

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Perceraian adalah putusnya ikatan dalam hubungan suami istri, yang mengakibatkan putusnya hukum perkawinan sehingga keduanya tidak lagi berkedudukan sebagai suami istri (Rijaya, 2021). Berdasarkan data Badan Pusat Statistik (BPS) terkait jumlah kasus perceraian di Indonesia pada tahun 2021 dan pada tahun 2022, terjadi peningkatan sebesar 15,3 persen. Salah satu wilayah dengan jumlah kasus perceraian terbanyak dalam beberapa tahun terakhir adalah Provinsi Jawa Tengah. Perceraian dapat terjadi karena berbagai faktor, seperti faktor ekonomi dan faktor pendidikan. Adapun dalam penelitian ini, data perceraian yang digunakan berbentuk data longitudinal. Data longitudinal merupakan data dengan beberapa subjek pengamatan yang saling bebas, dengan setiap subjek diamati secara berulang dalam kurun waktu tertentu (Liang dan Zeger, 1986). Oleh sebab itu, perlu sebuah metode analisis yang dapat diterapkan pada data longitudinal sehingga dapat digunakan untuk menjelaskan pengaruh faktor-faktor tersebut terhadap peningkatan dan penurunan jumlah perceraian di Provinsi Jawa Tengah. Salah satu metode yang dapat digunakan adalah metode analisis regresi.

Analisis regresi merupakan sebuah metode dalam ilmu statistika yang memberikan penjelasan tentang pola hubungan antara variabel independen dan variabel dependen (Drapper & Smith, 1992). Analisis regresi terbagi menjadi tiga model yaitu regresi parametrik, regresi semiparametrik, dan regresi nonparametrik. Jika bentuk kurva regresi diketahui, maka pendekatan model regresi yang digunakan adalah pendekatan model regresi parametrik (Islamiyati, dkk., 2017). Jika bentuk pola data diketahui sementara sebagian bentuk pola data lainnya tidak diketahui, maka pendekatan model regresi yang disarankan adalah pendekatan model regresi semiparametrik (Wahba, 1990). Jika bentuk kurva regresi tidak diketahui atau tidak terdapat informasi yang lengkap tentang bentuk pola datanya, maka pendekatan model regresi yang dapat digunakan adalah pendekatan model regresi nonparametrik (Lestari, dkk., 2018). Penelitian ini menggunakan metode regresi nonparametrik dengan mempertimbangkan bahwa pola hubungan antara variabel dependen dan variabel independen pada kurva regresi tidak diketahui.

Metode regresi nonparametrik tidak menggunakan bentuk asumsi tertentu melainkan dengan pendekatan fungsi dengan teknik *smoothing* pada data pengamatan untuk melakukan pendugaan terhadap pola sebaran pada data. Teknik *smoothing* bertujuan untuk menentukan dan mengestimasi bentuk fungsi regresi dalam model regresi nonparametrik (Chamidah & Lestari, 2022). Salah satu diantara beberapa teknik dalam regresi nonparametrik yang dapat digunakan adalah regresi kernel (Hardle, 1990). Regresi kernel memiliki keunggulan sifat yang lebih fleksibel dan memiliki perhitungan matematis yang mudah diselaraskan serta memiliki nilai rata-rata memusat yang relatif cepat. Selain itu, metode kernel mampu memodelkan data yang tidak memiliki pola tertentu dengan baik.

Estimator *Nadaraya-Watson* merupakan salah satu estimator yang paling sering digunakan di antara estimator regresi Kernel lainnya. Estimator *Nadaraya-Watson* memperkirakan nilai rata-rata suatu variabel dependen berdasarkan variabel independen. Estimator ini memberikan bobot yang berbeda pada titik data yang berbeda, dengan titik data yang lebih dekat dengan nilai independen yang ingin diprediksi memiliki bobot yang lebih besar. Selain estimator, regresi kernel memiliki model fungsi yang beragam, diantaranya fungsi *uniform*, fungsi *triangle*, fungsi *epanechnikov*, fungsi *quartik*, fungsi *triweight*, fungsi *gaussian* dan fungsi *cosinus*. Pada penelitian ini menggunakan fungsi *Epanechnikov*. Kelebihan fungsi kernel *epanechnikov* adalah memberikan bobot yang berkurang secara kuadratik pada setiap titik data dalam jangkauan *bandwidth* yang ditentukan. Pemilihan *bandwidth* yang tepat sangat penting karena akan mempengaruhi kemulusan kurva regresi serta tingkat kesalahan. Metode yang digunakan memperoleh nilai *bandwidth* optimal digunakan metode *Generalized Cross Validation* (GCV).

