ANALISIS REGRESI KOMPONEN UTAMA ROBUST DENGAN METODE *MINIMUM VOLUME ELLIPSOID* PADA DATA JUMLAH PENGANGGURAN DI SULAWESI SELATAN

SKRIPSI

NUR UKHTY HAKIM

H051171511

Pembimbing Utama : Dr. Anna Islamiyati, S.Si.,M.Si.

Pembimbing Pertama: Dr. Nirwan,M.Si.

Dosen Penguji : 1. Prof. Dr. Nurtiti Sunusi,S.Si.,M.Si.

2. Dr.Dr.Georgina Maria Tinungki,S.Si.,M.Si

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

DESEMBER 2023

Dipindai dengan CamScanner

ANALISIS REGRESI KOMPONEN UTAMA ROBUST DENGAN METODE *MINIMUM VOLUME ELLIPSOID* PADA DATA JUMLAH PENGANGGURAN DI SULAWESI SELATAN

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

NUR UKHTY HAKIM

H051171511

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

DESEMBER 2023

Dipindai dengan CamScanner

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

Analisis Regresi Komponen Utama *Robust*

Dengan Metode *Minimum Volume Ellipsoid*

Pada Data Jumlah Pengangguran

Di Sulawesi Selatan

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun

Makassar, 11 Desember 2023

NUR UKHTY HAKIM

NIM. H051171511

Dipindai dengan CamScanner

ANALISIS REGRESI KOMPONEN UTAMA ROBUST DENGAN METODE *MINIMUM VOLUME ELLIPSOID* PADA DATA JUMLAH PENGANGGURAN DI SULAWESI SELATAN

Disetujui Oleh:

Pembimbing Utama	Pembimbing Pendamping
Dr. Anna Islamiyati, S.Si., M.Si.	Dr. Nirwan, M.Si.
NIP. 19770808 200501 2 002	NIP. 19630306 198702 1 002

Ketua Program Studi

Dr. Anna Islamiyati, S.Si., M.Si.

NIP. 19770808 200501 2 002

Pada 11 Desember 2023

Dipindai dengan CamScanner

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama	:	Nur Ukhty Hakim
NIM	:	H051171511
Program Studi	:	Statistika
Judul Skripsi	:	Analisis Regresi Komponen Utama *Robust* Dengan Metode *Minimum Volume Ellipsoid* Pada Data Jumlah Pengangguran di Sulawesi Selatan

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1. Ketua : Dr. Anna Islamiyati, S.Si., M.Si. (.....)
2. Sekretaris : Dr. Nirwan, M.Si. (.....)
3. Anggota : Prof. Dr. Nurtiti Sunusi, S.Si., M.Si. (.....)
4. Anggota : Dr. Dr. Georgina Maria Tinungki, M.Si (.....)

Ditetapkan di : Makassar

Tanggal : 11 Desember 2023

Dipindai dengan CamScanner

KATA PENGANTAR

Segala puji bagi Allah *Subhanahu Wa ta'ala Rabb* semesta alam, shalawat serta salam semoga senantiasa tercurahkan kepada Nabi yang paling dimuliakan, pemimpin orang-orang bertakwa yakni Rasulullah Muhammad *Shallallahu Alaihi Wasallam* dan kepada para keluarga serta sahabat beliau yang senantiasa kita rindukan perjumpaan dengannya. Amma ba'du.

Alhamdulillah, berkat pertolongan Allah akhirnya skripsi dengan judul "Analisis Regresi Komponen Utama *Robust* Dengan Metode *Minimum Volume Ellipsoid* Pada Data Jumlah Pengangguran di Sulawesi Selatan" yang disusun sebagai salah satu syarat akademik untuk meraih gelar sarjana pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin ini dapat dirampungkan. Dalam penulisan skripsi ini, penulis dengan segala keterbatasan kemampuan dan pengetahuan dapat melewati segala hambatan dan masalah berkat bantuan dan dorongan dari berbagai pihak. Oleh karena itu penulis menyampaikan ucapan terima kasih yang tak terhingga kepada orang tua penulis, Ibunda **Hj. Naidah, S.Pd.** dan Ayahanda **H. Drs. Abd Haking, M.Si.**, yang telah membesarkan dan mendidik dengan penuh kesabaran, memberikan cinta dan limpahan kasih sayang, dukungan dan doa yang tulus tanpa henti kepada penulis. Rasa terima kasih juga kepada adikku tercinta **Nur Ilmi Hakim**, **Nur Shalshabilah Hakim**, dan **Nur Ikhwan Hakim** atas doa, dukungan, semangat, dan bantuannya kepada penulis. Tak lupa pula terima kasih untuk **Keluarga Besar** atas doa, dukungan, semangat, dan bantuannya kepada penulis.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada:

1. **Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin, M.Si.** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Anna Islamiyati, S.Si., M.Si.** selaku Ketua Departemen Statistika, segenap Dosen Pengajar dan Staf Departemen Statistika yang telah membekali ilmu dan kemudahan-kemudahan kepada Penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.

