

BAB I PENDAHULUAN

1.1 Latar Belakang

Pengembangan statistika pada era globalisasi telah berkembang dengan cepat. Perkembangan ini ditandai dengan semakin luasnya penerapan metode-metode statistik dalam berbagai bidang, baik untuk eksplorasi data maupun pengambilan keputusan berbasis data. Salah satu metode statistika yang paling sering digunakan dalam penarikan kesimpulan suatu penelitian adalah analisis regresi. Analisis regresi adalah salah satu metode statistika yang digunakan untuk mendapatkan pola hubungan antara prediktor dengan respon (Montgomery et al., 2021). Dalam analisis regresi, terdapat tiga pendekatan utama: parametrik, nonparametrik, dan semiparametrik. Pendekatan parametrik digunakan saat bentuk kurva regresi telah diketahui, nonparametrik digunakan ketika bentuknya tidak diketahui atau bervariasi di setiap segmen, sementara semiparametrik digunakan jika sebagian bentuk kurva diketahui dan sebagian lainnya belum diketahui (Frost, 2020). Pendekatan regresi semiparametrik dan nonparametrik saling melengkapi, di mana regresi semiparametrik digunakan saat sebagian bentuk kurva regresi telah diketahui, sedangkan regresi nonparametrik digunakan ketika bentuk kurva sama sekali tidak diketahui dan tidak tersedia informasi sebelumnya mengenai pola data.

Regresi nonparametrik merupakan salah satu pendekatan yang digunakan untuk mengetahui pola hubungan antara variabel respon dan variabel prediktor yang kurva regresinya tidak diketahui dan tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola data (Čížek & Sadıkoğlu, 2020). Regresi nonparametrik memiliki fleksibilitas dalam mencari sendiri bentuk kurva regresinya tanpa dipengaruhi subjektivitas peneliti (Mohammed Sadek & Mohammed Ali, 2023). Untuk mengidentifikasi pola data yang kompleks dalam regresi nonparametrik, diperlukan estimator yang fleksibel dan adaptif terhadap struktur data yang tidak diketahui sebelumnya (Amato et al., 2020). Estimator yang sering digunakan karena memiliki fleksibilitas tersebut adalah spline, kernel, dan fourier.

Estimator dalam model regresi nonparametrik penting digunakan untuk melakukan estimasi terhadap kurva regresi, diantaranya adalah spline. Spline merupakan suatu potongan-potongan polinomial yang memiliki sifat tersegmen yang terbentuk pada titik-titik knot (Araveeporn, 2024). Fungsi spline lebih khusus dan memberikan fleksibilitas yang lebih tinggi daripada fungsi polinomial pada umumnya, mempunyai interpretasi statistik yang sederhana, dan memiliki penyajian visual yang sangat baik (Beccari et al., 2017; Ratnasari et al., 2021). Salah satu bentuk khusus



spline truncated, di mana fungsi spline dikonstruksi dengan menenun fungsi truncated pada setiap titik knot. Fungsi ini bernilai nol dan mengikuti polinomial setelah melewati knot, sehingga tidak menambahkan secara lokal pada bagian-bagian tertentu dari antara, 2022). Regresi nonparametrik dengan estimator *spline*

truncated efektif untuk menangani pola data tak beraturan karena memberikan fleksibilitas lokal melalui fungsi yang nol sebelum knot dan polinomial setelahnya.

Metode regresi nonparametrik yang dapat mengatasi pola data yang tidak beraturan menggunakan regresi nonparametrik dengan estimator *spline truncated*. Regresi nonparametrik *spline truncated* dapat memodelkan data pada pola yang berubah-ubah pada sub-sub interval tertentu (Octavanny et al., 2020; Suriaslan et al., 2025). *Spline truncated* memiliki fleksibilitas karena berbantuan titik knot untuk mengikuti pergerakan pola data (Arifin et al., 2023). Telah dikembangkan berbagai estimator *spline truncated* untuk meningkatkan fleksibilitas dalam pemodelan nonparametrik (Islamiyati, Kalondeng, et al., 2022; Islamiyati, Raupong, et al., 2022; Islamiyati, Sunusi, et al., 2020; Yulianti et al., 2024). Metode *spline truncated* dalam analisis pemodelan regresi, diperoleh tingkat akurasi sebesar 87,5%. Pencapaian tingkat akurasi tersebut mengindikasikan bahwa model yang dibangun mampu merepresentasikan pola data dengan baik (Islamiyati et al., 2023). Temuan ini memperkuat validitas penggunaan *spline truncated* sebagai pendekatan alternatif dalam analisis regresi, khususnya untuk data yang tidak mengikuti pola linier sederhana.

Pola linier yang tidak sederhana terjadi pada regresi nonparametrik. Namun, ketika data melibatkan dua respons yang saling berkorelasi, regresi nonparametrik konvensional tidak mampu memodelkannya secara efektif (Islamiyati, Anisa, et al., 2022). Salah satu bentuk data yang kerap menampilkan permasalahan semacam ini adalah data birespons longitudinal, yaitu data yang mencatat dua variabel respons secara berulang dalam kurun waktu tertentu pada subjek yang sama (Islamiyati, Fatmawati, et al., 2020). Data ini menunjukkan adanya korelasi antar waktu dan antar respons, sehingga diperlukan metode statistik yang mampu menangani keduanya secara bersamaan (Islamiyati et al., 2018). Sehingga dibutuhkan model fleksibel seperti regresi nonparametrik birespons untuk data longitudinal (Islamiyati, 2022).

