kemiskinan. Berikut model regresi dengan metode OLS dari data pada yang tercantum pada (Lampiran 1):

$$\begin{aligned} Y = 179{,}712 - 0{,}008X_1 + 4{,}217X_2 + 0{,}438X_3 + 86{,}062X_4 \\ -18{,}613X_5 + 0{,}374X_6 - 0{,}058X_7 \end{aligned}, \qquad (4.28)$$

Persamaan (4.25) terdiri dari koefisien regresi kepadatan penduduk sebagai variable X_1 sebesar -0,008, yang artinya bahwa setiap peningkatan satu persen kepadatan penduduk maka akan menurunkan jumlah penduduk miskin sebesar 0,8%. Kedua yaitu koefisien tingkat pengangguran terbuka sebagai variabel X_2 sebesar 4,217, artinya bahwa setiap peningkatan satu persen tingkat pengangguran terbuka maka akan menurunkan jumlah penduduk miskin sebesar 421,7%. Ketiga yaitu koefisien produk domestik regional bruto sebagai variabel X_2 sebesar 0,438, artinya bahwa setiap peningkatan satu persen produk domestik regional bruto maka akan menurunkan jumlah penduduk miskin sebesar 0,438. Keempat yaitu koefisien indeks keparahan kemiskinan sebagai variabel X_4 sebesar 86,062, artinya bahwa setiap peningkatan satu persen jumlah penduduk miskin maka akan menurunkan jumlah penduduk miskin sebesar 8606,2%. Kelima yaitu koefisien rata-rata lama sekolah sebagai variabel X_5 sebesar -18,613, artinya setiap peningkatan satu persen untuk rata-rata lama sekolah maka akan menurunkan jumah penduduk miskin sebesar 1861,3%. Keenam yaitu koefisien indeks pembangunan manusia sebagai variabel X_6 sebesar 0,374, artinya setiap peningkatan satu persen untuk indeks pembangunan manusia maka akan menurunkan jumlah penduduk miskin sebesar 37,4%. Ketujuh yaitu koefisien pengeluaran per kapita riil sebagai variabel X_7 sebesar -0,058, artinya setiap peningkatan satu persen pengeluaran per kapita riil maka akan menurunkan jumlah penduduk miskin sebesar 5,8%. Sedangkan untuk konstanta sebesar 179,712 menyatakan jika ketujuh koefisien tersebut dalam keadaan konstanta nol maka jumlah penduduk miskin sebesar sebesar 179.712.000 jiwa.

4.2.3 Uji Asumsi Data

Uji asumsi dijalankan untuk menguji model regresi linier berganda yang sudah didapatkan pada persamaan (4.28), yaitu dengan mengidentifikasi terpenuhinya asumi klasik bagi model atau tidak untuk mendapatkan kevalidan model.

1. Uji Normalitas

   Pengujian normalitas dilakukan dengan uji *Kolmogorov-smirnov*. Berikut hipotesis untuk uji normalitas:

   H_0 : Residual berdistribusi normal,

   H_1 : Residual tidak berdistribusi normal.

   Hasil uji normalitas dengan uji *Kolmogorov-smirnov* adalah sebagai berikut:

Tabel 4.2 Uji *One-Sample Kolmogorov-Smirnov*

Nilai Statistik Uji	P_{value}	Keputusan
0,114	0,200	Terima H_0

Tabel (4.2) diperoleh nilai statistik uji sebesar 0,114 dengan nilai $D_{tabel(\alpha=0,05)}$ sebesar 0,210, maka $D \leq D_{tabel(\alpha=0,05)}$ dan diperoleh $P_{value} > \alpha = 0{,}05$. Sehingga dapat diambil kesimpulan bahwa tolak H_1 dan terima H_0, artinya *residual* berdistribusi normal.

2. Uji Multikolinieritas

Pengujian selanjutnya adalah pendeteksian adanya multikolinieritas dalam model. Berikut hasil dari pedeteksian multikolinieritas:

Tabel 4.3 Deteksi Multikolinieritas pada Model Regresi Linier Berganda

Variabel	Statistik Korelasi		Keterangan
	Toleransi	VIF	
X_1	0,354	2,826	Tidak terjadi multikolinieritas
X_2	0,558	1,793	Tidak terjadi multikolinieritas
X_3	0,632	1,582	Tidak terjadi multikolinieritas
X_4	0,478	2,092	Tidak terjadi multikolinieritas
X_5	0,033	29,861	Multikolinieritas
X_6	0,033	29,999	Multikolinieritas
X_7	0,494	2,025	Tidak terjadi multikolinieritas

Pada tabel (4.3) terlihat bahwa variabel yang terdeteksi multikolinieritas adalah variabel X_5 (Indeks pembangunan manusia) dan X_6 (Rata-rata lama sekolah) dengan nilai $VIF \geq 10$ atau nilai $tollerance < 0{,}1$, artinya terjadi multikolinieritas pada model.

3. Uji Heterokedastisitas

Uji heterokedastisitas bertujuan untuk menguji adanya perbedaan nilai varian dari residual pada suatu pengamatan ke pengamatan yang lain. Apabila nilai varian yang dihasilkan dari residual tetap maka memenuhi sifat homokedastisitas dan terbebaskan dari heterokedastisitas. Pengujian dilakukan menggunakan uji Gletser dengan hasil sebagai berikut:

Tabel 4.4 Hasil Uji Heterokedastisitas pada Model Regresi Linier Berganda

Variabel	P_{value}	Keputusan	Keterangan
X_1	0,112	Terima H_0	Homokedastisitas
X_2	0,591	Terima H_0	Homokedastisitas
X_3	0,001	Tolak H_0	Heterokedastisitas
X_4	0,037	Tolak H_0	Heterokedastisitas
X_5	0,400	Terima H_0	Homokedastisitas
X_6	0,957	Terima H_0	Homokedastisitas
X_7	0,760	Terima H_0	Homokedastisitas

Tabel (4.4) menunjukkan bahwa terdapat variabel yang mengandung heterekedastisitas dengan $P_{value} < \alpha = 0{,}05$, yaitu variabel X_3 (PDRB) dan variabel X_4 (Indeks keparahan kemiskinan). Untuk menghilangkan variabel yang mengandung heterokedastisitas maka akan dilakukan dengan metode WLS.

4.2.4 Uji Signifikansi Parameter dari Metode OLS

Uji signifikansi dilakukan estimasi parameter, untuk mengukur ketepatan model maka dilakukan uji signifikansi koefisien regresi secara individu seperti berikut:

1. Uji Signifikansi Model Regresi Secara Serentak (Uji F)

   Pengaruh variabel bebas secara serentak dapat ditentukan dengan menggunakan uji F. Adapun hipotesis uji parameter model secara serentak adalah sebagai berikut:

   $H_0 = \beta_1 = \beta_2 = \cdots = \beta_7 = 0,$

   $H_1 =$ Minimal ada satu $\beta_i \neq 0, \exists i = 1,2,3, \dots, 7,$

   didapatkan F_{hitung} dengan α sebesar 0,05 sebagai berikut:

Tabel 4.5 Hasil Uji Signifikansi Parameter Model Regresi Linier Berganda dengan Uji ***F***

Model	Sum of Squares	df	Mean Square	F_{hitung}	F_{tabel}
Regresi	117451,004	6	19575,167	13,127	2,41
Residual	46226,822	31	1491,188		
Total	163677,826	37			

   Pada tabel (4.5) diperoleh nilai F_{hitung} sebesar 13,127 dan F_{tabel} sebesar 2,41, sehingga $F_{hitung} > F_{tabel}$ maka kesimpulan hipotesis yang didapatkan adalah H_0 ditolak. Penolakan H_0 artinya secara signifikan variabel kepadatan penduduk, TPT, PDRB, IKK, rata-rata lama sekolah, IPM dan pengeluaran per kapita riil berpengaruh terhadap jumlah penduduk miskin di Jawa Timur tahun 2019.

2. Uji Signifikansi Model Regresi Secara Individual (Uji t)

   Pengujian parameter berikutnya dilakukan secara individual dengan uji t, yaitu untuk mengukur parameter yang berpengaruh secara signifikan terhadap variabel bebas dalam data. Adapun hipotesis uji parameter model secara individual adalah sebagai berikut:

   $H_0 = \beta_i = 0, \forall i = 1,2,3, \dots, 7,$

$H_1 = \beta_i \neq 0, \exists i = 1,2,3, \dots ,7.$

Statistika uji menggunakan uji t dan hasil dari pengujian signifikansi parameter secara individual adalah sebagai berikut:

Tabel 4.6 Hasil Uji Siginifikansi Parameter Model Regresi Linier Berganda dengan Uji $\boldsymbol{t}$

Variabel	t_{hitung}	P_{value}	Keputusan
X_1	-1,640	0.112	Tidak Signifikan
X_2	0,544	0.591	Tidak Signifikan
X_3	3,727	0.001	Signifikan
X_4	2,177	0.037	Signifikan
X_5	-0,854	0.400	Tidak Signifikan
X_6	0,054	0.957	Tidak Signifikan
X_7	-0,309	0.760	Tidak Signifikan

Berdasarkan tabel (4.6) dengan melihat kriteria uji untuk uji t yaitu apabila $|t_{hitung}| > t_{(\frac{\alpha}{2};n-k-1)}$ atau $P_{value} < \alpha$ maka tolak H_0, sebaliknya jika $|t_{hitung}| < t_{(\frac{\alpha}{2};n-k-1)}$ atau $P_{value} > \alpha$ maka terima H_0. Uji ini diperoleh t_{tabel} sebesar 2,045, maka dari hasil pada tabel (4.7) dapat diambil keputasan bahwa variabel yang memiliki pengaruh secara signifikansi terhadap variabel jumlah penduduk miskin yaitu variabel X_3 (PDRB) dan X_4 (Indeks keparahan kemiskinan) dengan $t_{hitung} > t_{tabel=2,045}$ atau $P_{value} < \alpha = 0,05$.