Dipindai dengan CamScanner

4. **Ibu Prof. Dr. Nurtiti Sunusi,S.Si.,M.Si.** selaku Penasehat Akademik atas saran, nasehat, motivasi, dan dukungan yang telah diberikan kepada penulis selama menjadi mahasiswa dan meluangkan waktu untuk membimbing dan memberikan masukan dalam penulisan skripsi ini.
5. **Ibu Dr. Anna Islamiyati, S.Si., M.Si.** selaku Pembimbing Utama dan **Bapak Dr. Nirwan, M.Si.** selaku Pembimbing Pertama yang telah ikhlas meluangkan waktu dan pemikirannya untuk memberikan arahan, pengetahuan, motivasi dan bimbingan ditengah kesibukan beliau serta menjadi tempat berkeluh kesah untuk penulis.
6. **Ibu Prof. Dr. Nurtiti Sunusi,S.Si.,M.Si.** dan **Ibu Dr. Dr. Georgina Maria Tinungki,S.Si.,M.Si**. selaku tim Penguji yang telah memberikann saran dan kritikan yang membangun dalam penyempurnaan penyusunan skripsi ini serta waktu yang telah diberikan kepada Penulis.
7. Sahabat seperjuangan dalam penyusunan skripsi, **Zulfa Putri Asmawi**, **Nurhidayatullah**, **Sriulan S.S. Abdullah**, dan **Zakiah Fitri** yang selalu setia membersamai penulis, memberikan dukungan satu sama lain dan menjadi tempat bertukar cerita.
8. Teman-teman **Statistika 2017**, terima kasih atas kebersamaan, suka, dan duka selama menjalani pendidikan di Departemen Statistika. Penulis senang mengenal kalian semua. Terkhusus untuk **Aqilah**, **Haura** dan **Ata**, terima kasih atas segala bantuan dan motivasinya untuk penulis dalam menyelesaikan skripsi ini.
9. Keluarga besar **DISKRIT 2017**, terima kasih telah memberikan pelajaran yang berharga dan arti kebersamaan selama ini kepada penulis. Pengalaman yang berharga telah penulis dapatkan dari teman-teman selama berproses.
10. **Keluarga Mahasiswa FMIPA Unhas** terkhusus anggota keluarga **Himatika FMIPA Unhas** dan **Himastat FMIPA Unhas**, terima kasih atas ilmu yang mungkin tidak bisa didapatkan di proses perkuliahan dan telah menjadi keluarga selama penulis kuliah di Universitas Hasanuddin.
11. Semua pihak yang telah banyak berpartisipasi, baik secara langsung maupun tidak langsung yang tak sempat penulis sebutkan satu per satu. Terima kasih untuk segala bantuan dan dukungannya.

Dipindai dengan CamScanner

Penulis menyadari bahwa masih banyak kekurangan dalam tugas akhir ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini memberikan manfaat untuk pembaca.

Makassar, 11 Desember 2023

Nur Ukhty Hakim

Dipindai dengan CamScanner

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama	: Nur Ukhty Hakim
NIM	: H051171511
Program Studi	: Statistika
Departemen	: Statistika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam
Jenis Karya	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-ekslusif (*Non-exclusive Royalty-Free Right*)** atas tugas akhir saya yang berjudul:

"Analisis Regresi Komponen Utama *Robust* Dengan Metode *Minimum Volume Ellipsoid* Pada Data Jumlah Pengangguran di Sulawesi Selatan"

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.
Dibuat di Makassar pada tanggal 11 Desember 2023.

Yang menyatakan

(Nur Ukhty Hakim)

Dipindai dengan CamScanner

ABSTRAK

Analisis Regresi Komponen Utama Robust merupakan perkembangan dari metode Analisis Komponen Utama (AKU) dengan menerapkan regresi *robust* pada metode Analisis Komponen Utama (AKU) dan Regresi Komponen Utama (RKU). Metode *robust* yang digunakan pada penelitian ini merupakan kombinasi antara analisis komponen utama robust menggunakan metode *Minimum Volume Ellipsoid* (MVE) dengan analisis regresi komponen utama robust menggunakan metode MM-*estimator*. Tujuan dari penelitian ini adalah untuk mendapatkan penaksir parameter model Analisis Regeresi Komponen Utama Robust yang digunakan untuk menganalisis data jumlah pengangguran di Sulawesi Selatan serta untuk mendapatkan model data jumlah pengangguran di Sulawesi Selatan berdasarkan analisis Regresi Komponen Utama Robust. Data yang digunakan pada penelitian ini adalah data jumlah pengangguran di Sulawesi Selatan tahun 2022 dengan data jumlah penganggura sebagai variabel terikat (Y), variabel bebas terdiri dari Jumlah penduduk miskin (X_1), Indeks Pembangunan Manusia (X_2), Angka Partisipasi Sekolah untuk umur 16-18 tahun (X_3), Angka Partisipasi Murni untuk umur 16-18 tahun (X_4), Rata-rata lama sekolah (X_5), dan Jumlah penduduk (X_6). Model analisis regresi komponen utama *robust* pada data jumlah pengangguran di Sulawesi Selatan menghasilkan nilai kebaikan model sebesar 40,75% dan nilai RSE sebesar 2485, ini menunjukkan bahwa diperoleh model yang efektif dalam menangani kasus multikolinieritas dan *outlier*.