Data longitudinal merupakan gabungan dari data *cross-section* dan *time-series*, yaitu data yang diperoleh dari subyek pengamatan yang saling independen dengan setiap subyek diamati secara berulang dalam kurun waktu dan antar pengamatan dalam subyek yang sama saling berkorelasi (Diggle et al., 2001). Data longitudinal diasumsikan bahwa setiap subyek yang diukur tidak saling berkorelasi, tetapi antar data pengamatan di dalam subyek yang sama saling berkorelasi (Islamiyati et al., 2017). Data longitudinal yang diasumsikan kurva regresi yang tidak diketahui bentuknya, direkomendasikan untuk menggunakan regresi nonparametrik (Octavanny et al., 2020). Dalam penerapannya, meskipun regresi nonparametrik mampu menangani ketidaktauan terhadap bentuk kurva regresi, tetap diperlukan



asumsi tertentu agar hasil analisis menjadi valid.

yang harus terpenuhi dalam analisis regresi nonparametrik adanya heteroskedastisitas. Untuk mengatasi masalah yaitu dengan menggunakan metode *Weighted Least Square* (20). Menurut (Basri, 2020; Dani et al., 2021; Jung & Kim, 2022) *Weighted Least Square* (WLS) merupakan metode yang memberikan nilai mengatasi masalah heteroskedastisitas. Metode WLS memiliki

kemampuan untuk mempertahankan sifat efisiensi estimatornya tanpa harus kehilangan sifat bias dan konsistensinya.

Penelitian yang menggunakan regresi nonparametrik birespon *spline truncated* yaitu Mustain Ramli, dkk (Ramli et al., 2020). Perbedaan penelitian Mustain Ramli dengan penelitian ini yaitu terletak pada variabel pembobotnya, dimana penelitian Mustain Ramli menggunakan multivariabel dan pembobot W yaitu matriks varians kovarians berukuran $nt \times nt$ dan berisi diagonal $(W_1, W_2, \dots, W_n)$, sedangkan penelitian ini menggunakan regresi nonparametrik birespon dan pembobot W yaitu matriks varians kovarians. Selain itu, Penelitian yang dilakukan oleh Atika Mayang Sari dkk (Mayang Sari et al., 2024) menggunakan pendekatan regresi nonparametrik *spline truncated* untuk memodelkan hubungan dua variabel respon secara simultan, yaitu inflasi dan Produk Domestik Regional Bruto (PDRB). Meskipun model yang digunakan bersifat birespon dan berbasis *Spline Truncated*, estimasi parameter dalam penelitian tersebut masih dilakukan dengan metode *Ordinary Least Squares* (OLS) tanpa mempertimbangkan pembobot terhadap varians residual. Berbeda dengan penelitian ini menggunakan metode *Weighted Least Squares* (WLS) dalam proses estimasi parameter model untuk menyesuaikan ketergantungan antara dua respon yang dianalisis secara simultan. Selanjutnya penelitian oleh Sifriyani dkk (Sifriyani et al., 2023) menggunakan data *cross-sectional* dari pasien diabetes mellitus, di mana setiap individu diobservasi sekali pada satu waktu pengukuran. Hal ini berbeda dengan penelitian ini yang menggunakan data longitudinal, di mana setiap individu diikuti dan diukur secara berulang dalam periode waktu tertentu.

Berdasarkan uraian tersebut, maka penulis akan membahas analisis data longitudinal dengan model regresi nonparametrik Birespon *Spline Truncated* Terboboti dan diaplikasikan pada data penimbangan balita. Data penimbangan balita yang digunakan adalah data yang berbentuk data longitudinal, yang pengukurannya dilakukan setiap bulan. Oleh karena itu, akan dilakukan estimasi model regresi nonparametrik birespon pada data longitudinal dan model pertumbuhan balita menggunakan regresi nonparametrik Birespon *Spline Truncated* Terboboti.

1.2 Rumusan Masalah

Berdasarkan latar belakang, disusun rumusan masalah sebagai berikut:

1. Bagaimana bentuk penaksir parameter model regresi nonparametrik birespon pada data longitudinal dengan *Spline Truncated Terboboti*?
2. Bagaimana model pertumbuhan balita berdasarkan regresi nonparametrik Birespon *Spline Truncated* Terboboti?



an

pada perumusan masalah dalam subbab sebelumnya, tujuan lam penelitian ini adalah:

anaksir parameter model regresi nonparametrik birespon pada dengan *Spline Truncated* Terboboti.

pentuk model pertumbuhan balita berdasarkan regresi irespon *Spline Truncated* Terboboti.

1.4 Batasan Masalah

Penelitian ini membatasi masalah sebagai berikut:

1. Analisis data longitudinal dibatasi pada data penimbangan balita usia 1 – 59 bulan tahun 2021.
2. Estimasi parameter menggunakan metode *Weighted Least Square* (WLS).
3. Titik knot yang digunakan adalah 12, 24, 36, dan 48

1.5 Manfaat Penelitian

Hasil penulisan ini dapat memberikan manfaat yaitu:

1. Hasil penelitian ini mampu menambah wawasan bagi pembaca serta dapat menjadi dasar penelitian yang dapat dikembangkan ke tingkat lebih lanjut terkait analisis data longitudinal dengan model regresi nonparametrik birespon *Spline Truncated* Terboboti.
2. Penelitian ini diharapkan dapat membuat peneliti mampu menganalisis data longitudinal dengan model regresi nonparametrik birespon *Spline Truncated* Terboboti.

