BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan salah satu metode statistika yang bertujuan untuk memodelkan pola hubungan antara variabel respon dan variabel prediktor. Pemilihan pendekatan regresi bergantung pada pola hubungan antarvariabel, yang dapat diamati melalui *scatter plot*. Apabila pola hubungan mengikuti bentuk kurva tertentu seperti linier, kuadratik, dan kubik, maka pendekatan regresi parametrik dapat digunakan (Tanaty & Sifriyani, 2023). Namun, dalam banyak kasus, hubungan antarvariabel bersifat kompleks dan tidak mengikuti pola parametrik tertentu, sehingga regresi nonparametrik menjadi alternatif yang lebih fleksibel karena bentuk kurva regresi ditentukan langsung oleh karakteristik data tanpa dipengaruhi subjektivitas peneliti (Dani & Adrianingsih, 2021). Pendekatan regresi nonparametrik telah banyak dikembangkan seperti *spline*, *kernel*, polinomial lokal, dan *fourier* (Syam dkk., 2020). Di antara pendekatan regresi nonparametrik yang banyak dikembangkan hingga saat ini adalah *spline*.

Spline merupakan generalisasi dari fungsi polinomial, yang mana proses optimasinya masih mengadopsi konsep dalam pendekatan regresi parametrik (Arifin dkk., 2020). Salah satu kelebihan *spline* adalah kemampuannya dalam menangkap perubahan pola data pada sub-interval tertentu melalui penggunaan titik-titik knot (Islamiyati, 2019). Salah satu basis fungsi yang biasa digunakan dalam pendekatan *spline* adalah *spline truncated. Spline truncated* terdiri dari dua komponen utama, yaitu fungsi polinomial yang memodelkan hubungan pada interval kontinu dan komponen *truncated* yang memungkinkan penyesuaian fungsi pada interval spesifik melalui keberadaan titik knot. Titik knot merupakan titik perpaduan dua kurva yang menunjukkan pola perubahan perilaku kurva pada interval yang berbeda (Islamiyati, 2019). Namun, pada pembentukan model regresi sering terjadi permasalahan korelasi yang kuat antara variabel prediktor satu dengan variabel prediktor lainnya yang dikenal dengan istilah multikolinearitas (Larasati dkk., 2020).

Permasalahan multikolinearitas tidak terbatas pada model regresi parametrik, tetapi juga dapat terjadi pada model regresi nonparametrik (Arifin dkk., 2020). Efek dari multikolinearitas adalah dapat mengakibatkan regresi memiliki bias dan variansi yang besar (Sungkono & Nugrahaningsih, 2017). Oleh sebab itu, penting untuk mengatasi masalah multikolinearitas agar analisis regresi dapat memberikan hasil yang valid. Beberapa metode statistika yang dapat digunakan untuk mengatasi masalah ini, antara lain mengumpulkan data tambahan, spesifikasi model dengan menghapus suatu variabel yang berkorelasi, metode regresi *ridge* dan metode analisis komponen utama (AKU) (Montgomery dkk., 2012).

AKU merupakan suatu teknik statistik multivariat yang bertujuan untuk mentransformasi sekumpulan variabel asal yang saling berkorelasi menjadi sejumlah komponen utama yang saling bebas dengan tetap mempertahankan sebagian besar informasi data (Sanusi & Saputro, 2020). Pembentukan komponen utama dilakukan dengan memanfaatkan vektor dan nilai eigen dari matriks variansi-kovariansi atau

matriks korelasi dari variabel asal (Amni dkk., 2024). Namun, secara umum sifat matriks variansi-kovariansi dalam AKU tidak bersifat *robust* terhadap keberadaan pencilan (Sanusi & Saputro, 2020). Oleh karena itu, dikembangkan Analisis Komponen Utama (AKU) *robust* berbasis *Minimum Covariance Determinant* (MCD). Metode MCD dilakukan dengan mengidentifikasi sebagian kecil observasi dari sampel utama yang memiliki nilai determinan matriks variansi-kovariansi terkecil. Proses ini bertujuan untuk menghasilkan matriks variansi-kovariansi yang lebih resisten terhadap pencilan (Leys dkk., 2018).

Penanganan pencilan juga perlu diperhatikan pada tahap estimasi parameter regresi. Estimasi menggunakan *Least Square* diketahui sangat sensitif terhadap pencilan karena setiap observasi diperlakukan dengan kontribusi yang sama dalam proses estimasi parameter, tanpa mekanisme pembobotan. Kondisi ini dapat menyebabkan estimasi parameter menjadi bias dan tidak stabil ketika data mengandung pencilan (Fitrianto & Xin, 2022). Oleh karena itu, diperlukan pendekatan regresi *robust*, yaitu metode regresi yang digunakan ketika distribusi dari nilai residual tidak normal dan adanya pencilan yang mempengaruhi model (Setiawan dkk., 2019). Penggunaan regresi *robust* akan menghasilkan model persamaan regresi yang bersifat resisten terhadap pencilan. Ada beberapa metode estimasi dalam regresi *robust*, yaitu estimator *Scale* (S), estimator *Maximum Likelihood* (M), estimator *Method of Moment* (MM), estimator *Least Median of Square* (LMS), dan estimator *Least Trimmed Square* (LTS) (Raza dkk., 2024). LTS merupakan metode estimasi parameter *robust* yang memiliki nilai *breakdown point* tinggi karena proses estimasinya dilakukan dengan meminimalkan jumlah kuadrat residual dari h observasi (Faizia dkk., 2019). Dengan hanya mempertimbangkan sebagian observasi tersebut, metode ini secara efektif membatasi pengaruh observasi ekstrem atau pencilan terhadap hasil estimasi parameter.

Penelitian mengenai penanganan multikolinearitas menggunakan metode AKU telah dilakukan oleh Hoffmann dkk. (2009) yang menggunakan AKU pada regresi nonparametrik berbasis *kernel*, sedangkan Arifin dkk. (2020) menggunakan AKU berbasis *spline* pada data diabetes melitus dan memperoleh satu komponen utama dengan model *spline* linier terbaik pada satu titik knot. Adapun penelitian terkait penerapan AKU *robust* dilakukan oleh Astuti & Yundari (2021) melalui perbandingan AKU klasik dan AKU *robust* berbasis MCD, yang terbukti lebih efektif dalam mereduksi multikolinearitas dan pencilan. Kemudian penelitian mengenai estimasi parameter *robust* dengan LTS telah dilakukan oleh Wulandari (2020) yang menunjukkan bahwa estimasi parameter *robust* dengan LTS memberikan kinerja terbaik dibandingkan estimator *robust* M dan S berdasarkan nilai koefisien determinasi dan *Mean Squared Error* (MSE). Penelitian selanjutnya oleh Larasati dkk. (2020) menunjukkan bahwa Regresi Komponen Utama (RKU) *robust* dengan kombinasi metode MCD-LTS lebih efektif dalam menangani multikolinearitas dan keberadaan pencilan dibandingkan RKU klasik.

Penelitian-penelitian sebelumnya telah mengkaji AKU pada regresi nonparametrik, AKU *robust* berbasis MCD, serta estimasi parameter *robust* dengan metode LTS. Namun, penelitian-penelitian tersebut umumnya dilakukan secara

terpisah dan belum mengintegrasikan penanganan pola hubungan nonparametrik, multikolinearitas, dan keberadaan pencilan dalam satu kerangka analisis. Padahal, dalam banyak kasus data sosial ekonomi, ketiga permasalahan tersebut sering muncul secara bersamaan dan saling berkaitan. Oleh karena itu, diperlukan suatu pendekatan yang mampu mengakomodasi fleksibilitas bentuk hubungan sekaligus menghasilkan estimasi yang stabil dan resisten terhadap karakteristik data yang kompleks.

Salah satu fenomena sosial ekonomi dengan karakteristik tersebut adalah kemiskinan. Kemiskinan umumnya dipengaruhi oleh banyak faktor yang saling berhubungan, seperti kondisi lingkungan, pendidikan, pendapatan, kesehatan, dan lain-lain (Sari dkk., 2025). Faktor-faktor tersebut tidak berdiri sendiri, melainkan saling berinteraksi dan membentuk struktur hubungan yang kompleks. Kompleksitas hubungan ini menghadirkan tantangan dalam analisis, karena keterkaitan yang kuat antarvariabel dapat menyebabkan kesulitan dalam mengidentifikasi pengaruh masing-masing variabel terhadap kemiskinan (Naufal dkk., 2025). Hubungan antarvariabel sosial ekonomi dan kemiskinan juga tidak selalu mengikuti pola parametrik tertentu, tetapi dapat mengalami perubahan pola pada sub-interval tertentu (Nasir dkk., 2024). Selain itu, data sosial ekonomi juga seringkali mengandung observasi ekstrem atau pencilan akibat ketimpangan antarwilayah (Lestari dkk., 2024).

