BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan salah satu analisis statistik yang bertujuan untuk memahami hubungan antara variabel respon sebagai variabel yang dipengaruhi dan beberapa variabel prediktor yang memengaruhi melalui estimasi kurva regresi (Puteri dkk., 2020). Ada dua pendekatan dalam analisis regresi untuk memperkirakan kurva regresi, yaitu regresi parametrik dan nonparametrik. Regresi parametrik membutuhkan asumsi tentang bentuk spesifik kurva regresi. Sebaliknya, regresi nonparametrik tidak memerlukan asumsi tersebut. (Sanusi dkk., 2020). Hasil analisis regresi disajikan dalam bentuk persamaan matematis yang dikenal sebagai model regresi (Walpole, 1982). Pengestimasian parameter dalam persamaan regresi secara umum menggunakan metode *Ordinary Least Square* (OLS). Metode OLS bekerja dengan prinsip meminimumkan jumlah kuadrat *error* untuk mendapatkan estimator terbaik (Wahyudi & Zain, 2014). Namun, penggunaan metode OLS membutuhkan beberapa asumsi seperti normalitas, homoskedastisitas, serta bebas dari autokorelasi dan multikolinearitas. Ketika asumsi-asumsi ini terpenuhi, analisis regresi dapat memberikan model yang baik sebagai alat prediksi. Meskipun demikian, kerap kali terdapat pencilan pada data yang menyebabkan pelanggaran pada asumsi tersebut (Usman dkk., 2024).

Pencilan atau nilai-nilai yang menyimpang dari pola umum dalam model regresi yang ada pada data dapat menimbulkan bias dalam mengestimasi parameter. Penanganan pencilan dapat diatasi dengan menggunakan metode regresi kuantil (Wahyudi & Zain, 2014). Sejalan dengan perkembangannya, penggunaan regresi kuantil telah dikembangkan terhadap data yang pola sebarannya tidak mengikuti pola parametrik seperti regresi kuantil dengan estimator *spline truncated.* Pengembangan metode tersebut memberikan model yang lebih fleksibel untuk menyesuaikan secara otomatis dengan pola data sehingga hasil estimasi yang dihasilkan lebih akurat. Penggunaan regresi kuantil dengan estimator *spline truncated* telah dilakukan oleh Anisa dkk. (2024) dalam memodelkan trombosit pasien demam berdarah berdasarkan suhu tubuh. Akan tetapi, regresi kuantil tidak dapat mengatasi pelanggaran asumsi multikolinearitas (Usman dkk., 2024).

Multikolinearitas merupakan suatu kondisi ketika terdapat hubungan linier yang kuat antara dua atau lebih variabel prediktor. Artinya, variabel-variabel prediktor tersebut saling berkorelasi satu sama lain. Hal ini dapat menyebabkan masalah dalam analisis regresi karena sulit untuk memisahkan pengaruh individual dari [illegible]bel prediktor terhadap variabel respon. Akibatnya, estimasi [illegible]enjadi kurang akurat dan interpretasi koefisien regresi menjadi [illegible] 2023). Penanganan multikolinearitas dapat dilakukan dengan [illegible]gai metode, salah satunya adalah metode *Least Absolute* [illegible]*ction Operator* (LASSO). Metode LASSO merupakan teknik [illegible]uan untuk meminimalkan jumlah kuadrat *error* dengan [illegible] kendala. Pembatasan tersebut memungkinkan LASSO untuk mengecilkan beberapa koefisien variabel hingga mendekati nol sekaligus melakukan proses seleksi variabel yang lebih relevan dalam model (Widjaja, 2015). Penelitian mengenai penggunaan metode LASSO untuk mengatasi multikolinearitas dilakukan oleh Jamco dkk. (2023) menggunakan regresi linier berganda terhadap distribusi persentase produk domestik regional bruto di Provinsi Maluku tahun 1999-2021.

Penerapan metode LASSO juga mengalami perluasan dalam pengaplikasiannya pada data dengan sebaran yang tidak mengikuti pola parametrik. Salah satunya ialah penggabungan metode LASSO dengan estimator nonparametrik seperti *spline*. Gabungan ini memungkinkan model tidak hanya melakukan seleksi variabel dan mengatasi multikolinearitas, tetapi juga menangkap pola dalam data. Metode ini dikenal sebagai regresi *spline* LASSO. Metode ini diyakini lebih efektif dalam mengontrol kompleksitas model dengan tetap mempertahankan fleksibilitas bentuk kurva regresi (Mullah dkk., 2021). Penelitian terkait kombinasi metode ini telah dilakukan oleh Guo dkk., (2016) yang menghasilkan estimasi parameter yang lebih akurat pada data dengan koefisien yang berubah secara halus, serta efektif dalam seleksi variabel pada data dengan multikolinearitas.

Belloni dan Chernozhukov (2011) mengembangkan pendekatan kombinasi regresi kuantil yang tangguh terhadap pencilan dengan LASSO dan dikenal sebagai regresi kuantil LASSO. Metode ini memungkinkan regresi lebih stabil dalam menghadapi data yang mengandung pencilan dan multikolinearitas. Randa dkk. (2022) melakukan penerapan LASSO dan regresi kuantil LASSO dalam mengidentifikasi faktor-faktor yang mempengaruhi tingkat kemiskinan di Jawa Tengah dan Usman dkk. (2024) melakukan pemodelan regresi kuantil grup LASSO pada jumlah kasus tuberkulosis di Provinsi Sulawesi Selatan. Namun, penelitian-penelitian tersebut masih menggunakan pendekatan parametrik dalam penerapan regresi kuantil LASSO sehingga perlu dilakukan analisis regresi kuantil LASSO menggunakan pendekatan nonparametrik salah satunya dengan estimator *spline truncated*.