4.2.5 Uji Ukuran Kebaikan Model

Tahap selanjutnya adalah pengujian kebaikan model untuk model regresi linier berganda pada persamaan (4.1). Pengujian dilakukan dengan melihat nilai koefisien determinasi yang disesuaikan dan nilai *residual standard error* (RSE). Melalui pembentukan model dengan software SPSS 26 diperoleh nilai koefisien determinasi yang disesuaikan (*Adjusted R-square*) sebesar 0,653 dan nilai *residual standard error* sebesar 39,192, artinya faktor-faktor yang menjadi penduga jumlah penduduk miskin memiliki nilai persentase sebesar 65,3% dan sisa sebesar 34,7% untuk faktor-faktor lain yang mempengaruhi jumlah penduduk miskin di luar model.

4.2.6 *Principal Component Analysis*

Langkah pertama dalam PCA adalah menghitung nilai *eigen* dan vektor *eigen* berdasarkan matriks korelasi. Sebelum langkah-langkah dalam proses PCA dijalankan, maka terlebih dahulu membentuk matriks untuk data kemiskinan Jawa Timur tahun 2019. Berikut matriks yang dibentuk untuk nila-nilai dari data tersebut dengan bentuk sesuai persamaan (4.1):

$$H = \begin{bmatrix} 399 & 0,95 & 11,041 & \cdots & 90,33 \\ 667 & 3,58 & 14,297 & \cdots & 98,83 \\ 607 & 3,43 & 12,78 & \cdots & 98,65 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1,507 & 2,48 & 11,787 & \cdots & 128,7 \end{bmatrix}$$

Selanjutnya adalah menghitung korelasi antar variabel. Berikut matriks korelasi yang diperoleh dari korelasi antar variabel dan sesuai dengan persamaan (4.8), yaitu:

$$r = \begin{bmatrix} 1 & 0,441 & 0,430 & -0,452 & 0,777 & 0,773 & 0,562 \\ 0,441 & 1 & 0,479 & 0,062 & 0.432 & 0,457 & 0,298 \\ 0,430 & 0,479 & 1 & -0.099 & 0,342 & -0,406 & 0,286 \\ -0,452 & 0,062 & -0,099 & 1 & -0,562 & -0,542 & -0,571 \\ 0,777 & 0,432 & 0,342 & -0,562 & 1 & 0,980 & 0,618 \\ 0,773 & 0,457 & 0,406 & -0,542 & 0,980 & 1 & 0,636 \\ 0,562 & 0,298 & 0,286 & -0,571 & 0,618 & 0,636 & 1 \end{bmatrix}$$

Langkah berikutnya uji asumsi untuk *principal component* dengan uji *Kaiser Mayer-Olkin* (KMO). Berikut hasil uji KMO dengan *software* SPSS 26:

Tabel 4.7 Hasil Uji *Kaiser Mayer-Olkin* (KMO) dan Uji *Bartlett's*

Uji KMO	Uji *Bartlett's*	P_{value}
0,770	208,338	0,000

Tabel (4.7) menunjukkan hasil dari uji KMO yaitu $0{,}770 > \alpha = 0{,}05$, maka terima H_0 artinya variabel-variabel memenuhi untuk dianalisis lebih lanjut dengan metode *Principal Component Anaysis*. Sedangkan untuk nilai uji *Bartlet* diperoleh $208{,}338 > \chi^2_{(52,1923)}$ dan $P_{value} < \alpha = 0{,}05$, maka tolak H_0 artinya terdapat korelasi antar variabel bebas. Maka dari hasil uji tersebut dapat diambil kesimpulan bahwa proses analisis untuk data penelitian dapat dilanjutkan.

Proses analisis berikutnya adalah menghitung nilai *eigen*, vektor *eigen* dan nilai proporsi. Nilai *eigen* dan nilai proporsi digunakan untuk ketepatan *principal component* yang akan dipilih. *Principal component* terpilih harus memenuhi nilai *eigen* λ lebih dari 1 ($\lambda_j \geq 1$). Berikut *output* nilai *eigen* dan nilai proporsi dari metode *principal component analysis* dengan *software* SPSS 26:

Tabel 4.8 Hasil untuk Nilai *Eigen* dan Nilai Proporsi

Komponen	Nilai *Eigen*	Proporsi Varian	Proporsi Kumulatif
1	4,041	57,732	57,732
2	1,286	18,369	76,101
3	0,601	8,587	84,688
4	0,477	6,821	91,509
5	0,308	4,394	95,903
6	0,270	3,853	99,756
7	0,017	0,244	100,000

Tabel (4.8) menunjukkan bahwa *principal component* terpilih ada pada komponen pertama dan kedua dengan nilai *eigen* sebesar 4,041 dan 1,286, artinya dengan 2 komponen bisa menjelaskan sebesar 4,041 dan 1,286 dari total *communalities* dengan persentase proporsi kumulatif sebagai berikut:

$$\frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \cdots + \lambda_7} \times 100\% = \frac{4{,}041 + 1{,}286}{4{,}041 + 1{,}286 + \cdots + 0{,}017} \times 100\% = 76{,}101\%.$$

Sehingga dapat disimpulkan bahwa dua *principal component* terpilih dapat menjelaskan proporsi data sebesar 76,101%, artinya mamiliki pengaruh yang sangat besar terhadap jumlah penduduk miskin di Jawa Timur pada tahun 2019. Berikut daftar kelompok *principal component* terpilih:

Tabel 4.9 *Principal Component* Terpilih

Komponen	m_1	m_2
C_1	0,864	0,043
C_2	0,544	0,690
C_3	0,530	0,573
C_4	-0,618	0,642
C_5	0,930	-0,099
C_6	0,942	-0,051
C_7	0,764	-0,234

Berdasarkan tabel (4.9) dengan C_j sebagai variabel terstandarisasi dan m_j merupakan vektor *eigen* dari matriks korelasi, sehingga dapat ditunjukkan komponen yang memiliki nilai koefisien tinggi untuk kelompok pertama, adalah variabel C_1 (Kepadatan penduduk), variabel C_2 (TPT), variabel C_3 (PDRB), variabel C_5 (Rata-rata lama sekolah), variabel C_6 (IPM) dan variabel C_7 (Pengeluaran per kapita riil). Sedangkan untuk kelompok kedua yang memiliki koefisien tertinggi adalah variabel C_2 (TPT), C_3 (PDRB) dan C_4 (IKK). Kesimpulan dari hasil tersebut yaitu, variabel C_1, C, C_3, C_5, C_6 dan C_7 memiliki nilai koefisien yang tinggi sebagai faktor pengaruh jumlah penduduk miskin di

Provinsi Jawa Timur tahun 2019, sedangkan untuk variabel C_4 memiliki nilai koefisien terkecil sebagai faktor pengaruh jumlah penduduk miskin di Provinsi Jawa Timur tahun 2019.

Kombinasi linier variabel asal yang terbentuk dari *principal component* terpilih sesuai dengan persamaan (4.13) adalah:

$$\begin{aligned} H_1 = {} & 0{,}864C_1 + 0{,}544C_2 + 0{,}530C_3 - 0{,}618C_4 + 0{,}930C_5 \\ & +0{,}942C_6 + 0{,}764C_7 \end{aligned}, \tag{4.27}$$

$$\begin{aligned} H_2 = {} & 0{,}043C_1 + 0{,}690C_2 + 0{,}573C_3 + 0{,}642C_4 - 0{,}099C_5 \\ & -0{,}051C_6 - 0{,}234C_7 \end{aligned}. \tag{4.28}$$

4.2.7 *Principal Component Regression* (PCR)

Principal component regression dibentuk dari *principal component* terpilih dari proses PCA sebelumnya. Pemodelan untuk PCR yaitu dengan membentuk model regresi linier berganda pada umumnya menggunakan metode *ordinary least square* (OLS), dimana *principal component* H_1 dan H_2 sebagai variabel bebas dengan Y sebagai variabel terikat. Berikut persamaan regresi linier untuk PCR:

$$Y_{PCR} = \beta_0 + \beta_1 H_1 + \beta_2 H_2. \tag{4.29}$$

Dari hasil perhitungan menggunakan *software* SPSS 26 (Lampiran 9), diperoleh model PCR sebagai berikut:

$$Y_{PCR} = 108{,}218 - 50{,}925H_1 + 12{,}095H_2. \tag{4.30}$$

Selanjutnya dilakukan uji signifikansi untuk model PCR menggunakan uji F dan uji t. Berikut hasil uji signifikansi dengan uji F:

Tabel 4.10 Uji Signifikansi Parameter Model PCR dengan Uji ***F***

Model	Sum of Squares	df	Mean Square	F	P_{value}
Regresi	101368,391	2	50684,195	28,470	0,000
Residual	62309,435	35	1780,270		
Total	163677,826	37			

Tabel (4.10) menunjukkan nilai F_{hitung} yang diperoleh sebesar 28,470 dan F_{tabel} sebesar 3,27, maka $F_{hitung} > F_{tabel}$. Sehingga kesimpulan yang dapat diambil adalah tolak H_0, artinya secara signifikan variabel H_1 dan H_2 sebagai variabel *principal component* memiliki pengaruh terhadap variabel terikat Y dalam model PCR.