Kondisi tersebut mengakibatkan upaya penanggulangan kemiskinan belum sepenuhnya optimal, termasuk di Provinsi Sulawesi Selatan. Meskipun Provinsi Sulawesi Selatan termasuk salah satu provinsi dengan pertumbuhan ekonomi yang cukup baik, kemiskinan tetap menjadi tantangan utama di wilayah ini (Nasir dkk., 2024). Berdasarkan data Badan Pusat Statistik (BPS), persentase penduduk miskin di Sulawesi Selatan pada Maret 2025 sebesar 7,60 persen atau sekitar 698,13 ribu jiwa. Kondisi ini mengindikasikan bahwa pertumbuhan ekonomi belum sepenuhnya berdampak pada penurunan kemiskinan secara merata di seluruh kabupaten/kota, sehingga diperlukan pemodelan yang mampu menjelaskan hubungan antarvariabel secara lebih komprehensif.

Berdasarkan uraian tersebut, penelitian ini mengkaji penerapan regresi nonparametrik *spline truncated* berbasis komponen utama *robust* MCD-LTS untuk memodelkan faktor-faktor yang memengaruhi Persentase Penduduk Miskin di Sulawesi Selatan. Pendekatan ini dirancang untuk mengakomodasi pola hubungan nonparametrik sekaligus mempertimbangkan adanya multikolinearitas dan pencilan dalam data. Dengan demikian, model yang dihasilkan diharapkan lebih stabil dan akurat sebagai dasar perumusan kebijakan penanggulangan kemiskinan yang lebih efektif dan terarah.

1.2 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian adalah data Persentase Penduduk Miskin di Sulawesi Selatan tahun 2023-2024.
2. Pemodelan *spline truncated* dibatasi pada orde satu (linier).

3. Identifikasi observasi yang termasuk pencilan dilakukan dengan menggunakan nilai jarak Mahalanobis dan *Studentized Deleted Residual* (TRES).
4. Jumlah titik knot yang diuji terdiri atas 1 hingga 3 titik knot, serta pemilihan model terbaik ditentukan berdasarkan nilai *Generalized Cross Validation* (GCV) minimum.

1.3 Tujuan dan Manfaat Penelitian

Tujuan penelitian ini adalah sebagai berikut:

1. Mengestimasi parameter model Regresi Nonparametrik *Spline Truncated* Komponen Utama *Robust* dengan *Minimum Covariance Determinant-Least Trimmed Square*.
2. Memperoleh model Regresi Nonparametrik *Spline Truncated* Komponen Utama *Robust* dengan *Minimum Covariance Determinant-Least Trimmed Square* pada data Persentase Penduduk Miskin di Sulawesi Selatan tahun 2023-2024.

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Menambah wawasan keilmuan dan pengetahuan mengenai model Persentase Penduduk Miskin di Sulawesi Selatan tahun 2023-2024 yang dihasilkan melalui metode Regresi Nonparametrik *Spline Truncated* Komponen Utama *Robust* dengan *Minimum Covariance Determinant-Least Trimmed Square*.
2. Memberikan informasi yang komprehensif mengenai Persentase Penduduk Miskin beserta faktor-faktor yang memengaruhinya, sehingga dapat dijadikan sebagai referensi ilmiah maupun dasar pertimbangan bagi peneliti, praktisi, serta pembuat kebijakan dalam merumuskan strategi pengendalian dan penurunan kemiskinan di Sulawesi Selatan.

1.4 Teori

1.4.1 Multikolinearitas

Multikolinearitas dalam analisis regresi adalah suatu kondisi terjadinya korelasi yang kuat antar variabel prediktor. Kondisi ini menyebabkan model regresi tidak mampu membedakan pengaruh masing-masing variabel secara efektif, sehingga menghasilkan estimasi koefisien yang bias, serta meningkatkan variansi dari koefisien tersebut, yang pada akhirnya membuat estimasi menjadi tidak stabil dan menimbulkan masalah dalam penafsiran koefisien (Shrestha, 2020). Untuk mendeteksi terjadinya multikolinearitas, salah satu caranya dengan melihat nilai koefisien korelasi antar variabel prediktor berdasarkan matriks korelasi (Jeng, 2023). Matriks korelasi dapat diperoleh melalui proses standarisasi matriks variansi-kovariansi (Manoppo dkk., 2025). Untuk itu, analisis multikolinearitas perlu diawali dengan memahami struktur hubungan antar variabel melalui matriks variansi-kovariansi dan matriks korelasi sebagai representasi matematis dari keterkaitan linier antar variabel.

1. Matriks Variansi-Kovariansi dan Matriks Korelasi

Matriks merupakan kumpulan bilangan-bilangan yang disusun secara khusus dalam bentuk baris dan kolom sehingga membentuk persegi panjang atau persegi. Bilangan-bilangan dalam susunan tersebut dinamakan elemen matriks yang dibatasi

oleh tanda kurung siku (Andani dkk., 2020). Matriks variansi-kovariansi merupakan matriks simetris yang menampilkan variansi pada diagonal utamanya dan kovariansi pada elemen-elemen lainnya (Sari, 2021). Secara matematis, pembentukan matriks variansi-kovariansi berawal dari deviasi vektor acak terhadap nilai harapannya (vektor rata-rata). Vektor rata-rata data populasi dapat didefinisikan sebagai berikut:

$$Mean = E(\boldsymbol{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu} \tag{1}$$

Dalam praktiknya, parameter dari suatu populasi seperti rata-rata populasi ($\boldsymbol{\mu}$) umumnya tidak diketahui, sehingga perlu diestimasi berdasarkan statistik sampel (Damanik & Simamora, 2019). Oleh karena itu, vektor rata-rata sampel didefinisikan pada Persamaan (2).

$$\bar{\boldsymbol{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{1i} \\ \frac{1}{n}\sum_{i=1}^{n} x_{2i} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} x_{pi} \end{bmatrix} \tag{2}$$

Matriks variansi-kovariansi data populasi dengan ordo $p \times p$ dapat dinyatakan dalam Persamaan (3).

$$Cov(\boldsymbol{X}) = E(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T$$

$$= E\left(\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{bmatrix} [X_1 - \mu_1 \quad X_2 - \mu_2 \quad \dots \quad X_p - \mu_p] \right)$$

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

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

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \tag{3}$$

dengan σ_{ju} menyatakan kovariansi antara X_j dan X_u, untuk $j = 1,2, \dots p$ dan $u = 1,2, \dots, p$. Adapun matriks variansi-kovariansi data sampel sebagai berikut:

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x_i} - \overline{\boldsymbol{x}})(\boldsymbol{x_i} - \overline{\boldsymbol{x}})^T$$

$$= \frac{1}{n-1}[(\boldsymbol{x_1} - \overline{\boldsymbol{x}})(\boldsymbol{x_1} - \overline{\boldsymbol{x}})^T + (\boldsymbol{x_2} - \overline{\boldsymbol{x}})(\boldsymbol{x_2} - \overline{\boldsymbol{x}})^T + \cdots + (\boldsymbol{x_n} - \overline{\boldsymbol{x}})(\boldsymbol{x_n} - \overline{\boldsymbol{x}})^T]$$

$$= \frac{1}{n-1}\left[\begin{pmatrix} x_{11}-\bar{x}_1 \\ x_{21}-\bar{x}_2 \\ \vdots \\ x_{p1}-\bar{x}_p \end{pmatrix}\begin{pmatrix} x_{11}-\bar{x}_1 \\ x_{21}-\bar{x}_2 \\ \vdots \\ x_{p1}-\bar{x}_p \end{pmatrix}^T + \begin{pmatrix} x_{12}-\bar{x}_1 \\ x_{22}-\bar{x}_2 \\ \vdots \\ x_{p2}-\bar{x}_p \end{pmatrix}\begin{pmatrix} x_{12}-\bar{x}_1 \\ x_{22}-\bar{x}_2 \\ \vdots \\ x_{p2}-\bar{x}_p \end{pmatrix}^T + \cdots \right.$$
$$\left. + \begin{pmatrix} x_{1n}-\bar{x}_1 \\ x_{2n}-\bar{x}_2 \\ \vdots \\ x_{pn}-\bar{x}_p \end{pmatrix}\begin{pmatrix} x_{1n}-\bar{x}_1 \\ x_{2n}-\bar{x}_2 \\ \vdots \\ x_{pn}-\bar{x}_p \end{pmatrix}^T \right]$$

