

BAB I PENDAHULUAN

1.1 Latar Belakang

Salah satu metode dalam ilmu statistika yang dapat digunakan untuk mendeteksi pola hubungan antara satu atau lebih variabel adalah analisis regresi (Budiantara.2019). variabel yang digunakan pada analisis regresi adalah variabel respon dan variabel prediktor. Terdapat beberapa pendekatan dalam pemodelan menggunakan analisis regresi, yaitu regresi parametrik, regresi nonparametrik, dan regresi semiparametrik. (Dani and Adrianingsih 2021). Pendekatan menggunakan regresi semiparametrik merupakan kombinasi antara pendekatan regresi parametrik dan regresi nonparametrik. Regresi parametrik yaitu bentuk regresi yang dimana bentuk pola hubungan antara variabel respon dan prediktor diketahui, sebagai contoh linear, kuadratik, kubik, atau yang lainnya. Sedangkan regresi nonparametrik adalah bentuk regresi yang dimana pola hubungannya tidak diketahui (Budiantara.2019).

Model regresi semiparametrik digunakan apabila variabel respon memiliki pola hubungan yang diketahui dengan salah satu atau beberapa variabel prediktor, akan tetapi dengan variabel prediktor lainnya tidak diketahui dengan pasti pola hubungannya. Kurva regresi nonparametrik diasumsikan mulus (*smooth*) yaitu termuat dalam suatu fungsi tertentu. Pendekatan regresi nonparametrik memiliki fleksibilitas yang sangat tinggi, sehingga diharapkan data mampu menyesuaikan sendiri bentuk estimasi kurva regresi tanpa terpengaruhi oleh subyektifitas peneliti (L.Eubank 1999). Pada pendekatan regresi nonparametrik, terdapat beberapa estimator yang dikembangkan oleh para peneliti, diantaranya Kernel (Ampa, Nyoman Budiantara, and Zain 2021; Ratnasari et al. 2016; Selingerova, Katina, and Horova 2021), *spline truncated* (Budiantara, 2009; Islamiyati, 2018; Lestari et al., 2018 Chamidah, 2024;), *deret fourier* (Kustianingsih et al. 2021; Saputro, Sukmayanti, and Widyaningsih 2019), *local polynomial* (Fibriyani, Chamidah, and Saifudin 2024; Suparti and Prahutama 2017). Pada bahasan ini, model regresi semiparametri yang akan dibahas adalah kombinasi komponen parametrik (regresi linear) dan komponen nonparametrik (*spline truncated*).

Regresi linear merupakan salah satu metode regresi yang sangat familiar dan sangat sering kita jumpai dalam prakteknya. Regresi linear digunakan untuk menggambarkan hubungan linear antara variabel prediktor dengan variabel respon (Draper & Smith, 1992). Sedangkan untuk regresi nonparametrik menggunakan fungsi *spline truncated*, *Spline truncated* merupakan potongan polinomial yang sifatnya tersegmen dan kontinu. Kelebihan dari *spline truncated* yaitu model ini



n *spline truncated* memiliki titik-titik knot yang menunjukkan pola data. Titik knot merupakan titik perpaduan yang menandai fungsi pada interval yang berbeda. Pemilihan titik knot yang menentukan model *spline* terbaik. Salah satu metode untuk optimal adalah dengan mencari nilai Generalized Cross σ minimum (Islamiyati et al. 2024).

Pada umumnya, metode estimasi yang sering digunakan adalah Ordinary Least Squares (OLS). Mengingat ada beberapa kelemahan dari metode OLS, maka peneliti tertarik untuk mengkaji bagaimana bentuk estimasi model regresi semiparametrik spline truncated menggunakan metode *Maximum Likelihood Estimation* (MLE). Metode MLE adalah metode estimasi parameter yang dapat digunakan ketika diketahui distribusi dari error-nya diketahui (Arisandi & Purhadi, 2014). Distribusi dari error diasumsikan mengikuti distribusi normal. Adapun kelebihan dari metode MLE adalah secara konsep pengerjaan mudah untuk dipahami, dan apabila diperoleh hasil estimasi yang eksplisit, maka dapat dilanjutkan dengan pendekatan numerik untuk memperoleh estimasi dari parameternya (Adrianingsih And Dani 2021).

Model regresi semiparametrik spline truncated telah banyak dilakukan oleh banyak peneliti. Pemodelan regresi yang dilakukan tersebut ada yang menggunakan data cross section dan data longitudinal dalam aplikasi penelitiannya. Peneliti yang menggunakan regresi semiparametrik spline truncated dalam penelitiannya antara lain oleh (Ramadan et al. 2019) yang melakukan pemodelan regresi semiparametrik spline truncated multivariabel pada faktor-faktor yang memengaruhi pada status gizi yang ada di Jawa Timur. (Dwi Fadlirhohim, Tri Rian Dani, and R Dani 2024) juga melakukan pemodelan regresi semiparametrik spline truncated pada faktor-faktor prevalensi stunting di Indonesia. (Rahma Khairunnisa, Prahutama, and Santoso 2020) melakukan pemodelan spline truncated campuran dengan deret fourier dalam regresi semiparametrik pada angka harapan hidup di Indonesia. Sedangkan penelitian mengenai regresi semiparametrik spline truncated yang menggunakan data longitudinal dilakukan antara lain oleh (Dwi Kartikasari 2020) melakukan pemodelan terhadap regresi semiparametrik penalized biresponse pada data longitudinal. (Setyawati, Chamidah, and Kurniawan 2021) melakukan pemodelan regresi semiparametrik spline truncated pada data longitudinal. Kemudian (Damaliana, Nyoman Budiantara, and Ratnasari 2019) mengkaji tentang pemilihan titik knot optimal dalam regresi semiparametrik spline truncated pada data longitudinal untuk kasus data angka harapan hidup di Jawa Timur.

Menurut Wu dan Zhang (2007) kelebihan menggunakan data longitudinal adalah dapat mengetahui perubahan yang terjadi pada individu, tidak membutuhkan subjek yang banyak karena pengamatannya berulang dan juga estimasinya lebih efisien karena dilakukan setiap pengamatan. Data longitudinal adalah Data longitudinal adalah jenis data yang diperoleh dengan melakukan pengamatan berulang terhadap subjek yang sama dalam rentang waktu tertentu. Dengan kata lain, data longitudinal mencatat perkembangan, perubahan, atau pola yang terjadi



yang sama dari waktu ke waktu., analisis hasil estimasi akan mencakup hal-hal yang lebih mendekati realita (Ekananda, penelitian mengenai regresi semiparametrik yang menggunakan subset terbatas pada estimasi titik (point estimation). Estimasi titik yaitu peluang kesalahan dalam menduga nilai sebenarnya (true parameter θ lebih besar. Sahoo (2013) mengatakan bahwa