Penggunaan regresi kuantil LASSO dengan estimator *spline truncated* dapat diaplikasikan pada data persentase penduduk miskin di Pulau Sulawesi. Penelitian yang dilakukan oleh Dani dan Ni'matuzzahroh (2021) menemukan plot data persentase penduduk miskin di Provinsi Jawa Barat tidak mengikuti pola parametrik sehingga dianalisis menggunakan regresi nonparametrik. Kemiskinan adalah salah satu tantangan global yang masih menjadi masalah di negara-negara berkembang maupun maju. Pada tahun 2020, tingkat kemiskinan di Indonesia mencapai 10,19%, artinya sekitar 27,55 juta orang hidup di bawah garis kemiskinan nasional (Badan Pusat Statistik, 2021). Indonesia sebagai salah satu negara dengan populasi
m menyebabkan masalah kemiskinan berbeda-beda di tiap
Pulau Sulawesi.

tatistik mencatat bahwa Pulau Sulawesi menjadi satu-satunya
yang mengalami peningkatan persentase penduduk miskin
penduduk miskin di Pulau Sulawesi mencapai 2,04 juta orang
9 ribu jiwa dibandingkan pada tahun 2022 (Meidyana, 2023).
h satu langkah penting dalam menurunkan angka kemiskinan

adalah mengidentifikasi faktor-faktor yang memengaruhi persentase penduduk miskin di Pulau Sulawesi. Faktor-faktor yang diduga berpengaruh terhadap kemiskinan di antaranya indeks pembangunan manusia (Alviyadi, 2023). Selain itu, angka harapan hidup, rata-rata lama sekolah, dan pengeluaran perkapita juga diduga memiliki pengaruh terhadap tingkat kemiskinan (Hasanah dkk., 2021). Menilik lebih jauh, Arief Prasetyo Adi selaku Kepala Badan Pangan Nasional mengatakan bahwa kerawanan pangan berkaitan erat dengan tingkat kemiskinan (Badan Pangan Nasional, 2023). Faktor-faktor yang diduga memengaruhi persentase penduduk miskin ternyata memiliki hubungan satu sama lain, khususnya indeks pembangunan manusia, rata-rata lama sekolah dan pengeluaran perkapita. Sebagai bagian dari dimensi indeks pembangunan manusia, rata-rata lama sekolah dan pengeluaran perkapita memainkan peran signifikan dalam membentuk keterkaitan ini (Badan Pusat Statistik, 2021). Interaksi antar faktor tersebut dapat memicu pelanggaran asumsi multikolinearitas, yang harus diperhatikan dalam analisis. Selain itu, perbedaan kondisi sosial ekonomi pada tiap daerah berpotensi akan kehadiran pencilan.

Oleh karena itu, pada penelitian ini dilakukan pemodelan terhadap persentase penduduk miskin dengan menggunakan metode regresi kuantil LASSO dengan estimator *spline* terhadap faktor-faktor yang diduga memengaruhinya yaitu indeks pembangunan manusia, rata-rata lama sekolah, angka harapan hidup, pengeluaran perkapita, dan prevalensi ketidakcukupan konsumsi pangan berdasarkan kabupaten/kota di Pulau Sulawesi pada tahun 2023.

1.2 Tujuan dan Manfaat Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Memperoleh hasil estimasi fungsi regresi kuantil *least absolute shrinkage and selection operator* dengan estimator *spline*.
2. Memperoleh model persentase penduduk miskin di Pulau Sulawesi pada tahun 2023 menggunakan regresi kuantil *least absolute shrinkage and selection operator* dengan estimator *spline*.

Hasil penelitian ini diharapkan dapat memberikan manfaat berupa menambah wawasann keilmuan dan pengetahuan dalam mengimplementasikan regresi kuantil *least absolute shrinkage and selection operator* dengan estimator *spline*.

1.3 Batasan Masalah

Berdasarkan permasalahan yang telah diuraikan sebelumnya, batasan masalah ... alah sebagai berikut.

... ınakan untuk *spline truncated* adalah orde 1.

... ot yang diuji antara 1 hingga 3 titik knot dengan nilai lambda ... LASSO berkisar antara 0,001 hingga 1.

... ng digunakan terdiri dari $\tau = 0{,}25$, $0{,}5$, dan $0{,}75$.

... knot dan lambda optimum menggunakan nilai GCV minimum.

1.4 Landasan Teori

1.4.1 Pencilan

Pencilan dapat didefinisikan sebagai pengamatan yang menyimpang secara signifikan dari pengamatan lainnya. Keberadaan pencilan dapat memiliki dampak yang besar dalam pengambilan kesimpulan atau keputusan dalam penelitian. Salah satu metode untuk mendeteksi pencilan ialah dengan menggunakan metode grafik seperti *boxplot.* Suatu observasi pada data dianggap pencilan jika data tersebut berada di luar *whisker boxplot* (Sihombing dkk., 2023). Batas *whisker boxplot* ditentukan berdasarkan kuartil pertama (Q_1), kuartil ketiga (Q_3), jangkauan interkuartil (IQR). Perhitungan IQR dinyatakan pada Persamaan (1).

$$IQR = Q_3 - Q_1 \tag{1}$$

Sedangkan *lower* dan *upper whisker* ditentukan dengan menggunakan Persamaan (2) dan (3)

$$upper\ whisker = Q_3 + 1{,}5(IQR) \tag{2}$$

$$lower\ whisker = Q_1 - 1{,}5(IQR) \tag{3}$$

dengan Q_1 dan Q_3 diperoleh berdasarkan posisi data yang telah diurutkan dari nilai terkecil hingga terbesar. Sehingga diperoleh letak Q_1 dan Q_3 menggunakan persamaan (4) dan (5).

$$Q_1 = \left[\frac{1(n+1)}{4}\right] \tag{4}$$

$$Q_3 = \left[\frac{3(n+1)}{4}\right] \tag{5}$$

Suatu observasi dikatakan pencilan apabila nilai tersebut lebih kecil dari *lower whisker* atau lebih besar dari *upper whisker* (Barnett & Lewis, 1944).

