

BAB I PENDAHULUAN

1.1 Latar Belakang

Analisis Regresi adalah metode analisis kuantitatif yang dirancang untuk melihat pengaruh variabel prediktor terhadap variabel respon, khususnya dalam konteks data dengan skala interval (Gonibala dkk., 2019). Analisis regresi dapat dilakukan melalui tiga pendekatan yaitu regresi parametrik, regresi semiparametrik, dan regresi nonparametrik (Hidayat dkk., 2018). Pendekatan regresi parametrik dan semiparametrik mengandalkan asumsi tertentu terkait bentuk fungsi hubungan antara variabel, sedangkan regresi nonparametrik merupakan pendekatan statistik yang digunakan untuk menemukan pola hubungan antara variabel prediktor dan variabel respon tanpa memerlukan asumsi awal mengenai bentuk fungsi atau struktur kurva regresinya (Dewi, 2019). Perkembangan regresi nonparametrik memperkenalkan berbagai pendekatan seperti *spline*, kernel, polinomial lokal, wavelet, serta fourier. Salah satu metode yang sering digunakan dalam mengestimasi kurva regresi adalah pendekatan regresi nonparametrik *spline* (Sanusi dkk., 2019).

Spline merupakan suatu fungsi yang terdiri atas beberapa segmen polinomial dengan derajat tertentu, setiap segmen disambungkan secara kontinu dan halus pada titik-titik sambung yang disebut knots (Azwarini, 2024). Keberadaan titik knot inilah yang membuat model *spline* mampu menyesuaikan estimasi terhadap pola pergerakan data (Fadhilah dkk., 2016). Regresi *spline* memiliki beberapa pendekatan seperti *truncated*, *smoothing*, dan *penalized* (Hajatin, 2021). Pada penelitian ini akan digunakan salah satu pendekatan yaitu *spline truncated*. Regresi *spline truncated* merupakan model nonparametrik berbasis *truncated power* dengan menambahkan suku $(x - k)_+$, sehingga memberikan fleksibilitas dalam menangkap perubahan pola data (Fauziah dkk., 2024). Penerapan model regresi dapat terjadi hubungan antara satu variabel prediktor dengan variabel prediktor lainnya yang disebut dengan multikolinearitas (Sartika dkk., 2020). Multikolinearitas yang terjadi antarvariabel prediktor dapat memengaruhi proses estimasi, karena perubahan satu variabel cenderung diikuti oleh variabel lain sehingga menurunkan ketepatan hasil analisis (Wahyuningsih, 2018). Salah satu pendekatan yang sering digunakan untuk mengatasi masalah multikolinearitas adalah metode *Principal Component Analysis* (Chairunnisa dkk., 2025).

Principal Component Analysis (PCA) merupakan metode statistik multivariat yang mentransformasikan sekumpulan variabel asal ke dalam variabel baru dengan dimensi yang lebih kecil dan tidak saling berkorelasi (Manullang dkk., 2023). Penelitian oleh Arifin dkk. (2020) menunjukkan efektivitas penggunaan PCA dalam proses regresi *spline* linear pada data diabetes. Selain itu, penelitian oleh Fadhilah dkk. (2025) juga



menyebutkan PCA sebagai persiapan data dalam memodelkan prevalensi. Pada proses estimasi parameter, metode umum yang digunakan (LS) atau Metode Kuadrat Terkecil (MKT), tetapi metode ini sensitif terhadap *outlier* yang dapat menurunkan efisiensi hasil estimasi. *Outlier* merupakan pengamatan yang nilainya menyimpang jauh dari sebagian besar data. Pendekatan yang efektif untuk mengatasi permasalahan *outlier* (Raihannabil, 2025).

Regresi *robust* merupakan suatu metode regresi yang digunakan ketika data mengandung *outlier*, sehingga model yang dihasilkan resisten terhadap *outlier* (Rositawati & Fitri, 2022). Regresi *robust* memiliki beberapa jenis estimasi, yaitu M-*estimator* (*Maximum Likelihood type*), LMS-*estimator* (*Least Median Squares*), LTS-*estimator* (*Least Trimmed Squares*), MM-*estimator* (*Method of Moment*), dan S-*estimator* (*Scale*) (Setiarini & Listyani, 2017). Salah satu *estimator robust* yang memiliki nilai *breakdown point* yang tinggi adalah S-*estimator* (Mubyarjati dkk., 2019). *Breakdown point* adalah ukuran kemampuan suatu penaksir untuk tetap resisten terhadap pengaruh sejumlah *outlier* (Ichsan, 2023). Penelitian sebelumnya menunjukkan efektivitas S-*estimator* dalam menangani *outlier*, misalnya pada data indeks pemberdayaan gender di Nusa Tenggara Timur (Raihannabil, 2025) dan data indeks pembangunan manusia di Indonesia (Setiarini & Listyani, 2017). Penelitian-penelitian tersebut rata-rata berorientasi pada kasus kependudukan di Indonesia, sehingga penelitian ini dilakukan pada kasus kesehatan, khususnya kasus angka kematian ibu (AKI) di Provinsi Sulawesi Selatan dengan menggunakan estimator yang sama.

Penelitian sebelumnya telah membahas regresi *spline truncated*, penerapan regresi *spline* dalam PCA, serta penggunaan *robust S-estimator* dalam mengatasi *outlier*. Namun, penelitian-penelitian tersebut diterapkan secara terpisah dan tidak mengintegrasikan ketiga pendekatan tersebut. Oleh karena itu, penelitian akan menerapkan regresi *spline truncated* dengan melibatkan beberapa variabel prediktor. Untuk mengatasi multikolinearitas, variabel tersebut akan direduksi menggunakan PCA, hasil reduksi dimodelkan menggunakan regresi *spline truncated* dan parameternya akan diestimasi dengan *robust S-estimator*. Pendekatan tersebut diharapkan menghasilkan estimasi model regresi *spline* PCA yang optimal dan *robust* terhadap *outlier*. Hasil penelitian ini dapat menjadi acuan dan referensi bagi pemerintah atau instansi dalam menentukan kebijakan untuk mengurangi kasus AKI di Indonesia khususnya Provinsi Sulawesi Selatan.