meskipun diambil banyak sampel random, estimasi titik dari suatu parameter θ jarang sama dengan nilai sebenarnya (true value) dari suatu parameter θ . Oleh sebab itu perlu sebuah kisaran nilai probabilitas untuk θ dengan tingkat kepercayaan tertentu yang disebut interval estimasi (interval estimation) atau selang kepercayaan (confidence interval). Selang kepercayaan merupakan salah satu bagian terpenting dari inferensi statistik. Selang kepercayaan memberi keyakinan bahwa nilai parameter yang benar masih berada dalam selang tersebut. Untuk mengontruksi selang kepercayaan pada parameter regresi semiparametrik digunakan metode pivotal quantity. Beberapa penelitian mengenai selang kepercayaan selang kepercayaan yaitu (Daflin Syaranamual and Budiantara 2011) menggunakan *pivotal quantity* dalam model regresi spline kuadrat, (Nur et al. 2017) menggunakan spline truncated dalam mengaplikasikan selang kepercayaan parameter nonparametrik, (Islamiyati et al. 2022) mengestimasi selang kepercayaan menggunakan *pivotal quantity* pada regresi nonparametrik multivariabel. (Hidayati, Chamidah, and Budiantara 2020) estimasi selang kepercayaan pada regresi semiparametrik multirespon, serta (Setyawati, Chamidah, and Kurniawan 2021) selang kepercayaan regresi semiparametrik spline pada data cross section.

Pada penelitian ini peneliti menggunakan data tentang Indeks Khusus Penurunan Stunting. Stunting adalah gangguan pertumbuhan dan perkembangan pada anak yang disebabkan oleh asupan gizi yang buruk, seringkali menderita penyakit, serta kurangnya stimulasi psikososial (Alda Fuadiyah Suryono et al. 2024). Berdasarkan Survei Status Gizi Indonesia tahun 2021, prevalensi stunting di Indonesia tercatat sebesar 24,4%, dengan target penurunan menjadi 14% pada tahun 2024. Beberapa faktor yang berkontribusi terhadap tingginya angka stunting di Indonesia antara lain pola asuh yang kurang memadai, keterbatasan akses terhadap layanan kesehatan terutama pelayanan antenatal care (ANC) bagi ibu hamil—ketersediaan pangan bergizi yang terbatas bagi keluarga, serta akses air bersih yang kurang memadai (Huriah and Nurjannah 2020). Beberapa penelitian yang dilakukan mengenai stunting, categorical principal component analysis (Islamiyati et al. 2025), support vector machine (Fadil, Islamiyati, and Thamrin 2025), nonparametric spline truncated regression (Islamiyati et al. 2024), spline binary logistic (Islamiyati et al. 2023), and ordinal logistic regression models (Arifin, Islamiyati, and Herdiani 2023). Penelitian ini menggunakan data Indeks khusus Penanganan Stunting (IKPS), dimana salah satu variabel predikturnya memiliki pola hubungan diketahui dengan variabel responnya, namun juga ada beberapa variabel predikturnya memiliki pola hubungan yang tidak diketahui antara variabel responnya. Maka berdasarkan uraian tersebut penulis tertarik meneliti tentang estimasi selang kepercayaan pada model regresi

pada data longitudinal dengan metode *Maximum Likelihood* pada data Indeks Khusus Penurunan Stunting (IKPS) di 8-2023.



1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana estimasi selang kepercayaan untuk parameter model dalam regresi semiparametrik *spline* pada data longitudinal dengan metode *Maximum Likelihood* ?
2. Bagaimana penerapan selang kepercayaan untuk parameter model dalam regresi semiparametrik *spline* pada data Penanganan Khusus Stunting (IPKS) di Indonesia pada Tahun 2018-2023?

1.3 Batasan Masalah

Berdasarkan rumusan masalah, maka ruang lingkup dalam penelitian ini akan dibatasi agar tidak menimbulkan penafsiran yang lebih luas, adapun ruang lingkup dalam penelitian ini dibatasi beberapa hal yaitu:

1. Jumlah titik knot yang digunakan dibatasi pada jumlah titik knot 1 hingga 3 serta orde yang digunakan linear, kuadrat, dan kubik
2. Pemilihan titik knot optimal menggunakan metode *Generalized Cross Validation* (GCV).
3. Pembentukan selang kepercayaan menggunakan *Pivotal Quantity*.
4. Data yang digunakan merupakan Indeks Khusus Penanganan Stunting (IPKS) di Indonesia pada tahun 2018-2023

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Mendapatkan estimasi parameter selang kepercayaan untuk parameter model dalam regresi semiparametrik *spline* pada data longitudinal dengan metode *Maximum Likelihood*
2. Memperoleh selang kepercayaan untuk parameter model regresi semiparametrik *spline* pada data Indeks Khusus Penanganan Stunting (IPKS) di Indonesia pada tahun 2018-2023.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Mengembangkan metode Selang Kepercayaan dengan menggunakan model regresi semiparametrik *spline* pada data longitudinal dengan menggunakan metode *Maximum Likelihood*.
2. Meningkatkan akurasi dalam penanganan Indeks Khusus Penanganan Stunting (IPKS) di Indonesia berdasarkan faktor-faktor yang mempengaruhinya guna memperkecil resiko kejadian stunting pada balita Indonesia.



1.6 Teori

Sub-bab teori ini membahas mengenai beberapa teori yang digunakan untuk membahas pokok pembahasan terkait metode selang kepercayaan regresi semiparametrik spline pada data longitudinal dengan metode MLE.

1.6.1 Regresi Parametrik

Regresi parametrik digunakan jika penyebaran pola data membentuk pola tertentu atau diketahui. Dengan mengetahui bentuk kurva atau fungsi regresi, maka analisis regresi parametrik akan lebih menguntungkan, khususnya metode inferensi dan interpretasi parameternya akan lebih sederhana. Oleh karena itu analisis regresi parametrik lebih sering digunakan apabila terdapat informasi tentang bentuk kurva regresinya (Eubank, 1999). Dilihat dari pandangan statistik inferensia, pendekatan model regresi parametrik memiliki sifat yang sangat baik, seperti misalnya sederhana, mudah interpretasinya, parsimoni, tak bias, estimator linier, efisien, konsisten, dan Best Linier Unbiased Estimator (BLUE) yang tidak dimiliki oleh pendekatan model regresi nonparametrik (Budiantara, 2009).