1.4.2 Multikolinearitas

Analisis regresi yang melibatkan beberapa variabel prediktor memerlukan pemenuhan asumsi-asumsi dasar agar hasil yang diperoleh akurat dan dapat diandalkan. Salah satu asumsi penting adalah tidak adanya multikolinearitas, yaitu kondisi di mana terdapat hubungan yang kuat antara dua atau lebih variabel prediktor dalam model. Korelasi mengacu pada hubungan timbal balik atau hubungan sebab akibat antara dua kejadian. Salah satu cara untuk menghitung koefisien korelasi ... ggunakan metode *pearson* yang digunakan untuk mengukur ... variabel. Nilai korelasi *pearson* antara variabel dapat dihitung ... maan (6) (Pendi, 2021).

$$r_{x_a x_b} = \frac{n\sum x_a x_b - (\sum x_a)(\sum x_b)}{\sqrt{\{n\sum x_a^2 - (\sum x_a)^2\}\{n\sum x_b^2 - (x_b)^2\}}} \tag{6}$$

... ien korelasi *pearson* di tampilkan pada Tabel 1.

Tabel 1. Klasifikasi *Koefisien Pearson*

Nilai Korelasi			Keterangan
0,000 sampai -0,199	atau	0,000 sampai 0,199	Hubungan sangat rendah
-0,200 sampai -0,399	atau	0,200 sampai 0,399	Hubungan rendah
-0,400 sampai -0,599	atau	0,400 sampai 0,599	Hubungan sedang
-0,600 sampai -0,799	atau	0,600 sampai 0,799	Hubungan kuat
-0,800 sampai -1,000	atau	0,800 sampai 1,000	Hubungan sangat kuat

1.4.3 Regresi Nonparametrik

Regresi nonparametrik adalah salah satu pendekatan yang digunakan untuk mengestimasi kurva regresi tanpa asumsi bentuk tertentu. Berbeda dengan regresi parametrik, dalam regresi nonparametrik, bentuk kurva tidak ditentukan sebelumnya, melainkan hanya diasumsikan mulus (*smooth*) dan berada dalam sebuah fungsi yang sesuai. Hal ini memberikan fleksibilitas yang lebih besar karena kurva regresi dapat menyesuaikan diri dengan pola data yang ada tanpa bergantung pada subjektivitas peneliti (Mariati dkk., 2019). Secara umum, model regresi nonparametrik dinyatakan pada Persamaan (7).

$$y_i = f(x_i) + \varepsilon_i; i = 1,2,\dots,n \tag{7}$$

dengan y_i adalah variabel respon pada pengamatan ke-i, sementara $f(x_i)$ merupakan persamaan kurva regresi yang tidak diketahui dengan x_i sebagai variabel prediktor pada pengamatan ke-i. Selain itu, ε_i merupakan error dengan rata-rata 0 dan variansi sebesar σ^2 (Eubank, 1999).

1.4.4 Estimator Spline Truncated

Spline truncated merupakan salah satu metode regresi nonparametrik yang paling umum digunakan karena kemampuannya dalam menangani perubahan perilaku data pada sub-interval tertentu. Metode ini memanfaatkan titik-titik knot untuk mendeteksi perubahan pola dan menyesuaikan kurva regresi secara lokal (Dani dkk., 2021). Fungsi kurva regresi $f(x_i)$ dalam model *spline truncated* dengan orde q yang mempunyai titik knot di $K_1, K_2, \dots, K_r$ secara umum dituliskan sebagai berikut.

$$f(x_i) = \sum_{l=0}^{q} \beta_l x_i^l + \sum_{h=1}^{r} \beta_{q+h}(x_i - K_h)_+^q \tag{8}$$

Model regresi *spline truncated* yang dinyatakan melalui Persamaan (8) dapat diperoleh melalui subtitusi Persamaan (8) ke dalam Persamaan (7), maka

$$y_i = \sum_{l=0}^{q} \beta_l x_i^l + \sum_{h=1}^{r} \beta_{q+h}(x_i - K_h)_+^q + \varepsilon_i, \qquad i = 1,2,\dots,n \tag{9}$$

dengan $(x_i - K_h)_+^q$ disebut sebagai fungsi *truncated,* yang didefinisikan sebagai:

$$(x_i - K_h)_+^q = \begin{cases} (x_i - K_h)_+^q, & x_i > K_h \\ 0 \quad , & x_i \leq K_h \end{cases}$$

koefisien polinomial, sedangkan β_{q+k} adalah koefisien pada

Model regresi *spline truncated* untuk n data pengamatan sesuai pada Persamaan (9) dapat dinyatakan dalam notasi matriks yang ditunjukkan sebagai berikut:

$$\boldsymbol{y} = \boldsymbol{X}[K]\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{10}$$

dengan bentuk eksplisit:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{11}^2 & \cdots & x_{11}^q & (x_{11}-K_{11})_+^q & \cdots & (x_{11}-K_{r1})_+^q \\ 1 & x_{21} & x_{21}^2 & \cdots & x_{21}^q & (x_{21}-K_{11})_+^q & \cdots & (x_{21}-K_{r1})_+^q \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n1}^2 & \cdots & x_{n1}^q & (x_{n1}-K_{11})_+^q & \cdots & (x_{n1}-K_{r1})_+^q \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \\ \beta_{(q+1)} \\ \vdots \\ \beta_{q+r} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Estimasi parameter $\boldsymbol{\beta}$ pada Persamaan (10) dapat dilakukan dengan menggunakan metode *Least Square* (LS) yang bertujuan untuk meminimalkan jumlah *error*. Dalam hal ini, *error* didefinisikan sebagai:

$$\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{X}[K]\boldsymbol{\beta} \tag{11}$$

