

BAB I

PENDAHULUAN

1.1 Latar Belakang

Untuk memodelkan hubungan antara variabel respon dan variabel prediktor dapat digunakan metode analisis regresi. Analisis regresi terdiri atas regresi parametrik, nonparametrik dan semiparametrik. Analisis regresi nonparametrik merupakan salah satu metode statistika yang digunakan saat bentuk kurva regresi tidak diketahui, atau saat bentuk pola hubungan variabel respon dan variabel prediktor tidak dapat diidentifikasi berdasarkan informasi masa lalu. Saat ini, beberapa kasus riil tidak memiliki informasi terkait pola hubungan variabelnya, sehingga penelitian dengan metode regresi nonparametrik banyak dikembangkan karena lebih relevan (Sihombing & Famalika, 2022). Pada regresi nonparametrik kurva regresinya diasumsikan *smooth* yang termuat dalam suatu ruang fungsi tertentu. Data diharapkan mencari bentuk estimasinya sendiri, sehingga lebih fleksibel (Purnaraga et al., 2020).

Terdapat beberapa teknik estimasi dalam regresi nonparametrik diantaranya *spline truncated*, *smoothing spline*, *penalized spline*, kernel, polinomial lokal, deret fourier, wavelets, dan lain-lain. Estimator yang banyak digunakan saat ini yaitu *spline truncated* (Dani & Adrianingsih, 2021). *Spline truncated* yaitu estimator yang memiliki karakteristik perubahan pola perilaku kurva yang berbeda pada sub-interval tertentu (Nurchayani et al., 2021). *Spline truncated* memiliki titik-titik knot yang menunjukkan adanya perubahan pola data, sehingga menjadikan model ini fleksibel (Dani & Adrianingsih, 2021). Penelitian yang menggunakan metode *spline truncated* diantaranya, yaitu penelitian Widyastuti et al. (2021) yang menghasilkan model dengan nilai koefisien determinasi sebesar 87,65%.

Selain itu, estimator lain yang menjadi sorotan yaitu kernel. Estimator kernel memiliki bentuk pembobot yang didasarkan pada fungsi kernel, dengan ukuran bobot yang ditentukan oleh parameter pemulus atau *bandwidth* (Wulandary & Purnama, 2020). Secara matematis perhitungan estimator kernel mudah dikerjakan (Rahasia et al., 2020). Beberapa fungsi kernel di antaranya, yaitu fungsi kernel *uniform*, kernel *triangular*, kernel *biweight*, kernel *triweight*, kernel *gaussian*, dan kernel *epanechnikov* (Farida et al., 2022). Salah satu fungsi kernel yang banyak digunakan adalah fungsi kernel *epanechnikov*. Fungsi ini mempunyai laju konvergensi yang lebih cepat menuju nilai yang diestimasi dibanding fungsi kernel yang lainnya (Moraes et al., 2021). Salah satu penelitian yang menggunakan metode regresi kernel *epanechnikov*, yaitu penelitian Emiliani et al. (2022), model yang diperoleh menghasilkan nilai koefisien 36,90%.



si nonparametrik melibatkan beberapa variabel prediktor, setiap nggap memiliki pola yang sama, sehingga digunakan satu bentuk ua variabel prediktornya, sementara setiap variabel prediktor gan data yang berbeda. Hal ini mengakibatkan estimator yang suai dengan pola data. Oleh karena itu, permasalahan ini diatasi an estimator kurva campuran.

Pengembangan estimator campuran telah dilakukan oleh beberapa peneliti, diantaranya penelitian yang dilakukan oleh R. Hidayat et al. (2019), yaitu estimasi regresi nonparametrik kernel-spline pada model aditif, dalam penelitian ini menggunakan beberapa fungsi kernel, metode ini diterapkan pada data *cross section*. Penelitian Sifriyani et al., (2023), yang melakukan pemodelan statistika untuk mengestimasi kurva regresi menggunakan estimator campuran spline *truncated* dan kernel *epanechnikov*, metode ini juga diaplikasikan pada data *cross section*. Kemudian penelitian Dewanti et al. (2020) yang juga memodelkan data *cross section*, yaitu data pencapaian *Sustainable Development Goals* di Jawa Timur dengan menggunakan metode regresi nonparametrik birespon dengan estimator campuran spline *truncated* dan kernel. Metode campuran estimator pada penelitian sebelumnya diterapkan pada data *cross section*, sementara dalam pengembangannya, regresi nonparametrik dapat diterapkan tidak hanya terbatas pada data *cross section* saja, tetapi juga pada data longitudinal, sehingga pada penelitian ini dikembangkan metode campuran yang diterapkan pada data longitudinal.

Data longitudinal adalah data yang diperoleh dari pengamatan berulang terhadap beberapa objek. Pengamatan dilakukan dari waktu ke waktu pada setiap subjek penelitian (Hidayat et al., 2022). Penggunaan data longitudinal dapat memberikan informasi terkait perubahan pada subjek. Pengamatan yang berulang menyebabkan subjek yang dibutuhkan tidak terlalu banyak, serta estimasinya menjadi lebih efisien karena dilakukan pada setiap pengamatan (Hulin & Jin-Ting, 2006). Penelitian regresi nonparametrik yang menggunakan data longitudinal diantaranya, yaitu penelitian dari Hermawan (2021) yang mengestimasi kurva regresi spline pada data longitudinal. Kemudian Penelitian dari Octavanny et al. (2020) yang memodelkan data longitudinal dengan metode spline dan deret fourier.

Berdasarkan uraian tersebut, dilakukan analisis regresi nonparametrik dengan melibatkan dua jenis estimator pada variabel prediktor yang digunakan. Penelitian ini fokus pada model regresi nonparametrik campuran spline *truncated* dan kernel, khususnya fungsi kernel *epanechnikov*, yang kemudian diterapkan pada data longitudinal. Bentuk estimasi parameter model diperoleh dengan metode *weighted least square* (WLS). Dalam proses pemodelan dievaluasi titik knot yang berbeda untuk setiap variabel pada regresi spline *truncated*, serta lebar *bandwidth* yang berbeda untuk setiap variabel pada regresi kernel *epanechnikov*. Hasil estimasi parameter secara teoritis kemudian diaplikasikan pada data riil yaitu keberhasilan program pengendalian kependudukan di Sulawesi Selatan.

