PERBANDINGAN METODE *MINIMUM COVARIANCE DETERMINANT* DAN *MINIMUM VECTOR VARIANCE* DALAM PEMODELAN REGRESI KOMPONEN UTAMA PADA DATA PRODUK DOMESTIK REGIONAL BRUTO

NURUL DWINILDA ZHALZHABILAH
H051191014

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
2024

PERBANDINGAN METODE *MINIMUM COVARIANCE DETERMINANT* DAN *MINIMUM VECTOR VARIANCE* DALAM PEMODELAN REGRESI KOMPONEN UTAMA PADA DATA PRODUK DOMESTIK REGIONAL BRUTO

NURUL DWINILDA ZHALZHABILAH
H051191014

Skripsi

Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Statistika

Program Studi Statistika

pada

PROGRAM STUDI STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024

iii

SKRIPSI

PERBANDINGAN METODE *MINIMUM COVARIANCE DETERMINANT* DAN *MINIMUM VECTOR VARIANCE* DALAM PEMODELAN REGRESI KOMPONEN UTAMA PADA DATA PRODUK DOMESTIK REGIONAL BRUTO

yang disusun dan diajukan oleh

NURUL DWINILDA ZHALZHABILAH
H051191014

Skripsi,

Telah dipertahankan di depan Panitia Ujian Sarjana Statistika pada 05 Agustus 2024
dan dinyatakan telah memenuhi syarat kelulusan

pada

Program Studi Statistika
Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Hasanuddin
Makassar

Mengesahkan: Pembimbing utama tugas akhir,	Mengesahkan: Pembimbing pertama tugas akhir,
Sitti Sahriman, S.Si., M.Si. NIP. 198810182015042002	Dr. Nirwan Ilyas, M.Si. NIP. 196303061987021002

Mengetahui:
Ketua Program Studi,

Dr. Anna Islamiyati, S.Si., M.Si.
NIP. 197706082005012002

iv

PERNYATAAN KEASLIAN SKRIPSI DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, skripsi berjudul "Perbandingan Metode *Minimum Covariance Determinant* Dan *Minimum Vector Variance* Dalam Pemodelan Regresi Komponen Utama Pada Data Produk Domestik Regional Bruto" adalah benar karya saya dengan arahan dari pembimbing Sitti Sahriman, S.Si., M.Si dan Dr. Nirwan, M.Si.. Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka skripsi ini. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini adalah karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut berdasarkan aturan yang berlaku.

Dengan ini saya melimpahkan hak cipta (hak ekonomis) dari karya tulis saya berupa skripsi ini kepada Universitas Hasanuddin.

Makassar, 05 Agustus 2024

10000 METERAI TEMPEL
497ALX325656322

Nurul Dwinilda Zhalzhabilah

H051191014

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi ini. Shalawat serta salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa Sallam* beserta keluarga dan para sahabatnya. *Alhamdulillahirobbil'alamin*, berkat umur yang panjang beserta kesehatan yang diberikan oleh Allah *Subhanahu Wa Ta'ala*, penulis dapat menyelesaikan skripsi yang berjudul "**Perbandingan Metode *Minimum Covariance Determinant* Dan *Minimum Vector Variance* Dalam Pemodelan Regresi Komponen Utama Pada Data Produk Domestik Regional Bruto**" atas bimbingan dan arahan Ibu **Sitti Sahriman, S.Si., M.Si.** dan Bapak **Dr. Nirwan, M.Si**. yang dengan penuh kesabaran dan dedikasi telah memberikan arahan, masukan, serta motivasi hingga skripsi ini dapat terselesaikan. Terimakasih atas bimbingan yang sangat berharga, waktu yang diluangkan, serta kesediaan untuk terus memberikan pandangan dan ilmu yang mendalam selama proses penyusuusnan skripsi ini. Saya juga mengucapkan terima kasih kepada Bapak **Siswanto S.Si,. M.Si.** dan Bapak **Drs. Raupong, M.Si.** selaku dosen penguji yang telah meluangkan waktu dalam memberikan kritik dan saran yang sangat berharga untuk penyempurnaan skripsi ini. Terima kasih atas evaluasi yang mendalam serta pandangan yang tajam dalam proses pengujian.

Penulis menyadari bahwa dalam penyelesaian skripsi ini tidak lepas dari bantuan dan dorongan dari berbagai pihak yang senantiasa turut membantu dalam bentuk moril maupun materil sehingga dengan segala keterbatasan kemampuan dan pengetahuan, penulis dapat menyelesaikan skripsi ini. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang setulus-tulusnya serta penghargaan yang setinggi-tingginya kepada Ayahanda **Arif** dan Ibunda **Herawati**, yang selalu mendoakan, mendukung, dan memberikan kasih sayang yang tiada henti. Tanpa doa dan pengorbanan kalian saya tidak akan sampai pada titik ini. Terimakasih juga kepada Kakanda **Muhammad Nurfadly** dan Adik **Zhahrani Chintyabella** yang telah memberikan dukungan, limpahan cinta dan kasih sayang, serta dengan ikhlas telah menemani setiap langkah penulis dengan doa dan restu mulianya.

Penulis juga berterima kasih kepada sahabat-sahabat yang telah membersamai selama kuliah, **Melda Fitriyani Azis, Stevania Saskia**, dan **Mellyana Masa** yang selalu menjadi tempat berbagi cerita dan memberikan dukungan sepanjang perjalanan ini. Terima kasih telah menjadi sahabat yang setia mendengar keluh kesah dan selalu ada di saat suka maupun duka. Ucapan terimakasih juga saya sampaikan kepada **Shafira Dwi Chaerunnisa, Sintia Afrisa Wulandari, Fadia Utami Putri,** dan **Syamsuar Vina** selaku sahabat yang selalu mendorong dan memberikan semangat kepada penulis untuk menyelesaikan skripsi ini. Terima kasih atas persahabatan, kebersamaan, dan dukungan yang tak ternilai.

vi

Ucapan terima kasih juga penulis sampaikan kepada **Nurfadhila Muliyadi, Asrini Budiyanti**, dan **Arwiny Budiarti** selaku sepupu yang selalu memberikan dukungan dan semangat kepada penulis. Terima kasih kepada kanda-kanda staff admin Departemen Statistika, kak **Andi Anugrah Adil** dan kak **Muhammad Nurfadli** yang selalu baik dalam membantu dan memberikan informasi yang diperlukan, serta memberikan dukungan kepada penulis. Terima kasih kepada teman-teman **KKNT Unhas 108 Desa Corawali** yang telah menjadi keluarga baru dan memberikan banyak pengalaman berharga. Terakhir, terimakasih kepada teman-teman **Statistika 2019**, terkhusus **Musfira Hidayah** dan **Fachraeni Ulfiana**, atas kebersamaan dan dukungan yang tak ternilai selama masa perkuliahan hingga penyelesaian skripsi ini.

Makassar, 05 Agustus 2024

Nurul Dwihilda Zhalzhabilah

ABSTRAK

Nurul Dwinilda Zhalzhabilah. **Perbandingan Metode *Minimum Covariance Determinant* dan *Minimum Vector Variance* dalam Pemodelan Regresi Komponen Utama pada Data Produk Domestik Regional Bruto** (dibimbing oleh Sitti Sahriman dan Nirwan Ilyas).