Secara matematis fungsi regresi parametrik dapat ditulis dengan Persamaan sebagai berikut:

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

Dengan, fungsi $f(x_i)$ sebagai fungsi parametrik atau kurva regresi parametrik yang memiliki *error* random ε_i dan diasumsikan berdistribusi normal independen dengan rata-rata nol dan variansi σ^2 . Fungsi $f(x_i)$ dapat dituliskan dalam bentuk :

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_i ; i = 1, 2, \dots, n \quad (2)$$

Dimana, $\mathbf{x}_i = [1 \quad x_{i1} \quad x_{i2} \quad \dots \quad x_{ip}]$, $i = 1, 2, \dots, n$; sedangkan n adalah banyaknya

data dan p adalah banyaknya variabel, sementara $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$, sehingga

Persamaannya menjadi:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{pi} + \varepsilon_i ; i = 1, 2, \dots, n \quad (3)$$

dengan:

- y_i : variabel respon pada pengamatan ke- i
- $x_{i1}, x_{i2}, \dots, x_{pi}$: variabel-variabel prediktor pengamatan ke- i
- β_0 : intersep dari model
- $\beta_1, \beta_2, \dots, \beta_p$: fungsi komponen parametrik
- ε_i : *error* random yang diasumsikan berdistribusi normal dengan mean nol dan variansi σ^2

aan (3), maka bentuk matriksnya.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (4)$$



$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ 1 & x_{31} & x_{32} & \cdots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

dimana y merupakan vektor dari variabel respon berukuran $n \times 1$, X merupakan matrik berukuran $n \times (p + 1)$ dan β adalah vektor parameter yang akan diestimasi berukuran $(p + 1) \times 1$, ε adalah vektor error random berukuran $n \times 1$.

1.6.2 Regresi nonparametrik spline truncated

Salah satu model regresi nonparametrik yang dapat digunakan untuk menduga kurva regresi adalah regresi spline. Regresi spline merupakan pendekatan yang mempertimbangkan kemulusan kurva dalam mencocokkan plot data. Kurva dibentuk berdasarkan titik-titik knot yang dihubungkan satu sama lain. Regresi spline memungkinkan penggunaan berbagai orde, sehingga dapat dibentuk regresi spline linier, kuadratik, kubik, maupun dengan orde lainnya. Spline adalah potongan-potongan polinomial (fungsi *truncated*) yang mulus namun tetap tersegmen (Wang, 2011). Penggunaan spline ini difokuskan pada pola data, di mana setiap daerah dapat memiliki karakteristik berbeda. Kemampuan ini ditunjukkan oleh fungsi *truncated* yang melekat pada estimator, dengan titik-titik potong polinomial yang disebut titik knot (Bintariningrum & Budiantara, 2014).

Pada spline *truncated* digunakan titik *knot*. *Knot* adalah nilai variabel prediktor yang menjadi batas antar segmen. *Knot* diartikan sebagai suatu titik fokus dalam fungsi spline, sehingga kurva yang dibentuk tersegmen pada titik tersebut (Dani & Ni'matuzzahroh, 2022).

Jika diberikan data berpasangan $(z_i, y_i), i = 1, 2, \dots, n$. Dimana y_i adalah variabel respon sedangkan z_i adalah variabel prediktor yang mengikuti pola nonparametrik. Pola hubungan y_i dan z_i dinyatakan dalam model regresi seperti persamaan dibawah ini:

$$y_i = g(z_i) + \varepsilon_i \quad i = 1, 2, \dots, n \quad (5)$$

Selanjutnya jika kurva regresi $g(z_i)$ pada persamaan (5), dihipotesis dengan kurva regresi spline *truncated* dengan knot $K_1, K_2, \dots, K_r$ maka:

$$g(z_i) = \sum_{s=1}^m \theta_s z_i^s + \sum_{k=1}^r \theta_k (z_i - K_k)_+^m \quad (6)$$

Dengan, θ adalah parameter-parameter model dan $\sum_{s=1}^m \theta_s z_i^s$ merupakan komponen polinomial dan $\sum_{k=1}^r \theta_k (z_i - K_k)_+^m$ merupakan komponen *truncated*:

$$(z_i - K_k)_+^m = \begin{cases} (z_i - K_k)^m, & z_i \geq K_k \\ 0, & z_i < K_k \end{cases}$$

bers pada Persamaan (5) menjadi:

$$y_i = \theta_s z_i^s + \sum_{k=1}^r \theta_{k+m} (z_i - K_k)_+^m + \varepsilon_i \quad (7)$$

dalam bentuk matriks:

$$y = Z\theta + \varepsilon \quad (8)$$



Dimana,

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & \cdots & z_{1m} & (z_1 - K_1)_+^m & (z_1 - K_2)_+^m & \cdots & (z_1 - K_r)_+^m \\ 1 & z_{21} & \cdots & z_{2m} & (z_2 - K_1)_+^m & (z_2 - K_2)_+^m & \cdots & (z_2 - K_r)_+^m \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & \cdots & z_{nm} & (z_n - K_1)_+^m & (z_n - K_2)_+^m & \cdots & (z_n - K_r)_+^m \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\theta = [\theta_0 \quad \theta_1 \quad \cdots \quad \theta_m \quad \theta_{m+1} \quad \theta_{m+2} \quad \cdots \quad \theta_{m+r}]^T, \text{ dan}$$

Kurva regresi $g(z_i)$ merupakan kurva regresi nonparametrik spline truncated derajat m dengan banyaknya titik knot r . Derajat m merupakan derajat pada persamaan polinomial. Titik-titik knot $K_1, K_2, \dots, K_r$ adalah titik-titik knot yang menunjukkan perubahan pola perilaku dari kurva pada sub-sub interval yang berbeda, dimana $K_1 < K_2 < \cdots < K_r$.

1.6.3 Regresi semiparametrik spline truncated

Regresi semiparametrik merupakan gabungan antara komponen parametrik dan komponen nonparametrik (Budiantara, 2009). Dalam beberapa kasus, dapat ditemukan hubungan antara variabel respon dengan salah satu variabel prediktor bersifat linier, tetapi hubungan dengan variabel prediktor yang lain tidak diketahui pola hubungannya. Variabel yang memiliki pola data yang diketahui atau terdapat informasi sebelumnya mengenai pola datanya digolongkan pada komponen parametrik. Sedangkan variabel yang tidak diketahui pola datanya digolongkan pada komponen nonparametrik.

Misalkan diberikan data berpasangan (y_i, x_i, z_i) dengan $i = 1, 2, \dots, n$ dan hubungan antara y_i, x_i , dan z_i dapat dinyatakan dalam model sebagai berikut:

$$y_i = f(x_i) + g(z_i) + \varepsilon_i \quad (9)$$

dengan, y_i adalah variabel respon pada pengamatan ke- i , $f(x_i)$ adalah fungsi regresi parametrik, $g(z_i)$ adalah fungsi regresi nonparametrik dan ε_i adalah *error* random berdistribusi normal dengan mean nol dan variansi σ^2 .