$$= \frac{1}{n-1}\left[\begin{pmatrix} x_{11}-\bar{x}_1 \\ x_{21}-\bar{x}_2 \\ \vdots \\ x_{p1}-\bar{x}_p \end{pmatrix}(x_{11}-\bar{x}_1 \quad x_{21}-\bar{x}_2 \quad \dots \quad x_{p1}-\bar{x}_p)\right.$$
$$+ \begin{pmatrix} x_{12}-\bar{x}_1 \\ x_{22}-\bar{x}_2 \\ \vdots \\ x_{p2}-\bar{x}_p \end{pmatrix}(x_{12}-\bar{x}_1 \quad x_{22}-\bar{x}_2 \quad \dots \quad x_{p2}-\bar{x}_p) + \cdots$$
$$\left. + \begin{pmatrix} x_{1n}-\bar{x}_1 \\ x_{2n}-\bar{x}_2 \\ \vdots \\ x_{pn}-\bar{x}_p \end{pmatrix}(x_{1n}-\bar{x}_1 \quad x_{2n}-\bar{x}_2 \quad \dots \quad x_{pn}-\bar{x}_p)\right]$$

$$= \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \quad (4)$$

Matriks korelasi adalah matriks simetris berukuran $p \times p$ yang memuat koefisien korelasi antar pasangan variabel, yang diperoleh melalui proses standarisasi dari matriks variansi-kovariansi. Matriks ini digunakan jika variabel-variabel memiliki variansi yang berbeda secara signifikan yang dapat dinyatakan dalam persamaan berikut (Khikmah, 2021):

$$\boldsymbol{R} = \begin{bmatrix} \frac{s_{11}}{\sqrt{s_{11}}\sqrt{s_{11}}} & \frac{s_{12}}{\sqrt{s_{11}}\sqrt{s_{22}}} & \cdots & \frac{s_{1p}}{\sqrt{s_{11}}\sqrt{s_{pp}}} \\ \frac{s_{21}}{\sqrt{s_{22}}\sqrt{s_{11}}} & \frac{s_{22}}{\sqrt{s_{22}}\sqrt{s_{22}}} & \cdots & \frac{s_{2p}}{\sqrt{s_{22}}\sqrt{s_{pp}}} \\ \vdots & \vdots & & \vdots \\ \frac{s_{p1}}{\sqrt{s_{pp}}\sqrt{s_{11}}} & \frac{s_{p2}}{\sqrt{s_{pp}}\sqrt{s_{22}}} & \ddots & \frac{s_{pp}}{\sqrt{s_{pp}}\sqrt{s_{pp}}} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \ldots & 1 \end{bmatrix} \quad (5)$$

Adapun hubungan matriks variansi-kovariansi dengan matriks korelasi adalah sebagai berikut (Johnson & Wichern, 2007):

$$\boldsymbol{R} = \boldsymbol{D}^{-1/2}\boldsymbol{S}\boldsymbol{D}^{-1/2} \quad (6)$$

dengan $\boldsymbol{S}$ menyatakan matriks variansi-kovariansi dan $\boldsymbol{D}$ menyatakan matriks diagonal yang elemen diagonalnya merupakan variansi masing-masing variabel. Selanjutnya, matriks $\boldsymbol{D}^{-1/2}$ didefinisikan pada Persamaan (7).

$$\boldsymbol{D}^{-1/2} = \begin{bmatrix} \frac{1}{\sqrt{s_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{s_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{s_{pp}}} \end{bmatrix} \quad (7)$$

2. Nilai Koefisien Korelasi

Salah satu pendekatan untuk mengidentifikasi adanya multikolinearitas adalah dengan melihat nilai korelasi pada variabel prediktor berdasarkan matriks korelasi. Metode ini mengukur tingkat hubungan antarvariabel menggunakan koefisien korelasi Pearson yang memiliki rentang nilai antara −1 sampai 1. Tanda positif menunjukkan adanya hubungan linier searah antara dua variabel, sedangkan tanda negatif menunjukkan hubungan linier berlawanan arah. Semakin mendekati nilai 1 atau −1, maka hubungan antar variabel semakin kuat, sedangkan nilai yang mendekati 0 menunjukkan hubungan yang semakin lemah. Dalam praktiknya, korelasi antar variabel prediktor X_j dan X_u dengan nilai absolut koefisien korelasi lebih besar dari 0,8 $\left(\left|r_{ju}\right| > 0{,}8\right)$ dianggap menunjukkan adanya korelasi yang kuat (Gujarati & Porter, 2009). Korelasi yang kuat ini dapat menjadi indikasi adanya masalah multikolinearitas, yang dapat mengurangi akurasi model regresi.

1.4.2 Pencilan

Pencilan merupakan observasi yang memiliki karakteristik unik yang terlihat sangat jauh berbeda dari observasi lainnya dan muncul dalam bentuk nilai-nilai ekstrem baik pada satu variabel tunggal maupun pada kombinasi beberapa variabel (Shodiqin dkk., 2018). Keberadaan pencilan dalam model regresi dapat menyebabkan nilai residual menjadi besar, meningkatnya variansi data serta menghasilkan interval estimasi yang lebih lebar, sehingga menurunkan ketepatan terhadap hasil estimasi parameter (Husain & Jamaluddin, 2023). Meskipun pencilan dapat merepresentasikan fenomena yang jarang terjadi, keberadaannya tidak selalu dapat diabaikan. Pencilan dapat dieliminasi apabila disebabkan oleh kesalahan pengumpulan data, pengukuran, atau analisis yang tidak merefleksikan fenomena yang diteliti (Draper & Smith, 1998). Oleh karena itu, pendeteksian pencilan menjadi

tahap penting sebelum penerapan metode statistika pada data (Utami dkk., 2024). Pendeteksian pencilan dapat dilakukan melalui beberapa metode, diantaranya:

1. Jarak Mahalanobis

Keberadaan pencilan pada variabel prediktor dapat dideteksi menggunakan jarak Mahalanobis. Metode ini bekerja dengan menghitung jarak setiap observasi terhadap pusat dari kumpulan data. Dengan menggunakan ukuran jarak, dapat diukur kedekatan antara dua objek statistik secara lebih terukur. Oleh karena itu, jarak Mahalanobis memiliki keunggulan yang signifikan dalam mendeteksi pencilan, terutama pada data multivariat (Ghorbani, 2019). Jarak Mahalanobis memanfaatkan vektor rata-rata dan matriks variansi-kovariansi untuk menilai seberapa jauh observasi dari pusat data. Suatu observasi dideteksi sebagai pencilan jika nilai kuadrat jarak Mahalanobisnya:

$$d_i^2 = (\boldsymbol{x_i} - \overline{\boldsymbol{x}})^T \boldsymbol{S}^{-1} (\boldsymbol{x_i} - \overline{\boldsymbol{x}}) > \chi^2_{p;1-\alpha} \tag{8}$$

dengan:

d_i^2 : nilai kuadrat jarak Mahalanobis pada observasi ke-i
$\boldsymbol{x_i}$: vektor observasi ke-i berdimensi $p \times 1$
$\overline{\boldsymbol{x}}$: vektor rata-rata setiap variabel
$\boldsymbol{S}^{-1}$: invers matriks variansi-kovariansi
p : banyaknya variabel prediktor

2. *Studentized Deleted Residual*

Pendeteksian adanya pencilan pada variabel respon dapat diperiksa menggunakan *Studentized Deleted Residual* (TRES) (Astuti dkk., 2023). Statistik uji TRES dapat dihitung dengan rumus pada Persamaan (9).

$$t_i = \varepsilon_i \left[\frac{n-p}{JKR(1-h_{ii}) - \varepsilon_i^2} \right]^{\frac{1}{2}} \tag{9}$$

dengan:

t_i : nilai *studentized deleted residual* ke-i
ε_i : residual ke-i
n : banyaknya observasi
JKR : jumlah kuadrat residual
h_{ii} : diagonal utama dari matriks hat ($\boldsymbol{H}$)

Jumlah Kuadrat Residual (JKR) merupakan ukuran penyimpangan antara nilai pengamatan dengan nilai hasil estimasi model regresi. Nilai JKR diperoleh dari penjumlahan kuadrat residual seluruh observasi, yang secara matematis dirumuskan sebagai berikut:

$$JKR = \sum_{i=1}^{n} \varepsilon_i^2 \tag{10}$$

dengan $\varepsilon_i = y_i - \hat{y}_i$ menyatakan residual ke-i yaitu selisih antara nilai pengamatan $y_i$dan nilai estimasi $\hat{y}_i$. Nilai JKR digunakan untuk mengukur besar kesalahan model regresi secara keseluruhan, sehingga semakin kecil nilai JKR menunjukkan bahwa model memiliki kecocokan yang lebih baik terhadap data.