Kata Kunci : Multikolinieritas, *Outlier*, Analisi Regresi Komponen Utama Robust, MM-estimator, *Minimum Volume Ellipsoid* (MVE), Pengangguran

Dipindai dengan CamScanner

ABSTRACT

Robust Principal Component Regression Analysis is a development of the Principal Component Analysis (PCA) method by applying *robust* regression to the Principal Component Analysis (PCA) and Principal Component Regression methods. The *robust* method used in this study is a combination of robust principal component analysis using the *Minimum Volume Ellipsoid* (MVE) method with robust principal component regression analysis using the *MM-estimator* method. The purpose of this research is to get the parameter estimator of Robust Principal Component Regression Analysis model which is used to analyze the data of unemployment number in South Sulawesi and to get the data model of unemployment number in South Sulawesi based on Robust Principal Component Regression analysis. The data used in this research is the data of the number of unemployment in South Sulawesi in 2022 with the data of the number of unemployment as the dependent variable (Y), the independent variable consists of the number of poor people (X_1), Human Development Index (X_2), School Participation Rate for 16-18 years old (X_3), Pure Participation Rate for 16-18 years old (X_4), Average length of schooling (X_5), and Total population (X_6). *The robust* principal component regression analysis model on the data of the number of unemployed in South Sulawesi produces a model goodness value of 40.75% and an RSE value of 2485, indicating that an effective model is obtained in handling cases of multicollinearity and *outliers*.

Keywords: Multicollinearity, *Outliers*, *Robust Principal Component Regression Analysis*, MM-estimator, *Minimum Volume Ellipsoid* (MVE), Unemployment.

Dipindai dengan CamScanner

DAFTAR ISI

HALAMAN JUDUL i
LEMBAR PERNYATAAN KEOTENTIKAN ii
LEMBAR PERSETUJUAN PEMBIMBIMBING iii
HALAMAN PENGESAHAN iv
KATA PENGANTAR v
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS viii
ABSTRAK ix
ABSTRACT x
DAFTAR ISI xi
DAFTAR GAMBAR xiii
DAFTAR TABEL xiv
DAFTAR LAMPIRAN xv
BAB I PENDAHULUAN 1
1.1. Latar Belakang 1
1.2. Rumusan Masalah 3
1.3. Batasan Masalah 4
1.4. Tujuan Penelitian 4
1.5. Manfaat Penelitian 4
BAB II TINJAUAN PUSTAKA 5
2.1. Analisis Regresi Linear Berganda 5
2.2. Multikolinieritas 5
2.3. Outlier 6
2.4. Jarak Mahalanobis 6
2.5. Analisis Komponen Utama 7
2.6. Regresi Komponen Utama 9
2.7. Analisis Regresi Komponen Utama Robust 9
2.8. Metode *Minimum Volume Ellipsoid* 10
2.8. Metode MM-Estimator 11
2.9. Pengangguran 12
BAB III METODOLOGI PENELITIAN 13

Dipindai dengan CamScanner

3.1. Sumber Data 13
3.2. Deskripsi Variabel 13
3.3. Metode Analisis 14
BAB IV HASIL DAN PEMBAHASAN 16
4.1. Estimasi Model Analisis Regresi Komponen Utama Robust 16
4.2. Standarisasi Data 17
4.3. Regresi Linear Beranda 17
4.4. Menghitung Penduga *Minimum Volume Ellipsoid* 21
4.5. Analisis Komponen Utama 22
4.6. Analisis Regresi Komponen Utama Robust dengan MM-*Estimator* .. 26
BAB V PENUTUP 31
5.1. Kesimpulan 31
5.2. Saran 32
DAFTAR PUSTAKA 33
LAMPIRAN 35

Dipindai dengan CamScanner

DAFTAR GAMBAR

Gambar 4.1 Q-Q Plot Pencilan dengan metode Mahalanobis 28

Dipindai dengan CamScanner

DAFTAR TABEL

Table 4.1 Uji One-Sample Kolmogorov Smirnov 19
Tabel 4.2 Nilai VIF 19
Table 4.3 Hasil Uji Siginifikansi Parameter Model Regresi Linier Berganda dengan Uji t 21
Table 4.4 Mean berdasarkan MVE 22
Tabel 4.5 Matriks Varian Kovarian Berdasarkan Metode MVE 22
Table 4.6 Nilai Eigen 22
Table 4.7 Vektor Eigen 22
Table 4.8 Proporsi Keragaman dan Keragaman Kumulatif 23
Table 4.9 Hasi Uji Signifikansi Parameter Model PCR dengan Uji ***t*** 25
Tabel 4.10 Nilai VIF 26
Table 4.11 Jarak Mahalanobis Robust Kuadrat Pada Data AKU 27
Table 4.12 Deteksi Outlier dengan Jarak Mahalanobis Robust 27
Table 4.13 *Hasil Iterasi Memperoleh Nilai Koefesien Parameter* Robust 29
Table 4.14 Hasi Uji Signifikansi Parameter Model ROBPCA dengan Uji t 29

Dipindai dengan CamScanner

DAFTAR LAMPIRAN

Lampiran 1. Data Jumlah Pengangguran di Sulawesi Selatan Tahun 2022..... 36
Lampiran 2. Deskriptif data sebelum standarisasi dan sesudah standarisasi 37
Lampiran 3. Hasil Standarisasi Data 38
Lampiran 4. Model Regresi Linear Berganda 39
Lampiran 5. Tabel distribusi Chi-Square 40
Lampiran 6. Tabel Uji F 41
Lampiran 7. Tabel Uji t 43
Lampiran 8. Output Penduga MVE Pada Rstudio 44
Lampiran 9. Skor Komponen Utama 45
Lampiran 10. Nilai Eigen dan Vektor Eigen 46
Lampiran 11. Output Regresi Komponen Utama..... 47
Lampiran 12. Hasil Transformasi dari Variabel baru ke Variabel asal pada Model PCR: 48
Lampiran 13. Penduga MVE Berdasarkan Skor Komponen Utama..... 49
Lampiran 14. Jarak Mahalanobis dan Q-Q Plot 50
Lampiran 15. Analisis Regresi Komponen Utama Robust dari software R..... 51
Lampiran 16. Hasil Transformasi dari Variabel baru ke Variabel asal pada Model ROBPCA..... 54