1.6 Teori

1.6.1 Analisis Regresi

Analisis regresi merupakan metode statistika yang digunakan untuk pemodelan antara variabel respon dan variabel prediktor (Fahrmeir et al., 2021). Analisis regresi adalah kajian terhadap hubungan satu variabel yang disebut sebagai variabel yang diterangkan dengan satu atau dua variabel yang menerangkan. Variabel yang diterangkan selanjutnya disebut sebagai variabel respon, sedangkan variabel yang menerangkan biasa disebut variabel bebas (Arkes, 2023). Model regresi digunakan untuk membuat taksiran mengenai variabel dependen disebut persamaan regresi estimasi, yaitu suatu formula matematis yang menunjukkan hubungan keterkaitan antara satu atau beberapa variabel yang nilainya sudah diketahui dengan suatu variabel yang nilainya belum diketahui (Roustaei, 2024).

Dalam perkembangannya regresi linier terbagi dua, yaitu regresi linier sederhana dan regresi linier berganda. Regresi linier sederhana merupakan pemodelan regresi yang digunakan untuk menggambarkan hubungan antara variabel bebas dengan variabel terikat. Bentuk umum persamaan regresi linier sederhana adalah sebagai berikut (Montgomery et al., 2012):

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, 2, \dots, n \quad (1.1)$$

Model regresi linier berganda merupakan pengembangan dari model regresi linier sederhana. Jika pada model regresi linier sederhana hanya terdiri dari satu variabel terikat, maka pada regresi linier berganda jumlah variabel terikat lebih dari satu dan satu variabel terikat. Dengan bertambahnya variabel terikat maka bentuk umum dari persamaan regresi linier berganda akan lebih variabel bebas adalah sebagai berikut (Montgomery et al., 2012):



$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_f x_n + e \quad (1.2)$$

Apabila menggunakan data sampel, maka dapat menggunakan model estimasi dari Persamaan (1.2) adalah sebagai berikut (Montgomery et al., 2012):

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_f x_{im} + e_i, \quad i = 1, 2, \dots, n \quad (1.3)$$

Berdasarkan Persamaan (1.3) maka bentuk matriks dari model regresi linier berganda dapat dituliskan sebagai berikut (Montgomery et al., 2012):

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_f \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \quad (1.4)$$

atau dapat ditulis dengan $\vec{y} = X\vec{\beta} + \vec{e}$, dimana (Montgomery et al., 2012):

- $\vec{y}$: vektor variabel respon
- X : matriks variabel prediktor
- $\vec{\beta}$: vektor parameter model
- $\vec{e}$: vektor random error

1.6.2 Regresi Nonparametrik

Regresi nonparametrik merupakan salah satu pendekatan yang digunakan untuk mengetahui pola hubungan antara variabel respon dan variabel prediktor yang kurva regresinya tidak diketahui dan tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola data (Chatterjee et al., 2024). Secara visual bentuk pola yang diberikan oleh variabel prediktor dan variabel respon tidak mempunyai pola yang pasti. Secara umum regresi nonparametrik memiliki bentuk fungsi sebagai berikut (Halconrui & Marie, 2021).

$$y_i = f(x_i) + e_i, \quad i = 1, 2, \dots, n \quad (1.5)$$

dengan:

- y_i : variabel respon ke- i
- $f(x_i)$: fungsi regresi nonparametrik yang tidak diketahui
- x_i : variabel prediktor ke- i
- e_i : *error* yang diasumsikan berdistribusi normal, identik, dan independen dengan *mean* nol dan variansi σ^2

Model regresi nonparametrik dapat berbentuk fungsi apa saja, baik linear ataupun nonlinear. Regresi nonparametrik sangat memperhatikan fleksibilitas, dan hanya diasumsikan bentuk fungsinya mulus (Čížek & Sadıkoğlu, 2020). Regresi nonparametrik merupakan salah satu model regresi selain regresi parametrik. Regresi nonparametrik sangat tepat digunakan untuk mendeteksi adanya sifat nonlinieritas dan hubungan nonlinier dalam model. Nonlinier tersebut dapat terlihat



berbentuk. Kurva regresi nonparametrik yang terbentuk tidak tertentu atau bisa dikatakan bahwa bentuk kurvanya tidak tertentu (Chatterjee et al., 2022).

Pendekatan dalam melakukan estimasi kurva regresi yaitu menggunakan pendekatan regresi nonparametrik. Pendekatan regresi nonparametrik akan digunakan apabila bentuk kurva regresi atau pola data tidak diketahui.

Regresi nonparametrik memiliki fleksibilitas dalam mencari sendiri bentuk kurva regresinya tanpa dipengaruhi subjektivitas peneliti (Araveeporn, 2024).