Fungsi regresi parametrik $f(x_i)$ dengan p merupakan banyaknya variabel prediktor dapat dinyatakan dalam bentuk persamaan sebagai berikut:

$$f(x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{pi} \quad (10)$$

Sehingga, untuk $i = 1$; $y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_p x_{p1} + \varepsilon_1$

$$i = 2$$
; $y_2 = \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \cdots + \beta_p x_{p2} + \varepsilon_2$

$$; y_n = \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \cdots + \beta_p x_{pn} + \varepsilon_n$$

apat dinyatakan dalam bentuk matriks sebagai berikut:

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$



selanjutnya, jika kurva regresi $g(z_i)$ pada Persamaan (9) didekati dengan fungsi spline *truncated orde m* dengan knot $K_1, K_2, \dots, K_r$. secara umum, fungsi *spline truncated* dapat disajikan dalam bentuk:

$$g(z_i) = \sum_{s=1}^m \theta_s z_i^s + \sum_{k=1}^r \theta_{m+k} (z_i - K_k)_+^m \quad (11)$$

Dengan, θ adalah parameter-parameter model dan $\sum_{s=1}^m \theta_s z_i^s$ merupakan komponen polinomial dan $\sum_{k=1}^r \theta_k (z_i - K_k)_+^m$ merupakan komponen *truncated*

$$(z_i - K_k)_+^m = \begin{cases} (z_i - K_k)^m, & z_i \geq K_k \\ 0, & z_i < K_k \end{cases} \quad (12)$$

Dimana θ_s dengan $k = 1, 2, \dots, m$ dan θ_{m+k} dengan $k = 1, 2, \dots, r$ merupakan parameter-parameter yang tidak diketahui. Serta $K_1, K_2, \dots, K_r$ merupakan titik-titik knot.

Berdasarkan model regresi komponen spline pada persamaan (11), maka dapat dibentuk atau dikembangkan bentuk model regresi semiparametrik spline *truncated* multivariabel. Jika variabel $z_1, z_2, \dots, z_q$ bersifat aditif, maka fungsi multivariabel g adalah:

$$g(z_1, z_2, \dots, z_q) = \sum_{l=1}^q g_l(z_l) \quad (13)$$

Apabila fungsi $g_l(z_l)$ dihipotesis dengan fungsi spline *truncated orde m* dengan titik-titik knot $K_{1l}, K_{2l}, \dots, K_{rl}$ dengan $l = 1, 2, \dots, q$ maka fungsi multivariabel $g(z_1, z_2, \dots, z_q)$ disajikan dalam bentuk sebagai berikut.

$$g(z_1, z_2, \dots, z_q) = \sum_{l=1}^q \left(\sum_{s=1}^m \theta_{sl} z_{il}^s + \sum_{k=1}^r \theta_{(m+k)l} (z_{il} - K_{kl})_+^m \right) \quad (14)$$

Dengan, titik-titik knot $K_1, K_2, \dots, K_r$ dan q merupakan banyaknya variabel prediktor pada persamaan, sehingga fungsi g merupakan model regresi nonparametrik spline *truncated* multivariabel seperti persamaan (15) sebelumnya

Apabila diasumsikan nilai $m = 1$ pada persamaan (15) maka disebut fungsi spline *truncated* linear multivariabel yang didefinisikan sebagai berikut:

$$g(z_1, z_2, \dots, z_q) = \sum_{l=1}^q \left(\theta_{1l} z_{il} + \sum_{k=1}^r \theta_{(1+k)l} (z_{il} - K_{kl})_+^1 \right) \quad (15)$$

Berdasarkan Persamaan (15), diatas, untuk $i = 1, 2, \dots, n$ dapat diubah dalam bentuk matriksnya sebagai berikut:

$$g(z) = Z\theta \quad (16)$$

Dengan $\theta = (\theta_{11} \ \theta_{21} \ \dots \ \theta_{(1+r)l} \ \dots \ \theta_{1q} \ \theta_{2q} \ \dots \ \theta_{(1+r)q})^T$

$$Z = \begin{bmatrix} (z_{11} - K_{11})_+ & \dots & \theta_{1q} & (z_{1q} - K_{1q})_+ & \dots & (z_{1q} - K_{rq})_+ \\ (z_{12} - K_{11})_+ & \dots & \theta_{2q} & (z_{2q} - K_{1q})_+ & \dots & (z_{1q} - K_{rq})_+ \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ (z_{nl} - K_{11})_+ & \dots & \theta_{nq} & (z_{nq} - K_{1q})_+ & \dots & (z_{nq} - K_{rq})_+ \end{bmatrix}$$



1.6.4 Data longitudinal

Terdapat tiga jenis data dalam analisis regresi, yaitu data cross-section, data time series dan data longitudinal. Data cross-section merupakan data yang hanya dilakukan satu kali pengamatan pada setiap subjek dan saling independen. Data time series adalah kumpulan data yang dikumpulkan dan dicatat secara berurutan berdasarkan waktu, biasanya pada interval waktu yang tetap, seperti harian, mingguan, bulanan, atau tahunan. Sedangkan data longitudinal merupakan gabungan dari data cross-section dan data time series, Unit *cross section* berupa individu, perusahaan dan wilayah atau Lokasi, sedangkan unit *time series* berupa harian, bulanan dan tahunan (Amaliah dkk, 2020). Data longitudinal merupakan data yang didapat dari pengamatan berulang dari setiap subjek pada interval waktu yang berbeda. Beberapa keuntungan studi mengenai data longitudinal antara lain dapat mengetahui perubahan individu, dan membutuhkan subjek yang tidak terlalu banyak karena pengamatan berulang (Wu & Zhang (2006). Data longitudinal memiliki keunggulan yaitu memiliki kemampuan untuk mengenali pengaruh terhadap respon, data longitudinal diasumsikan obyek-obyek saling independent satu sama lainnya, tetapi pengamatan di dalam obyek yang sama saling dependen, karena cenderung berkorelasi (Islamiyati, 2014).

1.6.4. Regresi semiparametrik *spline truncated* pada data longitudinal

Persamaan (9) merupakan suatu bentuk model regresi semiparametrik *spline truncated* pada data *cross section*. Model tersebut menggunakan n obyek pengamatan pada suatu waktu tertentu, jika n obyek tersebut terus diamati secara berulang dalam kurun waktu tertentu dalam hal ini t tahun pengamatan, maka model tersebut dapat dikembangkan menjadi model regresi semiparametrik *spline* multivariabel pada data longitudinal.