Dipindai dengan CamScanner

BAB I
PENDAHULUAN

1.1. Latar Belakang

Analisis regresi merupakan salah satu metode analisis data yang digunakan untuk menyelidiki hubungan antara variabel respon dengan satu atau beberapa variabel bebas. Apabila hubungan yang diselidiki terdiri dari satu variabel respon dan satu variabel bebas maka disebut dengan analisis regresi linear sederhana. Sedangkan analisis regresi linear berganda terdiri dari satu variabel respon dan lebih dari satu variabel bebas. Salah satu cara untuk mendapatkan koefisien regresi linear berganda adalah dengan menggunakan metode MKT (Metode Kuadrat Terkecil). Dalam MKT terdapat beberapa asumsi yang harus dipenuhi yaitu tidak adanya multikolienaritas. Multikolienaritas akan menyebabkan ragam penduga koefisien regresi menjadi besar sehingga selang kepercayaan untuk parameter regresi juga cenderung akan lebih lebar (Draper & Smith, 2011).

Untuk mengatasi masalah multikolinieritas dapat dilakukan dengan menggunakan metode Analisis Komponen Utama (AKU). Metode AKU merupakan salah satu metode yang digunakan untuk menyelesaikan pelanggaran asumsi multikolinieritas yang terjadi pada regresi linier berganda. Metode AKU 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 komponen utama. Hubert, dkk (2005) dalam Larasati, dkk (2020) menjelaskan bahwa proses terbentuknya metode AKU ini memiliki dua langkah yaitu, pertama komponen utama terbentuk berdasarkan matriks varian kovarian atau matriks korelasi variabel bebas dengan memanfaatkan vektor eigen dan yang kedua komponen utama 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 AKU memerlukan suatu metode yang kuat terhadap *outlier*.

Outlier merupakan pengamatan yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat data (Noeryanti, 2001). Keberadaan *outlier* pada data

Dipindai dengan CamScanner

dapat mengakibatkan penyimpangan terhadap hasil analisis data seperti penyimpangan terhadap hasil uji statistik berdasarkan parameter rata-rata dan kovariansi. Oleh sebab itu, perlu dilakukan penanganan terhadap keberadaannya dan salah satu metode yang dapat mengatasi masalah *outlier* yaitu dengan menggunakan metode *robust*. AKU *Robust* dapat mengatasi masalah *outlier* dan multikolinieritas dengan menggunakan matriks kovarian *robust* pada analisis komponen utama dan metode regresi *robust* pada tahap analisis regresi.

Salah satu metode robust bagi matriks kovarian sampel yaitu metode *Minimum Volume Ellipsoid* (MVE). MVE diperkenalkan pertama kali oleh Rousseeuw pada tahun 1985 sebagai penaksir *robust*. MVE merupakan penaksir *Robust* dengan *high breakdown* pada lokasi dan *scatter* multivariat yang sering digunakan (Aelst & Rousseeuw, 2009). MVE terkenal karena kepekaannya terhadap *outlier* yang membuatnya dapat diandalkan untuk mendeteksi *outlier*. Alasan mendasar menggunakan metode ini adalah karena metode ini sangat berguna dalam mendeteksi *outlier* pada data multivariat, karena memenuhi jumlah batas maksimum *outlier* yaitu 50%.

Pengangguran adalah sebuah golongan yang sudah masuk usia kerja, tetapi belum melakukan suatu kegiatan atau pekerjaan yang menghasilkan uang. Adanya pengangguran bisa menyebabkan produktivitas dan pendapatan masyarakat berkurang. Kondisi tersebut yang sering menimbulkan kemiskinan dan masalah-masalah sosial lainnya. Di Sulawesi Selatan sendiri jumlah pengangguran mengalami penurunan, hal tersebut didasarkan pada data BPS, pada bulan Agustus 2022 jumlah pengangguran 4,51%, turun sebesar 1,21% dibandingkan Agustus 2021 yakni 5,72%. Meski begitu, pemerintah provinsi Sulawesi Seatan masih perlu bekerja keras untuk lebih menekan angka jumlah pengangguran di Sulawesi Selatan. Studi kasus yang dilakukan adalah untuk mendapatkan model RKU Robust pada data Jumlah Pengangguran di Sulawesi Selatan berdasarkan metode MVE.

Penelitian mengenai RKU *Robust* telah banyak dilakukan oleh peneliti sebelumnya dengan menggunakan berbagai metode diantaranya Jeffri dkk (2019) melakukan analisis Regresi Komponen Utama Robust *S-Estimator* dengan pendekatan *Minimum Volume Ellipsoid* (MVE) dengan kesimpulan bahwa model