Untuk mengukur jumlah *error*, perlu menghitung jumlah kuadrat *error*nya sebagai berikut:

$$\begin{aligned} \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} &= (\boldsymbol{y} - \boldsymbol{X}[K]\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}[K]\boldsymbol{\beta}) \\ \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} &= (\boldsymbol{y} - \boldsymbol{X}[K]\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}[K]\boldsymbol{\beta}) \\ &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}'[K]\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'[K]\boldsymbol{X}[K]\boldsymbol{\beta} \end{aligned} \tag{12}$$

Selanjutnya, Persamaan (12) diturunkan terhadap vektor $\boldsymbol{\beta}$ dengan menyamakan ruas kanannya dengan nol, maka diperoleh:

$$\frac{\partial(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})}{\partial\boldsymbol{\beta}} = 0$$

$$\frac{\partial(\boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}'[K]\boldsymbol{y} + \boldsymbol{\beta}\boldsymbol{X}'[K]\boldsymbol{X}[K]\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = 0$$

$$-2\boldsymbol{X}'[K]\boldsymbol{y} + 2\boldsymbol{X}'[K]\boldsymbol{X}[K]\boldsymbol{\beta} = 0$$

$$-\boldsymbol{X}'[K]\boldsymbol{y} + \boldsymbol{X}'[K]\boldsymbol{X}[K]\boldsymbol{\beta} = 0$$

$$\boldsymbol{X}'[K]\boldsymbol{X}[K]\boldsymbol{\beta} = \boldsymbol{X}'[K]\boldsymbol{y}$$

Kedua ruas dikalikan dengan $(\boldsymbol{X}'[K]\boldsymbol{X}[K])^{-1}$, sehingga:

$$\begin{aligned} (\boldsymbol{X}'[K]\boldsymbol{X}[K])^{-1}(\boldsymbol{X}'[K]\boldsymbol{X}[K])\boldsymbol{\beta} &= (\boldsymbol{X}'[K]\boldsymbol{X}[K])^{-1}\boldsymbol{X}'[K]\boldsymbol{y} \\ \boldsymbol{I}\boldsymbol{\beta} &= (\boldsymbol{X}'[K]\boldsymbol{X}[K])^{-1}\boldsymbol{X}'[K]\boldsymbol{y} \\ \widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}'[K]\boldsymbol{X}[K])^{-1}\boldsymbol{X}'[K]\boldsymbol{y} \end{aligned} \tag{13}$$

Berdasarkan persamaan (10) dan (13) diperoleh estimator fungsi regresi nonparametrik *spline truncated* dengan K titik knot sebagai berikut:

$$\begin{aligned} \widehat{\boldsymbol{y}} &= \boldsymbol{X}[K](\boldsymbol{X}'[K]\boldsymbol{X}[K])^{-1}\boldsymbol{X}[K]'\boldsymbol{y} \\ \widehat{\boldsymbol{y}} &= \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}} \end{aligned} \tag{14}$$

aan (14) dapat dinyatakan sebagai berikut:

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

dengan $\boldsymbol{A}[K] = \boldsymbol{X}[K](\boldsymbol{X}'[K]\boldsymbol{X}[K])^{-1}\boldsymbol{X}'[K]$ merupakan matriks yang digunakan dalam perhitungan titik knot, $K = (K_1, K_2, \dots, K_r)$ adalah vektor yang berisi titik knot, dan $\boldsymbol{y}$ adalah vektor dari variabel respon (Devi dkk., 2018).

1.4.5 Regresi Kuantil

Koenker dan Bassett memperkenalkan metode regresi kuantil pada tahun 1978. Regresi kuantil merupakan salah satu metode dalam analisis regresi yang digunakan untuk menduga hubungan antara variabel respon dengan variabel prediktor yang tidak hanya berfokus pada ukuran pemusatan dari variabel respon, tetapi pada berbagai kuantil (Puteri dkk., 2020). Metode ini berfungsi saat data tidak berdistribusi homogen (*heterogenous*) atau tidak berbentuk standar, seperti ketidaksimetrisan atau ketidaknormalan distribusi serta terdapat ekor pada sebaran data atau *truncated distribution* (Santri & Hanike, 2020). Kelebihan dari regresi kuantil salah-satunya ialah dapat meminimumkan pengaruh dari pencilan (Idris dkk., 2018).

Misalkan diberikan data $x_{i1}, x_{i2}, \dots, x_{ik}, y_i$ dengan $i = 1,2,\dots,n$ dan $j = 1,2,\dots,k$ adalah himpunan berpasangan dari variabel acak yang berdistribusi secara independen dan tidak identik dengan kuantil $\tau \in (0,1)$. Variabel y memiliki fungsi distribusi kumulatif (CDF) $F_y(y) = P(Y \leq y)$. Dalam regresi kuantil, digunakan CDF bersyarat yang didefinisikan sebagai $F(y|x_i) = P(Y \leq y|x_i)$ yang merepresentasikan probabilitas kumulatif dari variabel acak Y dengan mempertimbangkan kondisi variabel prediktor x_i.