Latar Belakang. Regresi Komponen Utama (RKU) adalah salah satu model yang dapat digunakan untuk menagatasi masalah multikolinearitas. RKU diawali dengan Analisis Komponen Utama (AKU) dalam pembentukan komponen utama. Namun dalam AKU klasik, matriks varian kovarian yang terbentuk tidak *robust* terhadap pencilan. Metode *Minimum Covariance Determinant* (MCD) dan *Minimum Vector Variance* (MVV) adalah pendekatan yang dapat digunakan untuk membentuk matriks varian kovarian yang dapat menangani masalah pencilan. **Tujuan**. Penelitian ini bertujuan membandingkan metode MCD dan MVV dalam pemodelan regresi komponen utama. **Metode**. Penelitian ini menggunakan metode MCD dan MVV untuk membentuk matriks varian kovarian yang *robust* terhadap pencilan. Matriks varian kovarian yang diperoleh, diubah menjadi matriks korelasi yang kemudian digunakan untuk membentuk komponen utama pada AKU. **Hasil.** Pembentukan komponen utama berdasarkan proporsi keragaman kumulatif minimal sebesar 80% untuk metode MCD dan MVV menghasilkan masing-masing 2 komponen utama. Model RKU dengan MCD memiliki performa yang lebih bagus berdasarkan nilai R^2 tertinggi sebesar 90% dan nilai MSE terkecil, sebesar 150.456 dibandingkan dengan metode MVV yang memiliki nilai R^2 sebesar 89% dan nilai MSE sebesar 157.021. **Kesimpulan.** Model MCD lebih efektif dalam mengatasi pencilan dan multikolinearitas daripada MVV untuk pemodelan regresi komponen utama pada data PDRB tahun 2020, dengan estimasi parameter yang lebih akurat berdasarkan nilai R^2 yang terbesar dan nilai MSE yang terkecil.

Kata Kunci: *Minimum Covariance Determinant, Minimum Vector Variance,* Multikolinearitas, Pencilan, Regresi Komponen Utama.

ABSTRACT

Nurul Dwinilda Zhalzhabilah. **Perbandingan Metode *Minimum Covariance Determinant dan Minimum Vector Variance dalam Pemodelan Regresi Komponen Utama pada Data Produk Domestik Regional Bruto*** (supervised by Sitti Sahriman dan Nirwan Ilyas).

Background. Principal Component Regression (PCR) is a model that can be used to address multicollinearity issues. PCR begins with Principal Component Analysis (PCA) for the formation of principal components. However, in classical PCA, the covariance matrix formed is not robust to outliers. The Minimum Covariance Determinant (MCD) and Minimum Vector Variance (MVV) methods are approaches that can be used to form a covariance matrix that can handle outliers. **Aim**. This study aims to compare the MCD and MVV methods in Principal Component Regression modeling. **Method**. This study employs the MCD and MVV methods to form a covariance matrix that is robust to outliers. The resulting covariance matrix is converted into a correlation matrix, which is then used to form principal component in PCA. **Results**. The formation of principal components based on a cumulative variance proportion of 80% for both MCD and MVV methods results in two principal components each. The PCA model with MCD shows better performance with the highest R^2 value of 90% and the lowest MSE value of 150.456, compared to the MVV method with an R^2 value of 89% and an MSE value of 157.021. **Conclusion**. The MCD model is more effective in addressing outliers and multicollinearity than the MVV model for Principal Component Regression modeling on the 2020 GRDP data, with more accurate parameter estimates based on the highest R^2 value and the smallest MSE value.

Keywords: Minimum Covariance Determinant, Minimum Vector Variance, Multicolinearity, Outliers, Principal Component Regression,.

DAFTAR ISTILAH

Istilah	Arti dan Penjelasan
Algoritma	Prosedur atau langkah-langkah sistematis untuk menyelesaikan masalah dengan metode tertentu
Index	Indikator atau penunjuk posisi dalam data, berguna untuk mengidentifikasi elemen
Invers	Operasi matematika untuk menemukan kebalikan suatu matriks, berguna dalam berbagai analisis
Iterasi	Pengulangan langkah-langkah dalam proses algoritma untuk mencapai hasil optimal
Korelasi	Hubungan antara dua variabel dalam data set
Kovarian	Ukuran seberapa dua variabel berubah bersama, menunjukkan hubungan linier antara variabel
Max	Nilai terbesar dalam dataset, menunjukkan batas atas dari data yang dianalisis
Mean	Rata-rata dari nilai-nilai dalam dataset, memberikan representasi pusat data
Min	Nilai terkecil dalam dataset, menunjukkan batas bawah dari data yang dianalisis
Multikolinearitas	Kondisi ketika terdapat dua atau lebih variabel bebas yang saling berkorelasi
Multivariat	Melibatkan banyak variabel atau dimensi dalam analisis data atau model statistik
Observasi	Unit data individu dalam sampel penelitian
Pencilan	Data yang menyimpang jauh daro data lainnya dalam suatu rangkaian data.
Robust	Kemampuan metode untuk tetap akurat mesik terdapat data yang mengandung pencilan
Transformasi	Proses mengubah data dari satu bentuk ke bentuk lain

DAFTAR SINGKATAN DAN LAMBANG

LAMBANG/SINGKATAN	Arti dan Penjelasan
Y	Variabel terikat
X	Variabel data
ε	Nilai toleransi *error* terkecil
$\boldsymbol{d_i^2}$	Jarak Mahalanobis
$\bar{x}$	*Mean* atau rata-rata
$\boldsymbol{S}$	Matriks varian kovarian
$\boldsymbol{S^{-1}}$	Invers matriks varian kovarian
$\boldsymbol{R}$	Matriks korelasi
$\boldsymbol{D\left(\frac{1}{\sigma}\right)}$	Matriks diagonal
λ	Nilai eigen
n	Jumlah data
$\chi^{\mathbf{2}}$	*Chi square*
AKU	Analisis Komponen Utama
MCD	*Minimum Covariance Determinant*
MSE	*Mean Square Error*
MVV	*Minimum Vector Variance*
PDRB	Produk Domestik Regional Bruto
JKR	Jumlah Kuadrat Regresi
JKG	Jumlah Kuadrat Galat
VIF	*Variance Inflation Factor*

DAFTAR ISI

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan salah satu teknik statistik yang digunakan untuk menduga dan memodelkan hubungan antara satu variabel terikat dan satu atau lebih variabel bebas. Salah satu bentuk yang lebih kompleks dari metode analisis regresi linier adalah analisis regresi linier berganda. Analisis regresi linier berganda melibatkan beberapa variabel bebas dalam modelnya (Yusuf dkk, 2024). Namun, dalam analisis regresi linier berganda sering dihadapi masalah multikolinearitas.

Multikolinearitas merupakan hubungan linier atau korelasi yang tinggi antar variabel bebas dalam model regresi yang menyebabkan nilai penduga ragam bagi parameter regresi menjadi lebih besar (Larasati, 2019). Masalah multikolinearitas dapat menyebabkan penggunaan metode regresi menjadi kurang tepat karena estimasi koefisien regresi menjadi kurang stabil dan lebih bervariasi, yang dapat memengaruhi akurasi model (Azizah dkk., 2020). Oleh karena itu, diperlukan sebuah analisis statistik yang dapat mengatasi masalah multikolinearitas seperti analisis komponen utama.