Di Sulawesi Selatan, program pengendalian penduduk masih menjadi perhatian, hal ini karena laju pertumbuhan penduduk pada setiap kabupaten di Sulawesi Selatan terus mengalami peningkatan dari tahun ke tahun (BPS Provinsi Sulawesi Selatan,



akan program pengendalian kependudukan merupakan program

Penanganan permasalahan sosial ini merupakan peran dari keluarga Berencana Nasional (BKKBN). Melalui program keluarga berencana dapat mengurangi laju pertumbuhan penduduk agar tidak menjadi seimbang. Salah satu program utama dari BKKBN pasangan usia subur (PUS) program KB. Dalam rencana strategis dituangkan bahwa indikator strategis untuk mencapai dampak

penurunan laju pertumbuhan penduduk yaitu dengan angka prevalensi penggunaan kontrasepsi modern (*Modern Contraceptive Prevalence Rate/mCPR*). Presentase mCPR yang tinggi menjadi acuan untuk mengukur keberhasilan dari program KB (BKKBN, 2020).

Oleh karena itu, data yang digunakan dalam penelitian ini yaitu data prevalensi penggunaan kontrasepsi modern. Data yang digunakan mewakili setiap kabupaten/kota di Sulawesi Selatan pada tahun 2019 sampai dengan 2023. Data kemudian dianalisis untuk menaksir prevalensi penggunaan kontrasepsi modern, kemudian diinterpretasikan sebagai indikator keberhasilan program KB.

1.2 Rumusan Masalah

Rumusan masalah yang dikaji dalam penelitian ini sebagai berikut:

1. Bagaimana bentuk estimasi parameter model regresi nonparametrik campuran spline *truncated* dan kernel *epanechnikov* pada data longitudinal?
2. Bagaimana aplikasi regresi nonparametrik campuran spline *truncated* dan kernel *epanechnikov* pada data longitudinal dalam studi kasus prevalensi penggunaan kontrasepsi modern di Sulawesi Selatan?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini dirincikan sebagai berikut:

1. Estimasi model regresi nonparametrik campuran spline *truncated* dan kernel *epanechnikov* menggunakan metode *Weighted Least Square* (WLS).
2. Model regresi nonparametrik spline *truncated* yang digunakan adalah spline kubik.
3. Model regresi nonparametrik kernel *epanechnikov* menggunakan estimator *Nadaraya-Watson*.
4. Pemilihan titik knot optimum dan *bandwidth* optimum menggunakan metode *Generalized Cross Validation* (GCV).
5. Nilai *bandwidth* yang digunakan yaitu nilai antara 0,1 sampai 0,9 dengan perbedaan setiap nilai adalah 0,1.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan dari penelitian ini, yaitu:

1. Memperoleh bentuk estimasi parameter model regresi nonparametrik campuran spline *truncated* dan kernel *epanechnikov* pada data longitudinal.
2. Memperoleh model regresi prevalensi penggunaan kontrasepsi modern di Sulawesi Selatan dengan pendekatan nonparametrik campuran spline *truncated* dan kernel *epanechnikov* pada data longitudinal.

1.5 Manfaat Penelitian



penelitian ini diuraikan sebagai berikut:

dapat dijadikan sebagai bahan rujukan dan pengembangan atistika khususnya model regresi nonparametrik spline *truncated* *echnikov*.

2. Memberikan informasi kepada pemerintah daerah Sulawesi Selatan untuk menindaklanjuti tingkat keberhasilan program pengendalian kependudukan di Sulawesi Selatan.

1.6 Teori

1.6.1 Analisis regresi nonparametrik

Analisis regresi adalah metode statistik yang digunakan untuk menganalisis dan memodelkan hubungan antara variabel respon dan variabel prediktor. Manfaat utama dari analisis regresi yaitu memungkinkan untuk menghitung signifikansi hubungan antara satu atau sekumpulan variabel prediktor terhadap variabel respon, memperkirakan kekuatan pengaruh dari berbagai variabel prediktor terhadap variabel respon, serta membuat prediksi (Sarstedt & Mooi, 2019).

Analisis regresi terdiri atas pendekatan parametrik, nonparametrik, dan semiparametrik. Regresi parametrik digunakan saat bentuk kurva regresi diketahui, misalnya linier, kuadratik, kubik atau polinomial. Regresi nonparametrik digunakan ketika kurva regresi tidak diketahui, sementara regresi semiparametrik adalah model yang digunakan ketika terdapat sebagian kurva regresi yang diketahui dan sebagian lainnya tidak diketahui (Widyastuti et al., 2021).

Regresi nonparametrik merupakan metode analisis yang lebih fleksibel dalam memodelkan hubungan antara variabel respon dan variabel prediktor. Model regresi nonparametrik diformulasikan dalam Persamaan (1).

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

Dengan y_i adalah variabel respon, $f(x_i)$ adalah fungsi halus yang tidak diketahui, dan ε_i adalah error yang diasumsikan independen dengan rata-rata 0 dan varians σ^2 (Čížek & Sadıkoğlu, 2020).

1.6.2 Data longitudinal

Data longitudinal adalah data pengamatan yang dikumpulkan secara berulang untuk setiap subjek. Setiap subjek pengamatan diamati dalam kurun waktu yang berbeda. Subjek data longitudinal diasumsikan saling independen terhadap subjek yang lain, tetapi antar pengamatan dalam subjek saling dependen (Hulin & Jin-Ting, 2006).

Pada data longitudinal, variabel prediktor x_{ijl} menyatakan subjek ke- i pada pengamatan waktu ke- j dan untuk prediktor ke- l , sementara variabel respon y_{ij} diamati secara longitudinal pada subjek ke- i dan waktu pengamatan ke- j , sehingga data longitudinal dapat dituliskan sebagai berikut (Rachma et al., 2024):

$$(x_{ijl}, y_{ij}), j = 1, 2, \dots, t; i = 1, 2, \dots, n; l = 1, 2, \dots, p$$



t pengamatan pada subjek ke- j dan n menyatakan banyaknya subjek, p adalah jumlah variabel prediktor.