Dipindai dengan CamScanner

regresi komponen utama yang menerapkan metode robust MVE dan *S-estimator* merupakan model yang paling efektif jika dibandingkan dengan model lainnya. Hal ini dikarenakan model tersebut mempunyai nilai Adjusted R^2 yang paling besar dan nilai RSE yang paling kecil. Siska dkk (2020) menganalisis Regresi Komponen Utama Robust dengan Metode *Minimum Covariance Determinant – Least Trimmed Square* (MCD-LTS) dengan kesimpulan bahwa metode RKU robust menggunakan MCD-LTS memberikan hasil yang baik dan merupakan metode efektif dan efisien dalam mengatasi masalah multikolinearitas dan Outlier. Ayu Riski (2021) melakukan Pemodelan Statistical Downscaling Menggunakan Regresi Komponen Utama Dengan Metode *Minimum Vector Variance* Untuk Pendugaan Curah Hujan dengan kesimpulan bahwa terdapat 5 kompoen utama yang optimal untuk digunakan pada model SD dengan regresi komponen utama dan penambahan 3 peubah dummy berdasarkan k-means cluster sebagai peubah prediktor untuk mengatasi masalah heterogenan sisaan data serta curah hujan lokal yang diperoleh dari regresi komponen utama dummy (k-means) merupakan model yang terbaik berdasarkan nilai koefisien determinasi (R^2)sebesar 92% dengan root mean square error (RMSE) yaitu 84.33, serta validasi modelnya menghasilkan root mean square error prediction (RMSEP) yang lebih rendah yaitu 94.74 dengan korelasi curah hujan aktual dan dugaannya sebesar 0.9570.

Berdasarkan penjelasan tersebut penulis bermaksud untuk menggunakan metode *Minimum Volume Ellipsoid* (MVE) pada Analisis Regeresi Komponen Utama *Robust* (AKU *Robust*) dengan judul "**Analisis Regresi Komponen Utama *Robust* Dengan Metode *Minimum Volume Ellipsoid* Pada Data Jumlah Pengangguran di Sulawesi Selatan**"

1.2. Rumusan Masalah

Rumusan masalah berdasarkan pada latar belakang yang telah dikemukakan diatas adalah:

1. Bagaimana penaksir parameter model Analisis Regresi Komponen Utama Robust yang digunakan untuk menganalisis data jumlah pengangguran di Sulawesi Selatan?
2. Bagaimana model data jumlah pengangguran di Sulawesi Selatan berdasarkan Analisis Regresi Komponen Utama Robust?

Dipindai dengan CamScanner

1.3. Batasan Masalah

Berdasarkan permasalahan yang telah diuraikan, batasan masalah dalam penelitian ini yaitu hanya mengkaji mengenai data jumlah pengangguran di Sulawesi Selatan tahun 2020-2022. Data tersebut merupakan data yang mengandung Multikolienaritas dan *Outlier*. Menggunakan estimasi regresi robust MM-Estimator. Serta penelitian ini juga membahas tentang Analisis Regresi Komponen Utama *Robust* menggunakan metode *Minimum Volume Ellipsoid*.

1.4. Tujuan Penelitian

Tujuan penelitian ini berdasarkan pada rumusan masalah adalah:

1. Untuk mendapatkan penaksir parameter model Regeresi Komponen Utama Robust yang digunakan untuk menganalisis data jumlah pengangguran di Sulawesi Selatan.
2. Untuk mendapatkan model data jumlah pengangguran di Sulawesi Selatan berdasarkan Regresi Komponen Utama Robust.

1.5. Manfaat Penelitian

Manfaat dari peneletian ini adalah agar dapat menjadi referensi kepada pembaca mengenai Analisis Komponen Utama *Robust* dengan menggunakan metode *Minimum Volume Ellpsoid*. Selain itu, penelitian ini bermanfaat sebagai sumber informasi tentang jumlah pengangguran di Sulawesi Selatan.

Dipindai dengan CamScanner

BAB II
TINJAUAN PUSTAKA

2.1. Analisis Regresi Linear Berganda

Analisis regresi merupakan alat statistik yang bermanfaat untuk mengetahui hubungan antara dua variabel atau lebih sehingga salah satu variabel dapat diduga dari variabel lainnya. Analisis regresi digambarkan dalam model regresi yaitu suatu cara untuk mengekspresikan dua unsur penting suatu hubungan statistik, yaitu kecenderungan berubahnya variabel dependen (Y) sejalan dengan berubahnya variabel independen (X) dan berpencarnya titik-titik di sekitar kurva hubungan statistik itu. Jika analisis regresi dilakukan untuk satu varibel tidak bebas (Y) dengan lebih dari satu variabel bebas (X) maka regresi ini dinamakan regresi linier berganda dengan model (Kurniawati, 2011)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \varepsilon_i, i = 1,2, \dots, n \qquad (2.1)$$

dengan, Y_i adalah variabel dependen pada pengamatan ke-i, X_{pi} adalah variabel independen pada pengamatan ke-i, dan $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ adalah parameter regresi yang tidak diketahui nilainya dan akan dicari nilai estimasinya, ε_i adalah galat yang berdistribusi normal dengan mean nol dan variansi σ^2atau $\varepsilon_i \sim N(0, \sigma^2)$.

2.2. Multikolinieritas

Istilah multikolinearitas mula-mula ditemukan oleh Ragnar Frisch pada tahun 1943. Pada mulanya multikolinearitas berarti adanya hubungan linear yang sempurna atau pasti, diantara beberapa atau semua peubah bebas dari model regresi ganda (Gujarati, 1995). Menurut Montgomery (1992) *dalam* Agriska (2011:37), menyatakan bahwa salah satu ukuran yang dapat digunakan untuk menguji adanya multikolinearitas pada regresi linear berganda adalah *Variance Inflation Factors* (VIF). Adanya multikolinearitas dinilai dari nilai VIF yang dihasilkan. Besarnya nilai VIF ini bergantung pada nilai koefisien determinasi (R^2) yang dihasilkan. Semakin besar nilai (R^2) maka semakin besar pula nilai VIF yang dihasilkan.