Analisis komponen utama (AKU) merupakan metode statistik multivariat yang digunakan untuk membentuk variabel baru yang disebut dengan komponen utama yang saling bebas. Selain itu, AKU juga mampu mereduksi dimensi data dengan mentransformasi peubah asal yang saling berkorelasi menjadi peubah yang tidak saling berkorelasi (Greenacre dkk, 2022). Dalam membentuk komponen utama, salah satu cara yang dapat digunakan yaitu membentuk sebuah matriks varian kovarian. Namun secara umum, sifat matriks varian kovarian dalam analisis regresi komponen utama tidak bersifat *robust* terhadap keberadaan pencilan (Delsen dkk., 2017). Keberadaan pencilan pada data dapat mengakibatkan penyimpangan terhadap hasil uji statistik berdasarkan parameter rata-rata dan kovariansi. Oleh karena itu, perlu dilakukan penanganan terhadap keberadaannya.

Terdapat beberapa metode *robust* yang dapat mengukur jarak dan mengatasi data yang mengandung pencilan yaitu metode *minimum covariance determinant* (MCD) dan *minimum vector variance* (MVV). Metode MCD dilakukan dengan cara mengidentifikasi sebagian kecil observasi dari sampel utama yang memiliki nilai determinan kovariansi yang rendah. Proses ini bertujuan untuk menghasilkan matriks varian kovarian yang lebih *robust* terhadap pencilan (Leys dkk., 2018). Sedangkan metode MVV dilakukan dengan cara memilih sebagian kecil vektor dari sampel utama yang memiliki nilai varian paling rendah. Vektor-vektor ini kemudian digunakan untuk membentuk sebuah matriks yang *robust* terhadap pencilan (Polat & Hazlina, 2020). Hasil dari proses pembentukan komponen utama berdasarkan matriks varian kovarian yang *robust* terhadap pencilan dapat digunakan untuk membantu dalam membuat keputusan.

Pengambilan keputusan yang tepat dan sesuai dengan proses ilmiah merupakan upaya dalam meningkatkan pertumbuhan ekonomi dalam sektor pembangunan dan perkembangan perekonomian suatu negara. Hal ini dapat meningkatkan kemakmuran dan kesejahteraan masyarakat pada tingkat pendapatan per kapita (Hodijah & Angelina, 2021). Todaro dan Smith (2008) mengatakan salah satu tolak ukur pelaksanaan pembangunan dilihat dari bagaimana pertumbuhan ekonomi pada suatu wilayah. Pertumbuhan ekonomi dicerminkan dari Produk Domestik Regional Bruto (PDRB). Semakin tinggi pertumbuhan ekonomi suatu wilayah maka semakin baik kegiatan ekonominya. Infrastruktur menjadi sektor sangat penting dalam pertumbuhan ekonomi suatu negara karena diharapkan mampu mencapai pertumbuhan ekonomi yang tinggi (Fauziah, 2021).

Infrastruktur meliputi jalan, jalan raya, rel kereta api, pelabuhan, dan bandara. Infrastruktur yang baik dalam hal energi listrik, bahan bakar minyak (BBM), dan gas juga diperlukan untuk mendukung transportasi, industri dan rumah tangga. Selain itu, fasilitas publik seperti sekolah, rumah sakit, pasar, kantor polisi sangat penting . Serta fasilitas air yang mencakup air bersih, penanganan limbah, dan perngaturan banjir juga sangat dibutuhkan demi mendukung tercapainya kesejahteraan masyarakat suatu negara. Ketersediaan dan peningkatan jaringan telekomunikasi juga dapat memenuhi kebutuhan masyarakat terhadap informasi dan komunikasi dengan lebih baik (Putri & Wisudanto, 2017)

Penelitian yang dilakukan oleh Ramadani (2021), tentang pemodelan *statistical downscaling* pada regresi komponen utama dengan metode MVV untuk pendugaan curah hujan. Adapun penelitian yang dilakukan Nisa dkk. (2019) tentang analisis regresi komponen utama robust dengan metode *minimum covariance determinant-least trimmed square*. Penelitian yang dilakukan oleh Purnama (2021), tentang perbandingan analisis komponen utama dan *robust* PCA. Selanjutnya penelitian yang dilakukan oleh Pendi (2021), tentang analisis regresi dengan metode komponen utama dalam mengatasi masalah multikolinearitas.

Berdasarkan hal tersebut, maka penulis melakukan pemodelan regresi komponen utama dengan metode *minimum covariance determinant* dan *minimum vector variance* yang diharapkan dapat mengatasi masalah multikolinearitas dan pencilan pada data produk domestik regional bruto tahun 2020. Dengan demikian, dapat diperoleh estimasi dari model regresi komponen utama MCD dan MVV, serta membandingkan hasil estimasi parameter untuk menentukan metode yang lebih efektif dalam mengatasi masalah multikolinieritas dan pencilan pada kasus produk domestik regional bruto tahun 2020.

1.2 Batasan Masalah

1. Pemilihan model terbaik berdasarkan nilai R-square (R^2) dan mean square error (MSE).
2. Data yang digunakan dalam penelitian ini merupakan data produk domestik regional bruto di 34 provinsi pada Tahun 2020.

1.3 Tujuan Penelitian

1. Mendapatkan estimasi parameter pada model regresi kompoen utama MCD dan MVV pada kasus produk domestik regional bruto tahun 2020.
2. Membandingkan hasil model MCD dan MVV pada kasus produk domestik regional bruto tahun 2020.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini yaitu untuk menambah wawasan dan pengetahuan pembaca mengenai model regresi komponen utama pada metode *minimum covariance determinant* (MCD) dan *minimum vector variance* (MVV). Dan juga dapat menjadi acuan bagi pemerintah maupun masyarakat dalam upaya meningkatkan pertumbuhan ekonomi di Indonesia.

1.5 Teori

1.5.1. Analisis Regresi Linier Berganda

Analisis regresi adalah salah satu teknik statistik yang digunakan untuk tujuan peramalan atau pendugaan variabel terikat (Y). Dalam analisis regresi, suatu persamaan regresi ditentukan dan digunakan untuk menggambarkan pola atau bentuk fungsi hubungan yang terdapat antar variabel. Secara umum bentuk persamaan model regresi dapat ditulis pada persamaan (1) (Uyanik & Guler, 2013):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i$$

Persamaan (1) dalam bentuk matriks, dinyatakan sebagai berikut (Larasati, 2019):

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad (1)$$

dengan:

$\boldsymbol{Y}$ = Vektor pada variabel terikat yang berukuran $(n \times 1)$

$\boldsymbol{X}$ = Matriks pada variabel bebas yang berukuran $(n \times (p + 1))$

$\boldsymbol{\beta}$ = Vektor parameter yang berdimensi $(p + 1)$

$\boldsymbol{\varepsilon}$ = Vektor residual yang berukuran $(n \times 1)$

Regresi linier merupakan salah satu metode yang umum digunakan untuk memodelkan hubungan antara variabel terikat dan variabel bebas. Terdapat dua jenis regresi linier, yaitu regresi linier sederhana dan regresi linier berganda. Regresi linier sederhana adalah model yang menggambarkan hubungan antara satu variabel bebas dan satu variabel terikat, yang hubungannya dapat direpresentasikan dengan garis lurus. Sementara itu, regresi linier berganda merupakan model persamaan yang menggambarkan hubungan antara satu variabel terikat dengan dua atau lebih variabel bebas dengan tujuan untuk mengetahui seberapa besar pengaruh variabel bebas terhadap variabel terikat (Larasati, 2019).