Adapun matriks hat ($\boldsymbol{H}$) merupakan matriks proyeksi yang digunakan untuk menghasilkan nilai estimasi (*fitted values*) dari variabel respon. Secara matematis, matriks hat didefinisikan sebagai:

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \tag{11}$$

Statistik uji TRES mengikuti distribusi t dengan derajat bebas $n - p - 1$. Oleh karena itu, pendeteksian keberadaan pencilan dilakukan dengan membandingkan nilai $|t_i|$ dengan nilai kritis distribusi t. Observasi ke-i dideteksi sebagai pencilan jika $|t_i| > t_{\frac{\alpha}{2};n-p-1}$.

1.4.3 Analisis Komponen Utama *Robust*

Analisis Komponen Utama (AKU) merupakan teknik statistik yang digunakan untuk menjelaskan struktur matriks variansi-kovariansi dari sekumpulan variabel asal melalui pembentukan sejumlah komponen utama yang saling bebas dan merupakan kombinasi linier dari variabel asal (Sriningsih dkk., 2018). Pembentukan komponen utama tersebut dilakukan dengan mengekstraksi keragaman data melalui dekomposisi matriks variansi-kovariansi, sehingga diperoleh representasi data yang lebih ringkas namun tetap mampu menjelaskan sebagian besar variansi yang terkandung dalam data. Adapun arah dari komponen utama ditentukan oleh vektor eigen dari matriks variansi-kovariansi, sedangkan besarnya variansi yang dijelaskan oleh setiap komponen utama ditunjukkan oleh nilai eigen (Afifa dkk., 2024).

1. Nilai Eigen dan Vektor Eigen

Nilai eigen merupakan nilai karakteristik yang dimiliki oleh sebuah matriks persegi. Misalkan matriks $\boldsymbol{R}$ berukuran $p \times p$, kemudian terdapat skalar λ (nilai eigen), dan vektor tak nol $\boldsymbol{v}$ (vektor eigen) yang memenuhi Persamaan (12).

$$\boldsymbol{R}\boldsymbol{v} = \lambda\boldsymbol{v} \tag{12}$$

Skalar λ disebut sebagai nilai eigen dari $\boldsymbol{R}$, sementara $\boldsymbol{v}$ disebut vektor eigen dari $\boldsymbol{R}$ yang bersesuaian dengan nilai eigen λ (Abdi & Williams, 2010). Untuk menentukan nilai eigen pada Persamaan (12), dapat dituliskan menjadi:

$$\begin{aligned} \boldsymbol{R}\boldsymbol{v} - \lambda\boldsymbol{v} &= 0 \\ \boldsymbol{R}\boldsymbol{v} - \lambda\boldsymbol{I}\boldsymbol{v} &= 0 \\ (\boldsymbol{R} - \lambda\boldsymbol{I})\boldsymbol{v} &= 0 \end{aligned} \tag{13}$$

Agar Persamaan (13) memiliki solusi tak nol bagi $\boldsymbol{v}$, maka determinan dari $(\boldsymbol{R} - \lambda\boldsymbol{I})$ harus sama dengan nol, sehingga menghasilkan persamaan karakteristik seperti pada Persamaan (14).

$$det(\boldsymbol{R} - \lambda\boldsymbol{I}) = 0 \tag{14}$$

Penyelesaian dari persamaan karakteristik ini akan menghasilkan akar-akar karakteristik yang merupakan nilai eigen $\lambda_1, \lambda_2, \dots, \lambda_p$ dengan $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$. Setelah semua nilai eigen diperoleh, vektor eigen untuk masing-masing komponen utama yang sesuai dapat diperoleh dengan menyelesaikan sistem persamaan linier $(\boldsymbol{R} - \lambda\boldsymbol{I})\boldsymbol{v} = 0$ (Johnson & Wichern, 2007).

2. Pembentukan Komponen Utama

Komponen utama dibentuk sebagai kombinasi linier dari variabel prediktor asal dengan menggunakan vektor eigen yang bersesuaian. Secara matematis, komponen utama ke-j dapat didefinisikan pada Persamaan (15).

$$\boldsymbol{Z_j} = \boldsymbol{v_j^T X} = v_{1j}X_1 + v_{2j}X_2 + \cdots + v_{pj}X_p, j = 1,2,\ldots p \tag{15}$$

$$\boldsymbol{v_j^T X} = \begin{bmatrix} v_{1j} & v_{2j} & \cdots & v_{pj} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

Komponen utama juga dapat diperoleh dari variabel yang distandarisasi, khususnya ketika variabel-variabel yang dianalisis memiliki perbedaan satuan dan skala pengukuran (Siburian dkk., 2019). Dalam analisis komponen utama, penggunaan matriks korelasi umumnya dilakukan pada kondisi tersebut karena perbedaan skala antar variabel dapat menyebabkan variabel dengan variansi yang lebih besar akan mendominasi pembentukan komponen utama sehingga memengaruhi hasil analisis. Untuk menghilangkan pengaruh perbedaan skala tersebut, maka digunakan bentuk variabel baku melalui proses standarisasi. Proses standarisasi variabel X menjadi variabel baku X^* dinyatakan dalam Persamaan (16).

$$x_{ji}^* = \frac{\left(x_{ji} - \bar{x}_j\right)}{s_j} \tag{16}$$

dengan:

x_{ji} : nilai variabel prediktor ke-j pada observasi ke-i, $(i = 1, 2, \ldots, n)$

x_{ji}^* : nilai baku variabel prediktor ke-j pada observasi ke-i

$\bar{x}_j$: rata-rata variabel prediktor ke-j

s_j : standar deviasi variabel prediktor ke-j

Komponen utama ke-j yang dibentuk sebagai kombinasi linier dari variabel baku dapat didefinisikan pada Persamaan (17).

$$\boldsymbol{Z_j} = \boldsymbol{v_j^T X^*} = v_{1j}X_1^* + v_{2j}X_2^* + \cdots + v_{pj}X_p^*, j = 1,2,\ldots p \tag{17}$$

dengan $\boldsymbol{Z_1}$ adalah komponen pertama yang memenuhi maksimum nilai $\boldsymbol{v_1^T R v_1} = \lambda_1$, $\boldsymbol{Z_2}$ adalah komponen kedua yang memenuhi sisa variansi selain komponen pertama dengan memaksimumkan nilai $\boldsymbol{v_2^T R v_2} = \lambda_2$, dan seterusnya hingga komponen ke-p ($\boldsymbol{Z_p}$) tercapai.

Proporsi variansi yang dapat dijelaskan oleh komponen utama ke-j dinyatakan sebagai perbandingan antara nilai eigen ke-j terhadap total nilai eigen, sebagaimana dirumuskan pada Persamaan (18).

$$\frac{\lambda_j}{\sum_{j=1}^{p} \lambda_j} \times 100\% = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \times 100\% \tag{18}$$

Selanjutnya, penentuan jumlah komponen utama yang optimal dapat dilakukan berdasarkan proporsi variansi kumulatif yang mampu menjelaskan sebagian besar total variansi data. Secara umum, jumlah komponen utama dipilih sedemikian

sehingga proporsi variansi kumulatif yang diperoleh berada antara 70% dan 100% dari total variansi data (Jolliffe & Cadima, 2016).

3. AKU *Robust* dengan *Minimum Covariance Determinant*

Penggunaan AKU klasik memiliki kelemahan signifikan jika data mengandung pencilan. Perkembangan AKU selanjutnya dipengaruhi adanya kebutuhan suatu model AKU yang *robust* terhadap data pencilan (Astuti & Yundari, 2021). AKU klasik sangat dipengaruhi oleh kehadiran pencilan karena didasarkan pada matriks variansi-kovariansi biasa yang sensitif terhadap keberadaan data pencilan. Hal ini membuat interpretasi komponen utama menjadi bias, terutama jika pencilan tidak diidentifikasi dan diatasi.

Metode untuk mengatasi masalah sensitivitas terhadap pencilan adalah dengan menggunakan AKU *robust* berbasis *Minimum Covariance Determinant* (MCD). Metode MCD pertama kali diperkenalkan oleh Rousseeuw pada tahun 1984. Metode ini bertujuan untuk memperoleh estimasi vektor rata-rata dan matriks variansi-kovariansi yang *robust* dengan memilih sub-sampel berukuran h dari total n observasi yang menghasilkan determinan matriks variansi-kovariansi terkecil (Hubert & Debruyne, 2010). Secara matematis, estimator MCD dinyatakan sebagai pasangan $(\bar{\boldsymbol{x}}_{MCD}, \boldsymbol{S}_{MCD})$, dengan:

$$\bar{\boldsymbol{x}}_{MCD} = \frac{1}{h}\sum_{i=1}^{h} \boldsymbol{x_i} \tag{19}$$

$$\boldsymbol{S}_{MCD} = \frac{1}{h-1}\sum_{i=1}^{h} (\boldsymbol{x_i} - \bar{\boldsymbol{x}}_{MCD})(\boldsymbol{x_i} - \bar{\boldsymbol{x}}_{MCD})^T \tag{20}$$

dengan $\bar{\boldsymbol{x}}_{MCD}$ merepresentasikan vektor rata-rata *robust* yang dihitung berdasarkan sub-sampel terpilih berukuran h. Sementara $\boldsymbol{S}_{MCD}$ merupakan matriks variansi-kovariansi *robust* yang diperoleh dari sub-sampel tersebut. Adapun h menyatakan ukuran sub-sampel yang umumnya ditentukan sebagai berikut:

$$h = \frac{n+p+1}{2} \tag{21}$$

dengan n menyatakan banyaknya observasi dan p menyatakan banyaknya variabel prediktor. Pemilihan sub-sampel dengan determinan matriks variansi-kovariansi terkecil bertujuan untuk memperoleh struktur matriks variansi-kovariansi yang paling kompak dan *robust* terhadap pencilan.