Mengidentifikasi multikolinearitas pada data presipitasi menggunakan *variance inflation factors* (VIF) dengan Persamaan (2.2) :

Dipindai dengan CamScanner

$$VIF_j = \frac{1}{1-R_j^2} \tag{2.2}$$

dengan R_j^2 adalah koefisien determinasi dari peubah penjelas yang diregresikan terhadap peubah penjelas lainnya. Dari banyaknya penelitian yang dilakukan mengindikasikan bahwa jika ada nilai VIF yang melebihi 10, maka ini menandakan bahwa koefisien-koefisien regresi adalah estimasi yang kurang baik karena pengaruh multikolinearitas.

2.3. Outlier

Outlier merupakan pengamatan yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat data. Keberadaan *Outlier* pada data dapat mengakibatkan penyimpangan terhadap hasil analisis data seperti penyimpangan terhadap hasil uji statistik berdasarkan parameter rata-rata dan kovarians. Oleh karena itu, perlu dilakukan identifikasi terhadap keberadaannya (Noeryanti, 2001). Identifikasi *Outlier* pada data multivariat menggunakan jarak Mahalanobis. Pengamatan didefinisikan sebagai *Outlier* apabila jarak Mahalanobis kuadrat lebih besar dari nilai chi-square pada *p* peubah (Jhonson & Wichern, 2007)

2.4. Jarak Mahalanobis

Jarak mahalanobis merupakan metode untuk mendeteksi Outlier pada data multivariat. Jarak mahalanobis diperoleh dengan menghitung jarak tiap observasi terhadap pusat datanya. Kuadrat jarak mahalanobis dihitung dengan rumusan (Jhonson & Wichern, 2007) sebagai berikut:

$$d_i^{\,2} = (\boldsymbol{x_i} - \overline{\boldsymbol{x}})' S^{-1} (\boldsymbol{x_i} - \overline{\boldsymbol{x}}) > X^2_{p,(1-\alpha=0{,}05)} \tag{2.3}$$

dimana, $d_i^{\,2}$ = kuadrat jarak pengamatan ke-i, $\bar{x}$ = vector rata-rata dari pengamatan, dan x_i = nilai pengamatan ke-i .

Jarak Mahalanobis dievaluasi dengan menggunakan X^2 pada derajat kebebasan (df) sejumlah variabel yang digunakan dalam penelitian. Identifikasi data Outlier pada pengamatan ke-i didefinisikan sebagai Outlier apabila jarak Mahalanobis kuadrat lebih besar dari nilai chi-square (X^2) pada p peubah (Jhonson & Wichern, 2007).

Dipindai dengan CamScanner

2.5. Analisis Komponen Utama

AKU adalah analisis *multivariate* yang mentransformasi variabel-variabel asal yang saling berkorelasi menjadi variabel-variabel baru yang tidak saling berkorelasi dengan mereduksi sejumlah variabel tersebut sehingga mempunyai dimensi yang lebih kecil namun dapat menerangkan sebagian besar keragaman variabel aslinya. AKU muncul sebagai solusi bagi proses pengumpulan data dimana data tersebut terdiri dari variabel-variabel yang jumlahnya sangat banyak sehingga diperoleh variabel-variabel baru yang jumlahnya lebih sedikit tetapi tetap mampu menjelaskan variansi data (Jhonson & Wichern, 2007).

AKU diperkenalkan pertama kali oleh Karl Pearson pada tahun 1901. Selanjutnya, dikembangkan oleh Hotelling pada tahun 1933 dan Rao's pada tahun 1964 yang hingga saat ini Analisis Komponen Utama (AKU) masih terus mengalami perkembangan. Dimana perkembangan komputer dan ilmu matematik, Analisis Komponen Utama juga hingga saat ini terus mengalami perkembangan. Perkembangan Analisis Komponen Utama yang selanjutnya dipengaruhi dengan adanya kebutuhan suatu model Analisis Komponen Utama (AKU) yang *Robust* terhadap data yang mengandung *outlier*. Pada Analasis Komponen Utama (AKU) klasik sangat rentang akan adanya data *outlier* hal ini didasarkan pada matriks varian kovarian yang sensitif terhadap keberadaan *outlier*. Maka dari itu, diperkenalkan konsep Analisis Komponen Utama *Robust* (AKU *Robust*) (Hubert, Rousseeuw, & Branden, 2005)

Misalkan suatu data terdiri dari n observasi dan p variabel yang telah dilakukan Standarisasi data dinyatakan dengan matriks pada persamaan (2.4) berikut:

$$\boldsymbol{X}_{n\times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \tag{2.4}$$

Persamaan (2.4) dapat dinyatakan dalam bentuk vektor yaitu $X = [X_1, X_2, \dots, X_p]'$ dengan $X_i' = [x_{i1}, x_{i2}, \dots, x_{ip}]$ dengan $i = 1,2,\dots,n$, matriks korelasi $\mathbf{R}$ dan nilai eigen $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ dengan vektor eigen yang bersesuaian $\boldsymbol{e_1}, \boldsymbol{e_2}, \dots, \boldsymbol{e_p}$.

