BAB I
PENDAHULUAN

1.1 Latar Belakang

Angka Kematian Ibu (AKI) menjadi salah satu target yang telah ditetapkan pada pencapaian *Milenium Development Goals* (MDGs) di Indonesia untuk menurunkan AKI sebesar 102 per 100.000 kelahiran hidup pada tahun 2015 (Lisbet, 2013). Namun, Indonesia pada tahun 2015 tidak berhasil mencapai target MDGs dan sampai saat ini AKI masih tinggi di Indonesia. Hasil dari survei Profil Kesehatan Indonesia tahun 2023 menunjukkan secara umum terjadi penurunan kematian ibu selama periode 1991-2015 dari 390 menjadi 305 per 100.000 kelahiran hidup dan AKI periode 2015-2020 terjadi penurunan dari 305 menjadi 189 per 100.000 kelahiran hidup. Angka ini hampir mencapai target Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024 sebesar 183 per 100.000 kelahiran hidup (Kemenkes RI, 2023). Pemerintah Indonesia mendapatkan kesulitan dalam mencapai target yang ditetapkan karena masih kurang meratanya fasilitas kesehatan khususnya untuk ibu hamil. Sulawesi Selatan menjadi salah satu provinsi yang mengalami kenaikan AKI pada tahun 2021 dari tahun sebelumnya. Jumlah kematian ibu di Provinsi Sulawesi Selatan pada tahun 2015 tercatat sebanyak 149 orang, lalu tahun 2016 mengalami kenaikan sebanyak 153 orang, tahun 2017 mengalami penurunan sebanyak 115 orang, tahun 2018 kembali mengalami kenaikan sebanyak 139 orang, 2019 mengalami kenaikan sebanyak 144, tahun 2020 mengalami penurunan sebanyak 133 orang, dan pada tahun 2021 mengalami kenaikan sebanyak 198 orang atau 132 per 100.000 kelahiran hidup (Dinkes Sulsel, 2022).

Metode regresi umumnya digunakan untuk menganalisis data yang mempunyai variabel respon dan variabel prediktor. Analisis regresi merupakan salah satu metode statistik yang mengamati pola hubungan pasangan data antara variabel respon dan variabel prediktor. Ada dua pendekatan analisis regresi yang umum digunakan dalam mengestimasi kurva regresi, yaitu pendekatan regresi parametrik dan regresi nonparametrik. Regresi parametrik digunakan untuk menganalisis data yang memiliki pola hubungan parametrik antara variabel respon dan variabel prediktor. Adapun, pola data yang tidak diketahui atau tidak mengikuti bentuk pola parametrik dapat menggunakan pendekatan regresi nonparametrik. Data yang tidak mengikuti pola parametrik cenderung beragam sehingga memiliki pencilan yang akan menyebabkan nilai *Mean Square Error* (MSE) besar (Fransiska dkk., 2020).

Pendekatan standar untuk mendapatkan nilai estimasi parameter dari model regresi linier adalah metode *Ordinary Least Square* (OLS). Estimator OLS yang bersifat *Best Linier Unbiased Estimator* bisa didapatkan dengan memenuhi beberapa asumsi klasik yang diperlukan. Empat asumsi klasik yang harus terpenuhi yaitu, uji normalitas, uji heterokedastisitas, uji autokorelasi, dan uji multikolinearitas. Metode OLS berfungsi untuk meminimumkan jumlah kuadrat *error* dan rentan terhadap data yang memiliki pencilan (Mahmuda dkk., 2015). Koenker & Basset (1978) memperkenalkan regresi kuantil, yaitu salah satu metode regresi dengan pendekatan membagi data menjadi kuantil tertentu dengan meminimumkan *error* mutlak berbobot yang tidak simetris. Oleh karena itu, beberapa peneliti telah mengembangkan pendekatan statistik untuk mengatasi data pencilan dengan menggunakan regresi kuantil diantaranya Puteri dkk. (2020) menggunakan regresi kuantil pada data trombosit pasien Demam Berdarah *Dengue* (DBD), Muharromah dkk. (2023) menggunakan regresi kuantil pada data tingkat kejahatan di jabodetabek, Oktafia dkk. (2016) menggunakan regresi kuantil untuk mengestimasi parameter dengan metode *bootstrap*, dan Fransiska dkk. (2020) menggunakan regresi kuantil pada data kemiskinan di Bengkulu.

Analisis regresi yang mempunyai variabel prediktor lebih dari satu sering timbul masalah yaitu terjadinya hubungan tinggi antara variabel prediktor satu dengan yang lainnya. Variabel prediktor yang saling berkorelasi disebut multikolinearitas. Kasus multikolinearitas merupakan salah satu pelanggaran asumsi klasik dalam regresi parametrik yang terjadi pada variabel prediktor. Pada pendekatan parametrik, salah satu metode yang dapat digunakan yakni *Principal Component Analysis* (PCA) atau dikenal dengan Analisis Komponen Utama (AKU) (Pendi, 2021). AKU merupakan salah satu metode yang telah dikembangkan untuk mengatasi masalah mutikolinearitas untuk mengubah sebagian besar variabel asli yang digunakan dan saling berkorelasi antar variabel prediktor menjadi satu komponen variabel baru yang lebih kecil dan tidak saling berkorelasi (Supriyadi dkk., 2017). Beberapa penelitian yang telah menggunakan metode AKU diantaranya Haumahu & Lewaherilla (2020) menggunakan AKU untuk mereduksi faktor-faktor penyebab diare di Provinsi Maluku, Karomah (2022) menggunakan AKU untuk menentukan variabel yang berpengaruh terhadap citra program di Universitas Surakarta, Rufaidah & Afif Effindi (2017) menggunakan AKU pada penerapan aplikasi pembelajaran.

Regresi nonparametrik memiliki fleksibilitas yang tinggi dikarenakan data yang membentuk estimasi kurva regresi dan pola hubungan antara variabel respon dan variabel prediktor tidak diketahui. Terdapat beberapa estimator dalam regresi nonparametrik yaitu *spline,* kernel, polinomial lokal, dan deret fourier. *Spline* perlu dipertimbangkan sebagai salah satu estimator dalam regresi nonparametrik untuk menganalisis hubungan regresi. Adapun jenis estimator *spline* yang telah dikembangkan peneliti, diantaranya *spline truncated*, *spline smoothing*, *penalized spline*, dan *spline poisson* (Ramadhan dkk., 2023). Salah satu keunggulan *spline* adalah kemampuannya untuk mengatasi bentuk pola data yang cenderung naik dan turun secara tajam. Model *spline* dapat dilihat dari beberapa kriteria yaitu mempunyai nilai *Generalized Cross Validation* (GCV) minimum dan titik knot yang optimal (Utami, 2018).