1.6.3 Fungsi *Spline Truncated*

Regresi spline adalah regresi dimana kurva regresinya didekati oleh fungsi spline. Fungsi spline adalah fungsi polinomial yang lebih khusus dan memberikan fleksibilitas yang lebih tinggi daripada fungsi polinomial pada umumnya (Acharjee & Das, 2022). Menurut (A. Huang & Y. Huang, 2024), spline memiliki beberapa kelebihan yaitu:

1. Spline memiliki interpretasi statistik dan interpretasi visual yang sangat baik
2. Spline memiliki fleksibilitas yang tinggi
3. Spline memiliki kemampuan yang sangat baik untuk menangani data yang perilakunya berubah-ubah
4. Spline mempunyai kemampuan yang sangat baik untuk digeneralisasikan pada pemodelan statistika yang kompleks dan rumit.

Fungsi regresi *spline truncated* dapat dituliskan sebagai berikut (Steppeler & Li, 2022):

$$f(x_i) = \sum_{j=0}^u \beta_j x_i^j + \sum_{l=1}^p \beta_{(u+l)} (x_i - k_l)_+^u \quad (1.6)$$

dengan $(x_i - k_l)_+^u = \begin{cases} (x_i - k_l)_+^u; & x_i \geq k_l \\ 0; & x_i < k_l \end{cases}$

dimana:

$f(x_i)$: fungsi regresi spline

$k_1, k_2, \dots, k_l$: titik knot

x_i : variabel prediktor

β_j : parameter model

$\beta_{(u+l)}$: slope pada peubah x truncated knot ke- k pada spline berorde u (konstanta riil)

k_l : titik knot ke- l

p : banyaknya titik knot

Regresi spline adalah pendekatan ke arah plot data dengan tetap memperhitungkan kemulusan kurva. Spline merupakan suatu polinomial dimana segmen-segmen polinomial yang berbeda ditautkan pada knot $k_1, k_2, k_3, \dots, k_l$ dan kontinu sehingga bersifat fleksibel dibandingkan polinomial biasa. Spline merupakan model polinomial yang tersegmen (*piecewise polynomial*) dengan sifat kontinu pada knot-knotnya. Spline mempunyai titik knot yang merupakan titik perpaduan bersama arah perilaku kurva. Spline dapat menyesuaikan diri secara otomatis terhadap data yang ada, sehingga didapatkan hasil yang mendekati data sebenarnya (et al., 2025).

Hasil regresi spline (1.6) disubstitusi ke Persamaan (1.5) maka diperoleh model regresi nonparametrik *spline truncated* sebagai berikut:



$$y_i = \sum_{j=0}^u \beta_j x_i^j + \sum_{l=1}^p \beta_{(u+l)} (x_i - k_l)_+^u + e_i, \quad i = 1, 2, \dots, n \quad (1.7)$$

dengan:

- y_i : variabel respon pada pengamatan ke- i
 β_j : parameter polinomial pada pengamatan ke- j
 x_i : variabel prediktor pada pengamatan ke- i
 x_i^j : variabel prediktor pada pengamatan ke- i berorde j
 $\beta_{(u+l)}$: parameter truncated bernilai riil pada pengamatan ke- $(u + l)$
 k_l : titik knot ke- l
 e_i : *error* pada pengamatan ke- i
 p : banyaknya titik knot

1.6.4 Regresi Nonparametrik Birespon

Regresi birespon adalah salah satu model regresi yang memiliki variabel respon yang terdiri dari dua dan diantara variabel-variabel tersebut terdapat korelasi atau hubungan yang kuat, baik secara logika maupun matematis (Turner, 2021). Pendekatan nonparametrik digunakan apabila bentuk kurva regresi birespon tidak diketahui. Secara umum, model regresi nonparametrik birespon spline dapat dituliskan sebagai berikut (DiPasquale & Sottile, 2020):

$$y_i^{(r)} = \sum_{j=0}^u \beta_j^{(r)} x_i^j + \sum_{l=1}^p \beta_{(u+l)}^{(r)} (x_i - k_l)_+^u + e_i, \quad i = 1, 2, \dots, n; \quad r = 1, 2 \quad (1.8)$$

dengan:

- $y_i^{(r)}$: variabel respon r pada pengamatan ke- i
 $\beta_j^{(r)}$: parameter polinomial pada pengamatan ke- j

Untuk melihat tingkat korelasi atau kekuatan hubungan, maka disajikan tabel tingkat korelasi antar variabel berikut (Sugiyono, 2013).

Tabel 1.1. Tingkat Korelasi dan Kekuatan Hubungan

Nilai Korelasi (R)	Tingkat Hubungan
0	Tidak Berkorelasi
0,01 – 0,20	Sangat Rendah
0,21 – 0,40	Rendah
0,41 – 0,60	Agak Rendah
0,61 – 0,80	Cukup
0,81 – 0,99	Tinggi
	Sangat Tinggi



Idinal

Idinal merupakan data yang pengamatannya dilakukan secara tiap subjek pada beberapa subjek. Pada data longitudinal-subjek saling independen satu sama lain, akan tetapi antar

pengamatan didalam subjek saling dependen sehingga terdapat korelasi (Li et al., 2022). Model data longitudinal mengandung informasi yang lengkap dari kemungkinan-kemungkinan individu. Oleh karena itu, data longitudinal dapat digunakan untuk memprediksi periode baru untuk setiap subyek pengamatan (Cauchi et al., 2025).