(Poerwanto 2014), Misalkan diberikan data berpasangan (y_{ij}, x_{ij}, z_{ij}) dengan $i = 1, 2, \dots, n; j = 1, 2, \dots, t$ dimana y_{ij} adalah variabel respon sedangkan x_{ij} adalah variabel prediktor yang mengikuti pola parametrik dan z_{ij} adalah variabel prediktor yang mengikuti pola nonparametrik. Pola antara y_{ij} , x_{ij} , dan z_{ij} dapat dinyatakan dalam model regresi semiparametrik pada data longitudinal dapat dituliskan dalam bentuk Persamaan:

$$y_{ij} = \sum_{h=1}^p f(x_{ij}) + \sum_{l=1}^q g_{ilj}(z_{ilj}) + \varepsilon_{ij} \quad (17)$$

Dimana, $f(x_{ij})$ Menghampiri kurva regresi komponen parametrik, dimana $h = 1, 2, \dots, p$ dengan fungsi linear



$$f(x_{ij}) = \beta_0 + \sum_{h=1}^p \beta_{ih} x_{ihj}$$

amproxi kurva regresi komponen nonparametrik, dimana $l = 1, 2, \dots, q$ Spline Truncated linier dengan knot sebanyak r untuk masing-

$$\sum_{l=1}^q g_{ilj}(z_{ilj}) = \theta_{il}z_{ilj} + \sum_{k=1}^r \theta_{(1+k)il}(z_{ilj} - K_{kil})_+^m$$

Berdasarkan Persamaan (14), maka dibentuk model regresi semiparmetrik spline truncated multivariabel pada data longitudinal adalah sebai berikut

$$y_{ij} = \sum_{h=1}^p \beta_{ih}x_{ihj} + \sum_{l=1}^q \left(\theta_{il}z_{ilj} + \sum_{k=1}^r \theta_{(1+k)il}(z_{ilj} - K_{kil})_+^m \right) + \varepsilon_{ij} \quad (18)$$

Adapun bentuk matriksnya sebgai berikut:

$$\mathbf{y} = \mathbf{S}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \quad (19)$$

Dengan, $\mathbf{y}$ adalah vektor variabel respon berukuran $nt \times 1$, matriks $\mathbf{S} = [\mathbf{X} : \mathbf{Z}]$, berukuran $nt \times n[1 + p + q(1 + r)]$ adalah matriks yang memuat variabel prediktor komponen parametrik dan $\mathbf{Z}$ merupakan matriks yang memuat variabel prediktor komponen nonparametrik yang tergantung pada titik-titik knot yang diberikan sedangkan $\boldsymbol{\delta}$ vektor parameter berukuran $n[1 + p + q(r + 1)] \times$ yang memuat parameter komponen parametrik dan parameter komponen nonparametrik terdiri atas vektor-vektor parameter β dan θ . $\boldsymbol{\varepsilon}$ adalah vektor error dengan ukuran $nt \times 1$

1.6.8 Pemilihan titik knot

Dalam regresi nonparametrik maupun semiparametrik dengan pendekatan Spline, hal penting yang berperan dalam mendapatkan estimator Spline adalah pemilihan titik knot yang optimal. Salah satu metode yang digunakan dalam memilih titik knot adalah Generalized Cross Validation (GCV). Menurut Wahba (1990) dan Wang (1998), salah satu metode yang baik dan banyak digunakan karena kelebihan yang dimiliki untuk pemilihan titik knot yang optimal adalah Generalized Cross Validation (GCV).

Titik knot optimal diperoleh dari nilai GCV minimum. Fungsi GCV didefinisikan sebagai berikut (Sari, 2016):

$$GCV(K_1, K_2, \dots, K_r) = \frac{MSE(K_1, K_2, \dots, K_r)}{[n^{-1} \text{trace}(\mathbf{I} - \mathbf{A}(K_1, K_2, \dots, K_r))]^2} \quad (20)$$

Dimana, $MSE(K_1, K_2, \dots, K_r) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$, $\mathbf{I}$ adalah matriks identitas, n adalah banyaknya pengamatan, $(K_1, K_2, \dots, K_r)$ adalah titik-titik knot, dan $\mathbf{A}(K_1, K_2, \dots, K_r) = \mathbf{S}(\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T$.

Nilai titik knot yang optimal diberikan nilai GCV yang minimum (Wu & Zhang, 2006).

1.6.9 Koefisien determinasi

Koefisien Determinasi (R^2) merupakan ukuran ketelitian atau ketepatan modelnya kontribusi prediktor terhadap respon. Rumus Koefisien R^2 adalah sebagai berikut:

$$R^2 = \frac{SSR}{SST} = \frac{(\hat{Y} - \bar{Y})'(\hat{Y} - \bar{Y})}{(Y - \bar{Y})'(Y - \bar{Y})} \quad (21)$$

Nilai R^2 akan negatif dan batasannya adalah $0 \leq R^2 \leq 1$ (Gujarati dan



1.6.10 Estimasi selang kepercayaan

Estimasi selang merupakan pengembangan dari estimasi titik, dimana nilai estimasi tidak bergantung pada satu titik akan tetapi memuat suatu range tertentu yang memiliki batas bawah dan batas atas. Dibandingkan dengan estimasi, estimasi interval dapat memberikan informasi dengan akurasi yang lebih baik dalam menentukan estimasi suatu parameter populasi. Oleh karena itu dirumuskanlah suatu interval random, yang merupakan selang suatu parameter populasi misalnya β , yang batas-batasnya merupakan statistik. Interval ini disusun sedemikian hingga memiliki peluang yang sebesar mungkin. Montgomery et al (2012) menjelaskan tahapan untuk mendapatkan estimasi interval untuk suatu parameter dalam regresi dengan menggunakan metode pivotal quantity sebagai berikut.

Misalkan diberikan model regresi linear berganda pada persamaan (22) untuk mendapat selang kepercayaan untuk koefisien regresi β_p , diasumsikan ε_i berdistribusi normal independen dengan mean nol dan varians σ^2 . Oleh karena itu, observasi y_i berdistribusi normal dengan mean $\beta_0 + \sum_{j=1}^p \beta_j x_{ji}$ dan varians σ^2 .

Apabila persamaan (1) dinyatakan dalam bentuk matriks

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (22)$$

Salah satu metode yang digunakan untuk mendapatkan $\hat{\boldsymbol{\beta}}$ adalah dengan menggunakan metode *Maximum Likelihood* (MLE) yang diasumsikan error $\boldsymbol{\varepsilon}$ berdistribusi normal Independen dengan mean 0 dan varians σ^2 . Sehingga didapatkan distribusi probabilitas dari $\boldsymbol{\varepsilon}$ adalah

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma^2) &= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^2 \right\} \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \end{aligned}$$

$$\mathbf{W} = \begin{pmatrix} 1/\sigma_1^2 & 0 & \dots & 0 \\ 0 & 1/\sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\sigma_n^2 \end{pmatrix}$$

Dengan matriks $\mathbf{W}$ adalah matriks pembobot (matriks varians kovarians) dan berisi diagonal ($\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_n$) Maka penyelesaian optimasi dilakukan dengan penjabaran berikut,