1.2 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut.

1. Fungsi *spline* yang digunakan adalah *spline truncated* dengan pemilihan *spline* orde 1 (linear).
2. Jumlah komponen utama yang digunakan ditentukan berdasarkan persentase variansi kumulatif antara 80% dan 100%.
3. Titik knot yang digunakan pada penelitian ini yakni 1, 2, dan 3 serta pemilihan titik knot terbaik berdasarkan nilai *Generalized Cross Validation* minimum.
4. Nilai toleransi pada metode *Iteratively Reweighted Least Squares* yang digunakan sebesar 0,001.



akan adalah data kasus Angka Kematian Ibu di Provinsi Sulawesi Selatan 2019-2021.

1.3 Tujuan dan Manfaat Penelitian

Tujuan penelitian ini adalah sebagai berikut.

1. Memperoleh estimasi parameter model Regresi *Spline Principal Component Analysis* menggunakan *Robust S-Estimator*.
2. Memperoleh model Angka Kematian Ibu di Provinsi Sulawesi Selatan Tahun 2019-2021 berdasarkan Regresi *Spline Principal Component Analysis* dengan *Robust S-Estimator*.

Manfaat penelitian ini adalah sebagai berikut.

1. Penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan pengetahuan dan keilmuan terkait penerapan estimasi model Regresi *Spline Principal Component Analysis* Menggunakan *Robust S-Estimator* pada kasus Angka Kematian Ibu di Provinsi Sulawesi Selatan.
2. Hasil penelitian ini diharapkan menyediakan informasi yang komprehensif dan valid mengenai Angka Kematian Ibu beserta faktor-faktor penentunya, sehingga dapat dijadikan referensi bagi peneliti, praktisi, maupun pembuat kebijakan dalam upaya pengendalian dan penurunan angka kematian ibu di Provinsi Sulawesi Selatan.

1.4 Teori

1.4.1 Multikolinearitas

Multikolinearitas adalah kondisi ketika variabel-variabel prediktor dalam regresi saling berkorelasi tinggi sehingga melanggar asumsi dasar metode kuadrat terkecil. Keadaan ini menyebabkan pendugaan koefisien regresi menjadi tidak valid dan menurunkan akurasi prediksi model. Multikolinearitas juga berdampak pada koefisien regresi yang tidak signifikan, estimasi yang sangat sensitif terhadap perubahan data, serta simpangan baku koefisien yang besar (Saputro dkk., 2025). Multikolinearitas dapat diidentifikasi melalui tingginya korelasi antarvariabel prediktor (prediktor) dalam model regresi yang berpotensi menimbulkan permasalahan pada keakuratan estimasi parameter (Candra dkk., 2024). Korelasi yang tinggi ($\rho > 0,8$) dapat diidentifikasi dengan matriks korelasi yang diperoleh melalui standarisasi matriks variansi-kovariansi.

1. Matriks Variansi-kovariansi dan Matriks Korelasi

Matriks merupakan sekumpulan bilangan yang disusun dalam baris dan kolom sehingga membentuk suatu pola seperti persegi panjang. Bilangan-bilangan penyusunnya disebut sebagai elemen matriks, sedangkan elemen yang terletak pada baris ke- n dan kolom ke- p dinotasikan dengan $a_{n \times p}$. Pada konsep statistik matriks berperan penting dalam menyederhanakan permasalahan matematis yang kompleks (Maulida dkk., 2022). Matriks dinotasikan dengan huruf kapital seperti **A**, **B**, **C**, **D**. Sebuah matriks **A** dituliskan



$$A_{n \times p} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}$$

Matriks variansi-kovariansi adalah matriks simetris yang menempatkan nilai variansi pada diagonal utama, sementara elemen di luar diagonal berisi kovariansi antarpeubah. Kovariansi ini menggambarkan keeratan hubungan linear antara dua variabel. Secara matematis, pembentukan matriks variansi-kovariansi berawal dari deviasi vektor acak terhadap nilai harapannya (*mean vector*) (Sari, 2021). Adapun *mean vector* dapat ditulis sebagai berikut.

$$\text{Mean} = E(\mathbf{A}) = \begin{bmatrix} E(A_1) \\ E(A_2) \\ \vdots \\ E(A_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu}$$

sehingga, matriks variansi-kovariansi dengan ordo $p \times p$ ditulis dalam Persamaan (1).

$$\begin{aligned} \text{Cov}(\mathbf{A}) = \boldsymbol{\Sigma} &= E(\mathbf{A} - \boldsymbol{\mu})(\mathbf{A} - \boldsymbol{\mu})^T \\ &= E \left(\begin{bmatrix} A_1 - \mu_1 \\ A_2 - \mu_2 \\ \vdots \\ A_p - \mu_p \end{bmatrix} [A_1 - \mu_1, A_2 - \mu_2, \dots, A_p - \mu_p]^T \right) \\ &= E \begin{bmatrix} (A_1 - \mu_1)^2 & (A_1 - \mu_1)(A_2 - \mu_2) & \dots & (A_1 - \mu_1)(A_p - \mu_p) \\ (A_2 - \mu_2)(A_1 - \mu_1) & (A_2 - \mu_2)^2 & \dots & (A_2 - \mu_2)(A_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (A_p - \mu_p)(A_1 - \mu_1) & (A_p - \mu_p)(A_2 - \mu_2) & \dots & (A_p - \mu_p)^2 \end{bmatrix} \\ &= \begin{bmatrix} E(A_1 - \mu_1)^2 & E(A_1 - \mu_1)(A_2 - \mu_2) & \dots & E(A_1 - \mu_1)(A_p - \mu_p) \\ E(A_2 - \mu_2)(A_1 - \mu_1) & E(A_2 - \mu_2)^2 & \dots & E(A_2 - \mu_2)(A_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(A_p - \mu_p)(A_1 - \mu_1) & E(A_p - \mu_p)(A_2 - \mu_2) & \dots & E(A_p - \mu_p)^2 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix} \end{aligned} \quad (1)$$

dengan σ_{ij} merupakan nilai kovariansi dari A_i dan A_j ; $i = 1, 2, \dots, p$; $j = 1, 2, \dots, p$ dan μ merupakan *mean*. Keeratan antara variabel acak A_i dan A_j dapat diukur menggunakan koefisien korelasi ρ_{ij} yang dinyatakan oleh nilai variansi σ_{ii} dan kovariansi σ_{ij} pada Persamaan (2) (Helmavira, 2012).