1.5.2. Multikolinearitas

Multikolinearitas merupakan kondisi ketika terdapat dua atau lebih variabel bebas dalam model regresi saling berkorelasi (Daoud, 2017). Hal ini dapat menyebabkan hasil yang bias dan mengakibatkan beberapa variabel yang seharusnya signifikan secara statistik menjadi tidak signifikan. Selain itu, multikolinearitas meningkatkan varian koefisien regresi yang membuatnya tidak stabil dan menimbulkan masalah dalam menafsirkan koefisien (Shrestha, 2020).

Salah satu cara untuk mendeteksi adanya multikolinearitas pada model regresi yaitu dengan menggunakan nilai *variance inflation factor* (VIF). VIF digunakan sebagai kriteria dalam regresi linier yang melibatkan lebih dari dua variabel bebas. VIF untuk koefisien regresi-j diidentifikasi pada persamaan (2) (Daoud, 2017):

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

dengn R_j^2 merupakan koefisien determinasi antara variabel $ke-j$, $j = 1,2,\dots,p$. Apabila nilai VIF yang diperoleh > 10 maka, hal tersebut menunjukkan adanya indikasi multikolinearitas dalam model regresi.

1.5.3. Pencilan

Pencilan merupakan data yang menyimpang terlalu jauh dari data lainnya dalam suatu rangkaian data. Adanya satu atau lebih pencilan dalam data dapat menyebabkan tidak terpenuhinya satu atau lebih asumsi yang disyaratkan, sehingga mengakibatkan estimasi parameter tidak konsisten. Secara sederhana, pencilan dapat diartikan sebagai pengamatan yang memiliki galat yang besar (Kabes, 2021).

Wang dkk. (2019) menyatakan terdapat beberapa metode yang bisa digunakan untuk mendeteksi pencilan, salah satunya adalah jarak Mahalanobis dengan cara menghitung jarak setiap observasi ke pusat kumpulan data. Peran jarak statistik dalam menangani masalah pencilan sangatlah penting. Menggunakan ukuran jarak memungkinkan pengukuran kedekatan antara dua objek statistik. Oleh karena itu, jarak Mahalanobis memiliki keunggulan dalam mendeteksi pencilan pada data multivariat (Ghorbani, 2019).

Jarak Mahalanobis merupakan metode yang menggunakan vektor rata-rata dan matriks varian kovarian untuk mengukur seberapa jauh suatu observasi dari pusat data. Jarak mahalanobis diperoleh dengan menghitung jarak tiap observasi terhadap pusat data sesuai dengan Persamaan (3) berikut (Sari dkk, 2021):

$$d_i^2 = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})' \boldsymbol{S}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}) > \chi^2_{p(1-\alpha)} \tag{3}$$

dengan:

d_i^2 = Kuadrat jarak pengamatan ke-i

$\boldsymbol{x}_i$ = Nilai pengamatan ke-i

$\overline{\boldsymbol{x}}$ = Vektor rata-rata setiap variabel

$\boldsymbol{S}^{-1}$ = Invers matriks varian kovarian sampel

1.5.4. Analisis Multivariat

Analisis multivariat merupakan metode statistik yang mengolah beberapa pengukuran terkait individu atau objek secara bersamaan (*simultaneously*). Tujuan analisis multivariat adalah mengukur, menerangkan dan memprediksi tingkat hubungan antara variabel-variabel yang diamati (Simamora, 2005). Asumsi-asumsi dalam analisis multivariat meliputi variabel karakteristik yang harus saling berkorelasi dan berdistribusi normal multivariat. Adapun matriks data multivariat yang disusun dari sampel data, dapat ditulis dalam bentuk berikut (Johnson & Wichern, 2007):

$$\boldsymbol{X}_{nxp} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1\mathrm{p}} \\ x_{21} & x_{22} & \cdots & x_{2\mathrm{p}} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} X_1 \; X_2 \; X_3 \ldots X_p \end{bmatrix}$$

dengan:

n = jumlah observasi

p = jumlah variabel

a. Vektor Rata-rata

Vektor rata-rata populasi dapat ditulis dalam bentuk matriks sebagai berikut:

$$Mean = E(\boldsymbol{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu}$$

Vektor rata-rata sampel menggunakan persamaan (4) berikut:

$$\bar{\boldsymbol{x}} = (\bar{x}_{\mathrm{j}})$$

$$\bar{x}_{\mathrm{j}} = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \quad (4)$$

dengan:

$\bar{x}_{\mathrm{i}}$ = Rata-rata sampel untuk variabel ke-i

n = Jumlah observasi

x_{ij} = Nilai pengamatan untuk variabel ke-i pada observasi ke-j

b. Matriks Varian Kovarian dan Matriks Korelasi

Matriks merupakan susunan bilangan berbentuk persegi (Anton dan Rorres, 2005). Matriks varian kovarian merupakan matriks simetris yang menampilkan varian pada diagonal utamanya dan kovarian pada elemen-elemen lainnya. Koefisien varian menunjukkan indeks hubungan linier antara dua variabel (Larasati, 2019). Namun, matriks varian kovarian tidak *robust* terhadap pencilan, sehingga nilai pencilan dalam data dapat secara signifikan mempengaruhi hasil perhitungan. Matriks varian kovarian data populasi dapat dituliskan sebagai berikut (Johnson & Wichern, 2007):

$$Cov\ (\boldsymbol{X}) = \mathrm{E}[(\mathbf{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'] \quad (5)$$

$$Cov\ (\boldsymbol{X}) = E\left(\begin{bmatrix} X_1 & - & \mu_1 \\ X_2 & - & \mu_2 \\ & \vdots & \\ X_P & - & \mu_P \end{bmatrix}\left[X_1 - \mu_1, X_2 - \mu_2, \dots, X_p - \mu_P\right]\right)$$

$$= \begin{bmatrix} E(X_1-\mu_1)^2 & E(X_1-\mu_1)(X_2-\mu_2) & \cdots & E(X_1-\mu_1)(X_P-\mu_P) \\ E(X_2-\mu_2)(X_1-\mu_1) & E(X_2-\mu_2)^2 & \cdots & E(X_2-\mu_2)(X_P-\mu_P) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_p-\mu_p)(X_1-\mu_1) & E(X_p-\mu_p)(X_2-\mu_2) & \cdots & E(X_P-\mu_P)^2 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_{pxp} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

dengan matriks varian kovarian untuk data sampel sebagai berikut :

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_{ij} - \bar{X}_j\right)\left(X_{ij} - \bar{X}_j\right)' \qquad (6)$$

$$\boldsymbol{S} = \frac{1}{n-1}\left(\left[\begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix} - \begin{pmatrix} \bar{x}_1 \\ \bar{x}_1 \\ \vdots \\ \bar{x}_1 \end{pmatrix}\right]\left[\begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix} - \begin{pmatrix} \bar{x}_1 \\ \bar{x}_1 \\ \vdots \\ \bar{x}_1 \end{pmatrix}\right]' + \left[\begin{pmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{pmatrix} - \begin{pmatrix} \bar{x}_2 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_2 \end{pmatrix}\right]\left[\begin{pmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{pmatrix} - \begin{pmatrix} \bar{x}_2 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_2 \end{pmatrix}\right]' \right.$$

$$\left. + \cdots + \left[\begin{pmatrix} x_{1p} \\ x_{2p} \\ \vdots \\ x_{np} \end{pmatrix} - \begin{pmatrix} \bar{x}_\mathrm{p} \\ \bar{x}_\mathrm{p} \\ \vdots \\ \bar{x}_p \end{pmatrix}\right]\left[\begin{pmatrix} x_{1p} \\ x_{2p} \\ \vdots \\ x_{np} \end{pmatrix} - \begin{pmatrix} \bar{x}_\mathrm{p} \\ \bar{x}_\mathrm{p} \\ \vdots \\ \bar{x}_p \end{pmatrix}\right]'\right)$$