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma^2) &= \log(L(\boldsymbol{\beta}, \sigma^2)) \\ &= \log \left[\left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n + \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \right] \\ &= n \log \left(\frac{1}{\sigma\sqrt{2\pi}} \right) + -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &\quad n \log(\sigma) + n \log(\sqrt{2\pi}) - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &\quad \cdot \log(\sigma) + n \log(\sqrt{2\pi}) - \frac{1}{2} (\mathbf{y}^T \mathbf{W} \mathbf{y} - 2\mathbf{X}^T \mathbf{W} \mathbf{y} \boldsymbol{\beta}^T + \mathbf{X}^T \mathbf{W}^T \boldsymbol{\beta}^T \boldsymbol{\beta}) \end{aligned}$$

estimasi $\boldsymbol{\beta}$ dapat dilakukan derivatif parsial $\partial \ln L(\boldsymbol{\beta}, \sigma^2)$



$$\begin{aligned}
\frac{\partial \ln L(\beta, \sigma^2)}{\partial \beta} &= \frac{\partial}{\partial \beta} \left[n \cdot \log(1) - n \cdot \log(\sigma) + n \log(\sqrt{2\pi}) - \frac{1}{2} (\mathbf{y}^T \mathbf{W} \mathbf{y} - \right. \\
&\quad \left. 2 \mathbf{X}^T \mathbf{W} \mathbf{y} \beta^T + \mathbf{X}^T \mathbf{W}^T \beta^T \beta) \right] \\
&= 0 - 0 + 0 - 0 + \mathbf{X}^T \mathbf{W} \mathbf{y} - \frac{1}{2} \cdot 2 \mathbf{X}^T \mathbf{W} \mathbf{X} \hat{\beta} \\
&= \mathbf{X}^T \mathbf{W} \mathbf{y} - \mathbf{X}^T \mathbf{W} \mathbf{X} \hat{\beta}
\end{aligned}$$

Jika derivatif parsial diatas disamakan dengan nol, diperoleh persamaan:

$$\begin{aligned}
\mathbf{X}^T \mathbf{W} \mathbf{y} - \mathbf{X}^T \mathbf{W} \mathbf{X} \hat{\beta} &= 0 \\
\mathbf{X}^T \mathbf{W} \mathbf{X} \hat{\beta} &= \mathbf{X}^T \mathbf{W} \mathbf{y} \\
\hat{\beta} &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W} \mathbf{y})
\end{aligned} \tag{23}$$

Berdasarkan hasil estimasi $\hat{\beta}$ pada persamaan di atas, maka diperoleh:

$$\begin{aligned}
\hat{\mathbf{y}} &= \mathbf{X} \hat{\beta} \\
&= \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W} \mathbf{y})
\end{aligned} \tag{24}$$

Dengan memperhatikan bahwa $\hat{\beta}$ kombinasi linear dari $\mathbf{y}$, maka $\hat{\beta} \sim N(\beta, \sigma^2 (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1})$, sehingga untuk masing-masing estimator berdistribusi normal $\hat{\beta}_j \sim N(\beta_j, \sigma^2 C_{jj})$, dimana C_{jj} merupakan elemen diagonal ke- j dari matriks $(\mathbf{X}^T \mathbf{X})^{-1}$. Akibatnya diperoleh *pivotal quantity*.

$$T_j(x_1, x_2, \dots, x_n) = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 C_{jj}}}, \quad p = 0, 1, \dots, h \tag{25}$$

dimana σ^2 tidak diketahui, Berdasarkan hasil pada Persamaan (27), didapat selang kepercayaan terpendek berukuran $(1 - \alpha) \times 100\%$ untuk koefisien regresi $\beta_j, j = 1, 2, \dots, h$ adalah sebagai berikut

$$\hat{\beta}_j - t_{\frac{\alpha}{2}, n-h} \sqrt{\sigma^2 C_{jj}} \leq \beta_j \leq + t_{\frac{\alpha}{2}, n-h} \sqrt{\sigma^2 C_{jj}} \tag{26}$$

Dimana, C_{jj} merupakan elemen diagonal ke- j dari matriks $(\mathbf{X}^T \mathbf{X})^{-1}$, Berdistribusi t dengan derajat bebas $n - h$.

1.6.11 Pivotal quantity

Pivotal quantity atau besaran pivot merupakan kuantitas yang distribusinya tidak tergantung atau tidak memuat parameter. Misalkan diberikan suatu variabel random $X = (x_1, x_2, \dots, x_n)$, dan θ parameter populasi, maka suatu statistik $T(x_1, x_2, \dots, x_n, \theta)$, jika distribusi dari $T(x_1, x_2, \dots, x_n, \theta)$ tidak memuat atau tidak bergantung pada parameter θ . *Pivotal quantity* sangat penting dalam konstruksi statistik uji, dengan menggunakan prinsip dasar bahwa suatu statistik tidak memuat parameter dari populasinya, seperti konstruksi statistik t-student untuk kasus dengan varians populasi tidak diketahui.



juga dapat digunakan untuk mengkonstruksi estimasi interval. Untuk estimasi interval untuk parameter θ , maka langkah awal untuk mencari selang interval yaitu dengan mencari estimasi titik dari parameter θ , kemudian mencari distribusi sampling dari estimasi titik tersebut. Langkah selanjutnya adalah mencari *pivotal quantity* untuk parameter θ dan dilanjutkan dengan mencari

estimasi interval terpendek dengan menyelesaikan persamaan probabilitas seperti dibawah ini

$$P(b(x_1, x_2, \dots, x_n) \leq T(x_1, x_2, \dots, x_n, \theta) \leq a(x_1, x_2, \dots, x_n)) = 1 - \alpha \quad (27)$$

Dengan $T(x_1, x_2, \dots, x_n, \theta)$ merupakan pivotal quantity untuk parameter θ . Selanjutnya akan didapatkan estimasi interval $1 - \alpha$ untuk parameter θ sedemikian hingga dapat dituliskan dalam bentuk Persamaan:

$$P(u(x_1, x_2, \dots, x_n) \leq \theta \leq v(x_1, x_2, \dots, x_n)) = 1 - \alpha \quad (28)$$

Dimana $u(x_1, x_2, \dots, x_n)$ merupakan batas bawah dan $v(x_1, x_2, \dots, x_n)$ merupakan batas bawah.

1.6.12 Indeks khusus penanganan stunting

Percepatan pengurangan kasus stunting menjadi program prioritas pemerintah di bidang kesehatan. Stunting tidak hanya memengaruhi pertumbuhan secara fisik, tetapi berpengaruh juga terhadap fungsi penting lainnya pada tubuh, seperti perkembangan otak dan imun tubuh. Akibatnya, balita stunting berpotensi mengalami kecerdasan yang kurang maksimal, mudah terserang penyakit, dan mungkin berisiko mengalami penurunan produktivitas di masa depan.