Sedangkan untuk hasil uji signifikansi dengan uji t, yaitu:

Tabel 4.11 Hasi Uji Signifikansi Parameter Model PCR dengan Uji ***t***

Variabel	t_{hitung}	P_{value}	Keputusan
H_1	-7,342	0,000	Terima H_0
H_2	1,744	0,090	Tolak H_0

Tabel (4.11) diperoleh hasil t_{hitung} untuk variabel H_1 sebesar -7,342 dan variabel H_2 sebesar 1,744 dengan nilai t_{tabel} diperoleh sebesar 1,690. Kesimpulan yang dapat diambil adalah variabel H_1 memiliki nilai $t_{hitung} < t_{tabel}$ maka terima H_0, artinya variabel H_1 tidak memiliki pengaruh terhadap variabel terikat Y. Sedangkan untuk variabel H_2 memiliki nilai $t_{hitung} > t_{tabel}$ maka tolak H_0, artinya variabel H_2 memiliki pengaruh terhadap variabel terikat Y dalam model PCR.

Berdasarkan hasil uji signifikansi maka dapat disimpulkan bahwa variabel H_2 berupa variabel TPT, PDRB dan IKK memiliki pengaruh yang kuat terhadap jumlah penduduk miskin di Jawa Timur pada tahun 2019. Variabel H_2 yang sudah teruji, model PCR menjadi sebagai berikut:

$$Y_{PCR} = 108{,}218 + 12{,}095H_2, \qquad (4.31)$$

kemudian mensubtitusikan kombinasi linier pada persamaan (4.28) ke persamaan (4.31), yaitu:

$$\begin{aligned} Y_{PCR} &= 108{,}218 + 12{,}095(\&0{,}043C_1 + 0{,}690C_2 + 0{,}573C_3 + 0{,}642C_4 \\ &\quad -0{,}099C_5 - 0{,}051C_6 - 0{,}234C_7) \\ &= 108{,}218 + 0{,}520C_1 + 8{,}346C_2 + 6{,}930C_3 + 7{,}765C_4 - 1{,}197C_5 \\ &\quad 0{,}617C_6 - 2{,}830C_7 \end{aligned}$$

Maka diperoleh model persamaan sebagai berikut:

$$\begin{aligned} Y_{PCR} = {} & 108{,}218 + 0{,}520C_1 + 8{,}346C_2 + 6{,}930C_3 + 7{,}765C_4 - 1{,}197C_5 \\ & -0{,}617C_6 - 2{,}830C_7 \end{aligned}. \qquad (4.32)$$

Persamaan PCR (4.32) ditransformasikan kembali ke variabel asal (Lampiran 8), sehingga diperoleh model persamaan PCR untuk kemiskinan Jawa Timur pada tahun 2019 sebagai berikut:

$$\begin{aligned} Y_{PCR} = {} & 81{,}596 + 0{,}0002X_1 + 7{,}505X_2 + 0{,}1X_3 + 32{,}9X_4 \\ & - 0{,}741X_5 - 0{,}12X_6 - 0{,}058X_7 \end{aligned}. \qquad (4.33)$$

Selanjutnya dilakukan deteksi multikolinieritas pada model *principal component regression*, berikut hasilnya:

Tabel 4.12 Deteksi Multikolinieritas pada Model PCR

Variabel	Toleransi	VIF	Keterangan
H_1	1,000	1,000	Tidak terjadi multikolinieritas
H_2	1,000	1,000	Tidak terjadi multikolinieritas

Tabel (4.12) diperoleh nilai *tollerance* $1 > 0{,}1$ dan nilai VIF $1 < 10$, artinya tidak ada multikolinieritas pada model PCR atau model PCR terbebaskan dari kolinieritas.

Nilai koefisien determinasi (*R-square*) pada model diperoleh sebesar 0,787, artinya bahwa penelitian ini mampu menjelesakan sebesar 78,70% untuk varian Y dengan faktor-faktor yang mempengaruhinya di dalam model PCR dan nilai *residual standart error* (RSE) diperoleh sebesar 42,193.

4.2.8 Identifikasi *Outlier*

Identifikasi *outlier* untuk metode *principal component analysis* dapat dilakukan dengan menghitung nilai *DFFIT's* dan nilai *leverage.* Dengan menggunakan nilai *DFFIT's outlier* akan terdeteksi dalam data apabila nilai $DFFIT's > 2\sqrt{\frac{k}{n}}$, dimana k merupakan jumlah variabel dan n jumlah data maka diperoleh $2\sqrt{\frac{k}{n}} = 2\sqrt{\frac{7}{38}} = 0{,}8584$. Maka dengan melihat hasil pada (Lampiran 9), data yang terdeteksi *outlier* ada pada data ke-9 (Kab. Jember) dan data ke-35 (Kota Mojokerto).

Sedangkan untuk nilai dari *leverage*, terdapat *outlier* apabila nilai yang diperoleh lebih besar dari nilai *cutoff*. Nilai *cutoff* yang diperoleh untuk panelitian ini adalah $2 \times \frac{k}{n} = 2 \times \frac{7}{38} = 0{,}3684$, maka dengan melihat tabel pada (Lampiran 9) data yang terdeteksi *outlier* yaitu data ke- 25 (Kab. Gresik), ke-26 (Kab, Bangkalan), ke-32 (Kota Malang), ke-35 (Kota Pasuruan) dan ke-37 (Kota Madiun). Oleh karena itu, dapat disimpulkan bahwa data tentang kemiskinan pada penelitian ini mengandung mengandung *outlier* dan proses penyelesaian *outlier* dapat dilakukan dengan menggabungkan metode PCR dan metode *robust* yang dikenal dengan metode regresi *robust principal component analysis* (ROBPCA).

4.2.9 Estimasi Persamaan Regresi *Robust Principal Component Analysis* (ROBPCA) dengan *M-Estimator*

Metode regresi *robust principal component analysis* (ROBPCA) merupakan gabungan dari dua metode yaitu metode PCA dan metode *robust*. Metode PCA yang sudah dimodelkan dengan metode OLS dan dinamakan dengan model PCR akan dimodelkan kembali menggunakan metode WLS, dimana dengan metode ini akan ditambahkan pembobot untuk model PCR. Model PCR yang sudah terboboti dan terestimasi dengan *M-estimator* dikenal dengan model ROBPCA.

Pembobot yang digunakan adalah fungsi pembobot *Hubert* pada *M-estimator*, maka kriteria pembobot yang diguakan sesuai dengan persamaan (2.64) dan hasil perhitungan pembobot dengan *software* SPSS 26 terlampir pada (Lampiran 11) . Dengan pembobot yang sudah didapatkan, maka dapat dihitung koefisian regresi iterasi pertama dari metode WLS sesui dengan persamaan (4.25) sebagai berikut:

$$\hat{\beta}^{(0)} = \left(H'w^{(0)}\, H\right)^{-1} H'w^{(0)}Y$$

dimana:

$$H = \begin{bmatrix} 1 & 0,1472 & 1,78 \\ 1 & 0,172 & 0,371 \\ \vdots & \vdots & \vdots \\ 1 & 1,2814 & 1,272 \end{bmatrix}, \qquad Y = \begin{bmatrix} 75,86 \\ 83,97 \\ 76,44 \\ \vdots \\ 7,89 \end{bmatrix},$$

$$H' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0,1472 & 0,172 & \cdots & 1,2814 \\ 1,78 & 0,371 & \cdots & 1,272 \end{bmatrix},$$

$$w^{(0)} = \begin{bmatrix} 0,0031 & 0 & \cdots & 0 \\ 0 & 0,0013 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 0,0032 \end{bmatrix},$$

maka diperoleh:

$\hat{\beta}^{(0)} = \left(H'w^{(0)}\, H\right)^{-1} H'w^{(0)}Y$

$=$

$$\left(\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0{,}1472 & 0{,}172 & \cdots & 1{,}2814 \\ 1{,}78 & 0{,}371 & \cdots & 1{,}272 \end{bmatrix}\begin{bmatrix} 0{,}0031 & 0 & \cdots & 0 \\ 0 & 0{,}0013 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 0{,}0032 \end{bmatrix}\begin{bmatrix} 1 & 0{,}1472 & 1{,}78 \\ 1 & 0{,}172 & 0{,}371 \\ \vdots & \vdots & \vdots \\ 1 & 1{,}2814 & 1{,}272 \end{bmatrix}\right)^{-1}$$