Terdapat beberapa penelitian terdahulu terkait topik dan metode yang akan digunakan dalam penelitian ini diantaranya adalah penelitian yang dilakukan Lia (2021) tentang penentuan *bandwidth* optimal pada estimasi model regresi nonparametrik pada data *time series* dengan estimator *Nadaraya-Watson* menggunakan fungsi kernel *epanechnikov*. Terdapat juga penelitian oleh Armin (2022) tentang metode penentuan *bandwidth* terbaik berdasarkan nilai *Mean Square Error* dengan model regresi nonparametrik pada data *time series* dengan estimator *Priestly-Chao* dengan fungsi kernel *gaussian*. Selain kedua penelitian tersebut, Laode (2023) melakukan penelitian tentang regresi nonparametrik *B-spline* pada data longitudinal. Berdasarkan uraian di atas, maka peneliti mengambil judul “Estimasi Model Regresi Nonparametrik Kernel dengan Estimator *Nadaraya-Watson* pada Data Longitudinal”.

## 1.2 Rumusan Masalah

Rumusan masalah dalam penelitian yakni bagaimana hasil estimasi model dan *bandwidth* optimal dari data jumlah perceraian di Jawa Tengah menggunakan estimator *Nadaraya-Watson* dengan fungsi kernel *Epanechnikov*?

## 1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data longitudinal tahun 2017 hingga 2024 di Provinsi Jawa Tengah.
2. *Bandwidth* yang digunakan dibatasi dari 0,01 hingga 10 dengan kenaikan 0,01.
3. Pemilihan *bandwidth* optimal menggunakan nilai *Generalized Cross Validation*.
4. Fungsi kernel yang digunakan adalah fungsi kernel *epanechnikov*.

## 1.4 Tujuan Penelitian

Tujuan dalam penelitian ini adalah memperoleh hasil estimasi model dan *bandwidth* optimal dari data jumlah perceraian di Jawa Tengah menggunakan estimator *Nadaraya-Watson* dengan fungsi kernel *Epanechnikov*.

## 1.5 Manfaat Penelitian

Manfaat yang diharapkan dari hasil penelitian ini adalah sebagai berikut:

1. Menambah wawasan dan referensi terkait estimasi model regresi nonparametrik menggunakan estimator *Nadaraya-Watson* dengan fungsi kernel *epanechnikov* pada data longitudinal.
2. Memberikan gagasan baru bagi pembaca dalam mengembangkan metode pemodelan regresi nonparametrik kernel menggunakan estimator *Nadaraya-Watson* dengan fungsi kernel lainnya dalam berbagai kasus.

## 1.6 Teori

### 1.6.1 Data Longitudinal

Data longitudinal merupakan gabungan dari data *cross section* dan *time series*. Data *cross section* adalah data yang terdiri dari satu subjek namun memerlukan sub objek-objek lainnya yang berkaitan pada satu waktu. Adapun data *time series* adalah data yang terdiri dari beberapa waktu periode seperti harian, bulanan, tahunan, dll. (Diggle dkk., 2002). Struktur data longitudinal disajikan pada Tabel 2 berikut:

**Tabel 1.** Struktur Data Longitudinal

| Subjek<br>( <i>i</i> ) | Waktu Pengamatan ke- <i>i</i> ( <i>j</i> ) | $y_{ij}$   | $x_{ij1}$   | $x_{ij2}$   | ...      | $x_{ijl}$   |
|------------------------|--------------------------------------------|------------|-------------|-------------|----------|-------------|
| 1                      | 1                                          | $y_{11}$   | $x_{111}$   | $x_{112}$   | ...      | $x_{11l}$   |
|                        | 2                                          | $y_{12}$   | $x_{121}$   | $x_{122}$   |          | $x_{12l}$   |
|                        | $\vdots$                                   | $\vdots$   | $\vdots$    | $\vdots$    |          | $\vdots$    |
|                        | $m$                                        | $y_{1m_1}$ | $x_{1m_11}$ | $x_{1m_12}$ |          | $x_{1m_1l}$ |
| 2                      | 1                                          | $y_{21}$   | $x_{211}$   | $x_{212}$   | ...      | $x_{21l}$   |
|                        | 2                                          | $y_{22}$   | $x_{221}$   | $x_{222}$   |          | $x_{22l}$   |
|                        | $\vdots$                                   | $\vdots$   | $\vdots$    | $\vdots$    |          | $\vdots$    |
|                        | $m$                                        | $y_{2m_2}$ | $x_{2m_21}$ | $x_{2m_22}$ |          | $x_{2m_2l}$ |
| $\vdots$               | $\vdots$                                   | $\vdots$   | $\vdots$    | $\vdots$    | $\ddots$ | $\vdots$    |
| $n$                    | 1                                          | $y_{n1}$   | $x_{n11}$   | $x_{n12}$   | ...      | $x_{n1l}$   |
|                        | 2                                          | $y_{n2}$   | $x_{n21}$   | $x_{n22}$   |          | $x_{n2l}$   |
|                        | $\vdots$                                   | $\vdots$   | $\vdots$    | $\vdots$    |          | $\vdots$    |
|                        | $m$                                        | $y_{nm_n}$ | $x_{nm_n1}$ | $x_{nm_n2}$ |          | $x_{nm_nl}$ |

Dengan

- $y_{ij}$  : nilai variabel dependen untuk subjek ke-*i* dan waktu ke-*j*  
 $x_{ijk}$  : nilai variabel independen ke-*k* untuk subjek ke-*i* dan waktu ke-*j*  
*i* : subjek, ( $i = 1, 2, \dots, n$ )  
*j* : unit waktu, ( $j = 1, 2, \dots, m$ )  
*k* : variabel independent, ( $k = 1, 2, \dots, l$ )

Keunggulan dari data longitudinal yaitu memiliki kemampuan dalam mengenali pengaruh waktu pengukuran terhadap respon. Data longitudinal diasumsikan subjek-subjek saling independen satu sama lainnya, tetapi antara

pengamatan di dalam subjek yang sama saling dependen, karena cenderung berkorelasi (Wu dan Zhang, 2006).

### 1.6.2 Regresi Nonparametrik

Analisis regresi merupakan sebuah metode dalam ilmu statistika yang memberikan penjelasan tentang pola hubungan antara variabel independen dan variabel dependen (Drapper & Smith, 1992). Analisis regresi terbagi menjadi tiga model yaitu regresi parametrik, regresi semiparametrik, dan regresi nonparametrik. Jika bentuk kurva regresi diketahui, maka pendekatan model regresi yang digunakan adalah pendekatan model regresi parametrik (Islamiyati, dkk., 2017). Jika bentuk pola data diketahui sementara sebagian bentuk pola data lainnya tidak diketahui, maka pendekatan model regresi yang disarankan adalah pendekatan model regresi semiparametrik (Wahba, 1990). Jika bentuk kurva regresi tidak diketahui atau tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola datanya, maka pendekatan model regresi yang dapat digunakan adalah pendekatan model regresi nonparametrik (Lestari, dkk., 2018).

Regresi nonparametrik adalah suatu pendekatan dalam analisis regresi yang tidak mengasumsikan bentuk tertentu untuk kurva regresi. Kurva regresi hanya diasumsikan halus (*smooth*) dalam arti termuat di dalam suatu ruang fungsi tertentu. Data yang digunakan dalam regresi nonparametrik diberi keleluasaan untuk mencari pola kurva regresinya sendiri sehingga sangat fleksibel dan objektif. Beberapa model regresi nonparametrik yang banyak digunakan antara lain *histogram*, *kernel*, *spline*, *polinomial lokal*, *deret ortogonal*, *deret fourier*, *k-nearest neighbor*, *neural network*, *wavelets*, *MARS*, dan yang lain. Semua model memiliki kelebihan dan kekurangan serta motivasi sendiri dalam memodelkan pola (Budiantara, dkk., 2009).