Indeks Khusus Penanganan Stunting (IKPS) merupakan alat ukur pencapaian penanganan stunting di Indonesia. IKPS disusun oleh dimensi yang terdiri beberapa indikator dan mengikuti prinsip SMART (Specific, Measurable, Achievable, Realistic, Timely and Simplicity). Salah satu prinsip yang mudah dianalisis sebagai syarat indikator IKPS adalah prinsip measurable (dapat diukur) yang berarti indikator penyusun IKPS harus indikator yang dapat diukur (BPS, 2023). Berdasarkan prinsip tersebut, maka terbentuk 6 dimensi dan 11 indikator yang menjadi penyusun IKPS tingkat provinsi, yaitu (BPS, 2021):

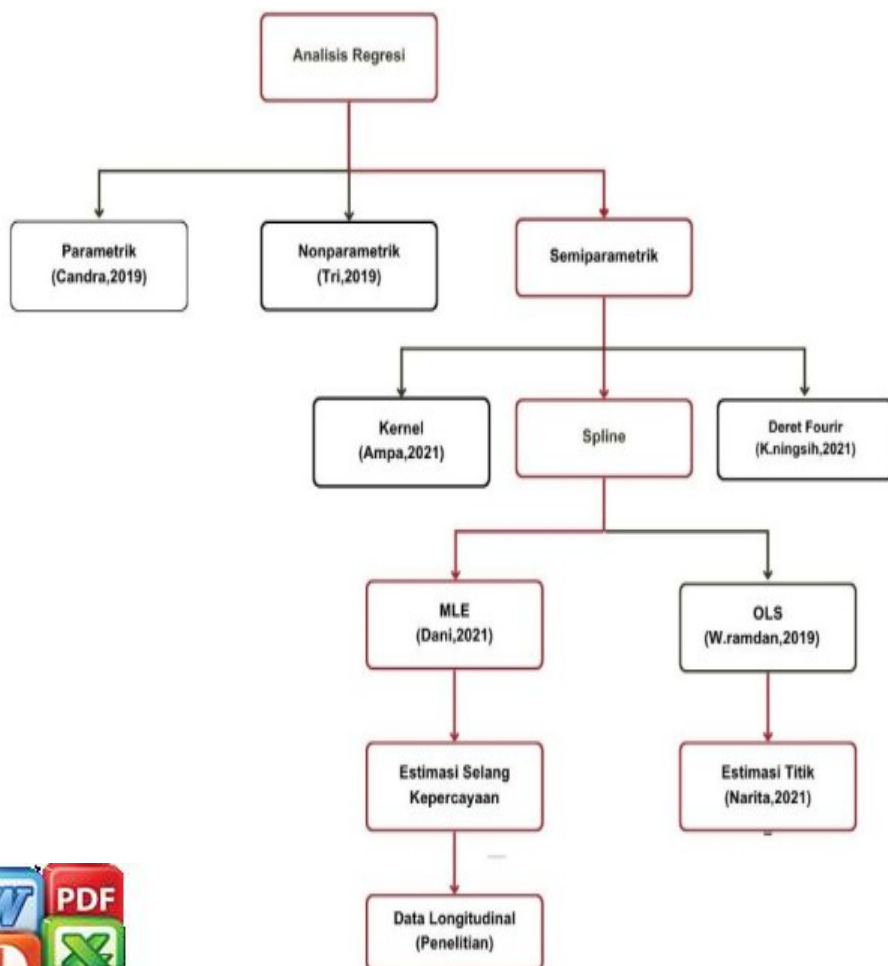
1. Dimensi kesehatan terbentuk dari tiga indikator, yaitu imunisasi, penolong persalinan oleh tenaga kesehatan di fasilitas kesehatan, dan keluarga berencana (KB) modern.
2. Dimensi gizi terbentuk dari dua indikator, yaitu ASI eksklusif dan makanan pendamping ASI.
3. Dimensi perumahan terbentuk dari dua indikator, yaitu air minum layak dan sanitasi layak.
4. Dimensi pangan terbentuk dari satu indikator, yaitu ketidakcukupan konsumsi pangan.
5. Dimensi Pendidikan terbentuk dari satu indikator, yaitu Pendidikan Anak Usia Dini (PAUD).
6. Dimensi perlindungan sosial terbentuk dari dua indikator, yaitu pemanfaatan an dan penerima Kartu Perlindungan Sosial (KPS) atau Kerti ra (KKS).



runan stunting merupakan program yang harus dilakukan oleh a diperlukan instrumen evaluasi guna mengetahui sejauh mana berhasil menurunkan angka prevelensi stunting. Penyusunan dilakukan pada tahun 2018, menggunakan data survei sosial

ekonomi nasional (Susenas) maret 2017 (BPS, 2023). Selanjutnya pada tahun 2020 dilakukan penyempurnaan IKPS menggunakan data Susenas Maret 2018-2019, penyempurnaan tersebut dilakukan dengan penambahan dimensi pendidikan dengan indikator pendidikan anak usia dini (PAUD) sehingga menjadi 6 dimensi dengan 12 indikator. 6 dimensi IKPS meliputi dimensi gizi, perumahan, pendidikan, perlindungan sosial, pangan dan Kesehatan (BPS, 2025). IKPS yang disajikan pada tingkat nasional dan provinsi berbeda dengan tingkat kabupaten/kota dimana pada tingkat nasional dan provinsi terdiri dari 6 dimensi sedangkan pada tingkat kabupaten/kota hanya tersusun dari 5 dimensi tanpa mengikutsertakan dimensi pangan dikarenakan ketidakterediaan data.

1.6.13 Kerangka Konseptual



Gambar 1. Kerangka Konseptual



BAB II METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari publikasi Kementerian Kesehatan Republik Indonesia dan Badan Pusat Statistik (BPS) yaitu Laporan Indeks Khusus Penanganan Stunting (IKPS) dari tahun 2018 sampai 2023. Data yang diambil dari Laporan IKPS meliputi presentase Indeks Khusus Penanganan Stunting (IKPS) pada 34 provinsi di Indonesia dan variabel-variabel yang diduga berpengaruh terhadap IKPS. Laporan Indeks Khusus Penanganan Stunting (IKPS) merupakan salah satu media publikasi data dan informasi yang berisi situasi dan kondisi kesehatan di Indonesia yang cukup komprehensif. Laporan Indeks Khusus Penanganan Stunting (IKPS) merupakan publikasi dari Kementerian kesehatan yang diterbitkan setiap tahun.

2.2 Variabel Penelitian

Pada penelitian ini variabel penelitian dibagi menjadi dua, yaitu variabel respon (dependen) (y) dan variabel prediktor (independen) (x). Dimana variabel respon pada penelitian ini (y) adalah Indeks Khusus Penanganan Stunting (IKPS) sedangkan variabel prediktor (x) yaitu terdiri Dimensi Kesehatan (x_1), Dimensi Pendidikan (x_2), Dimensi Gizi (x_3), Dimensi Perumahan (x_4), Dimensi Pangan (x_5), Dimensi Perlindungan Sosial (x_6). Berikut variabel penelitian yang ditunjukkan pada Tabel. 1.