Data longitudinal telah banyak dilakukan dalam beberapa dekade terakhir. Penelitian data longitudinal memiliki beberapa aplikasi praktis dalam berbagai bidang, seperti kedokteran, ekologi, epidemiologi, kesehatan, dan lain-lain. Pada data longitudinal, pengamatan terhadap individu atau subjek penelitian

dilakukan selama periode waktu tertentu, sehingga diperoleh pemahaman yang lebih baik dalam hal menganalisis hubungan antara variabel respon dan variabel prediktor.

Data longitudinal dapat dikumpulkan baik secara prospektif yaitu subjek mengikuti waktu yang bergerak maju ataupun retrospektif yaitu dengan mengumpulkan beberapa pengukuran setiap subjek di masa lalu. Untuk struktur data longitudinal multivariabel diberikan pada tabel di bawah ini.

Tabel 1. Struktur Data Longitudinal

Subjek ke- i	Pengamatan ke- j	Variabel respon y_{ij}	Variabel prediktor (x_{ijt})			
			x_{ij1}	x_{ij2}	...	x_{ijp}
1	1	y_{11}	x_{111}	x_{112}	...	x_{11p}
	2	y_{12}	x_{121}	x_{122}	...	x_{12p}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	t	y_{1t}	x_{1t1}	x_{1t2}	...	x_{1tp}
2	1	y_{21}	x_{211}	x_{212}	...	x_{21p}
	2	y_{22}	x_{221}	x_{222}	...	x_{22p}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	t	y_{2t}	x_{2t1}	x_{2t2}	...	x_{2tp}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	1	y_{i1}	x_{i11}	x_{i12}	...	x_{i1p}
	2	y_{i2}	x_{i21}	x_{i22}	...	x_{i2p}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	t	y_{it}	x_{it1}	x_{it2}	...	x_{itp}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	1	y_{n1}	x_{n11}	x_{n12}	...	x_{n1p}
	2	y_{n2}	x_{n21}	x_{n22}	...	x_{n2p}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	t	y_{nt}	x_{nt1}	x_{nt2}	...	x_{ntp}

1.6.3 Metode *Weighted Least Square*

Metode *Weighted Least Square* (WLS) merupakan salah satu bentuk pengembangan dari *Ordinary Least Square* (OLS). Metode OLS digunakan untuk mengestimasi koefisien regresi dengan meminimumkan jumlah kuadrat error. Metode WLS dikembangkan dengan menyertakan suatu variabel baru, yaitu pembobot W pada penyelesaian optimasi kuadrat terkecilnya (Virgantari et al., 2024). Suatu model dituliskan dalam notasi matriks seperti pada Persamaan (2).

$$Y = X\beta + \varepsilon, \quad (2)$$

diperoleh dengan meminimumkan Persamaan (3).

$$\varepsilon^T W \varepsilon = (Y - X\beta)^T W (Y - X\beta) \quad (3)$$



Metode WLS digunakan dalam regresi yang menggunakan data longitudinal karena antar pengamatan dalam subjek saling dependen sehingga terdapat korelasi. Salah satu matriks pembobot yang dapat digunakan dalam estimasi parameter data longitudinal yaitu $W_i = V^{-1}$ dengan $V^{-1} = \text{var}(y_i)$, sehingga memperhitungkan varians antar subjek dalam perhitungan (Hulin & Jin-Ting, 2006). Matriks pembobot W untuk data longitudinal dapat dituliskan dalam Persamaan (4) (Maulidia et al., 2019).

$$W = \begin{bmatrix} W_1 & 0 & \cdots & 0 \\ 0 & W_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_n \end{bmatrix} \text{ dengan } W_1 = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_1^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_1^2 \end{bmatrix} \quad (4)$$

1.6.4 Regresi spline *truncated* untuk data longitudinal

Spline *truncated* merupakan salah satu pendekatan metode regresi nonparametrik yang banyak digunakan. Spline adalah potongan polinomial yang memiliki sifat tersegmen pada titik knot. Metode spline sangat baik dalam memodelkan data yang polanya berubah pada sub-interval tertentu. Fungsi spline $f(x)$ dengan orde m dan k knot, dapat dituliskan dalam Persamaan (5).

$$s(x_i) = \beta_0 + \sum_{r=1}^{m-1} \beta_r X_i^r + \sum_{k=1}^N \beta_{((m-1)+k)} (X_i - K_k)_+^{m-1} \quad (5)$$

Dengan $i = 1, 2, \dots, n$ subjek, $k = 1, 2, \dots, N$ banyaknya titik knot, $r = 1, 2, \dots, m$ banyaknya orde dan $K_1, K_2, \dots, K_N$ adalah titik-titik knot. Jika nilai $m = 4$, maka diperoleh spline kubik. Fungsi *truncated* didefinisikan sebagai berikut.

$$(X_i - K_k)_+^{m-1} = \begin{cases} (X_i - K_k)^{m-1}, & \text{untuk } x_i \geq K_k \\ 0, & \text{untuk } x_i < K_k \end{cases}$$

Secara umum, model spline *truncated* dituliskan dalam Persamaan (6) (Ramli et al., 2020).

$$s_i = \beta_0 + \beta_1 X_i + \cdots + \beta_{m-1} X_i^{m-1} + \beta_{((m-1)+k)} (X_i - K_1)_+^{m-1} + \cdots + \beta_{((m-1)+N)} (X_i - K_N)_+^{m-1} + \varepsilon_i \quad (6)$$

Model regresi nonparametrik spline *truncated* multivariabel untuk data longitudinal yang memiliki subjek $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, t$ pengamatan untuk setiap subjeknya, serta sebanyak $l = 1, 2, \dots, p$ prediktor, dapat dirumuskan kembali dalam Persamaan (7).

$$s(x_{ij}) = \beta_0 + \sum_{l=1}^p \left(\sum_{r=1}^{m-1} \beta_{l,r} X_{ijl}^r + \sum_{k=1}^N \beta_{l,((m-1)+k)} (X_{ijl} - K_k)_+^{m-1} \right) + \varepsilon_{ij} \quad (7)$$

Dengan,



$$(x_{ijl} - K_k)_+^{m-1} = \begin{cases} (x_{ijl} - K_k)^{m-1}, & \text{untuk } x_{ijl} \geq K_k \\ 0, & \text{untuk } x_{ijl} < K_k \end{cases}$$

dalam bentuk matriks pada Persamaan (8) berikut:

$$s = X[K]\beta + \varepsilon \quad (8)$$

$$\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} = \begin{bmatrix} X_1[K] & 0 & \cdots & 0 \\ 0 & X_2[K] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_n[K] \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

orde matriks s adalah $(t \times n) \times 1$, orde matriks X adalah $(t \times n) \times ((m + N)np + n)$, orde matriks β adalah $((m + N)np + n) \times 1$ dan orde matriks ε adalah $(t \times n) \times 1$ (Dani et al., 2021).