Secara teoritis, untuk mendapatkan sub-sampel optimal diperlukan evaluasi seluruh kombinasi berukuran h dari n observasi, yaitu sebanyak $\binom{n}{h}$, yang jumlahnya sangat besar ketika n meningkat. Oleh karena itu, digunakan algoritma FAST-MCD yang dikembangkan oleh Rousseeuw dan Van Driessen pada tahun 1999. Algoritma ini menggunakan prosedur iteratif yang dikenal sebagai *Concentration Step* (C-step), sehingga tidak perlu mengevaluasi seluruh kombinasi yang mungkin.

Prosedur FAST-MCD dimulai dengan membentuk subset awal berukuran h, kemudian menghitung vektor rata-rata dan matriks variansi-kovariansi dari subset tersebut. Selanjutnya dihitung jarak *robust* untuk seluruh observasi menggunakan:

$$RD_i = \sqrt{(\boldsymbol{x_i} - \bar{\boldsymbol{x}}_{MCD})^T \boldsymbol{S}_{MCD}^{-1} (\boldsymbol{x_i} - \bar{\boldsymbol{x}}_{MCD})} \tag{22}$$

Nilai jarak *robust* tersebut diurutkan, dan dipilih kembali h observasi dengan jarak terkecil untuk membentuk subset baru. Dari subset baru ini dihitung ulang rata-rata dan matriks variansi-kovariansi, kemudian determinan yang dihasilkan dibandingkan dengan determinan sebelumnya. Proses ini diulang hingga determinan matriks variansi-kovariansi tidak lagi mengalami penurunan atau telah konvergen, sehingga diperoleh estimator MCD yang stabil. Pada setiap iterasi berlaku bahwa determinan matriks variansi-kovariansi tidak meningkat, sehingga algoritma secara sistematis bergerak menuju solusi dengan determinan minimum.

Setelah diperoleh matriks variansi-kovariansi *robust* $\boldsymbol{S}_{MCD}$, dilakukan dekomposisi matriks variansi-kovariansi untuk membentuk komponen utama *robust*. Vektor eigen menentukan arah komponen utama, sedangkan nilai eigen menunjukkan besarnya variansi yang dijelaskan oleh masing-masing komponen. Dengan demikian, AKU *robust* berbasis MCD mampu menghasilkan komponen utama yang lebih representatif terhadap mayoritas data dan *robust* terhadap adanya pencilan (Leys dkk., 2018).

1.4.4 Regresi *Robust*

Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi dari residual tidak normal atau adanya beberapa pencilan yang berpengaruh pada model (Nugrahani dkk., 2021). Berbeda dengan regresi klasik yang sangat sensitif terhadap data pencilan, metode regresi *robust* mampu memberikan estimasi parameter yang bersifat *robust* atau resisten terhadap pencilan. Suatu estimasi dikatakan resisten apabila estimasi tersebut relatif tidak terpengaruh ketika terjadi perubahan besar pada bagian kecil data atau sebaliknya, ketika terjadi perubahan kecil pada bagian besar data (Setiarini & Listyani, 2017). Metode yang telah dikembangkan dalam regresi *robust* untuk mengatasi masalah pencilan meliputi beberapa pendekatan seperti estimasi S, estimasi M, estimasi MM, estimasi LMS, dan estimasi LTS (Raza dkk., 2024).

Estimasi LTS (*Least Trimmed Square*) adalah metode dengan nilai *breakdown point* tinggi yang dikenalkan oleh Rousseeuw pada tahun 1984. Menurut Rousseeuw & Leroy (1987), prinsip dari estimasi LTS ini sama dengan metode OLS dalam mengestimasi parameter dengan meminimumkan jumlah kuadrat residual. Estimasi LTS didefinisikan sebagai berikut:

$$\hat{\beta}_{LTS} = min \sum_{i=1}^{h} \varepsilon_i^2 \tag{23}$$

fungsi tujuan dari metode LTS ini dituliskan sebagai berikut:

$$min \sum_{i=1}^{h} \varepsilon_i^2 = min \sum_{i=1}^{h} (y_i - \hat{y}_i)^2 \tag{24}$$

dengan meminimumkan jumlah kuadrat residual sebanyak h, maka:

$$h = \frac{n}{2} + \frac{k+1}{2} = \frac{n+k+1}{2}, h \leq n \tag{25}$$

dengan:

ε_i^2 : kuadrat residual ke-i

k : banyaknya parameter

n : banyaknya observasi

Jumlah h menunjukkan sejumlah sub-sampel data yang memiliki kuadrat residual terkecil (Willems & Aelst, 2005). Seperti halnya pada estimasi regresi *robust* lainnya, dalam metode LTS, data diberi pembobot (w_{ii}) untuk memastikan bahwa data pencilan tidak memberikan pengaruh terhadap parameter model yang diestimasi (Ade, 2024). Pembobot (w_{ii}) ini dituliskan dalam bentuk matriks, yang dinotasikan dengan $\boldsymbol{W}$,

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & 0 & \dots & 0 \\ 0 & w_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{nn} \end{bmatrix}$$

Pembentukan estimasi parameter dilakukan hingga *Final Weighted Least Square* dengan fungsi pembobot seperti pada persamaan berikut (Maharani dkk., 2014):

$$w_{ii} = \begin{cases} 1, & \dfrac{|\varepsilon_i|}{S_{LTS}} \leq 2{,}5 \\ 0, & lainnya \end{cases} \tag{26}$$

dengan

$$S_{LTS} = d_{h,n}\sqrt{\frac{1}{h}\sum_{i=1}^{h}\varepsilon_i^2}$$

$d_{h,n}$ merupakan faktor konsistensi yang digunakan untuk membuat estimasi skala awal LTS (S_{LTS}) menjadi konsisten dan tak bias di bawah asumsi residual berdistribusi Normal Standar, yaitu:

$$d_{h,n} = \frac{1}{\sqrt{1 - \dfrac{2n}{h.\,c_{h,n}}\emptyset\left(\dfrac{1}{c_{h,n}}\right)}};\ \text{dengan}\ \ c_{h,n} = \frac{1}{\Phi^{-1}\left(\dfrac{h+n}{2n}\right)}$$

dengan $\emptyset$ merupakan fungsi kepadatan peluang dari distribusi normal standar, dan Φ merupakan fungsi distribusi kumulatif normal standar. Estimasi parameter regresi *robust* LTS adalah sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_{LTS} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y} \tag{27}$$

dengan:

$\boldsymbol{y}$: vektor variabel respon

$\boldsymbol{X}$: matriks dari variabel prediktor

$\widehat{\boldsymbol{\beta}}_{LTS}$: vektor estimasi parameter regresi *robust* LTS

1.4.5 Regresi Nonparametrik *Spline Truncated*

Regresi nonparametrik merupakan pendekatan regresi yang digunakan untuk memodelkan hubungan antara variabel prediktor dan respon tanpa perlu menetapkan bentuk kurva tertentu sebelumnya. Berbeda dengan regresi parametrik yang memerlukan asumsi spesifik mengenai bentuk kurva, regresi nonparametrik menawarkan fleksibilitas lebih tinggi karena bentuk kurva regresinya ditentukan oleh data yang ada, bukan berdasarkan asumsi bentuk tertentu (Sifriyani dkk., 2023). Fleksibilitas ini sangat bermanfaat, khususnya dalam kasus yang melibatkan pola

hubungan yang kompleks. Secara matematis, model regresi nonparametrik dapat dinyatakan dengan persamaan berikut:

$$y_i = f(x_i) + \varepsilon_i; \quad i = 1,2,\dots,n \tag{28}$$

dengan

y_i : variabel respon pada observasi ke-i

$f(x_i)$: fungsi regresi yang diestimasi

x_i : variabel prediktor pada observasi ke-i

ε_i : residual pada observasi ke-i

Fungsi regresi nonparametrik pada Persamaan (28) tidak diketahui bentuk kurvanya dan hanya diasumsikan termuat dalam suatu ruang fungsi tertentu, pemilihan ruang fungsi tersebut didasarkan oleh sifat kemulusan (*smooth*) yang dimiliki fungsi. Salah satu pendekatan regresi nonparametrik adalah *spline* (Arifin dkk., 2020). *Spline* merupakan segmen polinomial dengan fleksibilitas yang lebih baik dari polinomial global sehingga memungkinkan penyesuaian pola data pada interval tertentu melalui titik knot (Davala dkk., 2024). Berbeda dengan polinomial global yang menggunakan satu fungsi tunggal untuk seluruh domain data, *spline* membagi domain tersebut menjadi beberapa sub-interval yang dihubungkan secara halus pada titik-titik knot. Salah satu bentuk khusus dari *spline* yang sering digunakan dalam analisis regresi nonparametrik adalah *spline truncated*. Pada dasarnya, regresi nonparametrik *spline truncated* terdiri dari dua komponen utama, yaitu fungsi polinomial yang memodelkan hubungan pada interval kontinu, dan komponen *truncated* yang memungkinkan penyesuaian fungsi pada interval spesifik. Secara umum, fungsi *spline truncated* berorde q dengan titik-titik knot pada $k_1, k_2, \dots, k_r$ dapat dinyatakan dalam bentuk persamaan berikut:

$$f(x_i) = \sum_{l=0}^{q} \beta_l x_i^l + \sum_{g=1}^{r} \beta_{q+g} (x_i - k_g)_+^q \tag{29}$$

dengan β_l adalah parameter polinomial, β_{q+g} adalah parameter pada komponen *truncated*, k_g adalah titik knot ke-g, dengan $g = 1, 2, \dots, r$ dan $(x_i - k_g)_+^q$ adalah fungsi *truncated* yang dijabarkan menjadi Persamaan (30).

$$(x_i - k_g)_+^q = \begin{cases} (x_i - k_g)^q, & x_i \geq k_g \\ 0 \quad , & x_i < k_g \end{cases} \tag{30}$$

Komponen fungsi *truncated* $(x_i - k_g)_+^q$ bertujuan untuk mengubah bentuk fungsi regresi pada titik-titik tertentu, yang dikenal sebagai titik knot (k_g), yang mencerminkan perubahan signifikan dalam data (Nurcahayani dkk., 2021). Fungsi *truncated* ini memberikan fleksibilitas yang diperlukan untuk menangkap perubahan lokal dalam data, yang biasanya tidak dapat diakomodasi oleh model regresi polinomial biasa (Pratiwi, 2020).

Apabila fungsi *spline truncated* pada Persamaan (29) disubstitusikan ke dalam Persamaan (28), maka diperoleh model regresi nonparametrik *spline truncated* berorde q dengan satu variabel prediktor seperti pada Persamaan (31).

$$y_i = \sum_{l=0}^{q} \beta_l x_i^l + \sum_{g=1}^{r} \beta_{q+g} (x_i - k_g)_+^q + \varepsilon_i \quad (31)$$

Jika data memuat variabel prediktor sebanyak p dengan respon tunggal, maka dapat dinyatakan dalam bentuk sebagai berikut:

$$y_i = f(x_{1i}) + f(x_{2i}) + \cdots + f(x_{pi}) + \varepsilon_i$$

$$= \sum_{j=1}^{p} f(x_{ji}) + \varepsilon_i; \quad i = 1,2, \dots, n \quad (32)$$

dengan

$$f(x_{ji}) = \beta_{0j} + \sum_{l=1}^{q} \beta_{jl} x_{ji}^l + \sum_{g=1}^{r} \beta_{j(q+g)} (x_{ji} - k_{jg})_+^q \quad (33)$$

Berdasarkan $f(x_{ji})$ pada Persamaan (33), model regresi nonparametrik *spline truncated* multiprediktor dapat dituliskan pada Persamaan (34).

$$y_i = \beta_0^* + \sum_{j=1}^{p} \left(\sum_{l=1}^{q} \beta_{jl} x_{ji}^l + \sum_{g=1}^{r} \beta_{j(q+g)} (x_{ji} - k_{jg})_+^q \right) + \varepsilon_i \quad (34)$$

dengan $\beta_0^* = \sum_{j=1}^{p} \beta_{0j}$ dan fungsi *truncated* seperti pada persamaan berikut (Sholicha dkk., 2018):

$$(x_{ji} - k_{jg})_+^q = \begin{cases} (x_{ji} - k_{jg})^q, & x_{ji} \geq k_{jg} \\ 0, & x_{ji} < k_{jg} \end{cases} \quad (35)$$

dengan

β_0^* : intersep global model regresi

β_{jl} : parameter polinomial pada prediktor ke-j dan orde ke-l

$\beta_{j(q+g)}$: parameter *truncated* pada prediktor ke-j dan titik knot ke-$(q+g)$

k_{jg} : nilai titik knot pada prediktor ke-j dan titik knot ke-g

r : banyaknya titik knot

q : orde polinomial

p : banyaknya variabel prediktor.

Model regresi nonparametrik *spline truncated* pada Persamaan (34) dapat ditulis dalam notasi matriks seperti pada Persamaan (36).

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \quad (36)$$

dengan uraian matriksnya ditunjukkan sebagai berikut:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \ \boldsymbol{\beta} = \begin{bmatrix} \beta_0^* \\ \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{1q} \\ \beta_{1(q+1)} \\ \vdots \\ \beta_{1(q+r)} \\ \vdots \\ \beta_{p1} \\ \beta_{p2} \\ \vdots \\ \beta_{pq} \\ \beta_{p(q+1)} \\ \vdots \\ \beta_{p(q+r)} \end{bmatrix}, \ \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\boldsymbol{X}=\begin{bmatrix} 1 & x_{11} & x_{11}^2 & \dots & x_{11}^q & (x_{11}-k_{11})_+^q & \dots & (x_{11}-k_{1r})_+^q & \dots & x_{p1} & x_{p1}^2 & \dots & x_{p1}^q & (x_{p1}-k_{p1})_+^q & \dots & (x_{p1}-k_{pr})_+^q \\ 1 & x_{12} & x_{12}^2 & \dots & x_{12}^q & (x_{12}-k_{11})_+^q & \dots & (x_{12}-k_{1r})_+^q & \dots & x_{p2} & x_{p2}^2 & \dots & x_{p2}^q & (x_{p2}-k_{p1})_+^q & \dots & (x_{p2}-k_{pr})_+^q \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{1n}^2 & \dots & x_{1n}^q & (x_{1n}-k_{11})_+^q & \dots & (x_{1n}-k_{1r})_+^q & \dots & x_{pn} & x_{pn}^2 & \dots & x_{pn}^q & (x_{pn}-k_{p1})_+^q & \dots & (x_{pn}-k_{pr})_+^q \end{bmatrix}$$

dengan $\boldsymbol{y}$ merupakan vektor variabel respon berukuran $n \times 1$, $\boldsymbol{X}$ merupakan matriks yang elemennya terdiri dari fungsi regresi dan fungsi *truncated* yang berukuran $n \times (1 + (q + r) \times p)$, dan $\boldsymbol{\beta}$ merupakan vektor parameter regresi berukuran $(1 + (q + r) \times p) \times 1$, serta $\boldsymbol{\varepsilon}$ merupakan vektor yang elemennya adalah residual pada observasi yang berukuran $n \times 1$.

1. Orde dalam *Spline Truncated*

Spline truncated dapat dibentuk dalam berbagai orde, seperti orde satu (linier), orde dua (kuadratik), dan orde tiga (kubik) (Dani & Ni'matuzzahroh, 2022). Secara umum, orde *spline* menentukan derajat polinomial pada setiap segmen serta tingkat kelengkungan fungsi yang dapat ditangkap oleh model (Likhachev, 2022). Semakin tinggi orde *spline* yang digunakan, semakin fleksibel fungsi dalam mengakomodasi perubahan pola dan kelengkungan data pada setiap sub-interval yang dibatasi oleh titik knot. Namun, peningkatan orde *spline* juga berdampak pada bertambahnya jumlah parameter yang harus diestimasi. Pada *spline truncated* berorde q dengan r titik knot, jumlah parameter meningkat seiring bertambahnya orde dan jumlah knot. Akibatnya, kompleksitas model menjadi lebih tinggi dan risiko *overfitting* meningkat, terutama ketika ukuran sampel terbatas.

Fleksibilitas yang berlebihan pada model *spline* dapat menyebabkan kurva terlalu mengikuti fluktuasi acak (*noise*) dalam data sehingga menurunkan kemampuan generalisasi model terhadap data baru (Ruppert dkk., 2009). Dari sisi stabilitas estimasi, penggunaan *spline* orde rendah cenderung menghasilkan estimasi yang lebih stabil dibandingkan orde yang lebih tinggi pada ukuran sampel yang terbatas. Oleh karena itu, dalam praktiknya pemilihan orde *spline* perlu mempertimbangkan bentuk model yang cukup fleksibel untuk merepresentasikan pola data namun tetap sederhana agar estimasi parameter stabil dan tidak rentan terhadap *overfitting*.