Misalkan  $x$  merupakan variabel independen dan  $y$  merupakan variabel dependen untuk  $n, m$  pengamatan berpasangan  $(x_{ij}, y_{ij})_{i=1, j=1}^{n, m}$ , maka hubungan antara variabel independen dan variabel dependen dapat dinyatakan dengan model regresi nonparametrik pada persamaan 1 (Lia, 2021):

$$y_{ij} = m(x_{ij}) + \epsilon_{ij} \quad (1)$$

Keterangan:

$m(x_{ij})$  : fungsi regresi yang memuat variabel independen dari data pada subjek ke- $i$  dan waktu ke- $j$

$\epsilon_{ij}$  : error pada subjek ke- $i$  dan waktu ke- $j$  yang diasumsikan saling bebas dengan mean 0 dan variansi  $\sigma^2$

### 1.6.3 Fungsi Kernel

Suatu fungsi kernel harus merupakan fungsi kontinu, bernilai riil, simetris, dan terbatas. Secara umum fungsi kernel didefinisikan sebagai berikut (Hardle dalam Putri, 2022):

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right) \quad (2)$$

Dengan  $K_h(x)$  adalah fungsi kernel untuk  $-\infty < x < \infty$ ,  $h > 0$

Fungsi kernel harus memenuhi beberapa syarat, sebagai berikut:

- (i)  $K(x) \geq 0$ , untuk semua  $x$
- (ii)  $\int_{-\infty}^{\infty} K(x)dx = 1$
- (iii)  $\int_{-\infty}^{\infty} x^2 K(x)dx = \sigma^2 > 0$
- (iv)  $\int_{-\infty}^{\infty} xK(x)dx = 0$

Terdapat beberapa fungsi kernel yang umum digunakan dalam melakukan estimasi data (Cheruiyot, Orwa, & Otieno, 2020), sebagai berikut:

**Tabel 2** Fungsi Kernel

| No | Nama Fungsi Kernel  | Persamaan Fungsi Kernel                                    |
|----|---------------------|------------------------------------------------------------|
| 1  | <i>Uniform</i>      | $\frac{1}{2}I$                                             |
| 2  | <i>Epanechnikov</i> | $\frac{3}{4}(1 - x^2)I$                                    |
| 3  | <i>Quartik</i>      | $\frac{15}{16}(1 - x^2)^2I$                                |
| 4  | <i>Triweight</i>    | $\frac{35}{32}(1 - x^2)^3I$                                |
| 5  | <i>Cosinus</i>      | $\frac{\pi}{4} \cos\left(\frac{\pi}{2}x\right)$            |
| 6  | <i>Gaussian</i>     | $\frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}(x^2)\right)}$ |

Dengan  $I$  adalah fungsi indikator

$$I(x) = \begin{cases} 1, & \text{jika } |x| \leq 1 \\ 0, & \text{jika } |x| > 1 \end{cases}$$

Fungsi kernel yang akan digunakan dalam penelitian ini adalah fungsi kernel *epanechnikov*. Fungsi kernel *epanechnikov* merupakan kernel yang mempunyai laju konvergensi lebih cepat menuju nilai yang diestimasi dibanding dengan kernel lainnya (Hardle, 1990). Bentuk umum fungsi kernel *epanechnikov* dinotasikan sebagai berikut:

$$K(x) = \frac{3}{4}(1 - x^2)I \quad (3)$$

#### 1.6.4 Estimator Nadaraya-Watson

Estimator kernel diperkenalkan oleh Rosenblatt (1956) dan Parzen (1962) sehingga disebut estimator densitas kernel Rosenblatt-Parzen (Hardle W., 1994). Estimator kernel merupakan suatu metode pendekatan terhadap fungsi densitas yang belum diketahui dengan menggunakan fungsi kernel. Penghalusan dengan pendekatan kernel dikenal sebagai pemulus kernel (*smoother*) yang sangat

bergantung pada fungsi kernel dan pemilihan *bandwidth* (Saputra & Listyani, 2016). Kelebihan estimator kernel dalam regresi nonparametrik adalah fleksibel, bentuk matematisnya mudah dan dapat mencapai tingkat kekonvergenan yang cepat (Budiantara & Mulianah, 2007).