1.6.5 Estimator kernel

Estimator kernel merupakan metode pemodelan dengan memperkirakan nilai ekspektasi bersyarat dari variabel acak untuk memperoleh hubungan non-linear antara variabel prediktor dan variabel respon (Wulandary & Purnama, 2020). Estimator kernel didasarkan pada fungsi densitas peluang dan tidak memerlukan asumsi terkait bentuk distribusi data, sehingga secara umum digunakan sebagai model pendekatan nonparametrik. Estimator kernel hampir sama dengan estimator linier lainnya. Akan tetapi, metode kernel lebih spesifik dalam penggunaan parameter penghalus yang disebut *bandwidth* (h). Estimator kernel terbagi menjadi tiga jenis, yaitu estimator Nadaraya-Watson, estimator Priestly-Chao, dan estimator Gasser-Muller (Saidi et al., 2021). Estimator Nadaraya-Watson menjadi estimator yang banyak digunakan.

Dalam mengestimasi model regresi nonparametrik menggunakan estimator kernel, dilakukan penaksiran nilai ekspektasi bersyarat dari suatu variabel acak, yang dapat dituliskan dalam Persamaan (9) (Wahyuliani et al., 2023).

$$y = m(x_i) + \varepsilon_i$$

$$m(x) = E(Y|X = x) = \frac{\int y f(x, y) dy}{f_x(x)} \quad (9)$$

dengan $f(x, y)$ = fungsi densitas (x, y)

$f_x(x)$ = fungsi densitas marginal (x)

Pembilang dan penyebut kemudian diestimasi sebagai berikut.

$$\hat{f}(x, y) = \frac{1}{nh_x h_y} \sum_{i=1}^n K_x \left(\frac{x - x_i}{h_x} \right) K_y \left(\frac{y - y_i}{h_y} \right) \quad (10)$$

$$\hat{f}_x(x) = \frac{1}{nh_x} \sum_{i=1}^n K_x \left(\frac{x - x_i}{h_x} \right) \quad (11)$$

Estimator $m(x)$ dapat didefinisikan dengan mensubstitusikan Persamaan (10) dan Persamaan (11) ke Persamaan (9).



$$\begin{aligned}
\frac{\int y f(x, y) dy}{f_x(x)} &= \frac{\int y \frac{1}{nh_x h_y} \sum_{i=1}^n K_x\left(\frac{x-x_i}{h_x}\right) K_y\left(\frac{y-y_i}{h_y}\right) dy}{\frac{1}{nh_x} \sum_{i=1}^n K_x\left(\frac{x-x_i}{h_x}\right)} \\
&= \frac{\frac{1}{nh_x} \sum_{i=1}^n K_x\left(\frac{x-x_i}{h_x}\right) \int \frac{y}{h_y} K_y\left(\frac{y-y_i}{h_y}\right) dy}{\frac{1}{nh_x} \sum_{i=1}^n K_x\left(\frac{x-x_i}{h_x}\right)} \\
&= \frac{\frac{1}{nh_x} \sum_{i=1}^n K_x\left(\frac{x-x_i}{h_x}\right) y_i}{\frac{1}{nh_x} \sum_{i=1}^n K_x\left(\frac{x-x_i}{h_x}\right)} \\
&= \frac{\sum_{i=1}^n K_x\left(\frac{x-x_i}{h_x}\right) y_i}{\sum_{i=1}^n K_x\left(\frac{x-x_i}{h_x}\right)} \quad (12)
\end{aligned}$$

Hasil estimasi pada Persamaan (12) disebut estimator Nadaraya-Watson. Untuk $h_x = h$, dan $K_x = K$ maka estimator dari fungsi regresi dapat dituliskan dalam Persamaan (13) (Wahyuliani et al., 2023).

$$m(x_i) = \frac{\sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) y_i}{\sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)} + \varepsilon_i \quad (13)$$

Dengan,

K = fungsi kernel.

h = bandwidth.

x = titik pengamatan yang dipilih dari variabel prediktor.

x_i = nilai variabel prediktor.

1.6.6 Estimator kernel *epanechnikov* untuk data longitudinal

Pada estimator kernel, terdapat beberapa fungsi kernel yang dapat digunakan untuk estimasi data, yaitu fungsi kernel *uniform*, *triangle*, *epanechnikov*, *biweight*, *triweight*, *gaussian*, *cosine*, *tricube*, *logistics*. Untuk fungsi kernel *epanechnikov* dinotasikan dalam Persamaan (14) (Saidi et al., 2021).

$$K\left(\frac{x-x_{ij}}{h}\right) = \frac{3}{4} \left(1 - \left(\frac{x-x_{ij}}{h}\right)^2\right) \quad (14)$$

$K(x) = \frac{3}{4} \left(1 - \left(\frac{x-x_{ij}}{h}\right)^2\right)$ merupakan fungsi dengan



$$K(x) = \begin{cases} \frac{3}{4} \left(1 - \left(\frac{x-x_{ij}}{h}\right)^2\right), & \text{jika } \left|\frac{x-x_{ij}}{h}\right| \leq 1 \\ 0, & \text{jika } \left|\frac{x-x_{ij}}{h}\right| > 1 \end{cases}$$

$$m(x_{ij}) = \frac{\sum_{i=1}^n \sum_{j=1}^t \frac{3}{4} \left(1 - \left(\frac{x - x_{ij}}{h} \right)^2 \right) y_{ij}}{\sum_{i=1}^n \sum_{j=1}^t \frac{3}{4} \left(1 - \left(\frac{x - x_{ij}}{h} \right)^2 \right)} + \varepsilon_{ij} \quad (15)$$