$$\boldsymbol{S} = \frac{1}{n-1}\left[\begin{pmatrix} x_{11} & - & \bar{x}_1 \\ x_{21} & - & \bar{x}_1 \\ & \vdots & \\ x_{n1} & - & \bar{x}_1 \end{pmatrix}(x_{11} - \bar{x}_1 \quad x_{21} - \bar{x}_1 \quad \dots \quad x_{n1} - \bar{x}_1)\right.$$

$$+ \begin{pmatrix} x_{12} & - & \bar{x}_2 \\ x_{22} & - & \bar{x}_2 \\ & \vdots & \\ x_{n2} & - & \bar{x}_2 \end{pmatrix}(x_{12} - \bar{x}_2 \quad x_{22} - \bar{x}_2 \quad \dots \quad x_{n2} - \bar{x}_2) + \cdots$$

$$\left. + \begin{pmatrix} x_{1p} & - & \bar{x}_\mathrm{p} \\ x_{2p} & - & \bar{x}_\mathrm{p} \\ & \vdots & \\ x_{np} & - & \bar{x}_p \end{pmatrix} x_{1p} - \bar{x}_p \quad x_{2p} - \bar{x}_p \quad \dots \quad x_{np} - \bar{x}_p)\right]$$

$$\boldsymbol{S} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

Matriks koefisien korelasi merupakan matriks simetris berukuran $p \times p$ dengan elemen diagonal utamanya berjumlah 1. Matriks korelasi digunakan jika variabel memiliki varian yang berbeda signifikan yang dapat dinyatakan dalam persamaan berikut (Khikmah (2021):

$$\boldsymbol{R} = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{21}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{11}}} & \cdots & \frac{\sigma_{p1}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{11}}} \\ \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{p2}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{22}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{bmatrix}$$

$$\boldsymbol{R} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix} \quad (7)$$

Hubungan matriks varian kovarian dengan matriks korelasi sebagai berikut :

$$\boldsymbol{R}_{pp} = D\left(\frac{1}{\sigma}\right)\boldsymbol{S}D\left(\frac{1}{\sigma}\right) \quad (8)$$

dengan:

$\boldsymbol{R}$ = Matriks korelasi (berukuran $p \times p$)

$D(1/\sigma)$ = Matriks diagonal

$\boldsymbol{S}$ = Matriks kovarians

c. Nilai Eigen dan Vektor Eigen

Nilai eigen dan vektor eigen adalah konsep penting dalam aljabar matriks, terutama dalam konteks matriks bujur sangkar. Misalkan $\boldsymbol{A}$ adalah sebuah matriks bujur sangkar berukuran $p \times p$. Jika suatu skalar λ dan vektor $\boldsymbol{v} \neq \boldsymbol{0}$ yang memenuhi persamaan (9), maka $\boldsymbol{v}$ disebut vektor eigen dari matriks $\boldsymbol{A}$ yang bersesuaian dengan nilai eigen λ. Misalkan $\boldsymbol{A}$ matriks bujur sangkar, dengan:

$$\boldsymbol{Av} = \lambda\boldsymbol{v}\text{ ; dimana } \boldsymbol{v} \neq \boldsymbol{0} \quad (9)$$

$$\boldsymbol{Av} - \lambda\boldsymbol{v} = \boldsymbol{Av} - \lambda\boldsymbol{I} = (\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}.$$

Karena $\boldsymbol{v} \neq \boldsymbol{0}$ maka $(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0$, sehingga :

$$\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = |\boldsymbol{A} - \lambda\boldsymbol{I}| = 0 \quad (10)$$

1.5.5. *Minimum Covariance Determinant*

Metode *Minimum Covariance Determinant* (MCD) pertama kali diusulkan oleh Rousseeuw pada tahun 1984. Ide dasarnya cukup sederhana, yaitu mencari sebagian kecil dari "*good observations*" yang tidak dianggap sebagai pencilan. Metode MCD dilakukan dengan cara memilih sebagian kecil observasi dari sampel utama, kemudian menghitung rata-rata kovarian dari subsampel tersebut. Prosedur ini diulang untuk semua kemungkinan subsampel dengan ukuran tertentu, dan pada akhirnya dipilih sub-sampel yang memiliki kovarian dengan determinan minimum. Dengan demikian, pengaruh dari observasi yang paling ekstrem termasuk pencilan dihilangkan. Hal ini membuat metode ini menjadi *robust* terhadap adanya pencilan (Leys dkk, 2018).

Metode MCD memiliki prinsip kerja menggunakan vektor rata-rata dan matriks varian kovarian dengan membentuk subsampel H yang berukuran h dari sampel berukuran n amatan yang matriks varian kovariannya memiliki determinan terkecil (Hubert & Debruyne, 2010). Nilai h diperoleh dari persamaan (10):

$$h = \left[\frac{n+p+1}{2}\right] \tag{11}$$

dengan:

h = sub-sampel

n = jumlah observasi

p = jumlah variabel

Penduga MCD merupakan pasangan $(\bar{\boldsymbol{X}}, \boldsymbol{S})$, dengan $\bar{\boldsymbol{X}}$ adalah vektor rata-rata dan $\boldsymbol{S}$ adalah matriks varian kovarian yang diestimasi dengan nilai determinan yang minimum pada subsampel data (Larasati, 2019). Oleh karena itu, secara berturut-turut estimasi MCD untuk vektor rata-rata dan matriks varian kovarian dapat ditunjukkan pada persamaaan (12) dan (13) berikut (Anggara & Supandi, 2021):

$$\bar{\boldsymbol{x}}_{MCD} = \frac{1}{h}\sum_{i \epsilon H} x_{ij} \tag{12}$$

$$\boldsymbol{S}_{MCD} = \frac{1}{h-1}\sum_{i \epsilon H}(\boldsymbol{x_i} - \bar{\boldsymbol{x}}_{MCD})\,(\boldsymbol{x_i} - \bar{\boldsymbol{x}}_{MCD})' \tag{13}$$

1.5.6. *Minimum Vector Variance*

Minimum Vector Variance (MVV) merupakan salah satu metode estimator *robust* yang pertama kali diperkenalkan oleh Herwindiati (2009) dengan memodifikasi algoritma FMCD menjadi lebih efektif dan tingkat kompleksitas yang lebih rendah berdasarkan ukuran vector variance. MVV merupakan salah satu metode estimator *robust* untuk mendeteksi pencilan dengan menggunakan kriteria vektor varian yang minimum (Juniardi dkk., 2014).