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0{,}1472 & 0{,}172 & \cdots & 1{,}2814 \\ 1{,}78 & 0{,}371 & \cdots & 1{,}272 \end{bmatrix}\begin{bmatrix} 0{,}0031 & 0 & \cdots & 0 \\ 0 & 0{,}0013 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 0{,}0032 \end{bmatrix}\begin{bmatrix} 1 & 0{,}1472 & 1{,}78 \\ 1 & 0{,}172 & 0{,}371 \\ \vdots & \vdots & \vdots \\ 1 & 1{,}2814 & 1{,}272 \end{bmatrix}$$

$$= \begin{bmatrix} 108{,}21766 \\ -50{,}92539 \\ 12{,}09509 \end{bmatrix},$$

sehingga nilai parameter untuk model estimasi awal yang diperoleh adalah $\hat{\beta}_0 = 108{,}218$, $\hat{\beta}_1 = -50{,}925$ dan $\hat{\beta}_2 = 12{,}095$. Berikut model ROBPCA untuk etimasi parameter awal dengan menggunakan WLS dan pembobot *Huber*:

$$Y = 108{,}218 - 50{,}925H_1 + 12{,}095H_2$$

Iterasi selanjutnya dari metode WLS dengan pembobot *Huber* diperoleh pada tabel sebagai berikut:

Tabel 4.13 Hasil Iterasi Estimasi Parameter menggunakan Fungsi Pembobot *Hubert* pada *M-estimator*

Iterasi	Koefisien	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
0		108,21766	-50,92539	12,09509
1	20,782	105,12286	-51,98637	14,60869
2	1,867	104,63437	-52,13879	14,33600
3	0,060	104,57150	-52,14797	14,33279
4	0,019	104,56285	-52,14753	14,33007
5	0,001	104,56178	-52,14736	14,33008
6	0,000	104,56164	-52,14732	14,33006

Tabel (4.13) menunjukkan iterasi berhenti pada iterasi keenam, artinya nilai parameter sudah konvergen dan nilai yang diperoleh mendekati nol atau kurang dari 0,001. Sehingga diperoleh model persamaan untuk regresi ROBPCA sebagai berikut:

$$Y_{ROBPCA} = 104{,}562 - 52{,}147H_1 + 14{,}330H_2, \tag{4.34}$$

Pengaruh variabel bebas pada model ROBPCA akan diukur dengan uji F atau uji secara serentak. Berikut hasil dari pengujian variabel bebas pada model ROBPCA dengan *software* NCSS:

Tabel 4.14 Hasil Uji Signifikansi Parameter Model ROBPCA dengan Uji $\boldsymbol{F}$

Model	Sum of Squares	df	Mean Square	F	P_{value}
Regresi	97365,56	2	48682,78	42,570	0,000
Residual	40025,56	35	1143,59		
Total	137391,12	37			

Berdasarkan tabel (4.14) diperoleh nilai F_{hitung} sebesar 42,570 dan untuk F_{tabel} sebesar 3,27, sehinngga $F_{hitung} > F_{tabel}$ maka tolak H_0 yang artinya secara signifikansi variabel bebas pada model ROBPCA berpengaruh terhadap tingkat jumlah kemiskinan di Jawa Timur pada tahun 2019. Variabel tersebut adalah kepadatan penduduk, TPT, PDRB, IKK, rata-rata lama sekolah, IPM dan pengeluaran perkapita riil.

Selanjutnya dilakukan uji signifikansi dengan uji t atau uji secara individual, berikut hasilnya:

Tabel 4.15 Hasi Uji Signifikansi Parameter Model ROBPCA dengan Uji $\boldsymbol{t}$

Variabel	t_{hitung}	Signifikansi	Keputusan
H_1	- 8,991	0,000	Terima H_0
H_2	2,527	0,022	Tolak H_0

Tabel (4.15) diperoleh hasil t_{hitung} untuk variabel H_1 sebesar -8,991 dan variabel H_2 sebesar 2,527 dengan nilai t_{tabel} diperoleh sebesar 1,690. Kesimpulan yang dapat diambil adalah variabel H_1 memiliki nilai $t_{hitung} < t_{tabel}$ maka terima H_0, artinya variabel H_1 tidak memiliki pengaruh terhadap variabel terikat Y. Sedangkan untuk variabel H_2 memiliki nilai $t_{hitung} > t_{tabel}$ maka tolak H_0, artinya variabel H_2 memiliki pengaruh terhadap variabel terikat Y dalam model ROBPCA.

Berdasarkan hasil uji signifikansi maka dapat disimpulkan bahwa variabel H_2 berupa variabel TPT, PDRB dan IKK memiliki pengaruh yang kuat terhadap jumlah penduduk miskin di Jawa Timur pada tahun 2019. Dikarenakan yang memiliki pengaruh signifikan terhadap model adalah H_2, maka model untuk regresi ROBPCA dapat ditulis lebih ringkas sebagai berikut:

$$Y_{ROBPCA} = 104{,}562 + 14{,}330H_2, \tag{4.35}$$

selanjutnya mensubtitusikan kombinasi linier pada persamaan (4.27) dan (4.28) ke persamaan (4.35), yaitu:

$$\begin{aligned} Y_{ROBPCA} &= 104{,}562 + 14{,}330(0{,}043C_1 + 0{,}690C_2 + 0{,}573C_3 \\ &\quad + 0{,}642C_4 - 0{,}099C_5 - 0{,}051C_6 - 0{,}234C_7) \\ &= 104{,}562 + 0{,}616C_1 + 9{,}888C_2 + 8{,}211C_3 \\ &\quad + 9{,}1999C_4 - 1{,}419C_5 - 0{,}731C_6 - 3{,}353C_7 \end{aligned}$$

sehinggga diperoleh model persamaan regresi ROBPCA:

$$\begin{aligned} Y_{ROBPCA} = 104{,}562 +\ 0{,}616C_1 + 9{,}888C_2 + 8{,}211C_3 + \\ 9{,}1999C_4 - 1{,}419C_5 -\ 0{,}731C_6 - 3{,}353C_7 \end{aligned} \quad (4.36)$$

Persamaan regresi ROBPCA (4.36) akan ditransformasikan kembali ke variabel asal (Lampiran 11), sehingga didaperoleh model persamaan regresi ROBPCA untuk kemiskinan Jawa Timur pada tahun 2019 sebagai berikut:

$$\begin{aligned} Y_{ROBPCA} = 73{,}025 + 0{,}0003X_1 + 8{,}892X_2 + 0{,}119X_3 \\ + 38{,}983X_4 - 0{,}878X_5 - 0{,}142X_6 - 0{,}068X_7 \end{aligned} \quad (4.37)$$

Nilai koefisien determinasi (*R-square*) pada model diperoleh sebesar 0,7087, artinya bahwa penelitian ini mampu menjelesakan sebesar 70,87% untuk variabel Y dengan *principal component robust* terbentuk sebagai faktor-faktor yang mempengaruhinya di dalam model regresi ROBPCA dan untuk sisa sebesar 29,13% sebegai penjelas untuk faktor-faktor lain yang mempengaruhi variabel Y di luar model. Sedangkan nilai *residual standart error* (RSE) diperoleh sebesar 33,817.

4.2.10 Perbandingan Hasil Estimasi PCR dan ROBPCA

Perbandingan dari hasil estimasi metode PCR dan ROBPCA dalam proses penyelesaian multikolinieritas dan *outlier* dapat dilihat kebaikannya dari nilai *Residual Standard Error* (RSE). Nilai RSE yang diperoleh dari uji kedua model tersebut, yaitu:

Tabel 4.16 Perbandingan Model PCR dan ROBPCA

Metode	RSE
Principal Component Regression (PCR)	42,193
Regresi *Robust Principal Component Analysis* (ROBPCA)	33,817

Tabel (4.16) dapat dilihat bahwa model tebaik ada pada model ROBPCA dengan nilai RSE lebih kecil dari model PCR. Artinya metode ROBPCA dengan *M*-

estimator lebih efisien dalam menyelesaikan multikolinieritas dan *outlier* pada data kemiskinan di Jawa Timur pada tahun 2019.

BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan uraian pada pemaparan sebelumnya, maka dapat diambil kesimpulan sebagai berikut:

1. Berikut *principal component* untuk metode regresi ROBPCA yang dinyatakan dalam bentuk kombinasi linier:

$$H_d = m_{1d}C_1 + m_{2d}C_2 + m_{3d}C_3 + m_{4d}C_4 + m_{5d}C_5 + m_{6d}C_6 + m_{7d}C_7$$

   dengan $d = 1,2,3, \dots ,7$.

2. Model yang diperoleh dari implementasi data kemiskinan di Jawa Timur tahun 2019 dengan metode Regresi ROBPCA *M-estimator* adalah:

$$\begin{aligned} Y_{ROBPCA} = {} & 73{,}025 + 0{,}0003X_1 + 8{,}892X_2 + 0{,}119X_3 + 38{,}983X_4 \\ & -0{,}878X_5 - 0{,}142X_6 - 0{,}068X_7 \end{aligned}$$

   dimana model Regresi ROBPCA ini menunjukkan nilai kebaikan model untuk kepadatan penduduk (X_1), TPT (X_2), PDRB (X_3), indeks keparahan kemiskinan (X_4), rata-rata lama sekolah (X_5), IPM (X_6) dan pengeluaran per kapita (X_7), sebesar 70,87% dan niali RSE sebesar 33,817.