2. Pemilihan Titik Knot Optimal

Titik knot merupakan titik perpaduan bersama yang menunjukkan perubahan pola hubungan antara variabel prediktor dan variabel respon dalam model regresi *spline*. Pemilihan jumlah dan posisi titik knot menjadi aspek yang sangat penting, karena titik knot yang tepat akan menghasilkan model *spline* yang optimal, sehingga mampu menggambarkan pola data secara akurat tanpa menyebabkan *overfitting* maupun *underfitting*. Salah satu kriteria yang umum digunakan adalah *Generalized Cross Validation* (GCV). Nilai GCV yang memberikan titik knot optimal adalah nilai GCV minimum (Davala dkk., 2024). Nilai GCV dapat diperoleh menggunakan persamaan berikut:

$$GCV(k) = \frac{MSE\ (k)}{[n^{-1} trace(\boldsymbol{I} - \boldsymbol{H}(k))]^2} \tag{37}$$

dengan

$$MSE\ (k) = n^{-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{38}$$

Mean Squared Error (MSE) merupakan rata-rata kuadrat residual yang digunakan untuk mengevaluasi tingkat kesalahan prediksi model. Pada persamaan (37), k menyatakan titik knot yang sedang diuji, sedangkan $\boldsymbol{H}(k)$ merupakan matriks hat yang bergantung pada struktur model *spline truncated* dengan k titik knot, yang didefinisikan sebagai:

$$\boldsymbol{H}(k) = \boldsymbol{X}(k)\left(\boldsymbol{X}(k)^T\boldsymbol{X}(k)\right)^{-1}\boldsymbol{X}(k)^T \tag{39}$$

dengan $\boldsymbol{X}(k)$ merupakan matriks desain yang dibentuk sesuai dengan kombinasi titik knot, sehingga $\boldsymbol{H}(k)$ berubah mengikuti struktur dari $\boldsymbol{X}(k)$. Adapun $\boldsymbol{I}$ merupakan matriks identitas, dan n merupakan jumlah observasi.

1.4.6 Persentase Penduduk Miskin

Kemiskinan didefinisikan sebagai ketidakmampuan memenuhi standar minimum kebutuhan dasar berupa kebutuhan makanan maupun non makanan (Syairozi, 2020). Penentuan penduduk miskin didasarkan pada ambang batas moneter yang disebut Garis Kemiskinan (GK), yang merupakan penjumlahan dari Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non Makanan (GKNM). GKM merupakan jumlah nilai pengeluaran dari 52 komoditi dasar makanan yang riil dikonsumsi penduduk yang setara dengan 2100 kilokalori perkapita perhari. Sementara GKNM adalah penjumlahan nilai kebutuhan minimum dari komoditi-komoditi non makanan terpilih yang meliputi perumahan, sandang, pendidikan dan kesehatan (BPS, 2025). Terdapat tiga faktor utama yang menyebabkan kemiskinan pada sisi ekonomi. Pertama, kemiskinan dapat terjadi akibat ketidakmerataan kepemilikan sumber daya produktif sehingga menghasilkan distribusi pendapatan yang tidak seimbang. Kedua, perbedaan kualitas sumber daya manusia yang berdampak pada rendahnya produktivitas dan pada akhirnya menghasilkan tingkat pendapatan atau upah yang rendah. Ketiga, kemiskinan juga dapat disebabkan oleh keterbatasan akses terhadap kepemilikan modal, yang menghambat kemampuan individu atau rumah tangga untuk meningkatkan kapasitas ekonomi mereka (Purnomo dkk., 2023).

Berdasarkan penelitian-penelitian terdahulu, terdapat sejumlah faktor yang diduga berpengaruh terhadap persentase penduduk miskin. Adapun faktor-faktor tersebut antara lain sebagai berikut:

1. Umur Harapan Hidup (UHH)
   UHH merupakan indikator demografis yang menggambarkan rata-rata tahun hidup yang diharapkan dapat dicapai oleh seseorang sejak lahir dengan asumsi bahwa pola mortalitas pada tahun tertentu tetap konstan sepanjang hidup individu tersebut. Menurut BPS, UHH dihitung berdasarkan tabel kehidupan (*life table*) yang mencerminkan tingkat kematian pada berbagai kelompok umur dalam suatu populasi. Nilai UHH yang tinggi mencerminkan kondisi kesehatan penduduk yang lebih baik, mencerminkan akses terhadap layanan kesehatan,

gizi yang memadai, sanitasi yang layak, serta tingkat kesejahteraan sosial dan ekonomi yang lebih baik. Sebaliknya, UHH yang rendah dapat mengindikasikan tingginya angka kematian bayi, penyakit menular atau tidak menular yang belum tertangani, serta ketimpangan dalam akses layanan kesehatan.

2. Tingkat Pengangguran Terbuka (TPT)
   TPT didefinisikan sebagai persentase jumlah pengangguran terhadap total angkatan kerja. Angkatan kerja sendiri mencakup penduduk usia kerja berusia 15 tahun ke atas yang sedang bekerja maupun yang memiliki pekerjaan tetapi sementara tidak bekerja, serta mereka yang tidak bekerja dan sedang aktif mencari pekerjaan (BPS, 2025). Tingginya tingkat pengangguran menunjukkan bahwa kesempatan kerja yang tersedia belum mampu mengimbangi jumlah penduduk usia kerja, sehingga berpotensi menurunkan produktivitas dan pendapatan masyarakat.
3. Persentase Penduduk Miskin yang Bekerja di Sektor Informal
   BPS mendefinisikan pekerja sektor informal sebagai individu yang bekerja pada kegiatan ekonomi berskala kecil dengan tingkat organisasi yang rendah, serta tidak memiliki perlindungan kerja yang memadai. Kategori ini mencakup pekerja berusaha sendiri, berusaha dibantu buruh tidak tetap/buruh tidak dibayar, pekerja bebas, serta pekerja keluarga. Penduduk miskin cenderung lebih banyak terserap dalam sektor informal karena keterbatasan akses terhadap pendidikan, keterampilan, dan peluang kerja di sektor formal. Pekerjaan informal umumnya tidak memberikan jaminan pendapatan yang stabil, tidak memiliki perlindungan sosial seperti BPJS Ketenagakerjaan, dan rentan terhadap risiko ekonomi.
4. Persentase Penduduk Miskin yang Bekerja di Sektor Pertanian
   Pekerja sektor pertanian dapat didefinisikan sebagai individu yang bekerja pada kegiatan ekonomi yang meliputi tanaman pangan, hortikultura, perkebunan, perikanan, peternakan, kehutanan, serta aktivitas pertanian lainnya (BPS, 2025). Kelompok pekerja di sektor ini umumnya bekerja pada usaha berskala kecil, berteknologi rendah, serta sangat dipengaruhi oleh faktor musim dan kondisi lingkungan. Pekerjaan di sektor ini memiliki tingkat produktivitas yang relatif rendah dan pendapatan yang fluktuatif, terutama pada komoditas yang bergantung pada harga pasar dan cuaca.
5. Rata-Rata Lama Sekolah (RLS)
   RLS merupakan indikator pendidikan yang menggambarkan rata-rata tahun sekolah yang telah diselesaikan oleh penduduk usia 25 tahun ke atas. RLS dihitung berdasarkan akumulasi lama pendidikan formal yang ditamatkan oleh penduduk dalam suatu wilayah, sehingga indikator ini mencerminkan tingkat pencapaian pendidikan masyarakat secara umum. Semakin tinggi nilai RLS, semakin baik kualitas modal manusia akibat dari kesempatan pendidikan yang lebih lama dan potensi keterampilan yang lebih baik. Sebaliknya, nilai RLS yang rendah menunjukkan keterbatasan akses pendidikan, tingginya angka putus sekolah, atau ketimpangan kualitas pendidikan antarwilayah.