Model regresi nonparametrik kernel *epanechnikov* untuk data longitudinal dengan subjek $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, t$ pengamatan untuk setiap subjeknya, serta variabel prediktor sebanyak $s = 1, 2, \dots, q$ dapat dituliskan dalam Persamaan (16) (R. Hidayat et al., 2019).

$$\mathbf{m} = \mathbf{Z}\mathbf{y} + \boldsymbol{\varepsilon} \quad (16)$$

Persamaan (16) dapat dituliskan dalam bentuk matriks:

$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 \mathbf{y}_1 \\ \mathbf{Z}_2 \mathbf{y}_2 \\ \vdots \\ \mathbf{Z}_n \mathbf{y}_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_n \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Dengan orde masing-masing adalah:

$$\mathbf{m}_{nt \times 1}; \mathbf{Z}_{nt \times nt} \text{ dan } \mathbf{y}_{nt \times 1}$$

Matriks $\mathbf{m}$ merupakan matriks dengan setiap unsurnya terdiri atas $\mathbf{m}_i = (m_{i1} \ m_{i2} \ \cdots \ m_{it})^T$. Matriks $\mathbf{y}$ merupakan matriks dengan tiap unsurnya terdiri atas $\mathbf{y}_i = (y_{i1} \ y_{i2} \ \cdots \ y_{it})^T$. Matriks $\mathbf{Z}$ merupakan matriks diagonal yang unsur diagonalnya adalah matriks $\mathbf{Z}_i$ yaitu matriks bujur sangkar dengan orde $(t \times t)$. Matriks $\mathbf{Z}_i$ dapat dituliskan dalam bentuk:

$$\mathbf{Z}_i = \begin{bmatrix} \sum_{s=1}^q n_1(x_{i1s}) & \sum_{s=1}^q n_1(x_{i2s}) & \cdots & \sum_{s=1}^q n_1(x_{its}) \\ \sum_{s=1}^q n_2(x_{i1s}) & \sum_{s=1}^q n_2(x_{i2s}) & \cdots & \sum_{s=1}^q n_2(x_{its}) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{s=1}^q n_t(x_{i1s}) & \sum_{s=1}^q n_t(x_{i2s}) & \cdots & \sum_{s=1}^q n_t(x_{its}) \end{bmatrix}$$

Untuk mencari nilai dari setiap unsur matriks $\mathbf{Z}_i$ digunakan Persamaan (17) berikut:

$$n_l(x_{ijs}) = \frac{\frac{3}{4} \left(1 - \left(\frac{x_l - x_{ij}}{h} \right)^2 \right)}{\sum_{j=1}^t \frac{3}{4} \left(1 - \left(\frac{x_l - x_{ij}}{h} \right)^2 \right)} \quad (17)$$



width.

amatan ke- l yang dipilih dari nilai variabel prediktor, dengan t .

bel prediktor pada subjek ke- i pengamatan ke- j .

nbobot kernel pada titik pengamatan ke- l , subjek ke- i an/perulangan ke- j dan variabel prediktor ke- s .

1.6.7 Estimator campuran spline *truncated* dan kernel *epanechnikov* untuk data longitudinal

Estimator campuran spline *truncated* untuk mengestimasi kurva regresi $s(x_{ij})$ dan kernel *epanechnikov* untuk mengestimasi kurva regresi $m(x_{ij})$ dapat dituliskan dalam matriks seperti pada Persamaan (18) (Maulidia et al., 2019).

$$\mathbf{y} = \mathbf{X}[K]\boldsymbol{\beta} + \mathbf{Z}(h)\mathbf{y} + \boldsymbol{\varepsilon} \quad (18)$$

Parameter error dapat dituliskan dalam Persamaan (19).

$$\boldsymbol{\varepsilon} = \mathbf{y} - [\mathbf{X}[K]\boldsymbol{\beta} + \mathbf{Z}(h)\mathbf{y}] \quad (19)$$

Estimasi parameter $\boldsymbol{\beta}$ dapat diperoleh dengan menggunakan metode *Weighted Least Square* (WLS).

$$\min_{\boldsymbol{\beta}} \left\{ \left((\mathbf{I} - \mathbf{Z}(h))\mathbf{y} - \mathbf{X}[K]\boldsymbol{\beta} \right)^T \mathbf{W} \left((\mathbf{I} - \mathbf{Z}(h))\mathbf{y} - \mathbf{X}[K]\boldsymbol{\beta} \right) \right\} \quad (20)$$

Sehingga diperoleh estimasi parameter $\boldsymbol{\beta}$ yang dapat dituliskan dalam Persamaan (21).

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}[K]^T \mathbf{W} \mathbf{X}[K])^{-1} \mathbf{X}[K]^T \mathbf{W} (\mathbf{I} - \mathbf{Z})\mathbf{y} \quad (21)$$

Estimasi kurva regresi untuk komponen spline *truncated* dapat dituliskan dalam Persamaan (22).

$$\begin{aligned} \hat{\mathbf{s}} &= \mathbf{X}[K](\mathbf{X}[K]^T \mathbf{W} \mathbf{X}[K])^{-1} \mathbf{X}[K]^T \mathbf{W} (\mathbf{I} - \mathbf{Z})\mathbf{y} \\ \hat{\mathbf{s}} &= \mathbf{A}[K, h]\mathbf{y} \end{aligned} \quad (22)$$

Estimasi nilai $\hat{\mathbf{y}}$ ditunjukkan dalam Persamaan (23).

$$\hat{\mathbf{y}} = \mathbf{M}(K, h)\mathbf{y} \quad (23)$$

Dengan $\mathbf{M}(K, h) = \mathbf{A}[K, h] + \mathbf{Z}(h)$.