5.2 Saran

Berikut adalah beberapa saran yang dapat diambil untuk penelitian lebih lanjut:

1. Penelitian ini bisa dilanjutkan dengan metode yang lain dalam menyelesaikan kasus multikolinieritas dan *outlier*, seperti metode *ridge*, metode LASSO, metode PLS, metode Akar Laten atau metode yang lainnya.

2. Selain itu penelitian bisa dilanjutkan dengan metode yang sama namun bisa mengganti metode estimasi dengan metode estimasi yang lain, seperti estimasi *LTS*, *LMS*, *S*, *MM* atau yang lainnya.

DAFTAR PUSTAKA

Alghifari. 1997. *Analisis Regresi, Teori, Kasus dan Solusi (Edisi pertama).* Yogyakarta: BPFE Universitas Gajah Mada.

Aziz, A. (2010). *Ekonometrika: Teori dan Praktik Eksperimen dengan MATLAB.* Malang: UIN-MALIKI PRESS.

Basyir, H. (2011). *Tafsir Al-Muyassar Jilid 3.* Solo: An-Naba'.

BPS. 2018. *Provinsi Jawa Timur dalam Angka 2017*. Surabaya: Badan Pusat Statistika.

BPS .2021. *Indeks Pembangunan Manusia*. https://jatim.bps.go.id/subject/26/indeks-pembangunan-manusia.html#subjekViewTab1. Diakses: 29 September 2021.

Chen, C. (2002). *Robust Regression and Outlier Detection with The ROBUSTREG Procedure.* Cary, NC: SAS Institute Inc.

Croux, C., & Haesbroeck, G. (2000). Principal Component Analysis Based on *Robust* Estimators of The Covariance or Correlation Matrix: Influence Function and Effeciencies. *Biometrika*, 87 (3): 603-618.

Draper, N. R., & Smith, H. (1992). *Applied Regression Analysis, 2nd (Analisis Regresi Terapan Edisi ke-2).* Jakarta: Penerjemah: Bambang Sumantri. PT. Gramedia Pustaka Utama.

Faizia, T., Prahutama, A., & Yasin, H. (2019). Pemodelan Indeks Pembangunan Manusia di Jawa Tengah dengan Regresi Komponen Utama *Robust*. *Jurnal Gaussin*, 8 (2) : 253-271.

Fox, J. 2011. *An R Companion to Applied Regression (Second Edition)*. USA: SAGE Publications, Inc.

Ghozali, Imam. 2013. *Aplikasi Analisis Multivariat dengan Program IBM SPSS 19 (Edisi Kelima).* Semarang: Badan Penerbit Universitas Diponegoro.

Gujarati, D. N. (2004). *Ekonometri Dasar*. Jakarta: Erlangga.

Hubert, M., Rousseeuw, P. J., & Branden, K. V. (2005). ROBPCA: A New Approach to *Robust* Principal Component Analysis. *Technometrics*, 47 (1): 64-79.

Larasati, S. D., Nisa, K., & Setiawan, E. (2020). Analisis Regresi Komponen Utama *Robust* dengan Metode Minimum Covariance Determinant - Least Trimmed Square (MCD-LTS). *Jurnal Siger Matematika*, 1 (1): 1-9.

Marisson, D. F. (1990). *Multivariate Statistical Methode: Third Edition.* New York: Mc Graw-Hill, Inc.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linier Regression Analysis (Fifth Edition).* New York: John Wiley and Sons, Inc.

Munawwir, Agus. 2014. *Estimasi Parameter Model Regresi Menggunakan Metode Weighted Least Square (WLS) dengan Fungsi Pemboobot Huber*. Skripsi dipublikasikan. Malang: UIN Maulana Malik Ibrahim.

Nasrum, A. (2018). *Uji Normalitas Data untuk Penelitian.* Bali: Jayapangus Press.

Nelwin, Jeffri J. O., Siburian, Rita Rahmawati, & Abdul Hoyyi. 2019. Regresi Komponen Utama *Robust* S-Estimator untuk Analisis Pengaruh Jumlah Pengangguran di Jawa Tengah. *Jurnal Gaussian*. Vol. 8. No. 4, Hal. 439-450.

Pratiwi, Yuliana Indah. 2016. *Penerapan Metode Regresi Ridge Robust Estimasi-M*. Skripsi dipublikasikan. Jakarta: UIN Syarif Hidayatullah.

Richard, J. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis (Sixth Edition).* United States of America: Pearson Education, Inc.

Safidah, Anis. 2014. *Analisis Dekomposisi Spektral dengan Metode Principal Component Analysis*. Skripsi dipublikasikan. Malang: UIN Maulanan Malik Ibrahim.

Sanusi, R. N., & S., D. R. (2019). Metode *Robust* Component Analysis (RPCA) dengan Algoritma Proyeksi dan Matriks Ragam Peragam. *Prisma: Prosding Seminar Nasional Matematika* , (2):52-57.

Sembiring, R. K. (1995). *Analisis Regresi: Edisi Kedua.* Bandung: Penerbit ITB.

Shihab, M. (2003). *Tafsir Al-Misbah: Pesan, Kesan dan Keserasian Al-Qur'an.* Jakarta: Penerbit Lentera Hati.

Shihab, M. (2005). *Tafsir Al-Misbah: Pesan, Kesan dan Keserasian Al-Qur'an.* Jakarta: Penerbit Lentera Hati.

Siagian, D., & Sugiarto. (2006). *Metode Statistika untuk Bisnis dan Ekonomi.* Jakarta: PT Gramedia Pustaka Umum.

Simamora, B. (2005). *Analisis Multivariat Pemasaran.* Jakarta: PT Gramedia Pustaka Utama.

Sumodiningrat, G. (2002). *Ekonometrika Pengantar.* Yogyakarta: BPFE.

Supranto, J. (1994). *Statistik: Teori dan Aplikasi Jilid II (Cetakan 5).* Jakarta: Erlangga.

Supriyadi, E. S. (2017). Perbandingan Metode Partial Least Square (PLS) dan Principal Component Regression (PCR) untuk Mengatasi Multikolinieritas pada Model Regresi Linier Berganda. *UNNES Journal of Mathematics*, 6 (2): 118-128.

Susilowati, Bekti Indar., & Pardomuan R. S. 2020. Metode ROBPCA (*Robust Principal Component Analysis*) dan Clara (*Clustering Large Area*) pada Data dengan *Outlier* (Studi Kasus: Data Laporan Indeks Kebahagiaan Dunia Tahun 2018). *Jurnal Ilmu Komputer*. Vol. 13, No. 2, Hal. 88-98.

Tulak, Egidia T., Raupong, & Nirwan Ilyas. 2017. Penerapan Regresi *Robust Principal Component Analysis* Pada Data yang Mengandung Multikolinieritas dan *Outlier*. Makassar: Departemen Matematika FMIPA Universitas Hasanuddin.

Verdian, E. (2019). Analisis Faktor yang Merupakan Intensi Perpindahan Merek Transportasi Online di Surabaya. *Jurnal AGORA*, 7 (1).

Waluyo, Herman J. 2001. *Drama Teori dan Pengajaran*. Yogyakarta: Hanindita Graha Widia.

Widarjono, A. 2005. *Ekonometrika: Teori dan Aplikasi untuk Ekonomi dan Bisnis (Edisi Pertama)*. Yogyakarta: Ekonisia.

Wohon, S. C., Hatidja, D., & Nainggolan, N. (2017). Penentuan Model Regresi Terbaik dengan Menggunakan Metode Stepwise (Studi Kasus: Impor Beras di Sulawesi Utara). *Jurnal Ilmiah Sains*, 17 (2): 80-88.