Tujuan dari analisis data longitudinal adalah untuk mempelajari atau mengetahui bagaimana perubahan subjek yang diamati dari waktu ke waktu. Pada data longitudinal apabila terdiri lebih dari satu respon, maka terdapat korelasi dari setiap respon pada subjek yang sama. Bentuk struktur data longitudinal birespon adalah sebagai berikut (Ouko et al., 2025).

Tabel 1.2. Struktur Data Longitudinal Birespon

Subjek ke- i	Pengamatan ke- t	Prediktor (x_{it})	Respon ke-1 ($y_{it}^{(1)}$)	Respon ke-2 ($y_{it}^{(2)}$)
Subjek ke-1	1	x_{11}	$y_{11}^{(1)}$	$y_{11}^{(2)}$
	2	x_{12}	$y_{12}^{(1)}$	$y_{12}^{(2)}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	m	x_{1m}	$y_{1m}^{(1)}$	$y_{1m}^{(2)}$
Subjek ke-2	1	x_{21}	$y_{21}^{(1)}$	$y_{21}^{(2)}$
	2	x_{22}	$y_{22}^{(1)}$	$y_{22}^{(2)}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	m	x_{2m}	$y_{2m}^{(1)}$	$y_{2m}^{(2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Subjek ke- n	1	x_{n1}	$y_{n1}^{(1)}$	$y_{n1}^{(2)}$
	2	x_{n2}	$y_{n2}^{(1)}$	$y_{n2}^{(2)}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	m	x_{nm}	$y_{nm}^{(1)}$	$y_{nm}^{(2)}$

dengan $i = 1, 2, \dots, n$ dan $t = 1, 2, \dots, m$.

Keuntungan dari analisis data longitudinal adalah tersedianya informasi perubahan nilai variabel respon terhadap waktu pada setiap subjek. Kelemahan analisis data longitudinal menurut (Lee et al., 2021) yaitu:

1. Respon antar satuan waktu berkorelasi, sehingga membutuhkan metode yang lebih canggih untuk menganalisis data.
2. Adanya data yang hilang yang sering disebabkan oleh berkurangnya jumlah subjek dengan bertambahnya waktu pengamatan.



parametrik Birespon untuk Data Longitudinal dengan *line Truncated*

si nonparametrik birespon untuk data longitudinal dinyatakan (Niyati et al., 2017):

$$\hat{y}_{it}^{(r)} = \sum_{j=0}^u \beta_j^{(r)} x_{it}^j + \sum_{l=1}^p \beta_{(u+l)}^{(r)} (x_{it} - k_l)_+^u \quad (1.9)$$

$$i = 1, 2, \dots, n; \quad t = 1, 2, \dots, m; \quad r = 1, 2;$$

dengan:

$\hat{y}_{it}^{(r)}$: variabel respon r pada subjek ke- i dalam waktu ke- t

x_{it}^j : variabel prediktor pada subjek ke- i dalam waktu pengamatan ke- t
berorde j

$\beta_j^{(r)}$: koefisien parameter regresi fungsi polinomial

$\beta_{(u+l)}^{(r)}$: koefisien parameter regresi fungsi truncated

1.6.7 Weighted Least Square (WLS)

Metode *Weighted Least Square* (WLS) merupakan salah satu metode yang dapat menyelesaikan masalah heteroskedastisitas. Heteroskedastisitas merupakan peristiwa ketidaksamaan varians dari vektor *error*, yakni $var(\varepsilon_i)$ tidak sama untuk setiap i (Nisa et al., 2020). Untuk mendeteksi ada atau tidaknya heteroskedastisitas dalam suatu model regresi linear berganda, maka dilakukan dengan melihat grafik *scatterplot* atau nilai prediksi variabel terikat yang disebut SRESID dengan *residual error* ZPRED. Pengujian heteroskedastisitas dapat dilakukan dengan menggunakan grafik *scatterplot* atau dari nilai prediksi variabel terikat yaitu SRESID dengan residual error yaitu ZPRED. Dasar pengambilan keputusan sebagai berikut:

1. Jika terdapat pola tertentu seperti titik-titik yang ada membentuk pola tertentu yang teratur (bergelombang, melebar kemudian menyempit), maka mengindikasikan bahwa terjadi heteroskedastisitas.
2. Jika tidak terdapat pola yang jelas, maupun titik-titik yang menyebar di atas dan di bawah angka 0 pada sumbu y , maka tidak terjadi heteroskedastisitas.

Solusi jika model menyalahi asumsi heteroskedastisitas adalah dengan mentransformasikan ke dalam bentuk logaritma, yang hanya dapat dilakukan jika semua data bernilai positif atau dapat juga dilakukan dengan membagi semua variabel dengan variabel yang mengalami gangguan heteroskedastisitas (Meiryani, 2021).