Pemilihan titik knot dan *bandwidth* optimum pada estimator dapat dilakukan dengan menggunakan nilai *Generalized Cross-Validation* (GCV). Nilai optimum GCV diperoleh pada $K = K^*$ dan $h = h^*$ yaitu titik knot optimum dan *bandwidth* optimum yang dihitung dengan Persamaan (24) (Sifriyani et al., 2023).

$$GCV(K^*, h^*) = \left(\frac{n \sum_{i=1}^n \sum_{j=1}^t (y_{ij} - \hat{y}_{ij})}{(\text{trace}[\mathbf{I} - \mathbf{M}(K^*, h^*)])^2} \right) \quad (24)$$

$\mathbf{M}(K^*, h^*)$ merupakan nilai matriks pada $K = K^*$ dan $h = h^*$.



an model ditinjau berdasarkan nilai *Mean Square Error* (MSE) atau alahan. Model terbaik yaitu model yang menghasilkan nilai MSE ri suatu model dapat diperoleh dari perhitungan Persamaan (25).

$$MSE = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^t (y_{ij} - \hat{y}_{ij})^2 \quad (25)$$

Untuk mengevaluasi model yang dihasilkan digunakan nilai koefisien determinasi (R^2) yaitu nilai yang menunjukkan seberapa jauh model dalam menerangkan variasi variabel respon yang dihitung dengan Persamaan (26).

$$R^2 = 1 - \left(\frac{\sum_{i=1}^n \sum_{j=1}^t (y_{ij} - \hat{y}_{ij})^2}{\sum_{i=1}^n \sum_{j=1}^t (y_{ij} - \bar{y}_{ij})^2} \right) \quad (26)$$

1.6.8 Prevalensi Penggunaan Kontrasepsi Modern

Program pengendalian penduduk merupakan program prioritas yang tertuang dalam Rencana Pembangunan Jangka Menengah Nasional Tahun 2020-2024. Berdasarkan acuan tersebut maka BKKBN menyusun Rencana Strategis (Renstra) 2020-2024. Dalam Renstra 2020-2024 BKKBN, *output* yang diharapkan adalah menurunnya laju pertumbuhan penduduk yang dapat dicapai dengan beberapa sasaran strategis. Salah satu sasaran strategisnya yaitu meningkatnya angka prevalensi penggunaan kontrasepsi modern (BKKBN, 2020).

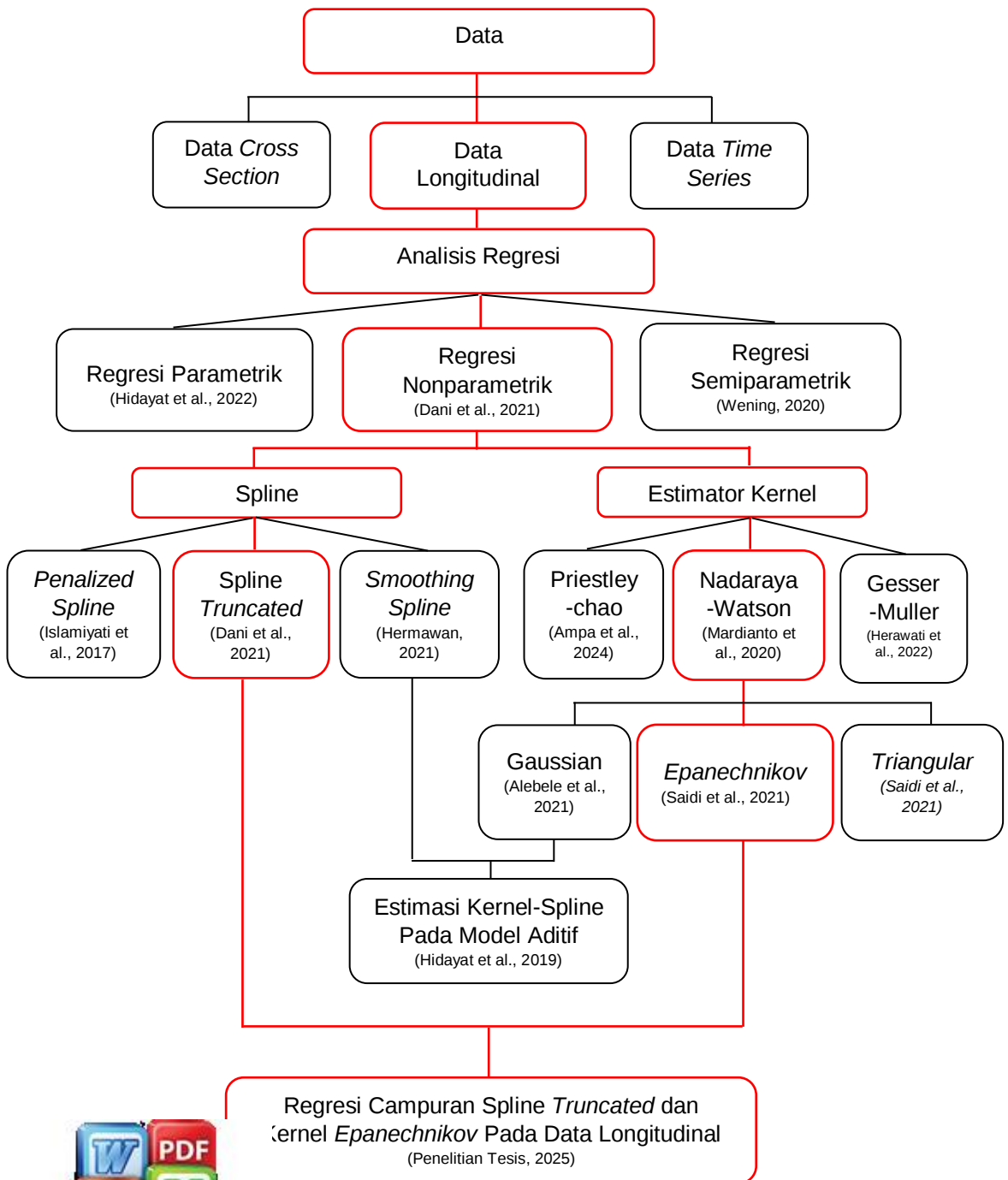
Prevalensi penggunaan kontrasepsi modern dipengaruhi oleh beberapa variabel prediktor. Beberapa penelitian telah dilakukan dalam memodelkan angka prevalensi kontrasepsi modern diantaranya, penelitian yang dilakukan oleh Fayon & Marsisno (2021), yang memodelkan penggunaan kontrasepsi modern dan memperoleh model terbaik yang terdiri dari variabel prediktor persentase penduduk miskin, umur kawin pertama dan pengetahuan pria kawin tentang kontrasepsi. Penelitian yang dilakukan Hardati' et al. (2017), menyimpulkan bahwa faktor-faktor yang mempengaruhi program KB dalam keberhasilan program pengendalian kependudukan adalah kepemilikan anak (2 anak, 1 anak, >2 anak), sikap, dukungan pasangan, keberadaan petugas penyuluh KB, ketersediaan biaya, jumlah anak hidup dan ketersediaan media.