Kuantil ke-τ dari $F_y(y)$ dapat diperoleh dengan mengambil nilai y terkecil yang memenuhi $F_y(y) \geq \tau$ atau dapat dituliskan dalam bentuk invers $F^{-1}(\tau) = \inf\{y: F_y(y) \geq \tau\}$ yang merupakan kuantil ke-τ dari variabel respon y merupakan fungsi invers dari $F_y(y)$. Secara umum, dalam regresi kuantil, fungsi kuantil bersyarat $Q_{y|x}(\tau)$ dituliskan sebagai $Q_y(\tau) = \inf\{y: F(y) \geq \tau\} = F^{-1}(\tau)$. Secara umum, model regresi kuantil linier khusus untuk kuantil bersyarat $Q_{y|x}(\tau)$ dari variabel respon y_i adalah sebagai berikut.

$$y_i(\tau) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \cdots + \beta_j(\tau)x_{ij} + \cdots + \beta_k(\tau)x_{ik} + \varepsilon_i(\tau) \qquad (15)$$

dengan

$y_i(\tau)$	:	variabel respon ke-i pada kuantil ke-τ
$\beta_0(\tau)$	:	koefisien konstanta atau intersep pada kuantil ke-τ
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

... i kuantil disajikan dalam bentuk matriks, Persamaan (15) dapat ... rikut.

$$\dots = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0(\tau) \\ \beta_1(\tau) \\ \vdots \\ \beta_k(\tau) \end{bmatrix} + \begin{bmatrix} \varepsilon_1(\tau) \\ \varepsilon_2(\tau) \\ \vdots \\ \varepsilon_n(\tau) \end{bmatrix} \qquad (16)$$

... n (16) dapat ditulis dalam bentuk model linier sebagai berikut.

$$\boldsymbol{y}(\tau) = \boldsymbol{X\beta}(\tau) + \boldsymbol{\varepsilon}(\tau) \tag{17}$$

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 $j = 1,2, \dots k$
$\boldsymbol{\beta}$	:	vektor kolom berukuran $(k+1) \times 1$ dari parameter β_j
$\boldsymbol{\varepsilon}$	:	vektor kolom berukuran $n \times 1$ dari *error* ε_i

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

$$\rho_\tau(\varepsilon) = \begin{cases} \tau\varepsilon & , \varepsilon \geq 0 \\ (1-\tau)\varepsilon & , \varepsilon < 0 \end{cases} \tag{18}$$

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

Loss function regresi kuantil bersifat asimetris, kecuali pada $\tau = \frac{1}{2}$. Nilai τ selain $\frac{1}{2}$ akan 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. Sehingga, ekspektasi *loss* dengan *error* adalah sebagai berikut.

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

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

$$E[\rho_\tau(y - \hat{y})] = \int_{-\infty}^{\hat{y}} (1-\tau)(\hat{y} - y) f(y) dy + \int_{\hat{y}}^{\infty} \tau(y - \hat{y}) f(y) dy \tag{19}$$

Kuantil ke-τ dari F_y dapat diperoleh dengan meminimumkan ekspektasi *loss* pada Persamaan (19). Selanjutnya, Persamaan (19) diminimumkan terhadap $\hat{y}$

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

diperoleh

$$F_y(\hat{y}) = \tau \tag{20}$$

sehingga kuantil ke-τ adalah solusi dari F_y (Davino dkk., 2013).

1.4.6 Estimasi Parameter Regresi Kuantil

Terdapat beberapa metode yang dapat digunakan untuk mengestimasi parameter dalam persamaan regresi. Salah satu pendekatan standar dalam penentuan model regresi klasik dan estimasi parameternya adalah metode kuadrat terkecil atau *Ordinary Least Square* (OLS). Prinsip utama OLS adalah meminimumkan jumlah kuadrat *error*. Namun, metode OLS sangat rentan dengan adanya pencilan dikarenakan dapat menyebabkan hasil dari estimasi parameter menjadi tidak stabil. Oleh karena itu, Koenker pada tahun 1978 mengembangkan metode alternatif, yaitu regresi kuantil. Dalam regresi kuantil, parameter diestimasi dengan meminimumkan jumlah *absolute error* yang lebih dikenal dengan *Least Absolute Deviation* (LAD) (Amyad dkk., 2018).

Pada regresi kuantil, *error* diberikan bobot yang berbeda. Bobot yang digunakan, yaitu τ untuk nilai *error* yang lebih besar atau sama dengan nol dan $1-\tau$ untuk nilai *error* yang kurang dari nol. Perkalian antara *error* dengan bobot yang diberikan disebut sebagai *loss function* (ρ_τ) yang dinyatakan sebagai berikut (Balami, 2017).

$$\rho_\tau(\varepsilon) = \sum_{i=1,\varepsilon_i\geq 0}^{n} \tau|\varepsilon_i| + \sum_{i=1,\varepsilon_i<0}^{n} (1-\tau)|\varepsilon_i| \tag{21}$$

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.

$$\min_{\beta\in\mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(\varepsilon) = \min_{\beta\in\mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau\big(y_i - Q_\tau(y|x)\big) \tag{22}$$

dengan $\rho_\tau(\varepsilon)$ sebagai *loss function*, di mana 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.

$$Q_\tau(y|x) = \boldsymbol{X}_i'\boldsymbol{\beta}(\tau) \tag{23}$$

Dalam regresi kuantil, pada kuantil ke-τ dari F_y meminimumkan *loss function* dari Persamaan (22) sebagai berikut.

$$\hat{\beta}(\tau) = \min_{\beta} \sum_{i=1}^{n} \rho_\tau(\varepsilon) = \min_{\beta} \sum_{i=1}^{n} \rho_\tau\big(y_i - \boldsymbol{X}_i'\boldsymbol{\beta}(\tau)\big) \tag{24}$$

ersamaan (24) didefinisikan sebagai berikut.