LAMPIRAN

Lampiran 1: Data Kemiskinan Kabupaten/Kota di Jawa Timur 2019

Kab./Kota	Y	X1	X2	X3	X4	X5	X6	X7
Pacitan	75,86	399	0,95	11,041	0,34	7,28	68,16	90,33
Ponorogo	83,97	667	3,58	14,297	0,4	7,21	70,56	98,83
Trenggalek	76,44	607	3,43	12,78	0,3	7,28	69,46	98,65
Tulungagung	70,01	984	3,36	27,3	0,17	8,07	72,62	108,91
Blitar	103,75	868	3,11	25,53	0,23	7,29	70,57	108,61
Kediri	163,95	1,136	3,68	29,194	0,22	8,01	71,85	111,46
Malang	246,6	738	3,82	68,38	0,4	7,27	70,35	102,7
Lumajang	98,88	582	2,81	22,563	0,2	6,22	65,33	92,74
Jember	226,57	792	3,8	54,2	0,24	6,18	66,69	95,25
Banyuwangi	121,37	279	4,08	55,274	0,34	7,13	70,6	122,64
Bondowoso	103,33	508	2,96	13,637	0,33	5,71	66,09	106,65
Situbondo	76,44	409	2,82	13,6	0,3	6,12	67,09	100,97
Probolinggo	207,22	689	3,88	23,395	0,85	5,77	65,6	109,72
Pasuruan	141,09	1,104	5,42	105,289	0,43	7,11	68,29	103,81
Sidoarjo	119,29	3,546	4,72	140,493	0,42	10,25	80,05	146,09
Mojokerto	108,81	1,557	3,68	58,467	0,27	8,49	73,53	128,6
Jombang	116,44	1,133	4,39	28,216	0,53	8,53	72,85	115,33
Nganjuk	118,51	861	3,22	18,304	0,23	7,63	71,71	12,2
Madiun	71,91	658	3,62	13,162	0,34	7,8	71,69	11,65
Magetan	60,43	913	3,08	13,238	0,36	7,96	73,49	117,79
Ngawi	119,43	641	3,7	13,711	0,72	6,98	70,41	114,68
Bojonegoro	154,64	568	3,7	69,986	0,43	7,09	68,75	102,65
Tuban	170,8	639	2,76	45,356	0,75	6,81	68,37	10,499
Lamongan	157,11	667	4	27,706	0,67	7,89	72,57	115,72
Gresik	148,61	1,102	5,54	101,347	0,61	9,29	76,1	13,295
Bangkalan	186,11	985	5,84	18,551	1,08	5,66	63,79	8,718
Sampang	202,21	794	2,81	13,995	0,7	4,55	61,94	8,76
Pamekasan	122,43	1,111	2,32	11,407	0,39	6,4	65,94	8,834
Sumenep	211,98	545	2,17	23,816	0,79	5,46	66,22	9,082
Kediri	20,54	4,533	4,22	90,002	0,32	9,92	78,08	124,4
Blitar	10,1	4,356	4,64	4,833	0,23	10,1	78,56	138,51
Malang	35,39	5,993	6,04	52,335	0,13	10,17	81,32	166,66
Probolinggo	16,37	4,186	4,41	8,339	0,26	8,69	73,27	122,8
Pasuruan	12,92	5,679	5,06	5,965	0,17	9,11	75,25	133,93
Mojokerto	6,63	7,833	2,65	4,986	0,14	10,24	77,96	137,1
Madiun	7,69	5,218	4,01	10,621	0,08	11,13	80,88	160,4
Surabaya	130,55	8,262	5,87	410,879	0,15	10,47	82,22	178,54
Batu	7,89	1,507	2,48	11,787	0,05	9,06	75,88	128,7

Sumber: Badan Pusat Statistik, 2020

Lampiran 2: Hasil Standarisasi Data

ZY	ZX1	ZX2	ZX3	ZX4	ZX5	ZX6	ZX7
-0,487	-0,699	-25,205	-0,478	-0,184	-0,321	-0,686	-0,125
-0,365	-0,576	-0,156	-0,431	0,070	-0,364	-0,219	0,048
-0,478	-0,603	-0,291	-0,453	-0,354	-0,321	-0,433	0,044
-0,575	-0,431	-0,354	-0,243	-0,905	0,168	0,182	0,253
-0,067	-0,484	-0,579	-0,268	-0,651	-0,315	-0,217	0,247
0,838	-0,361	-0,066	-0,215	-0,693	0,131	0,032	0,305
208,059	-0,543	0,0599	0,352	0,070	-0,327	-0,260	0,127
-0,140	-0,615	-0,848	-0,311	-0,778	-0,977	-12,361	-0,076
177,944	-0,519	0,042	0,147	-0,608	-10,015	-0,972	-0,025
0,198	-0,754	0,294	0,163	-0,184	-0,414	-0,211	0,533
-0,074	-0,649	-0,713	-0,440	-0,227	-12,923	-10,883	0,207
-0,478	-0,694	-0,839	-0,441	-0,354	-10,386	-0,894	0,092
148,851	-0,566	0,114	-0,299	197,919	-12,552	-11,836	0,270
0,494	-0,375	149,839	0,887	0,198	-0,426	-0,661	0,149
0,167	0,745	0,869	139,713	0,155	151,741	162,636	101,091
0,0089	-0,168	-0,066	0,209	-0,481	0,428	0,359	0,655
0,124	-0,362	0,572	-0,229	0,622	0,453	0,226	0,384
0,155	-0,487	-0,4796	-0,373	-0,651	-0,104	0,005	-171,734
-0,546	-0,58	-0,120	-0,448	-0,184	0,001	0,001	-172,855
-0,719	-0,463	-0,606	-0,446	-0,099	0,1002	0,351	0,434
0,169	-0,588	-0,048	-0,440	142,774	-0,506	-0,248	0,371
0,698	-0,621	-0,048	0,376	0,198	-0,438	-0,571	0,126
0,941	-0,589	-0,893	0,0189	1,555	-0,612	-0,645	-175,201
0,735	-0,576	0,222	-0,237	121,564	0,057	0,172	0,392
0,607	-0,376	160,628	0,830	0,961	0,923	0,858	-169,503
117,112	-0,43	1,876	-0,369	295,483	-13,233	-15,356	-17,883
141,318	-0,518	-0,848	-0,435	13,429	-20,102	-18,953	-178,744
0,214	-0,372	-128,876	-0,473	0,028	-0,865	-11,175	-178,593
156,008	-0,632	-142,362	-0,293	172,467	-14,471	-1,063	-178,088
-13,182	119,762	0,420	0,666	-0,269	131,318	124,327	0,569
-14,752	111,642	0,797	-0,568	-0,651	142,458	133,661	0,857
-1,095	186,739	205,581	0,120	-1,075	14,679	187,333	143,006
-13,809	103,844	0,590	-0,517	-0,524	0,552	0,308	0,536
-14,328	172,334	117,472	-0,552	-0,905	0,812	0,693	0,763
-15,274	271,147	-0,992	-0,566	-10,326	151,122	121,994	0,828
-15,114	151,186	0,231	-0,484	-12,871	206,203	178,776	130,251
0,336	290,828	190,297	531,418	-0,990	165,356	204,834	167,214
-15,084	-0,191	-114,491	-0,467	-14,144	0,781	0,816	0,657

Lampiran 3: Hasil Pembentukan Model Regresi Linier Berganda dengan metode OLS

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. *Error*	Beta			Tolerance	VIF
1	(Constant)	179.712	330.428		.544	.591		
	Kepadatan Penduduk	-.008	.005	-.267	-1.640	.112	.354	2.826
	TPT	4.217	7.756	.071	.544	.591	.558	1.793
	PDRB	.438	.117	.454	3.727	.001	.632	1.582
	Indeks Keparahan Kemiskinan	86.062	39.528	.305	2.177	.037	.478	2.092
	Rata-rata Lama Sekolah	-18.613	21.790	-.452	-.854	.400	.033	29.861
	IPM	.374	6.863	.029	.054	.957	.033	29.999
	Pengeluaran per Kapita Riil	-.058	.187	-.043	-.309	.760	.494	2.025

a. Dependent Variable: Jumlah Penduduk Miskin

Lampiran 4: Uji Asumsi pada Model Regresi Linier Berganda

1. Uji Normalitas

One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual
N		38
Normal Parameters[a,b]	Mean	.0000000
	Std. Deviation	.53059467
Most Extreme Differences	Absolute	.114
	Positive	.109
	Negative	-.114
Test Statistic		.114
Asymp. Sig. (2-tailed)		.200[c,d]

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

d. This is a lower bound of the true significance.

2. Hasil Uji Heterokedastisitas dari Variabel Transformasi

Coefficients[a]

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics	
		B	Std. *Error*			Tolerance	VIF
1	(Constant)	36.187	198.874	.182	.857		
	Kepadatan Penduduk	-.001	.003	-.243	.810	.354	2.826
	TPT	.896	4.668	.192	.849	.558	1.793
	PDRB	.014	.071	.192	.849	.632	1.582
	Indeks Keparahan Kemiskinan	-22.039	23.791	-.926	.362	.478	2.092
	Rata-rata Lama Sekolah	-6.288	13.115	-.479	.635	.033	29.861
	IPM	.500	4.130	.121	.904	.033	29.999
	Pengeluaran per Kapita Riil	.102	.112	.907	.372	.494	2.025

a. Dependent Variable: ABS

3. Hasil Uji ANOVA

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	117597.493	7	16799.642	10.937	.000[b]
	Residual	46080.334	30	1536.011		
	Total	163677.826	37			

a. Dependent Variable: Jumlah Penduduk Miskin

b. Predictors: (Constant), Pengeluaran per Kapita Riil, PDRB, TPT, Indeks Keparahan Kemiskinan, Kepadatan Penduduk, Rata-rata Lama Sekolah, IPM

4. Uji Kebaikan Model

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. *Error* of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.848[a]	.718	.653	39.19198	.718	10.937	7	30	.000

a. Predictors: (Constant), Pengeluaran per Kapita Riil, PDRB, TPT, Indeks Keparahan Kemiskinan, Kepadatan Penduduk, Rata-rata Lama Sekolah, IPM

b. Dependent Variable: Jumlah Penduduk Miskin

Lampiran 5: *Principal Component Analysis*

1. Matriks Korelasi

Correlation Matrix

		Zscore: Kepadatan Penduduk	Zscore: TPT	Zscore: PDRB	Zscore: Indeks Keparahan Kemiskinan	Zscore: Rata-rata Lama Sekolah	Zscore: IPM	Zscore: Pengeluaran per Kapita Riil
Correlation	Zscore: Kepadatan Penduduk	1.000	.441	.430	-.452	.777	.773	.562
	Zscore: TPT	.441	1.000	.479	.062	.432	.457	.298
	Zscore: PDRB	.430	.479	1.000	-.099	.342	.406	.286
	Zscore: Indeks Keparahan Kemiskinan	-.452	.062	-.099	1.000	-.562	-.542	-.571
	Zscore: Rata-rata Lama Sekolah	.777	.432	.342	-.562	1.000	.980	.618
	Zscore: IPM	.773	.457	.406	-.542	.980	1.000	.636
	Zscore: Pengeluaran per Kapita Riil	.562	.298	.286	-.571	.618	.636	1.000