Regresi kuantil *spline* telah dikembangkan penggunaannya pada kasus data pencilan yaitu Faeni (2021) menggunakan regresi kuantil dengan estimator *smoothing spline* pada data Indeks Kualitas Udara dan Anisa dkk. (2023) mengestimasi model regresi kuantil dengan *spline truncated* orde dua untuk menganalisis pola hubungan antara hematokrit dengan trombosit pasien DBD. Selanjutnya untuk masalah multikolinearitas, peneliti telah mengembangkan komponen utama dengan estimator *spline truncated* pada Arifin dkk. (2020) dengan satu respon dan Islamiyati dkk. (2022) dengan dua respon. Akan tetapi, beberapa kasus data memiliki permasalahan terhadap pencilan dan saling berkorelasi kuat antar variabel prediktor yang dapat memengaruhi keakuratan estimasi pada kasus multiprediktor. Oleh karena itu, penelitian ini akan mengkaji penggunaan regresi kuantil *spline* dengan melibatkan beberapa variabel prediktor. Data yang mengalami permasalahan pada multikolinearitas yaitu variabel prediktor yang saling berkorelasi akan direduksi melalui AKU sehingga menghasilkan estimasi model regresi kuantil *spline* yang optimal. Selanjutnya, metode akan diaplikasikan pada data angka kematian ibu berdasarkan faktor kesehatan ibu dan anak, faktor pelayanan persalinan, faktor pelayanan kesehatan pasca persalinan, dan faktor pelayanan kesehatan usia lanjut.

1.2 Batasan Masalah

Berdasarkan permasalahan yang telah diuraikan sebelumnya, Batasan masalah pada penelitian ini adalah sebagai berikut.

1. Jenis estimator *spline* yang digunakan adalah *spline truncated* dengan pemilihan *spline* orde 1 (linear).
2. Pemilihan banyaknya komponen utama yang digunakan berdasarkan persentase variansi kumulatif sebesar 85%.
3. Nilai kuantil yang digunakan terdiri dari τ = $0{,}25;\ 0{,}50$; dan $0{,}75$.
4. Titik knot yang digunakan yakni 1, 2 dan 3 dan pemilihan titik knot terbaik didasarkan pada nilai *Generalized Cross Validation* minimum.

1.3 Tujuan dan Manfaat Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Memperoleh hasil estimasi parameter model regresi kuantil komponen utama dengan estimator *spline* linear.
2. Memperoleh model data angka kematian ibu menggunakan regresi kuantil komponen utama dengan estimator *spline* linear.

Hasil penelitian ini diharapkan dapat memberikan manfaat sebagai berikut.

1. Menambah wawasan keilmuan dan pengetahuan dalam mengimplementasikan estimasi model regresi kuantil komponen utama dengan estimator *spline* linear pada data angka kematian ibu di Provinsi Sulawesi Selatan.
2. Memberikan informasi yang akurat dan lebih terperinci mengenai angka kematian ibu berdasarkan beberapa faktor, sehingga dapat menjadi acuan bagi akademisi maupun praktisi dalam penanganan angka kematian ibu di Provinsi Sulawesi Selatan.

1.4 Landasan Teori

1.4.1 Matriks Varians-Kovarians dan Matriks Korelasi

Matriks adalah sebuah susunan segi empat dari bilangan-bilangan yang disajikan dalam dua tanda kurung, yaitu ( ) atau [ ]. Bilangan-bilangan didalam susunan disebut anggota matriks (Marsudi & Marjono, 2012). Ukuran matriks diwakili oleh jumlah baris dan kolom yang memuat didalamny. Pada umumnya, matriks dinotasikan dengan huruf kapital seperti $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$, dll. Sebuah matriks $\mathbf{A}$ yang berukuran m baris dan n kolom dapat ditulis sebagai berikut.

$$\mathbf{A}_{m\times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Matriks $\mathbf{A}_{m\times n}$ dalam Persamaan 1 dapat dinotasikan dengan $[a_{ij}]_{m\times n}$ atau $[a_{ij}]$ dengan $i = 1,2,\ldots,m$ dan $j = 1,2,\ldots,n$. Matriks varians-kovarians yang didalamnya berisikan elemen-elemen seperti varian, kovarian, dan *mean*. Jika X merupakan variabel acak dengan *mean* (nilai ekspektasi $(E(\boldsymbol{A}) = \mu)$ dan matriks kovarians. Maka, *mean* vektor acak $\boldsymbol{A}$ dapat ditulis sebagai berikut.

$$E(\boldsymbol{A}) = \begin{bmatrix} E(A_1) \\ \vdots \\ E(A_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu}$$

sedangkan matriks kovarian dengan ordo $p \times p$ dapat dinyatakan dalam Persamaan 1 berikut.

$$cov(\boldsymbol{A}) = E(\boldsymbol{A}-\boldsymbol{\mu})(\boldsymbol{A}-\boldsymbol{\mu})'$$

$$= E\left[\begin{bmatrix} \boldsymbol{A_1}-\boldsymbol{\mu_1} \\ \vdots \\ \boldsymbol{A_p}-\boldsymbol{\mu_p} \end{bmatrix}(\boldsymbol{A_1}-\boldsymbol{\mu_1} \quad \cdots \quad \boldsymbol{A_p}-\boldsymbol{\mu_p})'\right]$$

$$= E\begin{bmatrix} (A_1-\mu_1)^2 & (A_2-\mu_2)(A_2-\mu_2) & \cdots & (A_1-\mu_1)(A_p-\mu_p) \\ (A_2-\mu_2)(A_1-\mu_1) & (A_2-\mu_2)^2 & \cdots & (A_2-\mu_2)(A_p-\mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (A_p-\mu_p)(A_1-\mu_1) & (A_p-\mu_p)(A_2-\mu_2) & \cdots & (A_p-\mu_p)^2 \end{bmatrix}$$

$$= \begin{bmatrix} E(A_1-\mu_1)^2 & E(A_2-\mu_2)(A_2-\mu_2) & \cdots & E(A_1-\mu_1)(A_p-\mu_p) \\ E(A_2-\mu_2)(A_1-\mu_1) & E(A_2-\mu_2)^2 & \cdots & E(A_2-\mu_2)(A_p-\mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(A_p-\mu_p)(A_1-\mu_1) & E(A_p-\mu_p)(A_2-\mu_2) & \cdots & E(A_p-\mu_p)^2 \end{bmatrix}$$

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

dengan σ_{ij} adalah kovarian dari A_i dan A_j; $i = 1,2,\ldots,p; j = 1,2,\ldots,p$ dan μ adalah *mean* (Verny Yolanda dkk., 2022). Ukuran keeratan hubungan linear antara variabel acak $\boldsymbol{A}_i$ dan $\boldsymbol{A}_j$ adalah koefisien korelasi $\boldsymbol{\rho}_{ij}$ yang dinyatakan dalam varians $\boldsymbol{\sigma}_{ii}$ dan kovarians $\boldsymbol{\sigma}_{ij}$ pada Persamaan 2 berikut.

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

Matriks koefisien korelasi adalah matriks simetri berukuran $n \times n$ yang dapat dinyatakan dalam Persamaan 3 berikut.

$$\rho_{ij} = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{1n}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{nn}}} \\ \frac{\sigma_{21}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{11}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{n1}}{\sqrt{\sigma_{nn}}\sqrt{\sigma_{11}}} & \frac{\sigma_{n1}}{\sqrt{\sigma_{nn}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{nn}}{\sqrt{\sigma_{nn}}\sqrt{\sigma_{nn}}} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{bmatrix} \quad (3)$$

1.4.2 Nilai Eigen dan Vektor Eigen

Nilai eigen dapat didefinisikan sebagai nilai sebenarnya atau nilai karakteristik karena kata eigen berasal dari bahasa Jerman dan Inggris. Jika $\mathbf{A}$ adalah matriks $n \times n$, maka vektor tak nol x di dalam R^n dinamakan vektor eigen dari $\mathbf{A}$ jika $\mathbf{A}x$ adalah kelipatan skalar dari x, yaitu $\mathbf{A}x = \lambda x$. Skalar λ disebut nilai eigen dari $\mathbf{A}$ dan x disebut vektor eigen dari $\mathbf{A}$ yang bersesuaian dengan λ. Untuk menentukan nilai eigen digunakan Persamaan (4) berikut.