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}} \quad (2)$$

Matriks korelasi merupakan matriks yang bersifat simetris berukuran $p \times p$ dan dinyatakan dalam Persamaan (3).

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



$$\rho = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix} \quad (3)$$

2. Nilai Koefisien Korelasi

Koefisien korelasi (ρ) menjelaskan arah dan kekuatan hubungan linear antara dua variabel pada rentang $[-1, 1]$, tanda $(+/-)$ menunjukkan arah, sedangkan besarnya menunjukkan kekuatan keterkaitan. Pada konteks pendeteksian multikolinearitas, fokus diarahkan pada nilai absolut koefisien korelasi antarprediktor, semakin mendekati 1, keterkaitan semakin meningkat dan dapat memicu indikasi inflasi ragam. Menurut Tichy (2012) nilai koefisien korelasi diklasifikasikan kedalam beberapa kategori, yaitu sebagai berikut.

- Nilai koefisien korelasi sebesar 0 – 0,3 menunjukkan korelasi rendah.
- Nilai koefisien korelasi sebesar 0,3 - 0,5 menunjukkan korelasi sedang.
- Nilai koefisien korelasi sebesar 0,5 - 0,7 menunjukkan korelasi yang dianggap penting.
- Nilai koefisien korelasi sebesar 0,7 - 0,9 menunjukkan korelasi yang kuat.
- Nilai koefisien korelasi sebesar 0,9 - 1,0 menunjukkan korelasi yang sangat kuat.

1.4.2 Principal Component Analysis

Principal Component Analysis (PCA) merupakan suatu metode statistik yang digunakan untuk menyederhanakan struktur data melalui transformasi linear, sehingga terbentuk sistem koordinat baru yang mampu memaksimalkan variansi data serta mereduksi dimensi tanpa kehilangan karakteristik utamanya. Metode ini mentransformasi variabel awal yang saling berkorelasi menjadi seperangkat variabel baru (komponen utama) yang bersifat tidak berkorelasi. Arah dari komponen ditentukan oleh vektor eigen dari matriks variansi-kovariansi dan besarnya variasi oleh tiap komponen dijelaskan oleh nilai eigen (Firlana dkk., 2015).

1. Nilai Eigen dan Vektor Eigen

Menurut Shofia dkk. (2021) nilai eigen merupakan nilai karakteristik yang dimiliki oleh sebuah matriks persegi berukuran $p \times p$. Nilai ini muncul dari hubungan antara matriks dengan suatu vektor tak nol yang disebut vektor eigen. Jika Σ adalah matriks variansi-kovariansi berukuran $p \times p$, maka sebuah vektor tak nol disebut sebagai vektor eigen (e) dari Σ apabila berlaku:

$$\Sigma e = \lambda e \quad (4)$$

dengan λ adalah skalar yang dinamakan nilai eigen dari Σ , sedangkan e merupakan vektor eigen dari Σ .

Untuk memperoleh nilai eigen dari suatu matriks Σ , Persamaan (4) kali menjadi $\Sigma e = \lambda I e$ dan ekuivalen dengan $(\lambda I - \Sigma)e = 0$. Agar λ an nilai eigen, Persamaan (4) harus memiliki solusi tak nol. Kondisi



$$\det(\lambda I - \Sigma) = 0 \quad (5)$$

ial sebagai persamaan karakteristik (*characteristic equation*) dari alar λ yang memenuhi persamaan karakteristik inilah yang disebut

nilai eigen. Setelah nilai eigen diperoleh, selanjutnya menentukan nilai eigen yang bersesuaian. Vektor eigen (e) diperoleh dengan mensubstitusikan setiap nilai eigen kedalam Persamaan (6).

$$(\lambda I - \Sigma)e = 0 \quad (6)$$

Persamaan (6) merupakan sistem persamaan linear dan vektor tak nol e yang menjadi solusi dari sistem persamaan linear pada Persamaan (6) disebut sebagai vektor eigen yang bersesuaian dengan nilai eigen (λ) dan Persamaan (6) dapat diselesaikan menggunakan metode aljabar linear seperti eliminasi Gauss. Untuk setiap nilai eigen dapat diperoleh satu vektor eigen yang membentuk ruang eigen (*eigenspace*). Vektor eigen yang diperoleh selanjutnya dinormalisasi dengan membagi setiap elemennya dengan panjang (norma) vektornya, sehingga dihasilkan vektor eigen dengan panjang satu.

2. Pembentukan Komponen Utama

PCA bertujuan untuk menjelaskan struktur variansi-kovariansi dari sekumpulan variabel dengan membentuk kombinasi linear dari variabel-variabel prediktor. Variabel yang saling berkorelasi diubah menjadi sekumpulan komponen utama yang bebas korelasi ($Z = (Z_1, Z_2, \dots, Z_p)$), sehingga komponen utama ke- j secara matematis dapat didefinisikan oleh Persamaan (7) (Exsight, 2023).

$$Z = e_j^T X = e_{j1}X_1 + e_{j2}X_2 + \dots + e_{jp}X_p, \quad j = 1, 2, \dots, p \quad (7)$$

$$e_j^T X = [e_{j1} \quad e_{j2} \quad \dots \quad e_{jp}]^T \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

Persamaan (7) jika dituliskan dalam bentuk matriks sebagai berikut.

$$e_j^T X = \begin{bmatrix} e_{j1} & e_{j2} & \dots & e_{jp} \\ e_{21} & e_{22} & \dots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{p1} & e_{p2} & \dots & e_{pp} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

dengan z_j merupakan komponen utama ke- j , $X = (X_1, X_2, \dots, X_p)^T$ yang merupakan vektor variabel prediktor dan $e_j = (e_{j1}, e_{j2}, \dots, e_{jp})^T$ yang merupakan vektor karakteristik (*eigenvector*), *eigenvector* e memenuhi Persamaan (8).

$$(\Sigma - \lambda I)e = 0 \quad (8)$$

dengan λ merupakan nilai eigen (*eigenvalue*) dari komponen utama.