Salah satu estimator yang digunakan pada metode kernel adalah estimator *Nadaraya-Watson*. Pada tahun 1964 Nadaraya dan Watson mengusulkan suatu estimator regresi kernel untuk memperkirakan  $m$  sebagai rata-rata tertimbang secara lokal dengan menggunakan kernel sebagai fungsi pembobotan (Saputra & Listyani, 2016). Estimator *Nadaraya-Watson* multivariat didefinisikan sebagai berikut:

$$\hat{m}(x_{ij}) = \frac{\sum_{i=1}^n \sum_{j=1}^m \prod_{k=1}^l K_h(x_{ij} - x) y_{ij}}{\sum_{i=1}^n \sum_{j=1}^m \prod_{k=1}^l K_h(x_{ij} - x)} \quad (4)$$

Keterangan:

- $\hat{m}(x_{ij})$  : fungsi penduga regresi  
 $x$  : nilai tertentu dari variabel independent  
 $n$  : jumlah subjek pengamatan  
 $m$  : periode waktu pada subjek pengamatan  
 $h$  : *Bandwidth*

### 1.6.5 Pemilihan Bandwidth Optimal

*Bandwidth* ( $h$ ) adalah parameter pemulus yang berfungsi untuk mengatur kemulusan dari kurva yang diestimasi. Jika *bandwidth* yang dipilih terlalu kecil, maka akan menghasilkan estimasi kurva kurang mulus (*under-smoothing*). Sebaliknya, jika *bandwidth* yang dipilih terlalu besar, maka akan menghasilkan estimasi kurva yang sangat mulus (*over-smoothing*) yang tidak sesuai dengan pola sebaran data. Estimasi kurva bertujuan untuk memperoleh kurva yang mulus namun memiliki tingkat kesalahan yang kecil. Berdasarkan hal tersebut, perlu dipilih nilai *bandwidth* optimal sehingga didapatkan kurva yang mulus dengan nilai *error* yang minimum (Hardle dalam Lia, 2021).

Salah satu metode yang digunakan untuk mendapatkan nilai *bandwidth* yang optimal adalah dengan menggunakan kriteria seperti *Generalized Cross Validation* (GCV). Nilai GCV ditentukan menggunakan persamaan 5, sebagai berikut (Herawati, dkk., 2022):

$$GCV = \frac{MSE}{\left(1 - \frac{\text{trace}(\mathbf{W})}{n}\right)^2} \quad (5)$$

dengan

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{m}(x_i))^2 \quad (6)$$

dan  $\mathbf{W}$  adalah matriks hat dengan ukuran  $n \times n$  yang memenuhi persamaan

$$[m_h(x_1), m_h(x_2), \dots, m_h(x_n)]^t = \mathbf{WY} \quad (7)$$

Keterangan:

$MSE$  : *mean square error* atau kuadrat rata-rata error

$tr$  : trace matriks  
 $I$  : matriks identitas  
 $W$  : matriks *hat*

Penentuan nilai *bandwidth* optimal berdasarkan dari nilai GCV terkecil yang diperoleh dari perhitungan nilai GCV untuk setiap *bandwidth*.