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{0} \quad (4)$$

dengan A adalah matriks $n \times n$ dan I adalah matriks identitas, Persamaan (4) disebut persamaan karakteristik. Untuk setiap vektor eigen x dari $\mathbf{A}, e = \frac{x}{\|x\|}$ merupakan vektor eigen dari $\mathbf{A}$ yang bersesuaian dengan λ yang sama dan $\|e\| = 1$. Proses ini dinamakan proses normalisasi dan vektor yang didapatkan adalah vektor yang telah distandarisasi.

1.4.3 Multikolinearitas

Salah satu asumsi klasik dari regresi linear adalah tidak terjadinya multikolinearitas diantara variabel-variabel prediktor yang terlibat dalam model regresi. Multikolinearitas terjadi ketika variabel prediktor dalam model regresi berganda memiliki hubungan linear yang signifikan. Hubungan linear yang kuat ini dapat menyebabkan estimasi parameter menjadi tidak stabil dan sulit untuk diinterpretasikan dengan baik (Amni dkk., 2024).

Matriks korelasi digunakan untuk mendeteksi adanya multikolineritas pada variabel prediktor dengan menghitung nilai koefisien korelasi sederhana (*simple* correlation). Terdapat beberapa klasifikasi nilai koefisien korelasi yaitu sebagai berikut (Tichy, 2012).

a. Nilai koefisien korelasi 0 – 0,3 artinya memiliki korelasi rendah
b. Nilai koefisien korelasi 0,3 – 0,5 artinya memiliki korelasi sedang
c. Nilai koefisien korelasi 0,5 – 0,7 artinya memiliki korelasi yang dianggap penting
d. Nilai koefisien korelasi 0,7 – 0,9 artinya memiliki korelasi yang kuat
e. Nilai koefisien korelasi 0.9 – 1 artinya memiliki korelasi sangat kuat

1.4.4 Analisis Komponen Utama

Analisis Komponen Utama (AKU) merupakan teknik statistik yang dapat digunakan untuk menjelaskan struktur varians-kovarians dari sekumpulan variabel baru yang menjadikan variabel baru saling bebas, dan merupakan kombinasi linier dari variabel asal (Sriningsih dkk., 2018). Komponen utama merupakan suatu kombinasi linear berdasarkan pada skala pengukuran variabel acak $X_1, \dots, X_p$ yang sama dan memiliki struktur matriks varians-kovarians yang kemudian akan dihasilkan pasangan nilai eigen dan vektor eigen yang saling ortonormal yaitu $(\lambda_1, \boldsymbol{e}_1), \dots, (\lambda_p, \boldsymbol{e}_p)$ dengan $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Maka, komponen utama ke-j didefinisikan pada Persamaan (5) berikut (Amni dkk., 2024).

$$\mathbf{W}_j = \boldsymbol{e}_j'\boldsymbol{X} = \boldsymbol{e}_{j1}\boldsymbol{X}_1 + \boldsymbol{e}_{j2}\boldsymbol{X}_2 + \dots + \boldsymbol{e}_{jp}\boldsymbol{X}_p, j = 1,2, \dots, p \tag{5}$$

$$\boldsymbol{e}_j'\boldsymbol{X} = [\boldsymbol{e}_{j1} \quad \boldsymbol{e}_{j2} \quad \cdots \quad \boldsymbol{e}_{jp}]' \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_p \end{bmatrix}$$

bentuk matriks pada Persamaan 4 adalah sebagai berikut.

$$\boldsymbol{e}_j'\boldsymbol{X} = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{p1} & e_{p2} & \cdots & e_{pp} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_p \end{bmatrix}$$

dengan $\mathbf{W}_1$ adalah komponen pertama yang memenuhi maksimum nilai $\boldsymbol{e}_1' \sum \boldsymbol{e}_1 = \lambda_1$, $\mathbf{W}_2$ adalah komponen kedua yang memenuhi sisa keragaman komponen pertama dengan memaksimumkan nilai $\boldsymbol{e}_2' \sum \boldsymbol{e}_2 = \lambda_2$, dan $\mathbf{W}_p$ adalah komponen ke-p yang memenuhi sisa keragaman selain dari $\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_p$ dengan memaksimumkan nilai $\boldsymbol{e}_p' \sum \boldsymbol{e}_p = \lambda_p$.

Ada tiga kriteria dalam pemilihan komponen utama salah satunya menggunakan persentase variansi kumulatif terhadap total variansi. Persentase variansi kumulatif yang dijelaskan komponen utama ke-j dapat dihitung menggunakan Persamaan (6) berikut.

$$\frac{\lambda_j}{\sum_{j=1}^{p} \lambda_j} \times 100\% = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \times 100\%; j = 1,2, \dots, p \tag{6}$$

Komponen utama dapat ditentukan dengan melihat proporsi kumulatif varians dan mampu menjelaskan total variansi data sekitar 70% sampai 80% (Putri & Imro'ah, 2021).

1.4.5 Regresi Kuantil

Regresi kuantil adalah metode analisis regresi yang digunakan untuk menduga hubungan antara variabel respon dan variabel prediktor pada fungsi kuantil bersyarat tertentu. Dalam hal ini, distribusi data tidak homogen (*heterogenus*) dan kurva tidak berbentuk standar, atau tidak simetris dan terdapat ekor pada sebaran (*truncated distribution*). metode merupakan suatu metode regresi dengan pendekatan memisahkan atau membagi data menjadi kuantil-kuantil tertentu yang kemungkinan memiliki nilai dugaan yang berbeda (Santri & Hanike, 2020).

Misalkan himpunan data berpasangan $\{x_{ij}, x_{ij}, \ldots, x_{np}, y_i\}$, dengan $i = 1,2,\ldots,n$ dan $j = 1,2,\ldots,p$ adalah himpunan berpasangan dari variabel acak yang berdistribusi secara independen dan tidak identik dengan kuantil $\tau \in (0,1)$. Data ini memiliki sebaran peluang bersyarat $F(y|x_i) = P(Y \leq y|x_i)$ dan fungsi invers $F^{-1}(\tau) = \inf\{y: F(y) \geq \tau\}$ yang merupakan kuantil ke-τ dari variabel respon y. Secara umum, model regresi kuantil linier khusus untuk kuantil bersyarat dari variabel respon y_i adalah sebagai berikut.

$$y_i(\tau) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \cdots + \beta_j(\tau)x_{ij} + \cdots + \beta_k(\tau)x_{ik} + \varepsilon_i(\tau) \qquad (8)$$

dengan $i = 1,2,\ldots,n$ dan $j = 1,2,\ldots,k$

$y_i(\tau)$: variabel respon ke-i pada kuantil ke-τ
β_0 : koefisien konstanta atau intersep pada sumbu y
τ : nilai kuantil
$\beta(\tau)$: penduga parameter pada kuantil ke-τ
x_{ij} : pengamatan ke-i pada variabel prediktor ke-j
$\varepsilon_i(\tau)$: error ke-i pada kuantil ke-τ