Metode WLS memiliki kemampuan untuk mempertahankan sifat efisiensi estimatornya tanpa harus kehilangan sifat tak bias dan konsistensinya. Metode WLS sama halnya seperti metode *Ordinary Least Square* (OLS) dengan meminimumkan jumlah sisaan, hanya saja pada metode WLS dilakukan pembobotan suatu faktor baru menggunakan metode OLS terhadap data yang telah (Setyaningsih et al., 2022). Metode WLS dilakukan dengan meminimumkan erboboti yang dapat dirumuskan sebagai berikut (Setyaningsih



$$e^T W e = (\vec{y} - X\vec{\beta})^T W (\vec{y} - X\vec{\beta}) \quad (1.10)$$

Parameter $\vec{\beta}$ adalah estimator WLS dan pembobot W merupakan matriks varians kovarians. Pembobot W merupakan gabungan dari varians setiap peubah dan kovarians dua peubah yang berbeda atau dapat ditulis dalam bentuk matriks sebagai berikut (Verny Yolanda et al., 2022):

$$W = \begin{bmatrix} \text{var}(e^{(1)}) & \text{cov}(e^{(1)}, e^{(2)}) \\ \text{cov}(e^{(1)}, e^{(2)}) & \text{var}(e^{(2)}) \end{bmatrix} \quad (1.11)$$

$$\text{dengan, } \text{var}(e^{(1)}) = \frac{1}{N-df^{(1)}} \sum_{i=1}^N (y_i^{(1)} - \hat{y}_i^{(1)})^2$$

$$\text{var}(e^{(2)}) = \frac{1}{N-df^{(2)}} \sum_{i=1}^N (y_i^{(2)} - \hat{y}_i^{(2)})^2$$

$$\text{cov}(e^{(1)}, e^{(2)}) = \frac{1}{N-df} \sum_{i=1}^N (y_i^{(1)} - \hat{y}_i^{(1)}) (y_i^{(2)} - \hat{y}_i^{(2)})$$

dimana:

$e^{(1)}$: *error* untuk respon 1

$e^{(2)}$: *error* untuk respon 2

$y_i^{(1)}$: nilai aktual variabel respon ke- 1 pada pengamatan i

$y_i^{(2)}$: nilai aktual variabel respon ke- 2 pada pengamatan i

$\hat{y}_i^{(1)}$: nilai prediksi variabel respon ke- 1 pada pengamatan i

$\hat{y}_i^{(2)}$: nilai prediksi variabel respon ke- 2 pada pengamatan i

i : observasi/pengamatan, $i = 1, 2, \dots, N$

N : banyaknya pengamatan secara keseluruhan

df : derajat bebas (jumlah parameter model)

Persamaan (1.10) dapat diturunkan terhadap β maka sedemikian sehingga dapat diperoleh Persamaan sebagai berikut:

$$\hat{\beta} = (X^T W X)^{-1} X^T W y \quad (1.12)$$

Persamaan (1.12) merupakan estimator dari *Weighted Least Square* (WLS) (Ayanlowo E. A et al., 2024).

1.6.8 Generalized Cross Validation (GCV)

Metode GCV merupakan pengembangan dari metode *Cross Validation* (CV) dimana perbedaannya terletak pada faktor-faktor yang membagi residual. Pada metode GCV faktor merupakan nilai rata-rata dari faktor-faktor tersebut. Nilai GCV kemudian diperoleh dengan menjumlahkan residual-residual kuadrat yang telah terkoreksi dengan kuadrat dari faktor-faktor ini (Fitriyani, 2014). Metode GCV merupakan metode yang paling unggul dari beberapa metode yang dapat digunakan untuk menentukan parameter smoothing karena aspek perhitungannya lebih efisien.

Frank (1988), *Generalized Cross Validation* (GCV) merupakan *Validation* (CV). Fungsi GCV didefinisikan sebagai (Hastie et



$$GCV = \frac{1}{N} \cdot \frac{\sum_{i=1}^N (y_i^{(r)} - \hat{y}_i^{(r)})^2}{\left[\frac{1}{N} \cdot Tr(\mathbf{I} - \mathbf{G}) \right]^2}$$

$$GCV = \frac{MSE}{\left[\frac{1}{N} \cdot Tr(\mathbf{I} - \mathbf{G}) \right]^2} \quad (1.13)$$

dengan, $\frac{1}{N} \cdot Tr(\mathbf{I} - \mathbf{G}) < N$ dan $\mathbf{G} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$.

Dimana:

GCV : vektor yang memuat nilai GCV dari titik-titik knot

MSE : *Mean Standar Error*

$y_i^{(r)}$: nilai aktual respon ke- r pada pengamatan i

$\hat{y}_i^{(r)}$: nilai prediksi respon ke- r pada pengamatan i

r : respon, $r = 1, 2$

i : observasi/pengamatan, $i = 1, 2, \dots, N$

N : banyaknya pengamatan secara keseluruhan

Tr : trace, jumlah bilangan pada diagonal utama matriks

$\mathbf{I}$: matriks identitas $N \times N$

$\mathbf{G}$: matriks yang memuat titik knot

1.6.9 Pertumbuhan Balita

Masa balita, yakni usia 0 hingga 5 tahun, merupakan tahap krusial dalam siklus kehidupan manusia karena pada periode ini berlangsung perkembangan pesat secara fisik, kognitif, dan perilaku. Anak pada rentang usia 1 hingga 3 tahun dikenal sebagai batita, sementara anak usia 3 hingga 5 tahun disebut sebagai anak pra-sekolah. Pada tahap batita, anak masih sangat bergantung pada orang tua dalam menjalani berbagai aktivitas sehari-hari. Tahapan balita menjadi fondasi penting bagi pertumbuhan dan perkembangan anak secara optimal di fase kehidupan selanjutnya. Pertumbuhan berat dan tinggi badan menunjukkan peningkatan yang signifikan, khususnya pada awal masa balita (Lubis et al., 2024).