Penelitian dari Setyowati et al. (2020) menyimpulkan bahwa variabel yang berpengaruh secara signifikan terhadap persentase CPR (penggunaan kontrasepsi) di Provinsi Sulawesi Selatan adalah jumlah penduduk miskin, persentase penduduk berumur 15 tahun ke atas dengan pendidikan tertinggi $\leq$ SMA, dan persentase pengguna BPJS yang tersedia di suatu daerah. Budiantara & Cristie (2015) memodelkan CPR dengan variabel prediktor persentase penduduk miskin, persentase wanita berumur 15 tahun ke atas dengan pendidikan tertinggi kurang atau sama dengan SLTP, persentase wanita berumur 10 tahun ke atas dengan usia perkawinan pertama 18 tahun ke bawah, persentase wanita berumur 10 tahun ke atas yang pernah kawin dengan anak lahir hidup kurang atau sama dengan dua, dan persentase wanita berumur 15 tahun ke atas yang bekerja, variabel tersebut terbukti berpengaruh signifikan.

Amelia (2014) memodelkan keberhasilan program KB di Indonesia dengan prevalensi penggunaan KB dan variabel prediktor indeks kedalaman persepsi KK dengan pendidikan SLTP ke bawah, persentase wanita berumur 15 tahun ke atas dengan usia perkawinan pertama 18 tahun ke bawah, dan persentase wanita berumur 10 tahun ke atas yang pernah kawin dengan anak lahir hidup kurang atau sama dengan dua.



1.6.9 Kerangka konsep



Gambar 1. Kerangka Konseptual



BAB II

METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang terkait dengan pengendalian kependudukan di Sulawesi Selatan. Data diperoleh dari portal Pemutakhiran PK – BKKBN pada situs web <https://portalpk.bkkbn.go.id/tabulasi> (diakses pada 29 Oktober 2024), dan data Badan Pusat Statistik (BPS) yang diperoleh dari situs web <https://sulsel.bps.go.id/id/publication>, yaitu pada publikasi buku Sulawesi Selatan Dalam Angka tahun 2020 sampai 2024, serta pada buku Statistik Kesejahteraan Rakyat Sulawesi Selatan tahun 2020 sampai 2024.

2.2 Variabel Penelitian

Variabel respon dan variabel prediktor yang digunakan dalam penelitian ini yaitu:

Tabel 2. Variabel Penelitian

Variabel		Deskripsi	Satuan
Respon	y	Prevalensi penggunaan kontrasepsi modern di Sulawesi Selatan	Persen
Prediktor	x_1	Indeks kedalaman kemiskinan	Indeks
	x_2	Fasilitas pelayanan KB	Unit
	x_3	Persentase wanita dengan usia perkawinan pertama kurang dari 19 tahun	Persen

Sumber: Data Olah 2025

Berikut ini merupakan definisi operasional variabel respon dan variabel prediktor yang digunakan.

1. Prevalensi penggunaan kontrasepsi modern (mCPR) adalah persentase atau angka yang menunjukkan proporsi PUS yang sedang memakai kontrasepsi modern terhadap jumlah seluruh PUS di suatu wilayah. Kontrasepsi modern terdiri atas pil kombinasi, suntikan, implan, kondom, AKDR (Alat Kontrasepsi Dalam Rahim), MOW (Metode Operasi Wanita), MOP (Metode Operasi Pria)/ Vasektomi.
2. Faktor ekonomi dapat berpengaruh terhadap persentase mCPR. Tingkat kesejahteraan yang rendah atau kemiskinan dapat berdampak negatif pada pemakaian kontrasepsi, salah satunya karena keterbatasan dalam mendapatkan akses pelayanan kontrasepsi. Untuk mengukur angka kemiskinan dapat



ndikator, yaitu:

iskinan (P_0)

iskinan merupakan proporsi penduduk yang mengeluarkan per kapita di bawah garis kemiskinan (GK), sementara GK

n jumlah rupiah minimum yang diperlukan seseorang untuk kebutuhan pokok makan dan minum yang setara dengan 2100

kilokalori per kapita per hari serta kebutuhan pokok non-makanan berupa perumahan, pendidikan dan kesehatan. Secara sederhana P_0 merupakan persentase penduduk miskin yang berada di bawah garis kemiskinan.

b. Kedalaman kemiskinan (P_1)

Kedalaman kemiskinan merupakan rata-rata selisih pengeluaran per kapita penduduk miskin dengan garis kemiskinan yang terjadi di masyarakat. Semakin tinggi indeks P_1 , maka semakin jauh rata-rata pengeluaran penduduk dari garis kemiskinan. Sebaliknya, semakin kecil indeks P_1 maka semakin mendekati garis kemiskinan. Nilai P_1 dapat dihitung dengan rumus berikut:

$$P_1 = \frac{1}{n} \sum_{i=1}^q \left[\frac{z - y_i}{z} \right]$$

Dengan

z = garis kemiskinan

y_i = rata-rata pengeluaran per kapita sebulan

q = banyaknya penduduk yang berada di bawah garis kemiskinan

n = jumlah penduduk

c. Keparahan kemiskinan (P_2)

Keparahan kemiskinan merupakan rata-rata dari kuadrat selisih pengeluaran per kapita penduduk miskin dengan garis kemiskinan. Keparahan kemiskinan memberikan gambaran mengenai penyebaran pengeluaran di antara penduduk miskin. Semakin tinggi nilai indeks P_2 , ketimpangan pengeluaran di antara penduduk miskin juga semakin tinggi. Nilai P_2 dapat dihitung dengan rumus berikut:

$$P_2 = \frac{1}{n} \sum_{i=1}^q \left[\frac{z - y_i}{z} \right]^2$$

Variabel prediktor yang digunakan dalam penelitian ini yaitu indikator kedalaman kemiskinan atau P_1 .