Tabel 1. Variabel Penelitian

Variabel Penelitian	Kode	Nama Variabel
Variabel Respon	y	Indeks Khusus Penanganan Stunting (IKPS)
Variabel Prediktor	x_1	Dimensi Kesehatan
	x_2	Dimensi Pendidikan
	x_3	Dimensi Gizi
	x_4	Dimensi Perumahan
	x_5	Dimensi Pangan
	x_6	Dimensi Perlindungan Sosial

2.3 Metode Analisis Data



Untuk mendapatkan tujuan dalam penelitian ini adalah estimasi selang kepercayaan untuk parameter model regresi *spline* pada data longitudinal dengan metode *Maximum*

a) Estimasi parameter model regresi semiparametrik spline pada data longitudinal:

- [1] Diberikan data berpasangan $(x_{i1j}, x_{i2j}, \dots, x_{ipj}, z_{i1j}, z_{i2j}, \dots, z_{iqj}, y_{ij})$, yang mengikuti model regresi semiparametrik untuk data longitudinal sebagai berikut.

$$y_{ij} = \sum_{u=1}^p f_{iu}(x_{iuj}) + \sum_{l=1}^q g_{ilj}(z_{ilj}) + \varepsilon_{ij}$$

Dimana p adalah jumlah variabel komponen parametrik yang dihampiri dengan fungsi linier, q adalah jumlah variabel prediktor yang dihampiri fungsi *spline truncated*. Kemudian $i = 1, 2, \dots, n$ merupakan subjek yang diamati sebanyak n dan $j = 1, 2, \dots, t$ merupakan pengulangan sebanyak t .

- [2] Menghampiri kurva regresi komponen parametrik $f_{iu}(x_{iuj})$ dengan fungsi linier dimana $u = 1, 2, \dots, p$

$$f_{iuj}(x_{ij}) = \beta_0 + \sum_{u=1}^p \beta_{iu} x_{iuj}$$

- [3] Menghampiri kurva regresi komponen nonparametrik $g_{ilj}(z_{ilj})$ dimana $l = 1, 2, \dots, q$. menggunakan fungsi *spline truncated linear* dengan knot sebanyak r untuk masing-masing x_1 yaitu.

$$\sum_{l=1}^q g_{ilj}(z_{ilj}) = \theta_{il} z_{ilj} + \sum_{k=1}^r \theta_{(1+k)il} (z_{ilj} - K_{kil})_+^1$$

- [4] Menuliskan model regresi semiparametrik spline pada data longitudinal sebagai berikut,

$$\begin{aligned} y_{ij} &= \sum_{u=1}^p f_{iu}(x_{iuj}) + \sum_{l=1}^q g_{ilj}(z_{ilj}) + \varepsilon_{ij} \\ &= \sum_{u=1}^p \beta_{iu} x_{iuj} + \sum_{l=1}^q \left(\theta_{il} z_{ilj} + \sum_{k=1}^r \theta_{(1+k)il} (z_{ilj} - K_{kil})_+^1 \right) + \varepsilon_{ij} \end{aligned}$$

- [5] Menuliskan model regresi semiparametrik untuk data longitudinal dalam bentuk matriks.

$$\mathbf{y} = \mathbf{S}\boldsymbol{\delta} + \boldsymbol{\varepsilon},$$

Dengan, $\mathbf{S} = [\mathbf{X} \quad \mathbf{Z}]$ dan $\boldsymbol{\delta} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\theta} \end{bmatrix}$

dapatkan estimasi untuk parameter $\boldsymbol{\delta}$ menggunakan metode *Maximum Likelihood Estimation* (MLE).

- Membentuk fungsi Likelihood $L(\boldsymbol{\delta})$
- Membuat transformasi $l(\boldsymbol{\delta}) = \log(L(\boldsymbol{\delta}))$



- Dengan menggunakan derivatif, menurunkan parsial $l(\delta)$ terhadap δ dan disamakan dengan 0, sehingga didapatkan estimasi δ .
- Didapatkan estimasi δ yaitu $\hat{\delta}$,

b) Selang Kepercayaan

1. Mencari distribusi dari $\hat{\delta}$.
2. Setelah didapatkan distribusi $\hat{\delta}$, kemudian dicari *pivotal quantity* untuk parameter $\hat{\delta}$.
3. Misalkan $T_j(x_1, x_2, \dots, x_p, z_1, z_2, \dots, z_q)$ adalah *pivotal quantity* untuk parameter $\hat{\delta}_j$ dengan $j = 1, 2, \dots, (1 + p + q(1 + r))$ Menyelesaikan persamaan dalam probabilitas

$$P(u \leq T_j(x_1, x_2, \dots, x_p, z_1, z_2, \dots, z_q) \leq v) = 1 - \alpha,$$

4. Menghitung panjang selang kepercayaan $1 - \alpha$ yaitu:

$$l = (u, v)$$

5. Membentuk fungsi Lagrange :

$$G(u, v, \lambda) = l(u, v) + \theta(\varphi(v) - \varphi(u))$$

Dimana

φ : fungsi distribusi kumulatif $N(0,1)$,

θ : parameter Lagrange.

λ : konstanta Lagrange

6. Melakukan optimasi terhadap fungsi Lagrange.
7. Dari langkah (6) diatas diperoleh selang kepercayaan terpendek $(1 - \alpha) \times 100\%$ untuk parameter δ .

(2) Mengaplikasikan selang kepercayaan untuk parameter model dalam regresi Semiparametrik *spline* pada data Indeks Penanganan Khusus Stunting (IKPS) di Indonesia pada tahun 2018-2023 dengan langkah-langkah sebagai berikut:

1. Melakukan eskplorasi data untuk mengetahui gambaran umum tentang data IKPS.
2. Membuat *Scatter plot* antara variabel respon dengan variabel prediktor untuk mengetahui pola hubungannya.
3. Mengidentifikasi variabel prediktor komponen parametrik, komponen nonparametrik Spline Truncated
4. Untuk menentukan variabel komponen parametrik dan nonparametrik dalam penelitian ini digunakan *scatter plot* dan nilai koefisien determinasi (R^2), dengan kriteria pemilihan variabel penelitian adalah jika nilai R^2 besar ($> 50\%$), maka variabel tersebut cenderung mengikuti pola parametrik,

jika nilai R^2 kecil dari ($< 50\%$) maka cenderung mengikuti pola rik.

an data Indeks Khusus Penanganan Stunting (IKPS) dengan kan regresi semiparametrik *spline truncated* pada data l dengan berbagai titik knot (satu knot, dua knot, dan tiga knot).



6. Mendapatkan titik knot menggunakan metode *Generalized Cross Validation* (GCV).
7. Menetapkan model terbaik dari nilai *Generalized Cross Validation* (GCV) terkecil.
8. Menghitung R^2 adj sebagai bagian dari kriteria kebaikan model.
9. Mendapatkan selang kepercayaan untuk parameter model semiparametrik dengan pendekatan spline pada data longitudinal.
10. Kesimpulan dan saran