Masa balita adalah masa pembentukan dan perkembangan manusia, usia ini merupakan usia yang rawan karena balita sangat peka terhadap gangguan pertumbuhan serta bahaya yang menyertainya. Masa balita disebut juga sebagai masa keemasan, dimana terbentuk dasar-dasar kemampuan keindraan, berfikir, berbicara serta pertumbuhan mental intelektual yang intensif dan awal pertumbuhan moral (Schoenbuchner et al., 2019). Menurut (Wibaek et al., 2019) pertumbuhan (*growth*) berkaitan dengan masalah perubahan ukuran, besar, jumlah, atau dimensi



an maupun individu. Pertumbuhan bersifat kuantitatif sehingga satuan berat (gram, kilogram), satuan Panjang (centimeter, dan keseimbangan metabolik.

merupakan salah satu perubahan secara fisik yang terjadi. Pertumbuhan adalah bertambahnya ukuran dan jumlah sel elular, berarti bertambahnya ukuran fisik dan struktur tubuh

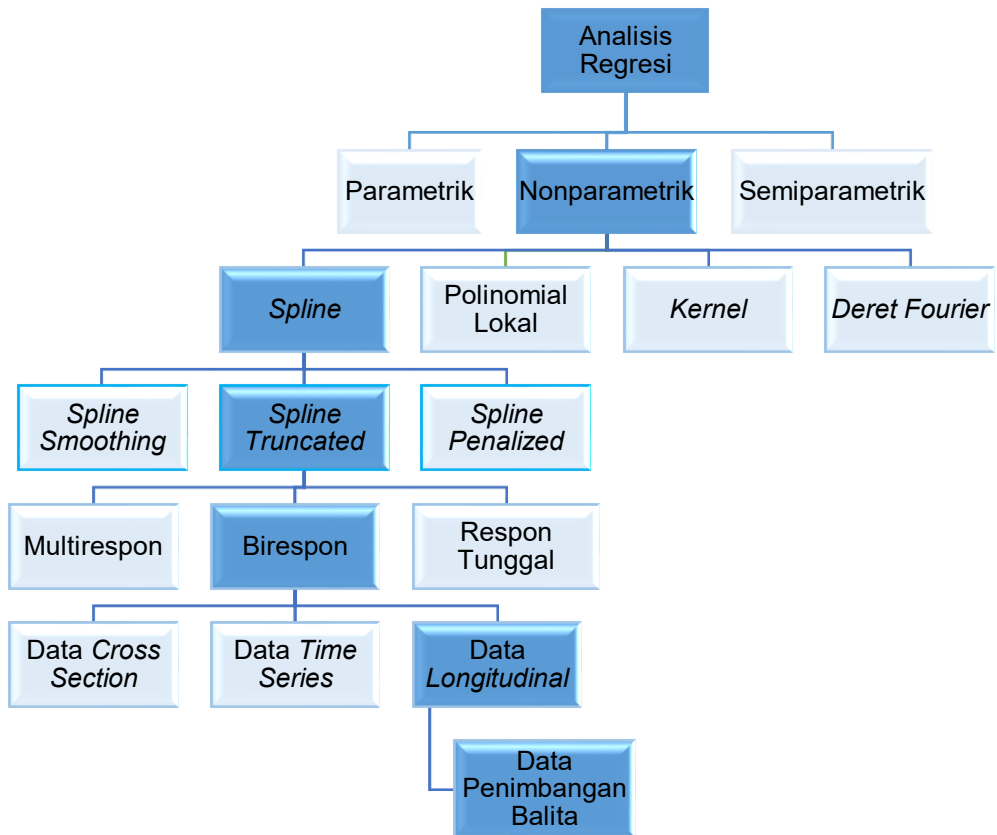
sebagian atau keseluruhan. Pertumbuhan berkaitan dengan perubahan dalam jumlah, ukuran dan fungsi tingkat sel, organ, maupun individu (Winarni et al., 2019). Pertumbuhan dapat dilihat dari berat badan, tinggi badan, dan lingkar kepala (Prastiwi, 2019). Pertumbuhan (*growth*) adalah suatu ukuran kematangan fisik. Hal ini ditandai dengan peningkatan ukuran tubuh dan organ-organ yang berbeda (Nurkholidah, 2020).

1.6.10 Kerangka Konseptual

Analisis regresi merupakan salah satu metode statistika yang digunakan untuk mendapatkan pola hubungan antara prediktor dengan respon. Teknik pendekatan yang dapat digunakan dalam analisis regresi yaitu regresi parametrik, nonparametrik, dan semiparametrik. Regresi nonparametrik dapat digunakan sebagai pendekatan pada pola yang tidak atau belum diketahui bentuknya. Teknik untuk mengestimasi kurva regresi dalam regresi nonparametrik, diantaranya yaitu regresi spline, polinomial lokal, kernel, deret fourier dan lain-lain. Salah satu estimator dalam model regresi nonparametrik yang sangat sering digunakan untuk melakukan estimasi terhadap kurva regresi adalah spline. Spline merupakan suatu potongan-potongan polinomial yang memiliki sifat tersegmen yang terbentuk pada titik-titik knot. Salah satu metode regresi nonparametrik yang dapat mengatasi pola data yang tidak beraturan adalah dengan menggunakan regresi nonparametrik *spline truncated*. Regresi nonparametrik *spline truncated* dapat memodelkan data pada pola data yang berubah-ubah pada sub-sub interval tertentu.