Apabila model regresi kuantil disajikan dalam bentuk matriks, Persamaan (8) dapat dituliskan sebagai berikut.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0(\tau) \\ \beta_1(\tau) \\ \vdots \\ \beta_k(\tau) \end{bmatrix} + \begin{bmatrix} \varepsilon_1(\tau) \\ \varepsilon_2(\tau) \\ \vdots \\ \varepsilon_n(\tau) \end{bmatrix} \qquad (9)$$

Kemudian Persamaan (9) dapat ditulis dalam bentuk model linier sebagai berikut.

$$\boldsymbol{y}(\tau) = \boldsymbol{X\beta}(\tau) + \boldsymbol{\varepsilon}(\tau) \qquad (10)$$

Dengan

$\boldsymbol{y}(\tau)$: vektor kolom berukuran $n \times 1$ dari variabel respon y
$\boldsymbol{X}$: matriks berukuran $n \times (k+1)$ dengan baris n merupakan observasi pada kolom k variabel prediktor ke-j dengan $j = 1,2,\ldots,k$
$\boldsymbol{\beta}$: vektor kolom berukuran $(k+1) \times 1$ dari parameter β_j dengan $j = 1,2,\ldots,k$
$\boldsymbol{\varepsilon}$: vektor kolom berukuran $n \times 1$ dari *error* ε_i

Menurut Koenker (2005), kuantil dapat dicari dengan menggunakan optimasi, yaitu dengan mendefinisikan *loss function*.

$$\rho_\tau(\varepsilon) = \begin{cases} \tau\varepsilon & , \varepsilon \geq 0 \\ (1-\tau)\varepsilon, & \varepsilon < 0 \end{cases} \qquad (11)$$

dengan $\rho_\tau(\varepsilon)$ merupakan *loss function* ke-τ dari ε, sehingga ε merupakan *error*, yaitu $y - \hat{y}$ dan τ merupakan konstanta dengan nilai $0 < \tau < 1$.

Loss function adalah fungsi asimetris, kecuali pada $\tau = \frac{1}{2}$. Sedangkan, untuk nilai τ lainnya, *loss function* memberikan bobot sebesar $(1-\tau)$ untuk *error* negatif dan bobot sebesar τ untuk *error* positif. Hal ini dilakukan agar kuantil yang diperoleh sesuai dengan nilai τ yang ditetapkan. Kuantil ke-τ dari F_y dapat diperoleh dengan meminimumkan ekspektasi *loss* pada Persamaan (12). Dengan demikian, ekspektasi *loss* dengan *error* $\varepsilon = y - \hat{y}$ adalah sebagai berikut.

$$E[\rho_\tau(\varepsilon)] = E[\rho_\tau(y-\hat{y}]$$
$$\int_{-\infty}^{\infty} \rho_\tau(y-\hat{y}) f(y) dy$$

perbedaan *error* membuat ekspektasi *loss* dibagi menjadi dua bagian, yaitu sebagai berikut.

$$E[\rho_\tau(\varepsilon)] = \int_{-\infty}^{\hat{y}} (1-\tau)(y-\hat{y}) f(y) dy + \int_{\hat{y}}^{\infty} \tau(y-\hat{y}) f(y) dy \tag{12}$$

Selanjutnya, Persamaan (12) diminimumkan terhadap $\hat{y}$

$$\begin{aligned}
\frac{\partial}{\partial \hat{y}} E[\rho_\tau(y-\hat{y})] &= 0 \\
\frac{\partial}{\partial \hat{y}} \left[\int_{-\infty}^{\hat{y}} (1-\tau)(y-\hat{y}) f(y) dy + \int_{\hat{y}}^{\infty} \tau(y-\hat{y}) f(y) dy \right] &= 0 \\
\frac{\partial}{\partial \hat{y}} \left[(1-\tau) \int_{-\infty}^{\hat{y}} (y-\hat{y}) f(y) dy + \tau \int_{\hat{y}}^{\infty} (y-\hat{y}) f(y) dy \right] &= 0 \\
(1-\tau) \left[(y-\hat{y}) f(y) |_{-\infty}^{\hat{y}} + \int_{-\infty}^{\hat{y}} \frac{\partial}{\partial \hat{y}} (y-\hat{y}) f(y) dy \right] & \\
+ \tau \left[(y-\hat{y}) f(y) |_{\hat{y}}^{\infty} + \int_{\hat{y}}^{\infty} \frac{\partial}{\partial \hat{y}} (y-\hat{y}) f(y) dy \right] & \\
= 0 & \\
(1-\tau) \left[(y-\hat{y}) f(y) |_{y=\hat{y}} + \int_{-\infty}^{\hat{y}} f(y) dy \right] & \\
+ \tau \left[(y-\hat{y}) f(y) |_{y=\hat{y}} + \int_{\hat{y}}^{\infty} f(y) dy \right] &= 0 \\
(1-\tau)\left[0 + F_y(\hat{y})\right] + \tau\left[0 - (1 - F_y(\hat{y}))\right] &= 0 \\
(1-\tau) F_y(\hat{y}) - \tau \left(1 - F_y(\hat{y})\right) &= 0 \\
(1-\tau) F_y(\hat{y}) - \tau + \tau F_y(\hat{y}) &= 0 \\
F_y(\hat{y}) - \tau &= 0
\end{aligned}$$

diperoleh

$$F_y(\hat{y}) = \tau \tag{13}$$

Sehingga kuantil ke-τ adalah solusi dari F_y(Saidah dkk., 2016).

1.4.6 Estimasi Parameter Regresi Kuantil

Terdapat beberapa pengembangan model regresi klasik, salah satu metode untuk menentukan estimasi parameter adalah Metode Kuadrat Terkecil (MKT) atau *Ordinary Least Square* (OLS). Tujuan Estimasi dengan OLS adalah meminimumkan jumlah kuadrat sisaan. Dengan kata lain, setiap parameter harus disamakan dengan nol. Metode ini akan menghasilkan nilai estimasi parameter yang bersifat *Best Linear Unbiased Estimator* (BLUE) jika memenuhi asumsi-asumsi klasik. Namun, OLS memiliki kelemahan yaitu dalam menghadapi pencilan yang dapat membuat estimasi parameter menjadi tidak stabil. Oleh karena itu, Koenker (1978) mengembangkan metode yaitu regresi kuantil. Regresi kuantil dilakukan dengan meminimumkan jumlah nilai mutlak dari *error* atau yang dikenal *Least Absolute Deviation* (LAD) Muharromah dkk., 2023).