3. Fasilitas pelayanan KB yaitu jumlah klinik keluarga berencana (KKB) dan pos pelayanan keluarga berencana desa (PPKBD). Fasilitas ini memberikan pelayanan KB terhadap masyarakat setempat. Ketersediaan fasilitas pelayanan KB yang mampu dijangkau penduduk akan mempengaruhi persentase CPR.
4. Persentase wanita dengan usia perkawinan pertama kurang dari 19 tahun. Usia 20-25 tahun merupakan usia ideal menikah bagi wanita secara fisik dan mental, sementara wanita yang menikah saat berusia di bawah 19 tahun terbilang masih muda. Malinda (2012) menyatakan bahwa penggunaan kontrasepsi merupakan konsekuensi dari fertilitas tinggi. Remaja yang menggunakan kontrasepsi adalah



rtilitasnya tinggi untuk mencegah fertilitas yang lebih tinggi lagi, iat remaja belum mempunyai anak, biasanya mereka cenderung ikan kontrasepsi. Selain alasan fertilitas, ketakutan akan efek B merupakan penyebab paling banyak sehingga wanita remaja nakan alat KB.

2.3 Metode Analisis

Tahapan analisis yang dilakukan untuk mencapai tujuan penelitian ini yaitu:

1. Menentukan bentuk estimator regresi nonparametrik campuran spline *truncated* dan kernel *epanechnikov* pada data longitudinal.

- a. Membuat model regresi nonparametrik dengan variabel respon y_{ij} serta variabel prediktor $x_{ij1}, \dots, x_{ijp}$ dan $x_{ij1}, \dots, x_{ijq}$:

$$y_{ij} = s(x_{ij}) + m(x_{ij}) + \varepsilon_{ij} \quad (27)$$

dengan $s(x_{ij})$ merupakan komponen variabel prediktor yang didekati dengan kurva regresi spline *truncated*, dan $m(x_{ij})$ merupakan kurva regresi yang didekati dengan fungsi kernel *epanechnikov*.

- b. Menyajikan model estimasi campuran regresi spline *truncated* dan kernel *epanechnikov* untuk data longitudinal dalam bentuk matriks, seperti pada Persamaan (18).

$$y = X[K]\beta + Z(h)y + \varepsilon$$

- c. Memperoleh nilai estimasi $\hat{\beta}$ sebagai komponen spline *truncated* menggunakan metode WLS dengan menyelesaikan optimasi:

$$\varepsilon = y - X[K]\beta - Z(h)y$$

$$\varepsilon = [I - Z(h)]y - X[K]\beta$$

Sehingga, diperoleh:

$$\varepsilon^T W \varepsilon = \left((I - Z(h))y - X[K]\beta \right)^T W \left((I - Z(h))y - X[K]\beta \right)$$

Sebagaimana yang disajikan pada Persamaan (20).

Syarat perlu untuk mendapatkan nilai optimal dari Persamaan (20) adalah:

$$\frac{\partial Q(\beta)}{\partial \beta} = 0$$

Sehingga diperoleh:

$$\hat{\beta} = (X[K]^T W X[K])^{-1} X[K]^T W (I - Z)y$$

- d. Memperoleh estimasi campuran regresi spline *truncated* dan kernel *epanechnikov* pada regresi nonparametrik multivariabel pada data longitudinal.

$$\hat{y}_{ij} = \hat{s}(x_{ij}) + \hat{m}(x_{ij}) \quad (28)$$



si nonparametrik campuran spline *truncated* dan kernel ada data longitudinal dalam studi kasus keberhasilan program pendudukan di Sulawesi Selatan.
eksplorasi data (deskripsi variabel).
scatter plot hubungan antara variabel respon dengan masing-masing variabel prediktor.

- c. Memilih titik knot pada estimator model spline *truncated* untuk variabel yang diestimasi dengan spline *truncated*.
- d. Memilih *bandwidth* pada estimator model kernel *epanechnikov* untuk variabel yang diestimasi dengan kernel.
- e. Memodelkan data dengan regresi nonparametrik menggunakan estimator campuran spline *truncated* dan kernel *epanechnikov*. Setiap variabel prediktor dimodelkan dengan beberapa kombinasi estimator (spline *truncated* atau kernel *epanechnikov*).
- f. Memilih titik knot dan *bandwidth* optimal dengan metode GCV yang dihitung berdasarkan Persamaan (24).

$$GCV(K^*, h^*) = \left(\frac{n \sum_{i=1}^n \sum_{j=1}^t (y_{ij} - \hat{y}_{ij})}{(\text{trace}[\mathbf{I} - \mathbf{M}(K^*, h^*)])^2} \right),$$

Dengan $\mathbf{M}(K, h) = \mathbf{A}[K, h] + \mathbf{Z}(h)$, dengan $\mathbf{A}[K, h] + \mathbf{Z}(h)$ adalah matriks komponen campuran spline *truncated* dan kernel *epanechnikov*.

- g. Menetapkan model terbaik dari nilai MSE terkecil yang dihasilkan oleh model. Nilai MSE dihitung berdasarkan Persamaan (25).

$$MSE = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^t (y_{ij} - \hat{y}_{ij})^2$$

- h. Mengevaluasi model dengan nilai koefisien determinasi yang dihitung berdasarkan Persamaan (26).

$$R^2 = 1 - \left(\frac{\sum_{i=1}^n \sum_{j=1}^t (y_{ij} - \hat{y}_{ij})^2}{\sum_{i=1}^n \sum_{j=1}^t (y_{ij} - \bar{y}_{ij})^2} \right)$$

- i. Menginterpretasikan dan menarik kesimpulan dari model terbaik.