Misalkan $\bar{\boldsymbol{x}}_{MVV}$ dan $\boldsymbol{S}_{MVV}$ merupakan estimasi MVV untuk parameter lokasi dan matriks varian kovarian. Estimasi ini diperoleh berdasarkan himpunan H. Jumlah lokasi elemen dari H adalah $h = \frac{(n+p+1)}{2}$ data yang memberikan matriks varian kovarian $\boldsymbol{S}_{MVV}$ berdasarkan vektor eigen yang memiliki nilai eigen value terbesar untuk semua kemungkinan himpunan yang mengandung h data (Herwindiati, 2009). Oleh karena itu, secara berturut-turut estimasi MVV untuk parameter lokasi dan matriks varian kovarian dapat ditunjukkan pada Persamaaan (14) dan (15) berikut:

$$\bar{\boldsymbol{x}}_{MVV} = \frac{1}{h}\sum_{i \epsilon H} x_i \tag{14}$$

$$\boldsymbol{S}_{MVV} = \frac{1}{h-1}\sum_{i \epsilon H}(\boldsymbol{x_i} - \bar{\boldsymbol{x}}_{MVV})\,(\boldsymbol{x_i} - \bar{\boldsymbol{x}}_{MVV})' \tag{15}$$

1.5.7. Analisis Komponen Utama

Analisis komponen utama pertama kali ditemukan oleh Karl Pearson pada tahun 1901 yang digunakan pada bidang biologi. Analisis komponen utama merupakan teknik statistik yang digunakan untuk mengubah sebagian besar variabel asli yang saling berkorelasi menjadi satu set variabel baru yang lebih kecil dan saling bebas. Tujuan dari analisis komponen utama ini yaitu untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan. Analisis komponen utama biasanya digunakan untuk (Delsen dkk., 2017):

1. Menemukan variabel baru yang mendasari data variabel ganda.
2. Mengurangi dimensi dari himpunan variabel yang umumnya terdiri dari banyak variabel yang saling berkorelasi, namun tetap mempertahankan sebanyak mungkin keragaman dalam data tersebut.
3. Menghapus variabel asli yang memberikan kontribusi informasi yang relatif kecil pada variabel baru yang disebut komponen utama, yang memiliki karakteristik sebagai berikut:
   a. Merupakan kombinasi linier dari variabel asli.
   b. Jumlah kuadrat koefisien dalam kombinasi linier tersebut adalah satu.
   c. Tidak saling berkorelasi dan memiliki varian yang diurutkan dari yang besar ke yang terkecil

Komponen utama merupakan kombinasi linier dari variabel X_{j} dengan $j \in \{1,2,\dots,p\}$. Analisis komponen utama lebih efektif jika variabel-variabel yang digunakan memiliki korelasi yang kuat. Jumlah komponen utama paling banyak sama dengan jumlah variabel asal (p), (Delsen dkk., 2017).

Menurut Johnson dan Wichern (2007), komponen utama merupakan salah satu metode yang digunakan untuk mengubah sebagian besar variabel asal yang saling berkorelasi menjadi variabel baru yang tidak saling berkoreasi. Bentuk kombinasi linier dapat diperoleh dengan menggunakan persamaan (16) berikut:

$$\begin{aligned} KU_1 &= \boldsymbol{a_1'}X = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\ KU_2 &= \boldsymbol{a_2'}X = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\ &\vdots \\ KU_p &= \boldsymbol{a_p'}X = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p \end{aligned} \tag{16}$$

Pada masalah regresi, biasanya data yang digunakan memiliki skala yang berbeda antar variabel yang dapat mempengaruhi hasil analisis. Sehingga, digunakan bentuk variabel baku untuk mengatasi perbedaan skala tersebut (Widiharih, 2001). Standarisasi dari variabel X menjadi variabel Z dilakukan degan menggunakan transformasi:

$$Z_p = \frac{X_p - \bar{X}_p}{\sqrt{s}} \tag{17}$$

Menurut Johnson dan Wichern (2007), komponen utama yang dibentuk sebagai kombinasi linier dari variabel yang dibakukan yaitu $Z_1, Z_2, \dots, Z_p$ adalah:

$$\begin{aligned} KU_1 &= \boldsymbol{a_1'}Z = a_{11}Z_1 + a_{21}Z_2 + \cdots + a_{p1}Z_p \\ KU_2 &= \boldsymbol{a_2'}Z = a_{21}Z_1 + a_{22}Z_2 + \cdots + a_{p2}Z_p \\ &\vdots \\ KU_p &= \boldsymbol{a_p'}Z = a_{p1}Z_1 + a_{p2}Z_2 + \cdots + a_{pp}Z_p \end{aligned} \tag{18}$$

1.5.8. Pengujian Hipotesis

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:

a. Uji Simultan

Uji simultan atau uji F merupakan uji yang digunakan dalam memastikan adanya hubungan atau tidak antara variabel terikat Y dengan variab444444el bebas $X_1, X_2, \dots, X_p$ secara serentak. Hipotesis yang sesuai adalah (Montgomery dkk, 2012).

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

$H_1 : \beta_j \neq 0$, untuk setidaknya satu j $(j = 1,2, \dots, p)$,

Penolakan hipotesis nol menunjukkan bahwa setidaknya satu dari variabel $X_1, X_2, X_3, \dots, X_p$ memberikan kontribusi yang signifikan terhadap model. Statistik uji untuk uji F adalah:

$$F_{hitung} = \frac{JKR/p}{JKG/(n-p-1)} \tag{19}$$

dengan:

JKR = Jumlah kuadrat regresi

JKG = Jumlah kuadrat galat

n = Jumlah observasi

p = Jumlah variabel bebas

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

b. Uji Parsial Parameter (Uji *t*)

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

$H_0 : \beta_j = 0$,
$H_1 : \beta_j \neq 0$,
Statistik uji untuk uji t yaitu:

$$t_{hitung} = \frac{\hat{\beta}_j}{se_{\hat{\beta}_j}}, (j = 1,2, \ldots, p) \tag{20}$$

dengan:
$\hat{\beta}_j$ = Estimasi koefisien regresi ke-j,
$se_{\hat{\beta}_j}$ = Standar error dari estimasi koefisien regresi ke-j.
Kriteria uji untuk uji t adalah jika $\left|t_{hitung}\right| > t_{\left(\frac{\alpha}{2};n-p-1\right)}$ atau $p - value < \alpha$ maka tolak H_0, sebaliknya jika $\left|t_{hitung}\right| < t_{\left(\frac{\alpha}{2};n-p-1\right)}$ atau $p - value > \alpha$ maka terima H_0.

1.5.9. Ukuran Kebaikan Model

Metode yang digunakan untuk menentukan ukuran kebaikan model dilakukan dengan menghitung nilai *Mean Square Error* (MSE). Perhitungan MSE dilakukan menggunakan persamaan (24) berikut (Montgomery dkk., 2012):

$$MSE = \frac{JKR}{n-p-1} \tag{21}$$

dengan JKR merupakan jumlah kuadrat galat, n adalah jumlah sampel, dan p adalah banyaknya variabel bebas. Jika nilai MSE semakin kecil hingga mendekati nol, maka dapat dikatakan bahwa model regresi semakin baik.

Selain metode tersebut, untuk menentukan ukuran kebaikan model juga dapat menggunakan koefisien determinasi (R^2). Kecocokan model lebih baik jika R^2 semakin mendekati satu (Gujarati, 2003). Kebanyakan peneliti lebih suka menggunakan koefisien determinasi yang disesuaikan (R^2) yang dapat diperoleh dengan menggunakan persamaan (25) (Montgomery dkk., 2012):

$$R^2 = 1 - \frac{JKR/(n-p-1)}{JKG/(n-1)} \tag{22}$$

dengan JKG merupakan jumlah kuadrat galat, JKR merupakan jumlah kuadrat residual, n adalah jumlah sampel, dan p adalah banyaknya variabel bebas.