Terdapat tiga sifat utama PCA yaitu komponen utama (Z_1) memuat variasi terbesar dari data, sedangkan komponen berikutnya menjelaskan variasi yang tersisa secara berurutan, selanjutnya komponen-komponen yang terbentuk bersifat ortogonal satu sama lain, dan proporsi variansi yang dijelaskan oleh j -dirumuskan sebagai berikut.

$$\frac{\lambda_j}{\sum_{j=1}^p \lambda_j} = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \dots + \lambda_p}; \quad j = 1, 2, \dots, p$$

komponen utama dapat ditentukan melalui proporsi variansi kumulatif antara 80% dan 100% dari total variansi (Irfan, 2024).



1.4.3 Regresi Nonparametrik *Spline*

Regresi nonparametrik adalah metode analisis yang tidak menetapkan bentuk fungsional tertentu dalam hubungan antara variabel respon dan prediktor, sehingga model dapat menyesuaikan diri secara fleksibel terhadap pola data. Pendekatan ini memungkinkan terbentuknya kurva regresi langsung dari data, sehingga efektif untuk menggambarkan hubungan yang kompleks dan nonlinear (Dani dkk., 2021). Apabila terdapat pasangan data (x_{ij}, y_i) dengan $i = 1, 2, \dots, n$ sebagai jumlah observasi pada masing-masing variabel dan $j = 1, 2, \dots, p$ sebagai variabel prediktor, serta pola hubungan diantara keduanya diasumsikan mengikuti regresi nonparametrik, maka model umum regresi nonparametrik dituliskan pada Persamaan (9).

$$y_i = \beta_0 + g(x_{ij}) + \varepsilon_i, \quad i = 1, 2, \dots, n; j = 1, 2, \dots, p \quad (9)$$

dengan y_i menyatakan variabel respon ke- i , x_{ij} sebagai variabel prediktor, $g(x_{ij})$ adalah fungsi dengan bentuk kurva regresi yang tidak diketahui, dan ε_i merupakan residual acak yang diasumsikan identik, independen, serta berdistribusi normal dengan rata-rata nol dan variansi σ^2 (Ni'matuzzahroh & Dani, 2024).

Spline adalah metode regresi nonparametrik berupa fungsi polinomial tersegmen dengan titik sambung (knot) yang memberikan fleksibilitas lebih tinggi (Dani & Ni'matuzzahroh, 2022). Knot merupakan titik sambung yang menandai adanya perubahan pola atau perilaku kurva pada interval yang berbeda (Akhyar dkk., 2025). Terdapat beberapa metode regresi nonparametrik seperti *spline truncated*, kernel, dan deret *fourier*, dan dari beberapa metode tersebut yang sering digunakan adalah *spline truncated* (Wongkar & Fahmuddin, 2023).

1. *Spline Truncated*

Regresi nonparametrik *spline truncated* merupakan metode yang memanfaatkan polinomial tersegmen dengan titik potong tertentu (knot) untuk menghasilkan kurva yang fleksibel dalam menangkap pola data (Dani & Ni'matuzzahroh, 2022). Metode ini menggunakan parameter orde *spline* yang dilambangkan dengan d untuk menunjukkan derajat polinomial pada setiap segmen, sedangkan jumlah titik knot dinyatakan dengan r dan p merepresentasikan banyaknya variabel prediktor yang digunakan, dan n menunjukkan jumlah observasi pada keseluruhan data. Keberadaan knot berfungsi sebagai batas perubahan arah fungsi polinomial, yang memungkinkan model *spline truncated* menangkap variasi pola data secara lebih akurat dan adaptif terhadap perubahan lokal. Secara umum, model regresi *spline* dapat dilihat pada Persamaan (10) (Mustafidah, 2024).

$$y_i = \beta_0 + \sum_{j=1}^p \sum_{l=1}^d \beta_{jl} x_{ij}^l + \sum_{j=1}^p \sum_{h=1}^r \beta_{j(d+h)} (x_{ij} - k_{jh})_+^d + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (10)$$



ted

$$(x_{ij} - k_{jh})_+^d = \begin{cases} (x_{ij} - k_{jh})_+^d, & ; x_{ij} \geq k_{jh}, \\ 0, & ; x_{ij} < k_{jh}. \end{cases}$$

spline truncated sebagaimana dinyatakan pada Persamaan (10) kali dalam bentuk matriks pada Persamaan (11).

$$y = X\beta + \varepsilon \quad (11)$$

dari Persamaan (11) diperoleh persamaan residual yang ditulis dengan:

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \quad (12)$$

Jumlah kuadrat residual dalam matriks dapat dijelaskan oleh Persamaan (13)

$$\sum_{i=1}^n \varepsilon_i^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} \quad (13)$$

dari Persamaan (11) dan jumlah kuadrat residual pada Persamaan (13) dapat diselesaikan persamaan matriksnya menjadi:

$$\begin{aligned} \sum_{i=1}^n \varepsilon_i^2 &= \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} \\ &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}^T \mathbf{y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} \end{aligned}$$