2. Uji KMO dan Uji *Bartlett's*

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.770
Bartlett's Test of Sphericity	Approx. Chi-Square	208.338
	df	21
	Sig.	.000

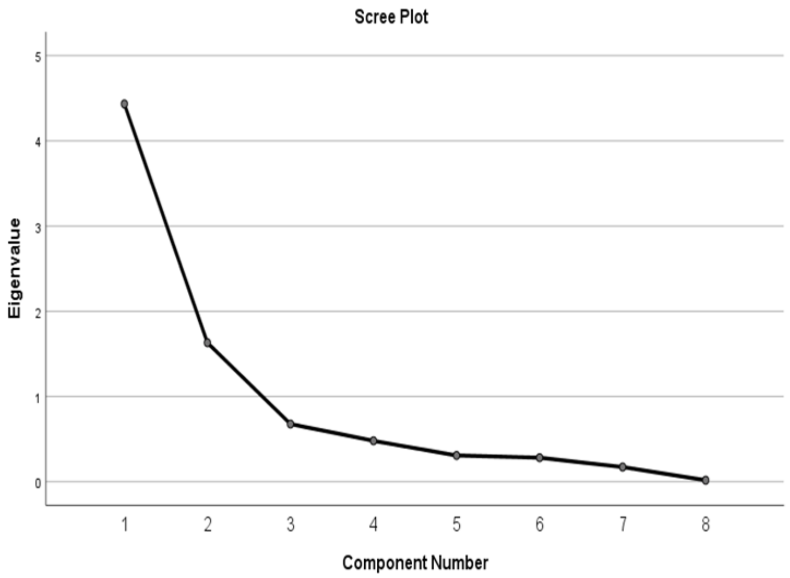
3. Nilai *Eigen* dan Nilai Kumulatif

Total Variance Explained

Component	Initial *Eigen*values			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.041	57.732	57.732	4.041	57.732	57.732	3.325	47.501	47.501
2	1.286	18.369	76.101	1.286	18.369	76.101	2.002	28.600	76.101
3	.601	8.587	84.688						
4	.477	6.821	91.509						
5	.308	4.394	95.903						
6	.270	3.853	99.756						
7	.017	.244	100.000						

Extraction Method: Principal Component Analysis.

4. *Scree Plot*



5. *Principal Component* Terpilih

Component Matrix[a]

	Component	
	1	2
Zscore: Kepadatan Penduduk	.864	.043
Zscore: TPT	.544	.690
Zscore: PDRB	.530	.573
Zscore: Indeks Keparahan Kemiskinan	-.618	.642
Zscore: Rata-rata Lama Sekolah	.930	-.099
Zscore: IPM	.942	-.051
Zscore: Pengeluaran per Kapita Riil	.764	-.234

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Lampiran 6: Variabel Baru Terbentuk

H1	H2
0,1472	-1,7799
-0,17197	-0,37109
-0,02682	-0,62121
0,50591	-0,66568
0,23304	-0,77225
0,3586	-0,4343
-0,27013	0,09413
-0,1569	-1,09808
-0,33357	-0,32867
-0,12818	-0,01056
-0,35187	-0,91607
-0,21686	-0,99769
-1,33249	0,36339
-0,68529	1,1252
0,74514	1,40084
0,49237	-0,12056
-0,10895	0,38824
-0,17451	-0,61929
-0,39717	-0,30509
0,31968	-0,61196
-0,67177	0,12882
-0,41326	0,04939
-1,29534	0,02058
-0,42183	0,35318
-0,80699	1,73388
-2,57637	1,71908
-1,77365	-0,39994
-0,72409	-0,96882
-1,54669	-0,42367
0,89307	0,73
1,22891	0,22579
1,51622	1,17034
0,67766	0,06065
1,05582	0,34722
1,91213	-0,64105
1,98526	-0,16256
1,23234	3,60961
1,28137	-1,27191

Lampiran 7: *Principal Component Regression* (PCR)

1. Model PCR

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. *Error*	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	108.218	6.845		15.811	.000		
	H1	-50.925	6.937	-.766	-7.342	.000	1.000	1.000
	H2	12.095	6.937	.182	1.744	.090	1.000	1.000

a. Dependent Variable: Jumlah Penduduk Miskin

2. Hasil Uji ANOVA Model PCR

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	101368.391	2	50684.195	28.470	.000[b]
	Residual	62309.435	35	1780.270		
	Total	163677.826	37			

a. Dependent Variable: Jumlah Penduduk Miskin

b. Predictors: (Constant), H2, H1

3. Hasil Uji Kebaikan Model PCR

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. *Error* of the Estimate	Change Statistics R Square Change	Change Statistics F Change	Change Statistics df1	Change Statistics df2	Change Statistics Sig. F Change
1	.787[a]	.619	.598	42.19324	.619	28.470	2	35	.000

a. Predictors: (Constant), H2, H1

b. Dependent Variable: Jumlah Penduduk Miskin

4. Hasil Uji Multikolinieritas

Collinearity Diagnostics[a]

Model	Dimension	*Eigen*value	Condition Index	Variance Proportions (Constant)	Variance Proportions H1	Variance Proportions H2
1	1	1.000	1.000	.00	1.00	.00
	2	1.000	1.000	1.00	.00	.00
	3	1.000	1.000	.00	.00	1.00

a. Dependent Variable: Jumlah Penduduk Miskin

Lampiran 8: Hasil Transformasi dari Variabel baru ke Variabel asal pada Model PCR:

$$C_j = \frac{X_j - \overline{X_j}}{\sigma_j},$$

$$C_1 = \frac{X_1 - 1922{,}3}{2179{,}868},$$

$$C_2 = \frac{X_2 - 3{,}8}{1{,}112},$$

$$C_3 = \frac{X_3 - 44{,}1}{69{,}028},$$

$$C_4 = \frac{X_4 - 0{,}4}{0{,}236},$$

$$C_5 = \frac{X_5 - 7{,}8}{1{,}616},$$

$$C_6 = \frac{X_6 - 71{,}7}{5{,}142},$$

$$C_7 = \frac{X_7 - 96{,}5}{49{,}075},$$

$$\begin{aligned}
Y_{PCR} &= 108{,}218 + 0{,}520\left(\frac{X_1 - 1922{,}3}{2179{,}868}\right) + 8{,}346\left(\frac{X_2 - 3{,}8}{1{,}112}\right) \\
&\quad + 6{,}930\left(\frac{X_3 - 44{,}1}{69{,}028}\right) + 7{,}765\left(\frac{X_4 - 0{,}4}{0{,}236}\right) - 1{,}197\left(\frac{X_5 - 7{,}8}{1{,}616}\right) \\
&\quad - 0{,}617\left(\frac{X_6 - 71{,}7}{5{,}142}\right) - 2{,}830\left(\frac{X_7 - 96{,}5}{49{,}075}\right) \\
&= 108{,}218 + 0{,}0002X_1 - 0{,}459 + 7{,}505X_2 - 28{,}521 \\
&\quad + 0{,}1X_3 - 4{,}427 + 32{,}9X_4 - 13{,}161 - 0{,}741X_5 \\
&\quad + 5{,}778 - 0{,}12X_6 + 8{,}603 - 0{,}058X_7 + 5{,}565 \\
&= 81{,}596 + 0{,}0002X_1 + 7{,}505X_2 + 0{,}1X_3 + 32{,}9X_4 - 0{,}741X_5 \\
&\quad - 0{,}12X_6 - 0{,}058X_7
\end{aligned}$$

Lampiran 9: Cek *Outlier*

Nilai Levarage	*DFFIT's*
0.345405337	-0.53299431
0.060429672	-0.21625949
0.061308633	-0.19349073
0.077655876	-0.09925287
0.070288648	0.018760338
0.067227066	0.56053338
0.052746913	0.761649559
0.188695893	-0.24700338
0.149966654	1.131221542
0.111903842	-0.12603546
0.153540034	-0.36873173
0.109758268	-0.47071361
0.251210056	0.382187728
0.260453612	-0.39676062
0.188809452	0.273657625
0.068815256	0.109285772
0.151194602	0.088500121
0.268684916	0.3643204
0.214309964	-0.50544109
0.130496314	-0.36073164
0.200153189	-0.40286066
0.092179954	0.071611702
0.193731709	0.028506521
0.164767766	0.277409813
0.410603935	0.079722358
0.460986823	-0.94051341
0.245062318	0.211149863
0.206918734	-0.02258467
0.344209406	0.60319611
0.09538451	-0.38872241
0.131050533	-0.0937499
0.407051578	0.365093637
0.131666191	-0.34416247
0.240510507	-0.1608983
0.468243428	1.1494602
0.205278171	0.40856676
0.854372419	-4.14837037
0.164927819	-0.39252041

Lampiran 10: *Robust Principal Component Analysis* (ROBPCA) dari *software* NCSS 21