Analisis regresi birespon merupakan regresi yang melibatkan dua variabel respon dan diantara variabel respon tersebut terdapat korelasi atau hubungan, baik secara logika maupun matematis. Jenis data yang digunakan dalam analisis regresi yaitu *Cross Section*, *time series*, dan longitudinal. Data longitudinal adalah data yang pengamatannya dilakukan secara berulang untuk setiap subjek pada beberapa subjek. Data longitudinal diasumsikan bahwa setiap subjek yang diukur tidak saling berkorelasi, tetapi antar data pengamatan di dalam subyek yang sama saling berkorelasi. Salah satu contoh data longitudinal yaitu data penimbangan balita. Penimbangan balita adalah kegiatan penimbangan berat badan dan pengukuran panjang atau tinggi badan pada balita usia 0-59 bulan yang dilakukan setiap bulan di Posyandu secara serentak.





Gambar 1.1. Kerangka Konseptual



BAB II METODOLOGI PENELITIAN

2.1 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder tentang usia (bulan), berat badan (kilogram), dan tinggi badan (centimeter) balita yang didapatkan dari hasil panimbangan Posyandu yang ada di Desa Bonto Sunggu Kec. Bontonompo Selatan, Kab. Gowa, Provinsi Sulawesi Selatan dari bulan Januari hingga Desember tahun 2021.

2.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri atas dua jenis, yaitu prediktor dan respon. Prediktor terdiri atas satu variabel dan respon terdiri atas dua variabel (bi-respon).

1. Variabel prediktor (x)
 x : usia (bulan)
2. Variabel respon (y)
 y_1 : berat badan (kg)
 y_2 : tinggi badan (cm)

2.3 Definisi Operasional Variabel

Definisi operasional variabel yang digunakan pada penelitian ini adalah sebagai berikut:

1. Variabel prediktor (x) adalah usia seorang balita dalam satuan waktu bulan.
2. Variabel respon pertama ($y^{(1)}$) adalah hasil pengukuran berat badan balita menggunakan alat ukur timbangan digital.
3. Variabel respon kedua ($y^{(2)}$) adalah hasil pengukuran tinggi badan balita yang diukur dengan microtoice bagi balita yang belum bisa berjalan, sedangkan pengukuran tinggi badan balita yang sudah bisa berjalan, menggunakan alat ukur langbord yang ditempel di dinding.

2.4 Metode Analisis

Model regresi yang digunakan yaitu model regresi nonparametrik bi-respon *spline truncated*. Tujuan penelitian ini adalah mendapatkan bentuk estimator *Spline Truncated* Terboboti dengan menentukan titik knot untuk mengestimasi kurva regresi nonparametrik bi-respon *spline truncated* pada data longitudinal.



langkah untuk mendapatkan estimator dari model *Spline Truncated* longitudinal adalah sebagai berikut:

longitudinal (x_{it}, y_{it})

regresi nonparametrik untuk data longitudinal

e_{it} , $i = 1, 2, \dots, n$; $t = 1, 2, \dots, m$; $r = 1, 2$

3. Mendekati $f^{(r)}(x_{it})$ dengan *spline truncated*

$$\hat{f}^{(r)}(x_{it}) = \sum_{j=0}^u \beta_j^{(r)} x_{it}^j + \sum_{l=1}^p \beta_{(u+l)}^{(r)} (x_{it} - k_l)_+^u$$

4. Menyatakan bentuk regresi spline untuk data longitudinal sebagai berikut:

$$\hat{y}_{it}^{(r)} = \sum_{j=0}^u \beta_j^{(r)} x_{it}^j + \sum_{l=1}^p \beta_{(u+l)}^{(r)} (x_{it} - k_l)_+^u$$

$$i = 1, 2, \dots, n; \quad t = 1, 2, \dots, m; \quad r = 1, 2;$$

5. Menentukan matriks pembobot W menggunakan matriks varians dan kovarians

$$W = \begin{bmatrix} \text{var}(e^{(1)}) & \text{cov}(e^{(1)}, e^{(2)}) \\ \text{cov}(e^{(1)}, e^{(2)}) & \text{var}(e^{(2)}) \end{bmatrix}$$

6. Melakukan estimasi parameter dengan metode *Weighted Least Square* (WLS)
7. Mendapatkan kurva estimasi regresi

Adapun Langkah-langkah untuk memodelkan pertumbuhan balita menggunakan regresi nonparametrik birespon berdasarkan estimator *Spline Truncated* Terboboti yaitu sebagai berikut:

1. Membuat statistik deskriptif dan *scatter plot* pada data pertumbuhan balita untuk mengetahui informasi dan bentuk pola hubungan antara usia terhadap berat dan tinggi balita
2. Mengestimasi model spline menggunakan titik knot 12, 24, 36, 48
3. Memodelkan data penimbangan balita menggunakan model regresi nonparametrik *spline truncated* pada data longitudinal
4. Melakukan estimasi parameter untuk model regresi nonparametrik birespon dengan penaksir *Spline Truncated* pada data longitudinal melalui metode *Weighted Least Square* (WLS)
5. Memodelkan persamaan menggunakan titik knot yang ditentukan
6. Menginterpretasi model dan menarik kesimpulan



2.5 Diagram Alir