Regresi kuantil ke-τ diperoleh dari *error* pembobot yang berbeda. Jumlah nilai mutlak dari *error* dengan pembobot (τ) digunakan untuk *error* positif atau lebih besar sama dengan nol dan pembobot $(1-\tau)$ digunakan untuk *error* negatif atau kurang dari nol (Matdoan dkk., 2017).

$$\rho_\tau(\varepsilon) = \sum_{i=1,\varepsilon_i\geq 0}^{n} \tau|\varepsilon_i| + \sum_{i=1,\varepsilon_i<0}^{n}(1-\tau)|\varepsilon_i| \qquad (14)$$

Dengan demikian, pada regresi kuantil terdapat fungsi kuantil ke-τ dari variabel y dengan syarat x mempertimbangkan penduga $\beta(\tau)$, sehingga diperoleh solusi untuk permasalahan tersebut dinyatakan sebagai berikut.

$$\min_{\beta\in\mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(\varepsilon) = \min_{\beta\in\mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - Q_\tau(y|x)) \qquad (15)$$

dengan $\rho_\tau(\varepsilon)$ sebagai *loss function,* sehingga τ adalah indeks kuantil dengan $\tau \in (0{,}1)$ dan $Q_\tau(y|x)$ adalah fungsi kuantil ke-τ dari variabel y dengan syarat x. Fungsi untuk kuantil $Q_\tau(y|x)$ dinyatakan sebagai berikut.

$$Q_\tau(y|x) = \boldsymbol{X}_i^T \boldsymbol{\beta}(\tau) \qquad (16)$$

Dalam regresi kuantil, pada kuantil ke-τ dari F_y meminimumkan *loss function* dari Persamaan (15) sebagai berikut.

$$\beta(\tau) = \min_{\beta} \sum_{i=1}^{n} \rho_\tau(\varepsilon) = \min_{\beta} \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}(\tau)) \qquad (17)$$

dengan $\rho_\tau(\varepsilon)$ pada Persamaan (17) didefinisikan sebagai berikut.

$$\rho_\tau(\varepsilon) = \begin{cases} \tau\varepsilon & , \varepsilon \geq 0 \\ (1-\tau)\varepsilon, & \varepsilon < 0 \end{cases}$$

Selanjutnya, akan dibuktikan bahwa *loss function* berbentuk asimetris dengan penjelasan sebagai berikut.

$$\rho_\tau = [\tau I(\varepsilon \geq 0) + (1-\tau) I((\varepsilon < 0))]|\varepsilon| = [\tau - I(\varepsilon < 0)]\varepsilon \qquad (18)$$

dengan

$$I(\varepsilon \geq 0) = \begin{cases} 1, \varepsilon \geq 0 \\ 0, \varepsilon < 0 \end{cases} \qquad (19)$$

dan

$$\rho_\tau(\varepsilon) = \begin{cases} \tau\varepsilon & , \varepsilon \geq 0 \\ (1-\tau)\varepsilon, & \varepsilon < 0 \end{cases}$$

Sehingga dapat dibuktikan sebagai berikut.

1. Untuk $\varepsilon \geq 0$

$\rho_\tau = \big[\tau I(\varepsilon \geq 0) + (1-\tau) I\big((\varepsilon < 0)\big)\big]|\varepsilon|$

$= [\tau I(\varepsilon \geq 0) + 1(-\tau) I((\varepsilon < 0))]\varepsilon$

$= [\tau 1 + (1-\tau) I(\varepsilon < 0)]\varepsilon$

$= [\tau + I(\varepsilon < 0) - \tau I(\varepsilon < 0)]\varepsilon$

$= \big[\tau + \big(1 - I(\varepsilon \geq 0)\big) - \tau(1 - I(\varepsilon \geq 0))\big]\varepsilon$

$= [\tau + (1-1) - \tau(1-1)]\varepsilon$

$= \tau\varepsilon$

Atau

$\rho_\tau = [\tau - I(\varepsilon < 0)]\varepsilon$
$= [\tau - (1 - I(\varepsilon \geq 0))]\varepsilon$
$= [\tau - (1 - 1)]\varepsilon$
$= \tau\varepsilon$

2. Untuk $\varepsilon < 0$

$\rho_\tau = \left[\tau I(\varepsilon \geq 0) + (1-\tau) I\big((\varepsilon < 0)\big)\right]|\varepsilon|$
$= \left[\tau I(\varepsilon \geq 0) + (1-\tau) I\big((\varepsilon < 0)\big)\right](-\varepsilon)$
$= [\tau 0 + (1-\tau) I(\varepsilon < 0)](-\varepsilon)$
$= [(\tau - 1) I(\varepsilon < 0)]\varepsilon$
$= \left[(\tau - 1)\big(1 - I(\varepsilon \geq 0)\big)\right]\varepsilon$
$= [(\tau - 1) - (1 - 0)]\varepsilon$
$= (\tau - 1)\varepsilon$

atau

$\rho_\tau = [\tau - I(\varepsilon < 0)]\varepsilon$
$= [\tau - (1 - I(\varepsilon \geq 0))]\varepsilon$
$= [\tau - (1 - 0)]\varepsilon$
$= (\tau - 1)\varepsilon$

Sehingga menjadi

$$\rho_\tau = [\tau I(\varepsilon \geq 0) + (1-\tau) I((\varepsilon < 0)]|\varepsilon| = [\tau - I(\varepsilon < 0)]\varepsilon, \forall_\varepsilon$$

Apabila y merupakan fungsi x yang diketahui dan memiliki fungsi probabilitas $F_{y|x}(y)$, maka kuantil ke-τ dari fungsi tersebut dapat dituliskan sebagai berikut.

$$\min_{\beta} \tau \int_{i=1;\varepsilon_i \geq 0}^{n} |\varepsilon_i| dF_y(y) + (1-\tau) \int_{i=1;\varepsilon_i < 0}^{n} |\varepsilon_i| dF_y(y) \quad (20)$$

dengan mempertimbangkan $\hat{\beta}(\tau)$, maka diperoleh solusi untuk permasalahan yang dinyakatan pada Persamaan (20) sebagai berikut.

$$\hat{\beta}(\tau) = \min_{\beta \in R^p} \left\{\tau \sum_{i=1,\varepsilon_i \geq 0}^{n} |y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}(\tau)| + (1-\tau) \sum_{i=1,\varepsilon_i < 0}^{n} |y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}(\tau)|\right\} \quad (21)$$

Solusi dari Persamaan (21) tidak dapat diperoleh secara analitik. Solusi yang dapat digunakan untuk menyelesaikan Persamaan (21) adalah secara numerik, yaitu dengan algoritma simpleks. Barrodale dan Robert (1974) mengembangkan metode yaitu algoritma simpleks yang dapat memberikan solusi permasalahan program linier yang melibatkan beberapa variabel keputusan dengan bantuan komputasi (Davino dkk., 2013). Adapun beberapa istilah yang terdapat dalam algoritma simpleks adalah sebagai berikut.

1. Variabel *Slack*
   Variabel *slack* berfungsi untuk menampung sisa kapasitas pada kendala yang berupa pembatas.
2. Variabel *Surplus*
   Variabel *surplus* berfungsi untuk menampung kelebihan nilai ruas kiri pada kendala yang berupa syarat.
3. Variabel *Artificial*
   Variabel *artificial* adalah variabel positif yang berfungsi untuk memulai penyelesaian dan harus dijadikan nol pada solusi akhir. Variabel ini digunakan untuk setiap yang tidak memiliki basis. Pada kasus regresi

kuantil dengan menggunakan metode LAD, variabel *artificial* adalah deviasi atas yang diboboti dengan τ.

4. Variabel Basis dan Nonbasis

   Variabel basis dan nonbasis merupakan dua terminologi penting yang akan selalu digunakan di dalam algoritma simpleks. Variabel basis adalah variabel yang bernilai positif dan variabel nonbasis adalah variabel yang bernilai nol.

Algoritma simpleks membutuhkan sebuah tabel atau yang biasa dikenal dengan tabulasi simpleks seperti pada Tabel 1.