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

ibuktikan bahwa *loss function* berbentuk asimetris dengan erikut

$$I(\varepsilon \geq 0) + (1-\tau)I\big((\varepsilon < 0)\big)\big]|\varepsilon| = [\tau - I(\varepsilon < 0)]\varepsilon \tag{25}$$

Adapun fungsi indikator yang digunakan.

$$I(\varepsilon \geq 0) = \begin{cases} 1\,, \varepsilon \geq 0 \\ 0\,, \varepsilon < 0 \end{cases}, \; I(\varepsilon \leq 0) = \begin{cases} 1\,, \varepsilon < 0 \\ 0\,, \varepsilon \geq 0 \end{cases} \text{ dan } |\varepsilon| = \begin{cases} \varepsilon \;\;\,, \varepsilon \geq 0 \\ -\varepsilon\,, \varepsilon < 0 \end{cases}$$

Sehingga dapat dibuktikan sebagai berikut.

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

1. Untuk $\varepsilon \geq 0$

$$\begin{aligned} \rho_\tau &= \left[\tau I(\varepsilon \geq 0) + (1-\tau) I\big((\varepsilon < 0)\big)\right]|\varepsilon| \\ &= [\tau(1) + (1-\tau)(0)]\varepsilon \\ &= [\tau]\varepsilon \\ &= \tau\varepsilon \end{aligned}$$

atau

$$\begin{aligned} \rho_\tau &= [\tau - 0]\varepsilon \\ &= [\tau]\varepsilon \\ &= \tau\varepsilon \end{aligned}$$

2. Untuk $\varepsilon < 0$

$$\begin{aligned} \rho_\tau &= [\tau(0) + (1-\tau)(1)]|\varepsilon| \\ &= [(1-\tau)](-\varepsilon) \\ &= (\tau - 1)\varepsilon \end{aligned}$$

atau

$$\begin{aligned} \rho_\tau &= [\tau - 1]\varepsilon \\ &= (\tau - 1)\varepsilon \end{aligned}$$

Sehingga menjadi

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

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.

$$\min_\beta \tau \int_{i=1;\varepsilon_i \geq 0}^{n} |\varepsilon_i| dF_y(y) + (1-\tau) \int_{i=1;\varepsilon_i < 0}^{n} |\varepsilon_i| dF_y(y) \tag{26}$$

dengan mempertimbangkan $\hat{\beta}(\tau)$, maka diperoleh solusi untuk permasalahan yang dinyatakan pada Persamaan (26) sebagai berikut.

$$\hat{\beta}(\tau) = \begin{matrix} min \\ \beta \in R^p \end{matrix} \left\{ \tau \sum_{i=1,\varepsilon_i \geq 0}^{n} |y_i - \boldsymbol{X}_i' \boldsymbol{\beta}(\tau)| + (1-\tau) \sum_{i=1,\varepsilon_i < 0}^{n} |y_i - \boldsymbol{X}_i' \boldsymbol{\beta}(\tau)| \right\} \tag{27}$$

Solusi dari Persamaan (27) tidak dapat diperoleh secara analitik. Solusi yang dapat digunakan untuk menyelesaikan Persamaan (27) adalah secara numerik, yaitu
ıpleks. Algoritma simpleks adalah metode yang dikembangkan
.obert pada tahun 1974. Algoritma simpleks dapat memberikan
program linier yang melibatkan beberapa variabel keputusan
nputasi (Davino dkk., 2013). Adapun beberapa istilah yang
tma simpleks adalah sebagai berikut.

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. Pada kasus regresi kuantil dengan menggunakan metode LAD, variabel *surplus* adalah deviasi bawah yang diboboti dengan $(1-\tau)$.

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 persamaan 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 2.

Tabel 2. Metode Simpleks untuk Kasus Regresi Kuantil

c_j			0	0	...	0	τ	...	τ	$(1-\tau)$	...	$(1-\tau)$
c_b	v_b	w_b	x_1	x_2	...	x_n	d_{11}	...	d_{1n}	d_{21}	$\dots$	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 2 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 $z_j = \sum d_i a_{ij}$, $j = 1,2,\dots,n$.

ritma simpleks.

salah optimasi linier ke dalam bentuk standar dan fungsi tujuan

-kendala ke dalam bentuk persamaan, yaitu dengan

variabel *slack*, *surplus*, dan *artificial* terhadap kendala yang

daksamaan.

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 $\left(a_{ij}\right)$. Perhitungan rasio dilakukan dengan $Rasio = \frac{b_i}{a_{ij}}$, di mana $rasio > 0$.
4. Menentukan pivot pada elemen kunci yang terletak pada perpotongan antara kolom kunci dan baris kunci dalam algoritma simpleks. Elemen kunci ini kemudian diubah nilainya menjadi 1.
5. Melakukan Operasi Baris 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, dimana:
   $baris\ kunci\ baru = \frac{baris\ kunci\ lama}{pivot}$
   $baris\ baru\ selain\ baris\ kunci = baris\ lama - unsur\ kolom\ kunci \times baris\ kunci\ baru$
6. Proses iterasi untuk masalah maksimum akan berhenti jika semua nilai pada baris $c_j - z_j \leq 0$ 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 (Khairunisa, 2015).

1.4.7 *Least Absolute Shrinkage and Selection Operator*

Least Absolute Shrinkage and Selection Operator (LASSO) merupakan metode yang menyusutkan estimasi parameter regresi menuju nol atau menghasilkan beberapa estimasi parameter dengan nilai tepat nol (Tibshirani, 1996). Metode ini beroperasi dengan menambahkan batasan pada metode kuadrat terkecil. Model regresi dengan LASSO dapat dinyatakan sebagai berikut:

$$\hat{\beta}^{lasso} = argmin\left\{\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{k} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{k}\left|\beta_j\right|\right\} \tag{28}$$

dengan λ adalah parameter penalti. Model pada Persamaan (28) dikenal sebagai norma L_1. Secara umum, nilai λ yang digunakan berkisar antara 0 hingga 1. Nilai λ yang lebih kecil akan mengurangi kekuatan penalti terhadap koefisien, sedangkan nilai λ yang lebih besar akan meningkatkan kekuatan penalti (Roozbeh dkk., 2023).