Dipindai dengan CamScanner

Komponen utama yang terbentuk dapat dituliskan sebagai kombinasi linear dari variabel-variabel asalnya yaitu sebagai berikut:

$$K_1 = e_1'\boldsymbol{X} = e_{11}X_1 + e_{12}X_2 + \cdots + eX_p$$
$$K_2 = e_2'\boldsymbol{X} = e_{21}X_1 + e_{22}X_2 + \cdots + e_{2p}X_p \quad (2.5)$$
$$\vdots$$
$$K_p = e_p'\boldsymbol{X} = e_{p1}X_1 + e_{p2}X_2 + \cdots + e_{pp}X_p$$

atau

$$\boldsymbol{K} = \begin{bmatrix} K_1 \\ K_2 \\ \vdots \\ K_p \end{bmatrix} = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{np} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

Langkah-langkah metode AKU sebagai berikut:

1. Menghitung matriks varian kovarians **R**
2. Menghitung nilai eigen dari matriks varian kovarians **R** dengan menggunakan rumus:

$$|\boldsymbol{R} - \boldsymbol{\lambda} I| = 0 \quad (2.6)$$

3. Menghitung vektor eigen dari matriks varian kovarians **R** dengan menggunakan rumus:

$$|\boldsymbol{R} - \boldsymbol{\lambda} I|x = 0 \quad (2.7)$$

4. Vektor eigen dinormalisasi dengan menggunakan rumus:

$$\boldsymbol{e} = \frac{x}{\sqrt{x'x}} \quad (2.8)$$

5. Menghitung proporsi varians dengan menggunakan rumus:

$$\frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i} \quad (2.9)$$

6. Menghitung proporsi kumulatif yang dapat dijelaskan oleh komponen utama ke-k dengan menggunakan rumus:

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} \lambda_i} \quad (2.10)$$

7. Membentuk komponen utama ke-1,2,…,k dengan menggunakan Persamaan (2.5)
8. Menghitung skor komponen utama berdasarkan pada poin 7

Dipindai dengan CamScanner

2.6. Regresi Komponen Utama

Regresi komponen utama merupakan teknik analisis regresi yang dikombinasikan dengan analisis komponen utama, dimana analisis komponen utama dijadikan sebagai tahap analisis. Salah satu asumsi yang harus dipenuhi untuk uji hipotesis pada parameter anallisis linier berganda adalah kebebasan multikolinieritas. Analisis komponen utama merupakan analisis yang memperkecil dimensi peubah tanpa kehilangan banyak informasi, dengan tujuan menyederhanakan peubah yang diamati dengan cara mereduksi dimensinya sehingga dihasilkan peubah-peubah baru yang merupakan kombinasi linier dari peubah-peubah prediktor asal dan antar peubah baru ini bersifat saling bebas (Ramadani, 2021).

Setelah dilakukan analisis komponen utama dan telah diperoleh skor komponen utama, langkah selanjutnya adalah meregresikan komponen-komponen utama terpilih dengan peubah respon. Bentuk Persamaan RKU dapat dilihat pada Persamaan (2.11) berikut :

$$\hat{y}_i = \beta_0 + \sum_{j=1}^{p} \beta_j K_j \tag{2.11}$$

dengan :

$\hat{y}_i$ =Penduga peubah respon pada pengamatan ke-i ;=1,2,… ,n.,

β_0 = nilai intersep,

β_j = koefisien regresi bagi komponen utama ke-j,

K_j = Komponen Utama ke-j; j=1, 2, … ,.,

p = banyaknya komponen utama.

2.7. Analisis Regresi Komponen Utama Robust

Menurut Larasati dkk, (2020), metode analisis regresi komponen utama *robust* adalah sebagai metode pengembangan dari metode AKU (Analisis Komponen Utama) 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 AKU dalam menyelesaikan masalah multikolinieritas pada model regresi linier berganda. Metode analisis regresi komponen utama *robust* dalam mengatasi masalah multikolieriatas yaitu

Dipindai dengan CamScanner

dengan menggunakan matriks varian kovarian *robust* pada PCA dan metode regresi *robust* pada tahap analisis regresi.

Metode analisis regresi komponen utama *robust* 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.

2.8. Metode *Minimum Volume Ellipsoid*

Metode *Minimum Volume Ellipsoid* (MVE) pertama kali diperkenalkan oleh Rousseeuw , merupakan metode penduga robust untuk vektor nilai tengah dan matriks peragam. Pada prinsipnya metode ini adalah mencari ellipsoid dengan volume paling minimum yang melingkupi suatu subhimpunan dari minimal h pengamatan. Subhimpunan berukuran h ini disebut halfset karena h sering dipilih lebih dari setengah n pengamatan. Penduga nilai tengah adalah pusat ellipsoid secara geometris dan penduga matriks peragam adalah matriks pembentuk ellipsoid. Dalam mencari solusi eksak untuk menyelesaikan masalah pada MVE ini dapat dikatakan sulit, maka alternatif lain adalah melalui pendekatan algoritma resampling yang disebut juga algoritma $(p + 1)$ subset.