### 1.6.6 Uji Kelayakan Model

Salah satu ukuran yang sering digunakan untuk mengetahui kebaikan suatu model adalah koefisien determinasi ( $R^2$ ). Koefisien determinasi ( $R^2$ ) adalah alat untuk mengukur proporsi keragaman atau variansi total disekitar nilai tengah  $y$  yang dapat dijelaskan oleh model regresi (Alwi & Musfirah, 2021). Adapun rumus  $R^2$  sebagai berikut:

$$R^2 = \frac{SSR}{SST} = \frac{\frac{1}{n \cdot m} \sum_{i=1}^n \sum_{j=1}^m (\hat{y}_{ij} - \bar{y}_{ij})^2}{\frac{1}{n \cdot m} \sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y}_{ij})^2} \quad (8)$$

Jika  $\hat{y}_{ij} = \hat{m}(x_{ij})$ , maka dapat ditulis

$$R^2 = \frac{\frac{1}{n \cdot m} \sum_{i=1}^n \sum_{j=1}^m (\hat{m}(x_{ij}) - \bar{y})^2}{\frac{1}{n \cdot m} \sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y})^2} \quad (8)$$

Keterangan:

$R^2$  : koefisien determinasi  
 $SSR$  : jumlah kuadrat regresi  
 $SST$  : jumlah kuadrat total  
 $\bar{y}$  : nilai rata-rata dari  $y_{ij}$

Nilai  $R^2$  terletak pada interval 0 sampai 1, sehingga saat nilai dari  $R^2$  semakin tinggi atau semakin mendekati 1 maka model yang dihasilkan akan semakin baik untuk digunakan. Sebaliknya, saat nilai dari  $R^2$  semakin rendah atau semakin mendekati 0 maka model yang dihasilkan akan semakin kurang baik (Chin, 1998).

### 1.6.7 Standarisasi Data

Standarisasi data merupakan salah satu metode transformasi data yang digunakan untuk menyeragamkan skala data sehingga mudah untuk dianalisis (Athifaturrofifah, dkk., 2020). Salah satu metode standarisasi data yang umum digunakan adalah standarisasi *Z-score*. *Z-score* merupakan suatu ukuran yang menentukan seberapa besar jarak suatu nilai data terhadap rata-ratanya dalam satuan standar deviasi data tersebut. Nilai *Z-score* akan berada pada suatu titik pada sumbu datar dari kurva normalnya. Standarisasi data *Z-score* dapat ditunjukkan pada persamaan berikut (Yousan, dkk. 2020):

$$X^* = \frac{X - \mu}{\sigma} \quad (9)$$

Keterangan:

- $X^*$  : Nilai standarisasi data dengan *Z-score*  
 $X$  : Data asli  
 $\mu$  : Rata-rata data asli  
 $\sigma$  : Standar deviasi data asli

### 1.6.8 Perceraian

Perceraian adalah putusnya ikatan dalam hubungan suami istri, yang mengakibatkan putusnya hukum perkawinan sehingga keduanya tidak lagi berkedudukan sebagai suami istri (Rijaya, 2021). Perceraian dapat terjadi karena talak atau berdasarkan gugatan perceraian. Perceraian diatur dalam Undang-Undang Nomor 1 Tahun 1974 Pasal 39 tentang Perkawinan (BPS, 2023). Jumlah perceraian merupakan jumlah kasus perceraian dalam suatu wilayah dalam rentang waktu tertentu. Jumlah perceraian dipengaruhi oleh beberapa faktor, menurut Irwan, dkk. (2023) jumlah perceraian dapat dipengaruhi oleh faktor ekonomi. Selain itu, menurut Nisva dan Ratnasari (2020) salah satu faktor yang mempengaruhi jumlah perceraian adalah faktor pendidikan. Variabel yang menjelaskan faktor-faktor tersebut berdasarkan beberapa penelitian sebelumnya adalah sebagai berikut:

#### 1. Tingkat Pengangguran Terbuka

Tingkat pengangguran terbuka (TPT) adalah persentase jumlah pengangguran terhadap jumlah angkatan kerja (BPS, 2023). Pengangguran merupakan keadaan dimana seseorang yang tergolong dalam angkatan kerja ingin mendapatkan pekerjaan tetapi belum dapat memperolehnya (Mukhtar, dkk., 2019). Beberapa kategori yang termasuk dalam pengangguran diantaranya:

- Penduduk yang tak punya pekerjaan dan mencari pekerjaan,
- Penduduk yang tak punya pekerjaan dan mempersiapkan usaha,
- Penduduk yang tak punya pekerjaan dan tidak mencari pekerjaan, karena merasa tidak mungkin mendapatkan pekerjaan.
- Penduduk yang sudah punya pekerjaan, tetapi belum mulai bekerja.