Agar nilai $\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = \mathbf{y}^T \mathbf{y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta}$ minimum, maka turunan pertamanya terhadap $\boldsymbol{\beta}$ harus sama dengan nol sehingga menjadi:

$$\begin{aligned} \frac{\partial(\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon})}{\partial \boldsymbol{\beta}} &= 0 \\ -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X}\hat{\boldsymbol{\beta}} &= 0 \\ \mathbf{X}^T \mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}^T \mathbf{y} \\ (\mathbf{X}^T \mathbf{X})^{-1}(\mathbf{X}^T \mathbf{X})\hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\ \mathbf{I}\hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\ \hat{\boldsymbol{\beta}}_{LS} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \end{aligned} \quad (14)$$

atau

$$\hat{\boldsymbol{\beta}}_{LS} = \begin{bmatrix} \beta_0 \\ \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{1d} \\ \beta_{1(d+1)} \\ \vdots \\ \beta_{1(d+r)} \\ \vdots \\ \beta_{p1} \\ \beta_{p2} \\ \vdots \\ \beta_{pd} \\ \beta_{p(d+1)} \\ \vdots \\ \beta_{p(d+r)} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Dengan, $\hat{\boldsymbol{\beta}}_{LS}$ merupakan vektor parameter dengan metode *least square*, $\mathbf{y}$ merupakan n , dan parameter $\mathbf{X}$ matriks berukuran $n \times (1 + pd + pr)$



$$\begin{bmatrix} (x_{11} - k_{11})_+^d & \cdots & (x_{11} - k_{1r})_+^d & \cdots & x_{1p}^1 & x_{1p}^2 & \cdots & x_{1p}^d & (x_{1p} - k_{p1})_+^d & \cdots & (x_{1p} - k_{pr})_+^d \\ (x_{21} - k_{11})_+^d & \cdots & (x_{21} - k_{1r})_+^d & \cdots & x_{2p}^1 & x_{2p}^2 & \cdots & x_{2p}^d & (x_{2p} - k_{p1})_+^d & \cdots & (x_{2p} - k_{pr})_+^d \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ (x_{n1} - k_{11})_+^d & \cdots & (x_{n1} - k_{1r})_+^d & \cdots & x_{np}^1 & x_{np}^2 & \cdots & x_{np}^d & (x_{np} - k_{p1})_+^d & \cdots & (x_{np} - k_{pr})_+^d \end{bmatrix}$$

2. Orde dalam Regresi *Spline Truncated*

Orde pada regresi *spline truncated* akan menyesuaikan pada derajat polinomial yang digunakan pada fungsi basis *spline*. Tingkat kehalusan (*smoothness*) dalam kurva *spline* akan ditentukan melalui orde *spline*. Secara umum, terdapat beberapa orde fungsi *spline* yaitu sebagai berikut (Mustafidah, 2024).

- Orde 1 yang akan menghasilkan fungsi linear (garis lurus) diantara titik knot.
- Orde 2 yang akan menghasilkan fungsi kuadrat antara knot
- Orde 3 yang akan menghasilkan fungsi kubik antara knot, dan seterusnya.

Semakin tinggi tingkatan orde, maka akan menghasilkan kurva yang lebih halus dan kompleksitas model juga akan semakin tinggi dan beresiko *overfitting*. Untuk pemilihan orde yang tepat dapat berdasarkan pada karakteristik data serta kebutuhan analisis.

1.4.4 Pemilihan Titik Knot Optimal

Titik knot merupakan titik potongan atau pertemuan yang menunjukkan perubahan pada pola perilaku data serta menjadi lokasi terjadinya pergeseran pola kurva pada interval tertentu. *Generalized Cross Validation* (GCV) merupakan salah satu metode untuk menentukan titik knot optimal, yang bertujuan memperoleh titik knot terbaik melalui nilai GCV terkecil. Adapun nilai GCV dapat ditentukan berdasarkan Persamaan (15) (Mustafidah, 2024).

$$GCV(k) = \frac{MSE(k)}{[n^{-1}trace(I - H(k))]^2} \quad (15)$$

dengan

$$MSE(k) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (16)$$

dengan *Mean Squared Error* (MSE) adalah rata-rata kuadrat residual yang digunakan untuk mengukur ketepatan prediksi model. Pada Persamaan (15) k merupakan titik knot, sedangkan $H(k)$ adalah matriks *hat* (Hessian) yang didefinisikan pada Persamaan (17).

$$\hat{y} = H(k)y$$

$$H(k) = X(k)(X(k)^T X(k))^{-1} X(k)^T \quad (17)$$

dengan $X(k)$ merupakan matriks desain yang dibentuk sesuai kombinasi knot, sehingga matriks *hat* $H(k)$ berubah mengikuti struktur $X(k)$. Oleh karena itu, prediksi $\hat{y}$ dan nilai GCV menjadi fungsi dari posisi knot yang sedang diuji.

1.4.5 Regresi *Spline Principal Component Analysis*

Regresi *spline* multiprediktor pada proses analisisnya sering terjadi masalah multikolinearitas, seperti korelasi yang tinggi antar variabel prediktor yang menyebabkan



di bias dan tidak stabil. Salah satu metode untuk mengatasinya entransformasikan variabel prediktor asal $(X_1, X_2, \dots, X_p)$ menjadi sebut komponen utama $(Z_1, Z_2, \dots, Z_p)$. Komponen utama tersebut berkorelasi) karena pembentukannya melalui kombinasi linear menggunakan nilai eigen dan vektor eigen dari matriks variansi-

kovariansi. Selanjutnya, komponen utama hasil PCA kemudian digunakan sebagai prediktor dalam model regresi nonparametrik pada Persamaan (18).

$$y_i = \beta_0 + \sum_{j=1}^p g(z_{ij}) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (18)$$

dengan z_{ij} merupakan komponen utama hasil PCA yang bebas dari multikolinearitas, kemudian *spline* memodelkan hubungan nonparametrik secara fleksibel melalui pembentukan kurva pada interval tertentu. Maka, diperoleh fungsi *spline* tiap komponen utama seperti pada Persamaan (19) (Arifin dkk., 2020).

$$g(z_{ij}) = \sum_{l=1}^d \beta_{jl} z_{ij}^l + \sum_{h=1}^r \beta_{j(d+h)} (z_{ij} - k_{jh})_+^d, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, p \quad (19)$$

berdasarkan Persamaan (19) diperoleh model regresi *spline* dengan komponen utama pada Persamaan (20).

$$y_i = \beta_0 + \sum_{j=1}^p \left[\sum_{l=1}^d \beta_{jl} z_{ij}^l + \sum_{h=1}^r \beta_{j(d+h)} (z_{ij} - k_{jh})_+^d \right] + \varepsilon_i; \quad i = 1, 2, \dots, n \quad (20)$$

Persamaan (19) dapat dinyatakan dalam bentuk matriks pada Persamaan (21).