Tabel 1. Tabel Metode Simpleks Untuk Kasus Regresi Kuantil

c_j			0	0	$\cdots$	0	τ	$\cdots$	τ	$(1-\tau)$	$\cdots$	$(1-\tau)$
c_b	v_b	w_b	x_1	x_2	$\cdots$	x_n	d_{11}	$\cdots$	d_{1n}	d_{21}	$\cdots$	d_{2n}
d_{11}^+	x_1	b_1	a_{ij}									
d_{21}^+	x_2	b_2										
$\vdots$	$\vdots$	$\vdots$										
d_{n1}^+	x_n	b_n										
z_j												
$c_j - z_j$												

Pengisian Tabel 1 diuraikan sebagai berikut.

a. Baris c_j diisi dengan koefisien fungsi tujuan.

b. Kolom c_b diisi dengan koefisien variabel yang menjadi basis.

c. Kolom v_b diisi dengan nama-nama variabel yang menjadi basis (variabel yang Menyusun matriks identitas), dalam hal ini diisi dengan variabel *artificial* yang merupakan deviasi atas.

d. Kolom w_b diisi dengan nilai ruas kanan dari kendala.

e. Baris z_j diisi dengan rumus $z_j = \sum d_i, a_i, j = 1,2, \dots, n$.

Berikut prosedur algoritma simpleks.

1. Mengubah masalah optimasi linier ke dalam bentuk standar dan fungsi tujuan serta kendala-kendala ke dalam bentuk persamaan, yaitu dengan menambahkan variabel *slack, surplus,* dan *artificial* terhadap kendala yang berbentuk pertidaksamaan.
2. Menentukan kolom kunci (variabel keluar). Untuk masalah maksimum, memilih nilai $c_j - z_j$ terbesar. Sedangkan untuk masalah minimum, memilih nilai $c_j - z_j$ terkecil.
3. Menentukan baris kunci (variabel keluar) dengan memilih nilai rasio terkecil antara nilai ruas kiri (b_i) dengan koefisien kolom kunci (a_{ij}). Perhitungan rasio dilakukan dengan $\text{Rasio} = \frac{b_i}{a_{ij}}$, dimana $\text{rasio} > 0$.
4. Menentukan pivot pada elemen kunci yang terletak pada pertolongan antara kolom kunci dan baris kunci dalam algoritma simpleks. Elemen kunci ini kemudian diubah nilainya menjadi 1.

5. Melakukan Operasi Besar Dasar (OBD) berdasarkan pada pivot untuk baris-baris lainnya, termasuk $c_j - z_j$. Elemen-elemen dalam kolom kunci pada baris ini dijadikan nol, kecuali elemen yang dijadikan pivot. $Baris\ kunci\ baru = \frac{Baris\ kunci\ lama}{pivot}$
$Baris\ baru\ selain\ baris\ kunci = Baris\ lama - Unsur\ kolom\ kunci\ \times$
$Baris\ kunci\ baru$
6. Proses iterasi untuk masalah maksimum akan berhenti jika semua nilai pada $c_j - z_j \leq 0$ terpenuhi, menandakan solusi telah optimal. Apabila masih terdapat nilai $c_j - z_j > 0$ (positif), maka iterasi algoritma simpleks perlu dilanjutkan. Adapun untuk masalah minimum, proses iterasi akan berhenti jika semua nilai pada baris $c_j - z_j \geq 0$ terpenuhi. Apabila masih terdapat nilai $c_j - z_j < 0$ (negatif), maka iterasi algoritma simpleks perlu dilanjutkan (Khairunnisa, 2015).

1.4.7 Regresi Nonparametrik *Spline*

Regresi nonparametrik merupakan metode pendekatan yang digunakan ketika bentuk hubungannya tidak diketahui polanya yang selanjutnya kurva regresinya diasumsikan mulus (*smooth*). Regresi nonparametrik memiliki fleksibilitas yang tinggi, estimasi kurva regresinya dapat disesuaikan tanpa dipengaruhi oleh faktor subyektivitas peneliti. Model regresi nonparametrik secara umum dapat dituliskan sebagai berikut.

$$y_i = f(x_i) + \varepsilon_i, i = 1,2, \dots, n \tag{22}$$

dengan y_i sebagai variabel respon pada pengamatan ke-i, $f(x_i)$ adalah persamaan kurva regresi pada titik x_i yang merupakan variabel prediktor pada pengamatan ke-i dan ε_i adalah nilai *error* pada pengamatan ke-i yang merupakan variabel acak prediktor dengan $mean$ 0 dan varians konstanta σ^2 (Nurhuda dkk., 2022).

Spline merupakan potongan polinomial tersegmen yang dihubungkan oleh titik-titik knot yang dapat menjelaskan karakteristik dari data. Knot merupakan titik fokus dalam fungsi *spline* sehingga kurva yang dibentuk tersegmen pada titik tersebut (Kurniasari dkk., 2019). Secara umum, fungsi *spline* berorde q dengan titik knot pada $K_1, K_2, \dots, K_r$ adalah sebagai berikut (Putra dkk., 2015).

$$f(x_i) = \sum_{l=0}^{q} \beta_l x_i^l + \sum_{h=1}^{r} \beta_{q+h}(x_i - K_h)_+^q \tag{23}$$

dengan fungsi *polynomial truncated* pada Persamaan (22) adalah sebagai berikut.

$$(x_i - K_h)_+^q = \begin{cases} (x_i - K_h)^q, x \geq K_h \\ \quad 0 \qquad\ , x < K_h \end{cases} \tag{24}$$

dengan

$f(x_i)$	: fungsi regresi kuantil
β_l	: parameter *polynomial* pada orde ke-l
x_i	: variabel prediktor pada pengamatan ke-i
$\beta_0, \beta_1, \dots, \beta_{q+h}$	: parameter truncated
K_h	: titik knot ke-h, $(h = 1,2, \dots, r)$
$(x_i - K_h)_+^q$	: fungsi *polynomial truncated*

Model yang menyatakan antara k variabel prediktor dengan satu variabel respon, jika dinyatakan dalam bentuk fungsi *spline* $f(x_i)$ dalam Persamaan (22), maka dapat dinyatakan sebagai berikut (Pratiwi, 2020).

$$f(x_{ij}) = \sum_{l=0}^{q} \beta_{lj} x_{ij}^{l} + \sum_{h=1}^{r} \beta_{(q+h)j} (x_{ij} - K_{hj})_{+}^{q} \quad (25)$$

dengan $i = 1,2,\dots,n$ dan $j = 1,2,\dots,k$

$f(x_{ij})$: fungsi regresi *spline* dari pengamatan ke-i pada variabel prediktor ke-j

β_{lj} : parameter *polynomial* pada orde ke-l dan prediktor ke-j

x_{ij} : pengamatan ke-i pada variabel prediktor ke-j

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

K_{hj} : nilai titik knot ke-h pada prediktor ke-j

Apabila model pada Persamaan (24) disajikan dalam bentuk matriks, maka dapat dituliskan sebagai berikut.