[illegible]ntil dengan *Least Absolute Shrinkage and Selection*

[illegible]ain dapat digunakan dalam regresi linier biasa, juga dapat [illegible]egresi lain seperti regresi kuantil. Secara umum, untuk [illegible]sir parameter metode LASSO dapat disederhanakan dengan meminimumkan fungsi *loss* dan menambahkan fungsi kendala LASSO sebagai berikut.

$$f(t) = L(t) + P(t) \tag{29}$$

dengan t menunjukkan parameter model, $f(t)$ sebagai fungsi tujuan dari model yang terpenalti, $L(t)$ menunjukkan fungsi dari model yang digunakan, sedangkan $P(t)$ menunjukkan fungsi kendala atau penalti yang ditambahkan dalam hal ini fungsi kendala LASSO.

Estimasi parameter dengan menggunakan penalti LASSO pada regresi kuantil dilakukan dengan meminimumkan fungsi berikut (Li & Zhu, 2008).

$$\underset{\beta(\tau)}{min} \sum_{j=1}^{k} \rho_\tau(\varepsilon_i) + \lambda \sum_{j=1}^{k} |\beta_j| \tag{30}$$

1.4.9 *Least Angle Regression*

Least Angle Regression (LAR) merupakan sebuah algoritma untuk menghasilkan model linier yang ditemukan Efron dkk. pada tahun 2002. Algoritma LAR membutuhkan p langkah untuk mendapatkan koefisien taksiran OLS. Pemodifikasian algoritma LAR dapat memberikan koefisien taksiran metode LASSO. Algoritma yang dimodifikasi ini memiliki langkah yang lebih efisien dibanding metode LASSO itu sendiri. Modifikasi Algoritma LAR memberikan jalan yang efisien dalam menyelesaikan regresi LASSO (Hastie dkk., 2011). Algoritma LAR di definisikan sebagai berikut.

1. Menetapkan semua koefisien parameter sama dengan nol.
2. Mencari variabel prediktor yang paling berkorelasi dengan ε.
3. Menduga koefisien nilai β_j untuk x_{ij} yang paling berkorelasi dengan ε.
4. Menggunakan GCV untuk menentukan nilai lambda dan titik knot yang optimal untuk setiap titik kuantil. Jika terdapat koefisien parameter yang memiliki nilai nol, maka koefisien parameter tersebut akan dikeluarkan dari model.

1.4.10 *Generalized Cros Validation*

Pemilihan model terbaik berdasarkan titik knot yang optimal merupakan hal penting yang berperan untuk mendapatkan estimator spline terbaik dalam analisis regresi nonparametrik dengan pendekatan *spline*. Titik knot merupakan titik perpotongan dimana terjadi perubahan pada pola perilaku fungsi. Salah satu metode yang dapat digunakan dalam pemilihan titik knot yang optimal adalah adalah metode *Generalized Cross Validation* (GCV) minimum (Marina & Budiantara, 2013). Titik knot yang optimal untuk model *spline* diperoleh dari nilai GCV paling kecil. Fungsi GCV ıonparametrik dituliskan pada Persamaan (31).

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

$_{12}, \dots, K_{1h}, \dots, K_{p1}, \dots, K_{pr}]$ adalah titik knot, $MSE(K) =$
n $\boldsymbol{A}[K] = \boldsymbol{X}[K](\boldsymbol{X}'[K]\boldsymbol{X}[K])^{-1}\boldsymbol{X}'[K]$ (Meimela, 2020).

1.4.11 Kemiskinan

Kemiskinan adalah ketidakmampuan ekonomi untuk memenuhi kebutuhan dasar makanan dan non-makanan yang diukur dari sisi pengeluaran. Ketersediaan data kemiskinan yang akurat dan tepat sasaran merupakan salah satu aspek penting untuk mendukung strategi penanggulangan kemiskinan. Pengukuran kemiskinan yang dapat dipercaya dapat menjadi instrumen tangguh bagi pengambil kebijakan dalam memfokuskan perhatian pada kondisi hidup orang miskin. Data kemiskinan yang baik dapat digunakan untuk mengevaluasi kebijakan pemerintah terhadap kemiskinan, membandingkan kemiskinan antar waktu dan daerah, serta menentukan target penduduk miskin dengan tujuan untuk memperbaiki kondisi yang dialami penduduk miskin.

Badan Pusat Statistik pertama kali melakukan penghitungan jumlah dan PPM pada tahun 1984. Nilai PPM diukur dengan menggunakan rumus *Head Count Index* (HCI-P0), yaitu persentase penduduk yang berada dibawah Garis Kemiskinan (GK). Penduduk yang memiliki rata-rata pengeluaran perkapita perbulan di bawah GK dikategorikan sebagai penduduk miskin (Badan Pusat Statistik, 2023). Berdasarkan berbagai penelitian terdahulu, terdapat sejumlah faktor yang diduga berpengaruh terhadap PPM, antara lain indeks pembangunan manusia yang dibentuk oleh indikator-indikator seperti rata-rata lama sekolah dan pengeluaran perkapita. Selain itu, faktor lain seperti angka harapan hidup dan prevalensi ketidakcukupan konsumsi pangan juga diperkirakan turut memengaruhi tingkat kemiskinan (Alviyadi, 2023; Hasanah dkk, 2021; Bizri dkk., 2024).

BAB II
METODE PENELITIAN

2.1 Sumber Data dan Variabel

Penelitian ini menggunakan data sekunder yang diperoleh dari website bps.go.id (Lampiran 1). Variabel penelitian yang digunakan terdiri dari satu variabel respon dan 5 variabel prediktor. Detail variabel yang digunakan pada penelitian ini dapat dilihat pada Tabel 3 berikut.