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (21)$$

dengan $\mathbf{y}$ merupakan vektor variabel respon, $\boldsymbol{\beta}$ merupakan vektor parameter yang diestimasi dengan metode *least square* dan $\mathbf{Z}$ merupakan matriks desain *spline* PCA dengan matriks berukuran $n \times (1 + pd + pr)$ yang bergantung pada titik knot dan $\boldsymbol{\varepsilon}$ adalah vektor residual.

1.4.6 Outlier

Outlier merupakan data yang memiliki jarak yang cukup jauh dari sebagian besar data lainnya dalam suatu kumpulan data. Keberadaannya dapat memberikan pengaruh yang signifikan terhadap hasil analisis, sehingga perlu ditangani secara tepat agar tidak menimbulkan bias atau penyimpangan dalam interpretasi data. Nilai *outlier* dapat muncul akibat berbagai faktor, seperti kesalahan pengukuran, kesalahan pengumpulan atau pemrosesan data, serta adanya variasi alami yang terjadi pada data yang diamati (Dash dkk., 2023).

Posisi dan sebaran *outlier* terhadap data lainnya perlu dianalisis secara mendalam untuk menentukan apakah nilai tersebut layak dipertahankan atau harus dieliminasi dari proses analisis. Sebelum keputusan tersebut ditetapkan, tahap identifikasi merupakan prosedur awal yang krusial guna memastikan keberadaan serta karakteristik penyimpangan data dalam suatu himpunan data. Terdapat beberapa metode untuk mendeteksi *outlier*, salah satunya adalah metode *residual studentized*, *residual studentized* dinyatakan pada Persamaan (22) (Cayo &



$$t_i = \frac{\varepsilon_i}{\sqrt{\hat{\sigma}_i^2(1 - h_{ii})}} \quad (22)$$

dengan $\hat{\sigma}_i^2$ adalah variansi yang dihitung tanpa menyertakan pengamatan ke- i , yaitu:

$$\hat{\sigma}_i^2 = \frac{(n-p)\sigma^2 - \frac{\varepsilon_i^2}{1-h_{ii}}}{n-p-1} \quad \text{dan} \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n \varepsilon_i^2}{n-p}$$

dengan h_{ii} merupakan nilai pengaruh (*leverage value*) dengan rentang antara 0 dan 1 yang diperoleh berdasarkan elemen-elemen diagonal dari matriks $\text{hat } H$, dimana $H = X(X^T X)^{-1} X^T$ dan X adalah matriks yang berisi variabel prediktor pengamatan, p merupakan banyaknya parameter pada persamaan regresi yang terbentuk dan n merupakan banyaknya observasi. Penentuan nilai *outlier* berdasarkan distribusi t dengan $df = n - p - 1$, jika nilai $|t_i| > t_{\alpha/2, df}$, maka pengamatan ke- i dikategorikan sebagai *outlier*.

1.4.7 Regresi Robust

Regresi *robust* menjadi alternatif dari regresi yang dirancang untuk mengatasi kelemahan metode *Least Square* yang sangat sensitif terhadap keberadaan *outliers* dan titik berpengaruh (*leverage points*), sehingga mampu memberikan estimasi parameter yang lebih stabil. Salah satu estimator *robust* yaitu *Scale (S)* atau dapat disebut *robust S-estimator* (Rositawati & Fitri, 2022).

1. Fungsi Objektif

Fungsi objektif merupakan fungsi pada regresi *robust* yang digunakan untuk mencari fungsi pembobot. Fungsi pembobot yang sering digunakan adalah fungsi pembobot *Tukey Bisquare* pada Persamaan (23).

$$w(u_i) = \frac{\psi_c(u_i)}{u_i} = \begin{cases} \left[1 - \left(\frac{u_i}{c}\right)^2\right]^2 & ; |u_i| \leq c \\ 0 & ; |u_i| > c \end{cases} \quad (23)$$

dengan c pada Persamaan (23) disebut *tuning constant*. Nilai *tuning constant* yang digunakan pada fungsi pembobot untuk *robust S-Estimator* adalah $c = 1,547$ (Fox & Weisberg, 2010).

2. Robust S-Estimator

Robust S-Estimator pertama kali diperkenalkan oleh Rousseeuw dan Yohai (1984) digunakan sebagai metode estimasi *robust* yang memiliki nilai *breakdown point* sebesar 50%. Nilai parameter regresi *robust* diperoleh dengan meminimalkan estimator skala *robust* $\hat{\sigma}_S$ yang didefinisikan pada Persamaan (24).

$$\hat{\beta}_S = \min \hat{\sigma}_S(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) \quad (24)$$

dengan $\varepsilon_i = y_i - Z_i^T \beta$ yang merupakan residual. Estimator skala *robust* $\hat{\sigma}_S$ ditentukan sebagai solusi dari fungsi objektif pada Persamaan (25)

$$\frac{1}{n} \sum_{i=1}^n \rho \left(\frac{y_i - \sum_{j=0}^p z_{ij} \beta_j}{\hat{\sigma}_S} \right) = B \quad (25)$$



h fungsi kendala *tukey bisquare*, dan B adalah konstanta yang i skala terhadap residual berdistribusi normal dan telah ditetapkan tingkat ke-*robust*-an seperti yang dijelaskan pada Tabel 1 (Setiawan

Tabel 1. Nilai *Tuning Constant* dan Konstanta *B* untuk setiap *Breakdown Point*

<i>Breakdown Point</i>	<i>c</i>	<i>B</i>
50%	1,547	0,199
45%	1,756	0,231
40%	1,988	0,263
35%	2,251	0,295
30%	2,560	0,327
25%	2,937	0,359
20%	3,420	0,389
15%	4,096	0,419
10%	5,182	0,447