$$\boldsymbol{y} = \boldsymbol{X}[K]\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (26)$$

dengan $\boldsymbol{y} = [y_1 \ \ y_2 \ \dots \ y_n]'$ adalah vektor berukuran $n \times 1$, $\boldsymbol{X}[K] = [\boldsymbol{1} \ \ \boldsymbol{X_1} \ \dots \ \boldsymbol{X_k}]$ adalah matriks $\boldsymbol{X}$ dalam bentuk *spline* dengan orde q dan r titik knot berukuran $n \times (1 + (q+r) \times k)$ dengan

$$1 = [1 \ \ 1 \ \dots \ 1]', \boldsymbol{X}_1 = \begin{bmatrix} 1 & x_{11} & (x_{11} - k_{11})_+ & \cdots & (x_{11-k_{r1}})_+ \\ 1 & x_{21} & (x_{12} - k_{11}) & \cdots & (x_{12-k_{r1}})_+ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & (x_{i1} - k_{11}) & \cdots & (x_{i1-k_{r1}})_+ \end{bmatrix},$$

$$\boldsymbol{X}_k = \begin{bmatrix} 1 & x_{11} & (x_{1k} - k_{1k})_+ & \cdots & (x_{1k-k_{rk}})_+ \\ 1 & x_{21} & (x_{1k} - k_{1k}) & \cdots & (x_{1k-k_{rk}})_+ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & (x_{ik} - k_{1k}) & \cdots & (x_{ik-rk})_+ \end{bmatrix}$$

$\boldsymbol{\beta}=[\beta_0 \ \dots \ \beta_{11} \ \dots. \ \beta_{q1} \ \dots \ \beta_{(q+1)1} \ \cdots \ \beta_{(q+r)1} \ \cdots \ \beta_{1k} \ \cdots \ \beta_{qk} \ \ \beta_{(q+1)k} \ \cdots \ \beta_{(q+r)k}]'$ adalah vektor kolom berukuran berukuran $(1 + (q+r) \times k)$ dan $\boldsymbol{\varepsilon} = [\varepsilon_1 \ \ \varepsilon_2 \ \dots \ \ \varepsilon_n]'$ adalah vektor kolom berukuran $n \times 1$ (Putra dkk., 2015).

1.4.8 Estimasi Parameter Regresi Nonparametrik *Spline*

Hidayat (2017) melakukan suatu kajian dengan kurva regresi f dinyatakan dengan fungsi *spline* f dengan titik knot K. Dalam bentuk matriks disajikan sebagai berikut.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (27)$$

Selanjutnya, estimasi parameter $\beta = [\beta_0, \beta_1, \dots, \beta_q, \dots, \beta_{q+r}]$ diperoleh melalui metode kuadrat kecil, yaitu meminimumkan jumlah kuadrat *error* sebagai berikut.

$$\min_{\beta\in\mathbb{R}^{q+1+h}}\{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\} = \min_{\beta\in\mathbb{R}^{q+1+h}}\left(\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{bmatrix}'\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{bmatrix}\right)$$

$$\min_{\beta\in\mathbb{R}^{q+1+h}}\{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\} = \min_{\beta\in\mathbb{R}^{q+1+h}}\{[\boldsymbol{y}-\boldsymbol{X}[K_{\boldsymbol{1}},K_{\boldsymbol{2}},\dots,K_{\boldsymbol{r}}]\boldsymbol{\beta}'][[\boldsymbol{y}-\boldsymbol{X}[K_{\boldsymbol{1}},K_{\boldsymbol{2}},\dots,K_{\boldsymbol{r}}]\beta]\}$$

dengan penyajian matriks diberikan sebagai berikut.

$$\begin{aligned}\sum_{i=1}^{n}\varepsilon_i^2 &= \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\\ &= ([\boldsymbol{y}-\boldsymbol{X}[K_{\boldsymbol{1}},K_{\boldsymbol{2}},\dots,K_{\boldsymbol{r}}]\boldsymbol{\beta}'][[\boldsymbol{y}-\boldsymbol{X}[K_{\boldsymbol{1}},K_{\boldsymbol{2}},\dots,K_{\boldsymbol{r}}]\boldsymbol{\beta}]) \\ &= ([\boldsymbol{y}-\boldsymbol{X}[K]\boldsymbol{\beta}]'[\boldsymbol{y}-\boldsymbol{X}[K]\boldsymbol{\beta}])\\ &= \boldsymbol{y}'\boldsymbol{y}-2\boldsymbol{\beta}'\boldsymbol{X}'[K]\boldsymbol{y}+\boldsymbol{\beta}'\boldsymbol{X}'[K]\boldsymbol{X}[K]\boldsymbol{\beta}\end{aligned} \tag{28}$$

Kemudian, Persamaan (27) diturunkan terhadap vektor $\boldsymbol{\beta}'$ dan disamakan dengan nol, maka diperoleh hasil sebagai berikut.

$$\frac{\partial(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})}{\partial\boldsymbol{\beta}} = \frac{\partial(\boldsymbol{y}'\boldsymbol{y}-2\boldsymbol{\beta}'\boldsymbol{X}'[K]\boldsymbol{y}+\boldsymbol{\beta}'\boldsymbol{X}'[K]\boldsymbol{X}[K]\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}$$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'[K]\boldsymbol{X}[K])^{-\boldsymbol{1}}\boldsymbol{X}'[K]\boldsymbol{y} \tag{29}$$

dengan

$$\boldsymbol{X}[K] = \boldsymbol{X}[K_1,K_2,\dots,K_r]$$

$$= \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^q & (x_1-K_1)_+^q & \cdots & (x_1-K_r)_+^q \\ 1 & x_2 & x_2^2 & \cdots & x_2^q & (x_1-K_1)_+^q & \cdots & (x_2-K_r)_+^q \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^q & (x_1-K_1)_+^q & \cdots & (x_1-K_1)_+^q \end{bmatrix}$$

Selanjutnya, diperoleh estimasi fungsi regresi *spline* sebagai berikut.

$$\widehat{\boldsymbol{f}}[K](x_i) = \boldsymbol{X}[K](\boldsymbol{X}'[K]\boldsymbol{X}[K])^{-1}\boldsymbol{X}'[K]\boldsymbol{y} \tag{30}$$

Persamaan (28) dapat dinyatakan ke dalam bentuk sebagai berikut.

$$\widehat{\boldsymbol{f}}[K](x_i) = \widehat{\boldsymbol{y}} = \boldsymbol{A}[K]\boldsymbol{y}$$

Dengan $\boldsymbol{A}[K] = \boldsymbol{X}[K](\boldsymbol{X}'[K]\boldsymbol{X}[K])^{-1}\boldsymbol{X}'[K]$ dan $\boldsymbol{X}[K]$ adalah matriks dari model yang bergantung pada titik knot dan $K = [K_1,K_2,\dots,K_r]'$ adalah titik knot. Dapat dilihat bahwa $\widehat{\boldsymbol{f}}[K](x_i)$ adalah estimator dalam observasi $\boldsymbol{y} = [y_1,y_2,\dots,y_n]'$.

1.4.9 Pemilihan Titik Knot Optimal

Salah satu tahapan penting dalam pendekatan *spline* adalah pemilihan titik knot yang optimal. Titik knot adalah titik perpaduan bersama, ketika terjadi perubahan dalam perilaku fungsi di interval yang berbeda. Pemilihan titik knot optimal diperlukan untuk menentukan model *spline* yang optimal. Salah satu metode pemilihan titik knot optimal adalah dengan menggunakan *Generalized Cross Validation* (GCV) minimum (Marina & Budiantara, 2013). Titik knot optimal untuk model *spline* terbaik diperoleh dari nilai GCV yang terkecil. Metode GCV dapat dituliskan sebagai berikut.

$$GCV(K) = \frac{MSE(K)}{(n^{-1}trace[\boldsymbol{I}-\boldsymbol{A}(K)])^2} \tag{31}$$

dengan $K = [K_{11},K_{12},\dots,K_{1h},\dots,K_{p1},\dots,K_{pr}]$ adalah titik knot, $MSE(K) = n^{-1}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$, dan $\boldsymbol{A}[K] = \boldsymbol{X}[K](\boldsymbol{X}'[K]\boldsymbol{X}[K]^{-\boldsymbol{1}}\boldsymbol{X}'[K]$ (Meimela, 2020).