1.5.10. Produk Domestik Regional Bruto

Indikator ekonomi makro untuk mengukur pertumbuhan ekonomi suatu negara yaitu Produk Domestik Bruto (PDB). PDB pada tingkat provinsi dan kabupaten/kota disebut Produk Domestik Regional Bruto (PDRB) (Panjaitan dkk, 2019). PDRB merupakan nilai tambah pada barang maupun jasa yang diperoleh dari kegiatan ekonomi pada periode tertentu. BPS Nasional mendefinisikan Produk Domestik Regional Bruto sebagai jumlah nilai tambah yang dihasilkan oleh seluruh unit usaha dalam suatu wilayah, atau merupakan jumlah seluruh nilai barang dan jasa akhir yang dihasilkan oleh seluruh unit ekonomi di suatu wilayah (Sembiring & Sasongko, 2019). Salah satu aspek penting dalam peningkatan pertumbuhan ekonomi yaitu pembangunan infrastruktur.

Infrastruktur merupakan komponen biaya tetap yang secara langsung mendukung produksi. Berdasarkan Peraturan Nomor 38 Tahun 2015 tentang Kerja Sama Pemerintah dengan Badan Usaha Dalam Penyediaan Infrastruktur,

Infrastruktur mencakup fasilitas teknis, fisik, sistem, perangkat keras, dan lunak yang diperlukan untuk memberikan pelayanan kepada masyarakat dan mendukung jaringan struktur agar pertumbuhan ekonomi dan sosial masyarakat dapat berjalan dengan optimal (Afriyana dkk, 2023). Beberapa jenis infrastruktur yang dapat mendorong pertumbuhan ekonomi antara lain:

1. Infrastruktur Jalan
   Perkembangan ekonomi Indonesia sangat bergantung pada peran jalan raya sebagai sarana transportasi darat. Investasi dalam infrastruktur jalan berdampak pada sektor ekonomi dengan berfungsi sebagai dorongan bagi perkembangan ekonomi wilayah atau sebagai pemenuhan kebutuhan pertumbuhan ekonomi suatu daerah (Marsus dkk., 2020).
2. Infrastruktur Listrik
   Listrik merupakan kebutuhan penting untuk kehidupan sehari-hari dan mendukung berbagai aktivitas manusia. Ketersediaan dan kualitas pasokan listrik yang stabil sangat penting bagi operasional industri, bisnis, dan rumah tangga. Infrastruktur distribusi listrik yang memadai memastikan semua sektor ekonomi dapat berfungsi dengan optimal, yang mendukung pertumbuhan PDRB secara keseluruhan (Prasetya dkk, 2021).
3. Infrastruktur Pendidikan
   Infrasttruktur pendidikan merupakan penunjang utama penyelenggaraan proses pendidikan. Fasilitas yang memadai diperlukan untuk meningkatkan kualitas sumber daya manusia yang dibutuhkan di pasar kerja. Dengan demikian, investasi dalam infrastruktur pendidikan dapat mendorong inovasi dan produktivitas, yang pada akhirnya meningkatkan PDRB (Afriyana dkk., 2023).
4. Infrastruktur Pariwisata
   Infrastruktur pariwisata berkontribusi signifikan dalam pembangunan ekonomi nasional. Pengembangan infrastruktur pariwisata yang berkualitas tidak hanya meningkatkan daya tarik suatu wilayah tetapi juga menciptakan peluang kerja dan mendorong pertumbuhan ekonomi lokal (Komuna dkk., 2021).
5. Infrastruktur Perumahan
   Infrastruktur perumahan yang memadai mendukung pertumbuhan ekonomi dengan menyediakan tempat tinggal yang layak bagi penduduk. Ketersediaan perumahan yang baik berhubungan langsung dengan kualitas hidup dan stabilitas sosial, yang pada gilirannya mempengaruhi produktivitas dan konsumsi. Dengan demikian, investasi dalam infrastruktur perumahan mendukung pertumbuhan ekonomi dengan cara meningkatkan kesejahteraan masyarakat (Sirega dkk., 2023).
6. Infrastruktur Industri
   Fasilitas industri yang efisien sangat penting untuk mendorong pertumbuhan ekonomi. Infrastruktur industri, termasuk pabrik dan pusat distribusi, memberikan dukungan yang diperlukan untuk kegiatan produksi dan manufaktur. Ketersediaan fasilitas industri yang baik memungkinkan peningkatan kapasitas produksi dan inovasi, yang berdampak positif pada PDRB (Sirega dkk., 2023).

BAB II
METODE PENELITIAN

2.1. Jenis dan Sumber Data

Penelitian ini menggunakan data sekunder, yaitu data Produk Domestik Regional Bruto menurut Provinsi di Indonesia tahun 2020. Data yang digunakan bersumber dari Badan Pusat Statistika, dan Kementrian Pekerjaan Umum dan Perumahan Rakyat tahun 2020.

2.2. Variabel penelitian

Variabel dalam penelitian ini terdiri dari satu variabel terikat (y) dan enam variabel bebas (x). Adapun variabel yang digunakan dalam penelitian ini yaitu:

Tabel 1 Variabel terikat dan variabel bebas

Variabel	Keterangan	Satuan
Y	Produk Domestik Regional Bruto	Miliar Rupiah
X_1	Panjang Jalan	KM
X_2	Distribusi Listrik	GWH
X_3	Infrastruktur Pendidikan	Jumlah Infrastruktur Pendidikan
X_4	Infrastruktur Parawisata	Jumlah Infrastruktur Parawisata
X_5	Infrastruktur Perumahan	Jumlah Infrastruktur Perumahan
X_6	Fasilitas Industri	Jumlah Fasilitas Industri

2.3. Struktur Data

Data yang digunakan dalam penelitian ini terdiri dari 34 observasi yang mencakup 34 provinsi di Indonesia. Tabel data penelitian lengkap dapat ditemukan pada Lampiran 1. Struktur data yang digunakan ditampilkan dalam Tabel 2:

Tabel 2 Struktur Data Penelitian

Provinsi	y	x1	x2	x3	x4	x5	x6
Aceh	131585	1781.72	2937.99	101597	4647806	3397	103740
Sumatra Utara	533746	3048.5	11192.85	244831	13045164	38949	120474
Sumatra Barat	169458	1525.2	3429.29	96669	6930832	9008	94514
Riau	490024	2799.81	4967.05	92611	5203647	34692	51151
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Papua	137678	2361.76	1132.4	27745	550392	6710	12138

2.4. Metode Analisis

Adapun langkah-langkah analisis data yang dilakukan sebagai berikut:

1. Melakukan eksplorasi data dengan menentukan statistika deskriptif dari setiap variabel Provinsi di Indonesia untuk melihat gambaran awal dan sebaran data.
2. Mengidentifikasi adanya multikolinearitas dengan memeriksa nilai *variance inflation facktor* (VIF) yang diperoleh menggunakan Persamaan (2).
3. Mendeteksi observasi yang termasuk sebagai nilai pencilan secara multivariat dengan melihat nilai jarak Mahalanobis dengan langkah-langkah pendeteksian sebagai berikut:
   a. Menghitung nilai rata-rata dari setiap vektor variabel ($\bar{\boldsymbol{x}}$)
   b. Menghitung nilai matriks varian kovarian dari himpunan data
   c. Menghitung nilai jarak Mahalanobis setiap titik pengamatan dengan vektor rata-ratanya menggunakan Persamaan (3)
   d. Mengurutkan nilai d_i^2 dari nilai minimum hingga nilai terbesar $d_1^2 \leq d_2^2 \leq \cdots \leq d_n^2$
   e. Mengevaluasi jarak Mahalanobis menggunakan nilai chi-square cut-off (χ^2) yaitu, jika $d_i^2 > \chi^2_{p,(1-\alpha)}$, maka titik pengamatan ke-i diidentifikasi sebagai pencilan.
4. Membentuk subsampel H pada metode *minimum covariance determinant* dan *minimum vector variance* untuk memperoleh subsampel h dengan menggunakan Persamaan (11) untuk prinsip kerja masing-masing penduga MCD dan MVV.
   a. Penduga MCD
      1) Mengambil subsampel pertama dari data pengamatan secara acak, misalkan subsampel tersebut H_1 dengan jumlah elemen sebanyak h.
      2) Menghitung vektor rata-rata $\bar{\boldsymbol{X}}_{MCD}$ pada H_1 yang selanjutnya dapat disebut $\bar{\boldsymbol{X}}_1$ dengan:
      $$\bar{\boldsymbol{X}}_{MCD} = \frac{1}{h}\sum_{i \epsilon H} x_i$$
      3) Menghitung matriks varian kovarian $\boldsymbol{S}_{MCD}$ pada H_1 yang selanjutnya disebut $\boldsymbol{S}_1$ dengan:
      $$\boldsymbol{S}_{MCD} = \frac{1}{h-1}\sum_{i \epsilon H} (x_i - \bar{\boldsymbol{x}}_{MCD})\,(x_i - \bar{\boldsymbol{x}}_{MCD})'$$
      4) M enghitung determinan $\boldsymbol{S}_1$. Jika $\det(\boldsymbol{S}_1) = 0$, maka berhenti. Namun apabila $\det(\boldsymbol{S}_1) \neq 0$, maka dilanjutkan dengan menghitung jarak *robust* (RD) untuk setiap pengamatan dengan:
      $$RD_i = \sqrt{(\boldsymbol{x}_i - \bar{\boldsymbol{X}}_{MCD})'\boldsymbol{S}_{MCD}^{-1}(\boldsymbol{x}_i - \bar{\boldsymbol{X}}_{MCD})}$$
      5) Mengurutkan nilai RD mulai dari yang terkecil hingga terbesar.
      6) Mengambil elemen sejumlah h pengamatan dengan jarak terkecil berdasarkan pada langkah 5 untuk menjadi himpunan bagian H_2, kemudian ulangi lagkah 2 sampai langkah 5 sehingga ditemukan himpunan bagian yang konvergen $\det(\boldsymbol{S}_{i+1}) = \det(\boldsymbol{S}_i)$.

7) Memilih subsampel H yang memiliki nilai determinan matriks varian kovarian terkecil dari subsampel yang ditemukan.
8) Mengambil subsampel yang terpilih (H), kemudian menghitung vektor rata-rata $\overline{\boldsymbol{X}}_{MCD}$ dan matriks varian kovarian $\boldsymbol{S}_{MCD}$.
9) Menghitung jarak *robust* kuadrat (RD_i^2) untuk setiap pengamatan dengan:
$$RD_i^2 = (\boldsymbol{x}_i - \overline{\boldsymbol{X}}_{MCD})'\boldsymbol{S}_{MCD}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{X}}_{MCD})$$
10) Menentukan pencilan data dengan kriteria $RD_i^2 > \chi^2_{p;\alpha}$, di mana $\chi^2_{p;\alpha}$ adalah nilai kritis distribusi chi-square dengan tingkat signifikansi α dan derajat kebebasan p.

b. Penduga MVV
1) Ambil himpunan data yang terdiri dari $h = \frac{(n+p+1)}{2}$ data, sebutlah himpunan data ini dengan H_{old}.
2) Hitung vektor mean $\overline{\boldsymbol{x}}_{H_{old}}$ dan matriks kovariansi $\boldsymbol{S}_{H_{old}}$ untuk semua data H_{old}. Selanjutnya untuk $i = 1,2,\dots,n$ hitung:
$$d^2_{H_{old}} = d^2_{H_{old}}\left(\boldsymbol{x_i}, \overline{\boldsymbol{x}}_{H\,old}\right) = \left(\boldsymbol{x_i} - \overline{\boldsymbol{x}}_{H\,old}\right)'\boldsymbol{S}^{-1}_{H_{old}}\left(\boldsymbol{x_i} - \overline{\boldsymbol{x}}_{H\,old}\right)$$
3) Urutkan hasil perhitungan dari yang terkecil hingga terbesar. Urutan ini menghasilkan indeks observasi π**.** Misalnya hasil pengurutan data tersebut adalah:
$$d^2_{H_{old}}(\pi_1) \leq d^2_{H_{old}}(\pi_2) \leq \cdots \leq d^2_{H_{old}}(\pi_n)$$
4) Bentuk suatu himpunan baru yang terdiri dari h observasi dengan indeks $\pi(1), \pi(2), \dots, \pi(h)$ dan berilah nama H_{new}
5) Hitung $\overline{\boldsymbol{x}}_{H_{new}}$,$\boldsymbol{S}_{H_{new}}$ dan $d^2_{H_{new}}\left(\boldsymbol{x}_i, \overline{\boldsymbol{x}}_{H_{new}}\right)$ seperti pada tahap 2
6) Jika $Tr\left(\boldsymbol{S}^2_{H_{new}}\right) = Tr\left(\boldsymbol{S}^2_{H_{old}}\right)$ maka proses dikerjakan. Jika $Tr\left(\boldsymbol{S}^2_{H_{new}}\right) < Tr\left(\boldsymbol{S}^2_{H_{old}}\right)$ maka proses dilanjutkan sampai iterasi ke-k mencapai $Tr\left(\boldsymbol{S}^2_{H_{new}}\right) = Tr\left(\boldsymbol{S}^2_{H_{old}}\right)$
7) Jika $\boldsymbol{S}_{Hi}$ adalah matriks kovariansi dari iterasi ke-k. Pada akhir iterasi ke-k diperoleh $(Tr(\boldsymbol{S}^2_{H1}) \geq Tr(\boldsymbol{S}^2_{H2}) \geq \cdots \geq Tr\left(\boldsymbol{S}^2_{H_{k-1}}\right) = Tr\left(\boldsymbol{S}^2_{H_k}\right)$
8) Jika nilai minimum trace sudah didapat pada matriks varian kovarian, langkah selanjutnya adalah mengurutkan data multivariat menggunakan jarak Mahalanobis.
9) Jarak Mahalanobis dari nilai trace minimum dievaluasi menggunakan distribusi χ^2 dengan derajat kebebasan sebanyak jumlah variabel penelitian. Data pada observasi ke-i dianggap sebagai outlier jika $Rd_{MVV} \geq \chi^2_{(p:1-\alpha)}$

5. Membentuk matriks varian kovarian berdasarkan sub sampel yang diperoleh dari Langkah (4) yang *robust* terhadap pencilan dengan menggunakan Persamaan (13) dan Persamaan (15).
6. Mentransformasi matriks varian kovarian yang terbentuk dari Langkah (5) hingga terbentuk matriks korelasi menggunakan Persamaan (8).

7. Menghitung nilai eigen dan vektor eigen berdasarkan matriks korelasi yang terbentuk dari langkah (6) dengan Persamaan (9) dan (10).
8. Membentuk komponen utama menggunakan metode analisis komponen utama berdasarkan matriks korelasi dengan Persamaan (18).
9. Menentukan dan mengevaluasi ukuran kebaikan model yang diperoleh berdasarkan nilai *Mean Square Error* (MSE) dan *R-squared* dengan Persamaan (21) dan Persamaan (22).
10. Menginterpretasikan model dan membuat kesimpulan.