Sumber: Setiawan dkk., 2019

Berdasarkan Tabel 1, nilai konstanta *B* yang digunakan untuk breakdown point 50% adalah 0,199. Pada tahap awal, skala *robust* di awal iterasi dihitung menggunakan *Median Absolute Deviation* (MAD), seperti pada Persamaan (26).

$$\hat{\sigma}_s^{(0)} = \frac{\text{median} \{ |\varepsilon_i^{(0)} - \text{median}(\varepsilon_i^{(0)})| \}}{b} \quad (26)$$

dengan $b = 0,6745$ yang digunakan agar MAD konsisten dan $\varepsilon_i^{(0)}$ merupakan residual awal yang diperoleh dari metode *least square*. Setelah skala awal diperoleh, estimasi parameter regresi dilanjutkan menggunakan metode kuadrat terkecil berbobot yang dilakukan secara iteratif, yang dikenal sebagai *Iteratively Reweighted Least Squares* (IRLS). Pada proses ini, skala *robust* diperbarui menggunakan pembaruan skala *pseudo-S* berdasarkan fungsi ρ , sehingga estimasi yang dihasilkan tidak hanya *robust* terhadap *outlier* pada variabel respon, tetapi juga terhadap *leverage points* pada variabel prediktor.

1.4.8 Angka Kematian Ibu

Angka kematian ibu (AKI) merupakan salah satu indikator utama dalam menilai kualitas sistem kesehatan di tingkat global. Walaupun telah terjadi kemajuan dalam beberapa dekade terakhir, permasalahan ini masih menjadi tantangan serius bagi dunia. Salah satu unsur kesejahteraan yang perlu diwujudkan adalah terpenuhinya hak asasi manusia, khususnya dalam bidang kesehatan. AKI didefinisikan sebagai jumlah ibu yang meninggal akibat komplikasi selama kehamilan, persalinan, maupun masa nifas per 100.000 kelahiran hidup. Berdasarkan laporan *World Health Organization* (WHO) tahun 2017, AKI di dunia mencapai sekitar 817 kasus setiap harinya. Pada kawasan Asia Tenggara, Indonesia masih termasuk negara dengan tingkat AKI yang relatif tinggi.

Kementerian Kesehatan memproyeksikan bahwa pada tahun 2024 AKI di Indonesia mencapai 183 per 100.000 kelahiran hidup, dan menurun menjadi 131 per 100.000 kelahiran hidup pada tahun 2030. Angka tersebut menunjukkan bahwa Indonesia masih jauh dari target yang ditetapkan dalam *Sustainable Development Goals* (SDGs). Salah satu penyebab utama tingginya morbiditas dan mortalitas ibu hamil di Indonesia adalah anemia. WHO melaporkan bahwa prevalensi anemia pada ibu hamil mencapai 14% di negara maju dan meningkat hingga 20% di negara berkembang, dengan India menjadi negara dengan prevalensi anemia tertinggi.



tertinggi. Selain anemia, faktor penyebab kematian ibu lainnya antara lain perdarahan sebesar 28%, eklampsia 24%, dan infeksi 11% (Sari dkk., 2024). Selain itu, AKI juga dipengaruhi oleh faktor nonmedis seperti tingkat pendidikan ibu, status ekonomi keluarga, dan akses layanan kesehatan yang memengaruhi perilaku pencarian layanan serta kualitas perawatan ibu hamil (Say dkk., 2014).

Tingginya kasus kematian ibu di Provinsi Sulawesi Selatan menunjukkan bahwa beberapa wilayah kabupaten/kota masih memiliki kualitas pelayanan kesehatan yang rendah. Berbagai faktor turut berkontribusi, seperti komplikasi kebidanan, proses persalinan yang tidak ditangani oleh tenaga kesehatan, serta terbatasnya layanan kesehatan bagi ibu hamil dan masa nifas. Upaya penurunan angka kematian ibu dapat dilakukan dengan memahami dan mengidentifikasi berbagai faktor yang memengaruhi risiko kematian pada ibu selama kehamilan, persalinan, dan pasca persalinan (Sauddin, 2020).



BAB II METODE PENELITIAN

2.1 Sumber Data dan Variabel

Data yang digunakan dalam penelitian ini adalah data sekunder mengenai jumlah kematian ibu di Provinsi Sulawesi Selatan pada Tahun 2019-2021. Data diperoleh dari buku Profil Kesehatan Sulawesi Selatan tahun 2020-2022 yang diterbitkan oleh Dinas Kesehatan Provinsi Sulawesi Selatan pada *website* dinkes.provsulsel.go.id. Adapun variabel yang digunakan dalam penelitian ini sebagai berikut.

Tabel 2. Variabel Penelitian

Variabel	Keterangan	Definisi Operasional	Satuan
y	Jumlah kematian ibu	Kematian ibu yang terjadi selama masa kehamilan, persalinan, atau dalam 42 hari setelah berakhirnya kehamilan akibat penyebab terkait kehamilan, persalinan, atau penanganannya, bukan karena kecelakaan	Orang
X_1	Persentase cakupan pelayanan K1 ibu hamil	Persentase ibu hamil yang mendapatkan pelayanan dan kunjungan pertama oleh tenaga kesehatan.	Persen (%)
X_2	Persentase cakupan pelayanan K4 ibu hamil	Persentase ibu hamil yang memperoleh pelayanan dan kunjungan lengkap minimal empat kali sesuai standar pelayanan kehamilan.	Persen (%)
X_3	Persentase persalinan ditolong tenaga kesehatan	Persentase seluruh proses persalinan yang ditolong oleh tenaga kesehatan yang memenuhi standar kompetensi.	Persen (%)
X_4	Persentase peserta KB pasca persalinan	Persentase ibu yang menggunakan metode kontrasepsi modern dalam masa 42 hari setelah persalinan untuk mencegah kehamilan berikutnya.	Persen (%)
X_5	Persentase cakupan kunjungan nifas	Persentase ibu yang melakukan kunjungan pemeriksaan kesehatan pada masa nifas minimal tiga kali sesuai standar pelayanan nifas.	Persen (%)
	se pelayanan in pra usia n usia lanjut	Persentase pra-lansia dan lansia yang memperoleh pelayanan kesehatan sesuai standar, meliputi pemeriksaan rutin dan deteksi dini penyakit kronis.	Persen (%)



2.2 Analisis Data

Pendekatan analisis dalam penelitian ini menerapkan model regresi *spline* PCA menggunakan *robust S-estimator*. Prosedur analisis yang digunakan untuk mencapai tujuan penelitian melalui tahapan-tahapan berikut.