6. Pengeluaran per Kapita Disesuaikan
   Pengeluaran per Kapita Disesuaikan merupakan indikator yang menggambarkan tingkat kemampuan ekonomi penduduk dengan memperhitungkan kebutuhan dasar dan perbedaan struktur konsumsi antarwilayah. Pengeluaran per kapita disesuaikan dihitung dengan membagi total pengeluaran rumah tangga untuk seluruh komoditas, baik makanan maupun non-makanan dengan jumlah anggota rumah tangga, kemudian dilakukan penyesuaian menggunakan *equivalence scale* untuk mencerminkan perbedaan kebutuhan antara anggota keluarga dewasa dan anak.
7. Indeks Pembangunan Manusia (IPM)
   IPM merupakan indikator untuk mengukur capaian pembangunan manusia melalui tiga dimensi dasar, yaitu umur panjang dan hidup sehat, pengetahuan, serta standar hidup layak. Nilai IPM yang tinggi mencerminkan peningkatan kualitas sumber daya manusia dan kemampuan masyarakat dalam mengakses layanan dasar. Dalam konteks kemiskinan, IPM sering digunakan untuk menggambarkan kemampuan suatu wilayah dalam menyediakan kesempatan hidup yang lebih baik. Wilayah dengan IPM rendah cenderung memiliki persentase penduduk miskin yang lebih tinggi, akibat keterbatasan akses pada pendidikan, kesehatan, dan pendapatan yang memadai.
8. Angka Melek Huruf (AMH) Penduduk Usia 15 Tahun ke Atas
   AMH didefinisikan sebagai persentase penduduk berusia 15 tahun ke atas yang dapat membaca dan menulis kalimat sederhana dengan huruf Latin maupun huruf lainnya dibandingkan total penduduk dalam kelompok usia tersebut (BPS, 2024). Indikator ini mencerminkan sejauh mana sistem pendidikan dasar dan program keaksaraan berjalan efektif dalam memberikan keterampilan literasi dasar. Angka melek huruf yang tinggi menunjukkan bahwa masyarakat memiliki kemampuan literasi yang memadai sebagai modal penting dalam peningkatan kualitas hidup, pengembangan sumber daya manusia, dan partisipasi dalam aktivitas ekonomi. Sebaliknya, rendahnya angka melek huruf berpotensi menghambat kegiatan sosial ekonomi karena individu dengan keterampilan literasi terbatas cenderung menghadapi hambatan dalam memperoleh pekerjaan yang layak, mengakses informasi publik, serta mengikuti perkembangan teknologi.

BAB II
METODE PENELITIAN

2.1 Sumber Data dan Variabel

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa data Persentase Penduduk Miskin di Provinsi Sulawesi Selatan pada tahun 2023-2024. Data tersebut bersumber dari Badan Pusat Statistik (BPS) Provinsi Sulawesi Selatan yang dapat diakses melalui laman resmi https://sulsel.bps.go.id/id. Variabel yang digunakan dalam penelitian ini terdiri atas satu variabel respon (Y) dan delapan variabel prediktor (X) yang rinciannya ditampilkan pada Tabel 1. Seluruh data yang digunakan dalam penelitian ini dapat dilihat pada Lampiran 1.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Satuan
Y	Persentase Penduduk Miskin	Persen
X_1	Umur Harapan Hidup	Tahun
X_2	Tingkat Pengangguran Terbuka	Persen
X_3	Persentase Penduduk Miskin yang Bekerja di Sektor Informal	Persen
X_4	Persentase Penduduk Miskin yang Bekerja di Sektor Pertanian	Persen
X_5	Rata-rata Lama Sekolah	Tahun
X_6	Pengeluaran per Kapita	Ribu Rupiah/ Orang/Tahun
X_7	Indeks Pembangunan Manusia	Persen
X_8	Angka Melek Huruf Penduduk Usia 15 Tahun ke Atas	Persen

2.2 Metode Analisis

Langkah – langkah analisis data dalam penelitian ini adalah sebagai berikut.

1. Melakukan estimasi parameter model Regresi Nonparametrik *Spline Truncated* Komponen Utama *Robust* menggunakan *metode Minimum Covariance Determinant-Least Trimmed Square*.
2. Melakukan pemodelan Regresi Nonparametrik *Spline Truncated* Komponen Utama *Robust* menggunakan metode *Minimum Covariance Determinant-Least Trimmed Square* pada faktor-faktor yang memengaruhi persentase penduduk miskin di Sulawesi Selatan, dengan tahapan sebagai berikut:
   a. Melakukan analisis statistik deskriptif terhadap variabel respon dan prediktor untuk memberikan gambaran dasar mengenai variabel yang terlibat dalam penelitian.
   b. Membuat *scatter plot* data untuk mengetahui pola hubungan antara variabel respon terhadap masing-masing variabel prediktor.
   c. Melakukan uji multikolinearitas dengan matriks korelasi pada variabel prediktor.

d. Mendeteksi observasi yang termasuk sebagai pencilan dengan melihat nilai jarak Mahalanobis pada Persamaan (8) dan nilai *Studentized Deleted Residual* (TRES) pada Persamaan (9).
e. Menerapkan Analisis Komponen Utama (AKU) *Robust* menggunakan metode *Minimum Covariance Determinant* (MCD) dengan algoritma FAST-MCD melalui langkah-langkah sebagai berikut:
   1) Mengambil sub-sampel awal dari data observasi secara acak menggunakan Persamaan (21). Misalkan sub-sampel tersebut H_1 dengan jumlah elemen sebanyak h.
   2) Menghitung vektor rata-rata $\bar{\boldsymbol{x}}_{MCD}$ pada H_1 yang selanjutnya dapat disebut dengan $\bar{\boldsymbol{x}}_1$ menggunakan Persamaan (19).
   3) Menghitung matriks variansi-kovariansi $\boldsymbol{S}_{MCD}$ pada H_1 yang selanjutnya dapat disebut dengan $\boldsymbol{S}_1$ menggunakan Persamaan (20).
   4) Menghitung 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 observasi menggunakan Persamaan (22).
   5) Mengurutkan nilai RD dari yang terkecil hingga terbesar.
   6) Mengambil elemen sejumlah h observasi dengan jarak *robust* terkecil kemudian menjadikan h observasi tersebut sebagai H_2.
   7) Mengulangi langkah 2 sampai langkah 5 hingga diperoleh himpunan bagian yang konvergen, yaitu ketika $\det(\boldsymbol{S}_{i+1}) \leq \det(\boldsymbol{S}_i)$.
   8) Memilih sub-sampel H yang memiliki nilai determinan $\boldsymbol{S}_{MCD}$ terkecil dari sub-sampel yang ditemukan.
   9) Menghitung vektor rata-rata $\bar{\boldsymbol{x}}_{MCD}$ dan matriks variansi-kovariansi $\boldsymbol{S}_{MCD}$ dari sub-sampel yang terpilih.
   10) Mentransformasi matriks variansi-kovariansi yang terbentuk dari langkah sebelumnya menjadi matriks korelasi menggunakan Persamaan (6).
   11) Menghitung nilai eigen dan vektor eigen berdasarkan matriks korelasi yang terbentuk.
   12) Membentuk komponen utama *robust* berdasarkan proporsi variansi kumulatif antara 70% dan 100%.
   13) Menggunakan komponen utama *robust* $(Z_1, Z_2, \dots, Z_m)$ sebagai variabel prediktor baru yang tidak saling berkorelasi.
f. Membentuk matriks desain *spline* dengan basis fungsi *truncated* untuk tiap komponen Z_c dengan jumlah titik knot $g = 1, 2, 3$ dan orde $q = 1$.
g. Melakukan pemodelan data menggunakan regresi nonparametrik *spline truncated* komponen utama *robust*.
h. Memilih titik knot optimal berdasarkan kriteria *Generalized Cross Validation* (GCV) minimum yang diperoleh melalui persamaan (37).
i. Melakukan estimasi parameter model *robust* menggunakan LTS-*estimator*. Adapun tahapannya sebagai berikut:
   1) Menentukan nilai h awal menggunakan Persamaan (25)

2) Mengestimasi parameter $\hat{\beta}_{LS}$ dari h awal menggunakan metode *Least Square*.
3) Menentukan nilai kuadrat residual $\varepsilon_i^2 = (y_i - \hat{y}_i)^2$, kemudian mengurutkan nilai ε_i^2 dari yang terkecil hingga terbesar kemudian menghitung $\sum_{i=1}^{h} \varepsilon_i^2$.
4) Menentukan nilai h_{baru} menggunakan rumus $h_{baru} = \frac{n_h + k + 1}{2}$
5) Mengestimasi parameter $\hat{\beta}_{baru}$ dari h_{baru} observasi menggunakan metode *Least Square*.
6) Menentukan nilai kuadrat residual ε_i^2 yang baru dan mengurutkannya dari yang terkecil hingga terbesar yang bersesuaian dengan nilai $\hat{\beta}_{baru}$ dan menghitung $\sum_{i=1}^{h_{baru}} \varepsilon_i^2$.
7) Mengulangi langkah (4) sampai (6) hingga diperoleh selisih antara jumlah kuadrat residual sebelum dan sesudah telah mendekati atau sama dengan nol (konvergen).
8) Menerapkan *Final Weighted Least Square* pada Persamaan (26) dan menentukan estimasi parameter model *robust* LTS.

j. Memperoleh model regresi nonparametrik *spline truncated* komponen utama *robust* menggunakan metode MCD-LTS.
k. Melakukan interpretasi model.