Berikut merupakan langkah-langkah algoritma resampling untuk mendapatkan elipsoid dengan volume minimum (Aelst & Rousseeuw, 2009):

1. Untuk setiap subset ukuran $(p + 1)$ disimbolkan dengan indeks $J = \{i_1, \dots, i_{p+1}\} \subset \{1, \dots, n\}$
2. Menentukan rata-rata (*mean*) sampel dengan rumus:

$$\bar{x}_J = \frac{1}{p+1}\sum_{i=1}^{p+1} x_{ij} \tag{2.12}$$

Selanjutnya menentukan matriks kovarian dari sampel dengan rumus:

$$S_J = \frac{1}{p}\sum_{i=1}^{p+1}(x_{ij} - \bar{x}_J)'\left(x_{ij} - \bar{x}_J\right) \tag{2.13}$$

3. Menghitung jarak mahalanobis untuk semua pengamatan dengan rumus:

$$D_J^2 = [(x_{ij} - \bar{x}_J)'S^{-1}\left(x_{ij} - \bar{x}_J\right)]_{h:n} \tag{2.14}$$

4. Sebagai penentu yang sangat penting dalam MVE adalah volume ellipsoid yang proporsional dengan nilai:

Dipindai dengan CamScanner

$$V_J = \left(\frac{D_J}{c}\right)^k \sqrt{\det(S_J)} \tag{2.15}$$

5. Setelah selesai langkah keempat, maka ulangi langkah pertama sampai keempat di atas untuk subsampel kedua. Proses diulang sampai sebanyak $\binom{n}{p+1}$ subsampel.
6. Pilih subsampel yang elipsoidnya memiliki volume paling minimum.
7. Hitung $\boldsymbol{T}(\boldsymbol{X})$ dan $\boldsymbol{S}(\boldsymbol{X}) = c^2(n,k)\left(X^2_{k,a}\right)^{-1} D_J^2 \boldsymbol{S}_J$ dari V_J yang akan digunakan sebagai estimasi mean dan matriks varians dan kovarians, dengan $c^2(n,k) = \left[1 + \frac{15}{n-k}\right]^2$ yang disebut *correction term*.

2.8. Metode MM-Estimator

MM-Estimasi merupakan estimasi yang menggabungkan *high breakdown point* dan efisiensi statistika yang diperkenalkan oleh Yohai (1987). Langkah pertama dalam estimasi ini adalah mencari estimator S, kemudian menetapkan parameter-parameter regresi menggunakan estimasi M. Estimasi S menjamin nilai *breakdown point* tinggi dan estimasi M membuat estimator mempunyai efisiensi tinggi. Pada umumnya digunakan fungsi *Tukey Bisquare* baik pada estimasi S maupun estimasi M (Nurdin, Raupong, & Islamiyati, 2014). Estimasi MM dinyatakan dalam persamaan (2.17) berikut:

$$\hat{\beta}_{MM} = \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}}\right) = \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}}\right) \tag{2.17}$$

di mana: ρ = fungsi obyektif dan $\hat{\sigma}$ = estimator skala robust.

Metode MM juga menggunakan IRLS (Iteratively Reweighted Least Square) untuk mencari estimasi parameter regresi (Kurniawati, 2011). Prosedur estimasi parameter pada model regresi linier ganda dengan regresi *robust* estimasi MM:

1. Menghitung estimator awal koefisien β_j dan residual e_i dari regresi *robust* dengan *high breakdown point* (estimasi S) dengan bobot huber / bisquare (dilihat sebagai bentuk estimasi M).
2. Residual e_i pada langkah pertama digunakan untuk menghitung skala estimasi $\hat{\sigma}_s$ dan dihitung pula pembobot awal w_i.

Dipindai dengan CamScanner

3. Residual e_i dengan skala estimasi $\hat{\sigma}_s$ pada langkah kedua digunakan dalam iterasi awal sebagai penaksir WLS untuk menghitung koefisien regresi. $\sum_{i=1}^{n} w_i \left(\frac{e_i}{\hat{\sigma}_s}\right) x_i = 0$, w_i merupakan pembobot Huber/bisquare.
4. Menghitung bobot baru w_i dengan skala estimasi dari iterasi awal WLS.

Mengulang langkah 2, 3, 4 (dengan skala estimasi tetap konstan) sampai mendapatkan $\sum_{i=1}^{n} \left| e_i^{(m)} \right|$ konvergen (selisih $\hat{\beta}_j^{(m+1)}$ dan $\hat{\beta}_j^{(m)}$ mendekati 0, dengan m banyaknya iterasi).

2.9. Pengangguran

Secara umum, pengertian pengangguran adalah orang yang tidak mempunyai pekerjaan. Pengertian lainnya, pengangguran adalah sebutan untuk orang yang tidak bekerja sama sekali, sedang mencari kerja, bekerja kurang dari 2 hari selama seminggu, atau sedang berusaha mendapatkan pekerjaan yang layak.

Pengangguran adalah seseorang yang tergolong dalam angkatan kerja (usia 15-65 tahun) dan ingin mendapatkan pekerjaan, namun belum berhasil memperolehnya. Umumnya penyebab pengangguran adalah ketidakseimbangan lapangan kerja yang tersedia dengan laju pertumbuhan penduduk. Artinya jumlah tenaga kerja lebih banyak dibandingkan jumlah lapangan kerja, sehingga menyebabkan beberapa orang tidak mendapatkan pekerjaan.

Menurut International Labor Organization (ILO) tedapat dau bentuk pengangguran, yaitu pengangguran terbuka dan setengan pengangguran terpaksa. Pengangguran Terbuka adalah seseorang dalam kelompk penduduk usia kerja yang tidak bekerja selama periode tertentu dan bersedia menerima atau mecari pekerjaan. Setengah Pengangguran Terpaksa adalah seseorang yang bekerja sebagai buruh karyawan dan pekerja mandiri yang selama periode tertentu terpaksa bekerja kurang dari jam kerja normal.

Dipindai dengan CamScanner