1. Melakukan analisis eksploratori terhadap variabel respon y dan variabel prediktor $(X_1, X_2, \dots, X_6)$ untuk memperoleh statistik deskriptif masing-masing variabel.
2. Menyusun spesifikasi model dengan *scatterplot* antara variabel respon dan masing-masing variabel prediktor sehingga pola hubungan dapat teridentifikasi.
3. Melakukan pengujian multikolinearitas antar variabel prediktor dengan memanfaatkan matriks korelasi.
4. Menerapkan PCA pada variabel prediktor $(X_1, X_2, \dots, X_6)$ sehingga diperoleh komponen utama $(Z_1, Z_2, \dots, Z_p)$ yang tidak saling berkorelasi.
5. Membentuk matriks desain *spline* dengan basis fungsi *truncated* untuk tiap komponen Z_j dengan jumlah titik knot $r = 1, 2, 3$ dan orde $d = 1$ menggunakan Persamaan (19).

$$g(z_{ij}) = \sum_{l=1}^d \beta_{jl} z_{ij}^l + \sum_{h=1}^r \beta_{j(d+h)} (z_{ij} - k_{jh})_+^d, \quad i = 1, 2, \dots, n; j = 1, 2, \dots, p \quad (19)$$

6. Memodelkan data menggunakan regresi *spline* PCA sebagaimana telah didefinisikan pada Persamaan (20).

$$y_i = \beta_0 + \sum_{j=1}^p \left[\sum_{l=1}^d \beta_{jl} z_{ij}^l + \sum_{h=1}^r \beta_{j(d+h)} (z_{ij} - k_{jh})_+^d \right] + \varepsilon_i; \quad i = 1, 2, \dots, n \quad (20)$$

7. Pada point 6, metode estimasi parameter yang digunakan adalah metode *least square* dengan komponen utama $(Z_1, Z_2, \dots, Z_p)$ sebagai variabel prediktor, sehingga diperoleh:

$$\hat{\mathbf{y}} = \mathbf{Z}\hat{\boldsymbol{\beta}}$$

$$\hat{\boldsymbol{\beta}}_{LS} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{y}$$

8. Menentukan titik knot optimal berdasarkan kriteria GCV minimum menggunakan Persamaan (15).

$$GCV(k) = \frac{MSE(k)}{[n^{-1} \text{trace}(\mathbf{I} - \mathbf{H}(k))]^2}, \quad (15)$$

9. Melakukan pendeteksian *outlier* menggunakan metode *residual studentized* pada residual model dengan parameter hasil metode *least square* menggunakan Persamaan (22).

$$t_i = \frac{\varepsilon_i}{\sqrt{\hat{\sigma}_i^2 (1 - h_{ii})}} \quad (22)$$

Observasi ke- i dinyatakan *outlier* ketika $|t_i| > t_{\alpha/2, n-p-1}$



outlier, maka digunakan *robust S-estimator* dalam mengestimasi *spline* PCA sebagai berikut.

residual menggunakan nilai parameter awal $\hat{\boldsymbol{\beta}}$ hasil dari metode *least square* sebagai berikut.

$$\boldsymbol{\varepsilon}^{(0)} = \mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\beta}}$$

dan skala awal $\hat{\sigma}$ dengan *Median Absolute Deviation* (MAD) menggunakan Persamaan (26).

$$\hat{\sigma}_s^{(0)} = \frac{\text{median} \{ |\varepsilon_i^{(0)} - \text{median}(\varepsilon_i^{(0)})| \}}{b} \quad (26)$$

dengan nilai $b = 0,6745$. Untuk iterasi ke- m , selanjutnya skala estimasi $\hat{\sigma}_s^{(m)}$ diperbarui menggunakan dengan menemukan solusi $\hat{\sigma}$ dari fungsi objektif pada Persamaan (25).

$$\frac{1}{n} \sum_{i=1}^n \rho \left(\frac{y_i - \sum_{j=0}^p Z_{ij} \beta_j}{\hat{\sigma}} \right) = B \quad (25)$$

dengan nilai $B = 0,199$

- c. Membentuk residual ternormalisasi yang distandarisasi terhadap skala pada iterasi ke- m .

$$u_i^{(m)} = \frac{\varepsilon_i^{(m)}}{\hat{\sigma}^{(m)}}$$

- d. Menghitung bobot menggunakan fungsi pembobot *Tukey Bisquare* sebagaimana telah didefinisikan pada Persamaan (23), dengan konstanta $c = 1,547$.

$$w(u_i)^{(m)} = \begin{cases} \left[1 - \left(\frac{u_i^{(m)}}{c} \right)^2 \right]^2 & ; |u_i^{(m)}| \leq c \\ 0 & ; |u_i^{(m)}| > c \end{cases} \quad (23)$$

- e. Mengestimasi parameter $\hat{\beta}_s^{(m)}$ dengan metode IRLS menggunakan pembobot $w(u_i)$.
- f. Langkah c, d, dan e diulang dengan menggunakan residual $\varepsilon^{(m)} = \mathbf{y} - \mathbf{Z}\hat{\beta}_s^{(m)}$ dan skala estimasi $\hat{\sigma}_s^{(m)}$ sampai diperoleh konvergensi. Iterasi dilakukan sampai $\sum_{i=1}^n |\varepsilon_i^{(m)}|$ konvergen atau $|\hat{\beta}_s^{(m+1)} - \hat{\beta}_s^{(m)}|$ mendekati 0, dengan m adalah banyak iterasi.
11. Memperoleh model regresi *spline* PCA menggunakan *robust S-estimator* yang konvergen.
12. Menginterpretasi, mentransformasi model serta membuat keputusan.