#### 2. Penduduk Miskin

Kemiskinan didefinisikan sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran. Penduduk dikategorikan sebagai penduduk miskin jika memiliki rata-rata pengeluaran per kapita per bulan di bawah garis kemiskinan (BPS, 2023).

#### 3. Pengeluaran Per Kapita Disesuaikan

Pengeluaran per kapita disesuaikan adalah kemampuan masyarakat dalam membelanjakan uangnya dalam bentuk barang maupun jasa. Pengeluaran per kapita disesuaikan ditentukan dari nilai pengeluaran per kapita dan paritas daya beli. Pengeluaran per kapita disesuaikan digunakan untuk menggambarkan tingkat kesejahteraan yang dinikmati oleh penduduk sebagai dampak semakin membaiknya ekonomi (BPS, 2023).

#### 4. Rata-rata Lama Sekolah



Rata-rata lama sekolah (RLS) didefinisikan sebagai jumlah tahun belajar penduduk usia 15 tahun keatas yang telah diselesaikan dalam pendidikan formal (tidak termasuk tahun yang mengulang). Tingginya angka RLS menunjukkan jenjang pendidikan yang pernah/sedang diduduki oleh seseorang. Semakin tinggi angka RLS, maka semakin lama/tinggi jenjang pendidikan yang ditamatkannya (BPS, 2023).

## BAB II METODE PENELITIAN

### 2.1 Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari Badan Pusat Statistik (BPS) Provinsi Jawa Tengah, yaitu publikasi Provinsi Jawa Tengah Dalam Angka. Data yang digunakan adalah data longitudinal sebanyak 280 observasi berupa data *time series* dari tahun 2017-2024 dan data *cross section* meliputi 35 kabupaten/kota di Provinsi Jawa Tengah. Adapun data variabel dependen dan variabel independen yang digunakan pada penelitian ini dapat dilihat pada Tabel 3.

**Tabel 3.** Variabel Dependen dan Variabel independen

| Variabel | Keterangan                         | Satuan |
|----------|------------------------------------|--------|
| $Y$      | Jumlah perceraian                  | Kasus  |
| $X_1$    | Tingkat pengangguran terbuka (TPT) | Persen |
| $X_2$    | Persentase penduduk miskin         | Persen |
| $X_3$    | Rata-rata lama sekolah (RLS)       | Tahun  |
| $X_4$    | Pengeluaran per kapita disesuaikan | Juta   |

### 2.2 Metode Analisis

Penelitian ini menggunakan pendekatan regresi nonparametrik dengan pendekatan fungsi kernel *epanechnikov* menggunakan estimator *Nadaraya-Watson* dalam mengestimasi kurva regresi. Pengolahan data dilakukan dengan bantuan *software* R-Studio. Tahapan analisis data yang dilakukan dalam penelitian ini yaitu sebagai berikut:

1. Membuat statistik deskriptif setiap variabel.
2. Membuat *scatter plot* antara variabel dependen dan variabel independen.
3. Melakukan standarisasi data dengan *Z-score*.
4. Membentuk estimasi model regresi kernel *epanechnikov* dengan *Nadaraya-Watson* berdasarkan persamaan 3 dan persamaan 4.
5. Menentukan *bandwidth* yang digunakan dalam model regresi nonparametrik kernel menggunakan estimator *Nadaraya-Watson* dengan fungsi kernel *epanechnikov* berdasarkan persamaan 5.
6. Menghitung nilai GCV dari setiap *bandwidth*.
7. Memilih *bandwidth* optimal berdasarkan nilai GCV minimum.
8. Memodelkan regresi nonparametrik kernel menggunakan estimator *Nadaraya-Watson* dengan fungsi kernel *epanechnikov*.
9. Melakukan uji kelayakan model menggunakan koefisien determinasi  $R^2$ .
10. Membuat kesimpulan.