1.4.10 Angka Kematian Ibu

Kematian ibu merupakan salah satu masalah kesehatan yang dihadapi oleh seluruh negara di dunia. Tolak ukur keberhasilan intervensi bidang kesehatan yang dilakukan oleh pemerintah dapat dilihat dari tingkat Angka Kematian Ibu (AKI). Unsur kesejahteraan yang harus diwujudkan salah satunya adalah pemenuhan hak asasi manusia berupa kesehatan. AKI adalah jumlah kematian ibu sebagai akibat dari komplikasi kehamilan, persalinan, dan masa nifas setiap 100.000 kelahiran hidup. AKI di dunia berdasarkan data *World Health Organization* (WHO) pada tahun 2017 mencapai 817 jiwa setiap harinya. AKI di Indonesia masih terkategori tinggi untuk cakupan Asia Tenggara.

Kementerian Kesehatan pada tahun 2020 memperkirakan pada tahun 2024 AKI di Indonesia akan mencapai 183 per 100.000 kelahiran hidup dan pada tahun 2030 sebesar 131 per 100.000 kelahiran hidup yang artinya masih jauh dari target SDGs. Faktor utama penyebab morbiditas dan mortalitas ibu di negara berkembang adalah anemia. WHO menyatakan prevalensi anemia pada ibu hamil sebesar 14% di negara maju dan 51% di negara berkembang. Diantara beberapa negara berkembang, India merupakan negara yang paling tinggi prevalensi anemianya. Beberapa faktor lainnya penyebab kematian ibu yaitu perdarahan 28%, eklamsia 24% dan infeksi 11% (Sari dkk., 2023).

Kasus kematian ibu yang masih cukup tinggi di Provinsi Sulawesi Selatan menunjukkan bahwa beberapa daerah di kabupaten/kota kualitas kesehatannya cukup rendah. Tingginya jumlah kematian ibu disebabkan oleh beberapa faktor seperti komplikasi kebidanan, proses persalinan tidak dilakukan oleh tenaga kesehatan, kurang memadainya pelayanan kesehatan ibu hamil maupun nifas. Menurunkan jumlah kasus kematian ibu dapat dilakukan dengan memahami dan mempelajari faktor-faktor apa saja yang dapat menyebabkan kematian pada ibu baik pada masa 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 yang merupakan data kuantitatif mengenai jumlah kematian ibu di Provinsi Sulawesi Selatan berdasarkan pada Tahun 2019 - 2021. Data diperoleh dari buku Profil Kesehatan Sulawesi Selatan tahun 2020 – 2022 yang diterbitkan oleh Dinas Kesehatan Provinsi Sulawesi Selatan melalui laman *website* dinkes.provsulsel.go.id. Variabel yang digunakan pada penelitian disajikan sebagai berikut.

Tabel 2. Variabel Penelitian

Variabel	Keterangan	Satuan
y	Jumlah kematian ibu	orang
x_1	Persentase cakupan pelayanan K1 ibu hamil	%
x_2	Persentase cakupan pelayanan K4 ibu hamil	%
x_3	Persentase persalinan ditolong tenaga kesehatan	%
x_4	Persentase peserta KB pasca persalinan	%
x_5	Persentase cakupan kunjungan nifas	%
x_6	Persentase pelayanan kesehatan pra usia lanjut dan usia lanjut	%

2.2 Metode Analisis

Pendekatan analisis yang digunakan dalam penelitian ini menggunakan model regresi kuantil komponen utama dengan estimator *spline* linear. Adapun prosedur analisis yang digunakan untuk mencapai tujuan penelitian adalah sebagai berikut.

2.2.1 Estimasi Parameter Model Regresi Kuantil Komponen Utama dengan Estimator *Spline* Linear

1. Misalkan himpunan data berpasangan $\{x_{ij}, x_{ij}, \dots, x_{np}, y_i\}$, dengan $i = 1,2,\dots,n$ dan $j = 1,2,\dots,p$ yang berdistribusi secara independen dan tidak identik serta mengikuti model regresi kuantil linier dengan kuantil $\tau \in (0,1)$.
$$y_i(\tau) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \cdots + \beta_j(\tau)x_{ij} + \cdots + \beta_k(\tau)x_{ik} + \varepsilon_i(\tau)$$
2. Mengasumsikan fungsi regresi kuantil dengan estimator *spline* linear yang bersifat aditif.
$$y_i(\tau) = f(x_{i1}) + f(x_{i2}) + \cdots + f(x_{ip)} + \varepsilon_i(\tau)$$
3. Menyajikan model regresi kuantil dengan estimator *spline* linear

$$y_i(\tau) = \sum_{j=1}^{p} f(x_{ij}) + \varepsilon_i(\tau)$$

4. Membentuk model AKU pada regresi kuantil dengan estimator *spline* linear.

$$y_i(\tau) = \sum_{j=1}^{k} f(w_{ij}) + \varepsilon_i(\tau); i = 1,2, \dots, n$$

5. Menyajikan model regresi kuantil dengan estimator *spline* linear dalam AKU dapat dinyatakan dalam bentuk matriks.

$$\boldsymbol{y}(\tau) = \boldsymbol{W}[K]\boldsymbol{\beta}(\tau) + \boldsymbol{\varepsilon}(\tau)$$

6. Memperoleh estimasi parameter $\boldsymbol{\beta}$ dengan meminimumkan persamaan berikut menggunakan algoritma simpleks.

$$\hat{\beta}(\tau) = \frac{min}{\beta(\tau)}\left\{\tau \sum_{i=1}^{n} |\varepsilon_{i1}(\tau)| + (1-\tau)\sum_{i=1}^{n} |\varepsilon_{i2}(\tau)|\right\}$$

7. Memperoleh estimasi model regresi kuantil dengan estimator *spline* linear.

$$\hat{\boldsymbol{y}}(\tau) = \boldsymbol{W}[K]\hat{\boldsymbol{\beta}}(\tau)$$

2.2.2 Pemodelan Regresi Kuantil Komponen Utama dengan Estimator *Spline* Linear

1. Melakukan eksplorasi data dari variabel respon dan variabel prediktor untuk mengetahui statistik deskriptif dari masing-masing variabel penelitian.
2. Melakukan spesifikasi model menggunakan *scatter plot* untuk mengetahui pola hubungan antara variabel respon terhadap masing-masing variabel prediktor.
3. Mendeteksi pencilan pada data.
4. Menguji multikolinearitas pada data melalui matriks korelasi pada variabel prediktor.
5. Menerapkan AKU pada data angka kematian ibu.
6. Membentuk variabel prediktor baru $(W_1, W_2, \dots, W_p)$ yang bebas linear dan merupakan hasil reduksi dari AKU.
7. Memodelkan data menggunakan regresi kuantil dengan estimator *spline* linear multivariabel dengan nilai kuantil, yaitu $\tau = 0.25, 0.50, \text{dan } 0.75$ pada satu, dua, tiga titik knot.
8. Memperoleh model regresi kuantil dengan estimator *spline* linear multivariabel yang optimal melalui nilai GCV minimum.
9. Menginterpretasikan model regresi kuantil dengan estimator *spline* linear multivariabel.