Tabel 3. Variabel Penelitian

Variabel	Keterangan	Definisi	Satuan
y	Persentase Penduduk Miskin (PPM)	PPM adalah ukuran proporsi penduduk yang hidup di bawah garis kemiskinan.	%
x_1	Indeks Pembangunan Manusia (IPM)	IPM adalah ukuran yang digunakan untuk menilai kualitas hidup dan pembangunan manusia di suatu kabupaten/kota.	Indeks
x_2	Rata-rata Lama Sekolah (RLS)	RLS adalah rata-rata jumlah tahun pendidikan formal yang telah dijalani oleh penduduk.	Tahun
x_3	Angka Harapan Hidup (AHH)	AAH adalah rata-rata perkiraan banyak tahun yang dapat ditempuh oleh seseorang sejak lahir.	Tahun
x_4	Pengeluaran Perkapita (PP)	PP adalah biaya konsumsi bulanan setiap anggota rumah tangga setelah disesuaikan dengan paritas daya beli.	Rupiah
x_5	Prevalensi Ketidakcukupan Konsumsi Pangan (PKKP)	Prevalensi ketidakcukupan konsumsi pangan adalah proporsi penduduk yang mengkonsumsi pangan di bawah standar energi untuk hidup sehat, aktif, dan produktif.	%

is

yang digunakan dalam penelitian ini menggunakan regresi an estimator *spline*. Adapun prosedur analisis yang digunakan n penelitian adalah sebagai berikut.

2.2.1 Estimasi Model Regresi Kuantil LASSO dengan Estimator *Spline*

1. Diberikan data $x_{i1}, x_{i2}, \dots, x_{ik}, y_i$, dengan $i = 1,2,\dots,n$ dan $j = 1,2..,k$, yang merupakan himpunan berpasangan dari variabel acak 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. Menyajikan model regresi kuantil dengan estimator *spline*.
$$y_i(\tau) = \sum_{j=1}^{k} f(x_{ij}) + \varepsilon_i(\tau)$$
3. Menyajikan model regresi kuantil dengan estimator *spline* dalam bentuk matriks.
$$\boldsymbol{y}(\tau) = \boldsymbol{X}[K]\boldsymbol{\beta}(\tau) + \boldsymbol{\varepsilon}(\tau)$$
4. Memperoleh estimasi parameter β.
$$\hat{\beta}(\tau) = \frac{min}{\beta(\tau)}\left\{\tau \sum_{i=1;\varepsilon_i \geq 0}^{n} |\varepsilon_i(\tau)| + (1-\tau) \sum_{i=1;\varepsilon_i < 0}^{n} |\varepsilon_i(\tau)|\right\}$$
5. Menambahkan regulasi LASSO kedalam estimasi parameter β sehingga diperoleh $\hat{\beta}(\tau)^{lasso}$
$$\hat{\beta}(\tau)^{lasso} = \frac{min}{\beta(\tau)} \sum_{j=1}^{k} \rho_\tau(\varepsilon_i) + \lambda \sum_{j=1}^{k} |\beta_j|$$
6. Meminimumkan $\hat{\beta}(\tau)^{lasso}$ dengan menggunakan algoritma simpleks dan modifikasi algoritma LAR.

2.2.2 Pemodelan Regresi Kuantil LASSO dengan Estimator *Spline*

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 bentuk pola hubungan dari variabel respon terhadap masing-masing variabel prediktor.
3. Mendeteksi pencilan pada data menggunakan *boxplot* menggunakan Persamaan (2) dan (3).
4. Mendeteksi multikolinearitas pada data dengan menggunakan Persamaan (6).
5. Melakukan standarisasi data pada variabel prediktor menggunakan *scalling* dan *centering* dengan persamaan:
$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{S_{x_{ij}}}$$

$\sqrt{\frac{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}{n-1}}$.

nilihan titik knot dengan tahapan sebagai berikut:

a. Menentukan jumlah titik knot yang mungkin berdasarkan perubahan pola perilaku data yang terjadi pada sub-sub interval. Jumlah titik knot yang digunakan pada variabel prediktor adalah satu hingga tiga titik knot.

b. Menentukan kemungkinan letak titik knot dalam interval data prediktor dapat dilakukan dengan cara membagi rentang data menjadi beberapa titik, misalnya r titik, sehingga diperoleh $x_{ij_{min}} < K_{1j} < K_{2j} < \cdots < K_{rj} < x_{ij_{max}}$.

7. Membentuk matriks *spline* X berdasarkan titik knot yang diperoleh.
8. Menentukan nilai lambda yang digunakan, yaitu $0{,}001 < \lambda < 1$.
9. Memperoleh nilai titik knot dan lambda optimum pada masing-masing kuantil, yaitu $\tau = 0{,}25,\ 0{,}5,$ dan $0{,}75$ pada satu hingga tiga titik knot berdasarkan Persamaan (31).
10. Memperoleh estimasi parameter model regresi kuantil LASSO dengan estimator *spline* pada data PPM di Pulau Sulawesi pada satu hingga tiga titik knot.
11. Melakukan perbandingan nilai GCV minimum dari satu hingga tiga titik knot untuk memperoleh model terbaik regresi kuantil LASSO dengan estimator *spline* pada masing-masing kuantil.
12. Melakukan transformasi balik estimasi parameter model ke parameter awal menggunakan persamaan:

$$K_{rj} = K_{rj}^{*} S_{x_{ij}} + \bar{x}_j;\ \ \beta_j = \frac{\beta_j^{*}}{S_{x_{ij}}};\ \beta_j = \frac{\beta_{(1+h)j}^{*}}{S_{x_{ij}}}$$

dan $\beta_0 = \bar{y} - \sum_{j=1}^{k} \beta_j \bar{x}_j + \sum_{h=1}^{r} \beta_{(1+h)j} \bar{x}_j$

13. Menginterpretasikan model yang diperoleh.