Run Summary Report

Item	Value	Rows
Value		
Dependent Variable	Y	Number
Processed	38	
Number Ind. Variables	2	Number Used in
Estimation	38	
Weight Variable	None	Number Filtered
Out	0	
Robust Method	Huber's Method	Number with X's
Missing	0	
Tuning Constant	1.345	Number with Weight
Missing	0	
MAD Scale Factor	0.6745	Number with Y
Missing	0	
Sum of *Robust* Weights	35.756	
Run Information	**Value**	
Iterations	6	
Max % Change in any Coef	0.000	
R^2 after *Robust* Weighting	0.7087	
S using MAD	34.76	
S using MSE	33.82	
Completion Status	Normal Completion	

Correlation Matrix

	H1	**H2**	**Y**
H1	1.0000	0.0506	-0.8096
H2	0.0506	1.0000	0.1893
Y	-0.8096	0.1893	1.0000

Robust Iterations - Coefficients

***Robust* Iteration**	**Max Percent Change in Coefficients**	**b(0)**	**b(1)**	**b(2)**
0		108.21766	-50.92539	12.09509
1	20.782	105.12286	-51.98637	14.60869
2	1.867	104.63437	-52.13879	14.33600
3	0.060	104.57150	-52.14797	14.33279
4	0.019	104.56285	-52.14753	14.33007
5	0.001	104.56178	-52.14736	14.33008
6	0.000	104.56164	-52.14732	14.33006

Regression Coefficients Confidence Intervals Assuming Random Weights

Independent Variable	Regression Coefficient b(i)	Standard *Error* Sb(i)	Lower 95% Conf. Limit of β(i)	Upper 95% Conf. Limit of β(i)
Intercept	104.56164	5.895053	92.59405	116.52924
H1	-52.14732	5.974186	-64.27557	-40.01908
H2	14.33006	5.974184	2.20182	26.45830

Note: The T-Value used to calculate these confidence limits was 2.030.

Estimated Equation

Y =
104.561641868382 - 52.1473246692994 * H1 + 14.330058329311 * H2

Analysis of Variance

Source	DF	R^2	Sum of Squares	Mean Square	F-Ratio	Prob Level
Intercept	1		377591.91	377591.91		
Model	2	0.7087	97365.56	48682.78	42.570	0.0000
Error	35	0.2913	40025.56	1143.59		
Total(Adjusted)	37	1.0000	137391.12	3713.27		

Nilai RSE untuk Model Regresi ROBPCA:

$$\begin{aligned} RSE &= \sqrt{MSE} \\ &= \sqrt{1143{,}59} \\ &= 33{,}817 \end{aligned}$$

Multicollinearity Report

Independent Variable	Variance Inflation Factor	R^2 Versus Other I.V.'s	Tolerance	Diagonal of X'X Inverse
H1	1.0026	0.0026	0.9974	0.0294170678988219
H2	1.0026	0.0026	0.9974	0.0281175633029468

***Eigen*values of Centered Correlations**

No.	*Eigen*value	Incremental Percent	Cumulative Percent	Condition Number
1	1.0506	52.528	52.528	1.000
2	0.9494	47.472	100.000	1.107

All Condition Numbers less than 100. Multicollinearity is NOT a problem.

Lampiran 11: Nilai Pembobot untuk metode WLS dari perhitungan dengan SPSS 26

Res_1	w
-3.33359	0.0031
-28.5168	0.0013
-25.6299	0.0015
-4.39225	0.0018
16.74059	0.0017
79.24725	0.0015
123.4873	0.001
-4.04635	0.0019
105.3403	0.0012
6.75267	0.0011
-11.7268	0.0016
-30.7543	0.0017
26.7496	0.0008
-15.6358	0.0007
32.0757	0.0008
27.12462	0.0013
-2.02192	0.001
8.89562	0.0015
-52.8438	0.0012
-24.1062	0.0016
-24.556	0.001
24.77968	0.001
-3.63211	0.0009
23.13888	0.0009
-21.6755	0.0005
-74.1026	0.0004
8.5057	0.001
-10.9442	0.0015
30.1207	0.001
-51.027	0.001
-38.266	0.0014
-9.76913	0.001
-58.0709	0.0013
-45.7294	0.0012
3.54178	0.0026
2.5386	0.002
41.43121	0.0004
-19.6895	0.0032

Lampiran 12: Hasil Transformasi dari Variabel Baru ke Variabel Asal pada Model Regresi ROBPCA

$C_j = \frac{X_j - \overline{X_j}}{\sigma_j}$,

$C_1 = \frac{X_1 - 1922{,}3}{2179{,}868}$,

$C_2 = \frac{X_2 - 3{,}8}{1{,}112}$,

$C_3 = \frac{X_3 - 44{,}1}{69{,}028}$,

$C_4 = \frac{X_4 - 0{,}4}{0{,}236}$,

$C_5 = \frac{X_5 - 7{,}8}{1{,}616}$,

$C_6 = \frac{X_6 - 71{,}7}{5{,}142}$,

$C_7 = \frac{X_7 - 96{,}5}{49{,}075}$,

$$\begin{aligned} Y_{ROBPCA} &= 104{,}562 + 0{,}616\left(\frac{X_1 - 1922{,}3}{2179{,}868}\right) + 9{,}888\left(\frac{X_2 - 3{,}8}{1{,}112}\right) \\ &\quad + 8{,}211\left(\frac{X_3 - 44{,}1}{69{,}028}\right) + 9{,}1999\left(\frac{X_4 - 0{,}4}{0{,}236}\right) - 1{,}419\left(\frac{X_5 - 7{,}8}{1{,}616}\right) \\ &\quad - 0{,}731\left(\frac{X_6 - 71{,}7}{5{,}142}\right) - 3{,}353\left(\frac{X_7 - 96{,}5}{49{,}075}\right) \\ &= 104{,}562 + 0{,}0003X_1 - 0{,}543 + 8{,}892X_2 - 33{,}79 \\ &\quad + 0{,}119X_3 - 5{,}246 + 38{,}983X_4 - 15{,}593 - 0{,}878X_5 \\ &\quad + 6{,}849 - 0{,}142X_6 + 10{,}193 - 0{,}068X_7 + 6{,}593 \\ &= 73{,}025 + 0{,}0003X_1 + 8{,}892X_2 + 0{,}119X_3 + 38{,}983X_4 \\ &\quad - 0{,}878X_5 - 0{,}142X_6 - 0{,}068X_7 \end{aligned}$$

RIWAYAT HIDUP

Elma Al Husna, lahir di Trenggalek tanggal 20 Maret 1997. Alamat asal dari Desa Pakis, Kecamatan Durenan, Kabupaten Trenggalek. Merupakan putri pertama dari pasangan Abah Nuryasin dan Ibu Muzdalipah, serta kakak perempuan dari kedua adik laki-laki Akmal Abid Nabawi dan Engga Habiburrohman.

Pendidikan dasar penulis di tempuh di MI Miftahul Huda Pakis dan lulus pada tahun 2009, kemudian melanjutkan pendidikan menengah pertama di MtsN Bandung Tulungangung dan lulus pada tahun 2012 dan melanjutkan pendidikan menengah atas di MAN Tambakberas Jombang dan lulus pada tahun 2015. Pendidikan selanjutnya di tempuh di UIN Maulana Malik Ibrahim Malang pada tahun 2015 melalui jalur SBMPTN. Penulis mengambil Program Studi Matematika di Fakultas Sains dan Teknologi.

Selama di Malang, penulis menjadi santri dan mukim di PPTQ Nurul Furqon Malang dan ditahun terakhir di Malang dilanjutkan mukim di cabang PP Nurul Furqon III yang beralamatkan di desa Bunulrejo Blimbing kota Malang. Bagi pembaca dapat menghubungi penulis melalui email: elma.alchusna@gmail.com.

KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No. 50 Dinoyo Malang Telp./Fax.(0341)558933

BUKTI KONSULTASI SKRIPSI

Nama : Elma Alhusna
NIM : 15610080
Fakultas/Jurusan : Sains dan Teknologi / Matematika
Judul Skripsi : Implementasi Regresi *Robust Principal Componant Analysis* Pada Data Kemiskinan Jawa Timur
Pembimbing I : Dr. Sri Harini, M.Si
Pembimbing II : Mohammad Nafie Jauhari, M.Si

No	Tanggal	Hal	Tanda Tangan
1	23 Februari 2021	Konsultasi BAB I dan BAB II	1.
2	3 Mei 2021	Konsultasi Integrasi Bab I dan Bab II	2.
3	7 April 2021	Revisi Bab I dan Bab II	3.
4	26 November 2021	Revisi Integrasi Bab I dan Bab II	4.
5	10 Februari 2022	Konsultasi Bab III dan Bab IV	5.
6	24 Februari 2022	Revisi Bab IV	6.
7	17 Maret 2022	Revisi Integrasi Bab IV	7.
8	25 Maret 2022	ACC Seminar Hasi	8.
9	18 Mei 2022	Konsultasi Bab IV dan Bab V	9.
10	25 Mei 2022	ACC Integrasi	10.
11	14 Juni 2022	ACC Keseluruhan	11.

Malang, 23 Juni 2022
Mengetahui,
Ketua Program Studi Matematika

Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005