

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Analisis regresi adalah salah satu metode statistik yang mempelajari hubungan antara variabel respon dengan variabel prediktor. Estimasi parameter pada regresi linear klasik menggunakan pendekatan *Ordinary Least Square* (OLS). Pada metode OLS koefisien regresi yang diduga berlaku global untuk keseluruhan unit observasi. Model persamaan global akan memberikan informasi yang akurat untuk wilayah lokal jika tidak ada atau hanya ada sedikit keragaman antar wilayah lokalnya (Fotheringham dkk., 2002). Pada kenyataannya, seringkali analisis yang menghasilkan model yang berbasis lokal kewilayahan sangat diperlukan karena bias hasil dari kasus yang diteliti akan beragam dari satu wilayah ke wilayah lain yang disebut heterogenitas spasial (Nadya, 2017).

Salah satu metode regresi yang dapat menangani data yang memiliki heterogenitas spasial adalah *Geographically Weighted Regression* (GWR). Model GWR merupakan teknik regresi lokal yang memungkinkan parameter model bervariasi di setiap lokasi. Model GWR dibangun dari metode pendekatan titik, yaitu berdasarkan posisi koordinat garis lintang (*latitude*) dan garis bujur (*longitude*). Parameter untuk model regresi di setiap lokasi akan menghasilkan nilai yang berbeda-beda. Model GWR dapat mengakomodasi efek spasial, sehingga dapat menjelaskan hubungan antara variabel respon dan prediktor dengan lebih baik (Rahayu, 2017). Namun pada saat pengujian parameter variabel prediktor, terkadang diperoleh beberapa variabel yang tidak signifikan atau tidak mempunyai pengaruh lokasi sehingga variabel tersebut berpengaruh secara global terhadap model (Junus, 2021). Oleh karena itu, model *Geographically Weighted Regression* (GWR) kemudian dikembangkan menjadi model *Mixed Geographically Weighted Regression* (MGWR).

Model *Mixed Geographically Weighted Regression* (MGWR) merupakan gabungan dari model regresi linier global dengan model GWR. Sehingga dengan model MGWR akan dihasilkan estimasi parameter yang sebagian bersifat global dan sebagian yang lain bersifat lokal sesuai dengan lokasi pengamatan data (Purhadi dan Yasin, 2012). Biasanya metode yang digunakan dalam menghasilkan estimasi dari parameter yang memodelkan hubungan variabel respon dengan variabel prediktor adalah metode *Weighted Least Square* (WLS) (Mahmuda, 2015).

Penaksiran parameter memerlukan adanya matriks pembobot, yaitu pemberian pembobot pada data sesuai kedekatan dengan lokasi pengamatan ke- $i$  (Fotheringham dkk., 2002). Untuk membentuk matriks pembobot, digunakan suatu fungsi pembobot, fungsi tersebut dipengaruhi oleh suatu ukuran atau biasa disebut *bandwidth*. Secara teoritis, *bandwidth* adalah ukuran jarak fungsi pembobot dan sejauh mana pengaruh lokasi terhadap lokasi lain. Peran pembobot sangat penting karena nilai pembobot tersebut mewakili letak data observasi satu dengan lainnya sehingga sangat dibutuhkan ketepatan cara pembobotan (Dewi, 2014). Fungsi

pembobot kernel merupakan fungsi pembobot umum yang sering digunakan untuk membentuk matriks pembobot, dan pada penelitian ini menggunakan fungsi pembobot *fixed tricube* kernel.

Adapun data yang dikumpulkan dari setiap lokasi pengamatan sering memuat pencilan. Pencilan adalah nilai ekstrim yang memiliki karakteristik unik dan berbeda dari observasi lainnya (Nurbaroqah dkk., 2022). Pencilan juga merupakan pengamatan yang kemungkinan dapat berpengaruh secara signifikan terhadap koefisien regresi, sehingga membuat estimasi parameter menggunakan WLS dan OLS menjadi bias. Sehingga diperlukan sebuah metode *robust* terhadap keberadaan pencilan untuk mengestimasi parameter model MGWR (Junus, 2021). Metode *MM-estimator* merupakan salah satu metode regresi *robust* yang penting dan luas digunakan. Salah satu kelebihan metode *MM-estimator* yaitu memiliki efisiensi tinggi ketika kesalahan terdistribusi secara normal dan juga memiliki nilai kerusakan (*breakdown point*) yang tinggi. Metode *MM-estimator* pertama kali diperkenalkan dengan menggabungkan *S-estimator* dengan *M-estimator* (Prahutama dan Rusgiyono, 2021). Pada umumnya, pencilan dapat terjadi karena beberapa alasan seperti kesalahan pengukuran, kesalahan sistem, dan kesalahan-kesalahan lainnya.

Penelitian tentang perbandingan metode regresi *robust S-estimator*, *M-estimator*, dan *MM-estimator* oleh Almetwally dan Almongy (2018) dengan hasil yang diperoleh bahwa regresi *robust MM-estimator* lebih efisien dibandingkan metode *robust* lainnya. Adapun penelitian kasus lain oleh Junus (2021) tentang estimasi parameter model *mixed geographically weighted regression* menggunakan metode *GM-estimator* dengan fungsi pembobot *tukey bisquare* yang diterapkan pada data persentase penduduk miskin di Provinsi Sulawesi Selatan. Serta penelitian lain oleh Monalisa (2022) tentang perbandingan *geographically weighted regression* (GWR) dan *mixed geographically weighted regression* (MGWR) pada angka kematian bayi di Indonesia dengan hasil yang diperoleh model MGWR merupakan model terbaik dibandingkan dengan model GWR.

Data spasial dapat ditunjukkan pada data kemiskinan yang diamati pada setiap kabupaten/kota di Provinsi Sulawesi Selatan tahun 2023. Tingginya angka kemiskinan di Indonesia secara keseluruhan tersebar diberbagai provinsi, salah satunya di Provinsi Sulawesi Selatan, jumlah penduduk miskin di Sulawesi Selatan pada Maret 2023 mencapai 788.85 ribu orang, naik 6.5 ribu orang terhadap jumlah penduduk miskin pada September 2022. Persentase penduduk miskin pada Maret 2023 tercatat sebesar 8.70 persen, naik 0.04 persen terhadap persentase kemiskinan pada September 2022. Adapun variabel yang digunakan pada penelitian ini terdiri dari variabel respon yaitu persentase penduduk miskin dan variabel prediktor yaitu persentase penduduk, tingkat pengangguran terbuka, persentase penduduk miskin usia 15 tahun ke atas tidak tamat sekolah dasar, harapan lama sekolah, persentase rumah tangga dengan sumber penerangan listrik non-PLN, persentase pengeluaran per kapita sebulan makanan, indeks kedalaman kemiskinan, tingkat setengah pengangguran, dan persentase rumah tangga yang memiliki akses terhadap sumber air minum layak.

Berdasarkan uraian di atas, maka pada penelitian ini akan dilakukan penerapan model *mixed geographically weighted regression* menggunakan metode *MM-estimator* yang diaplikasikan pada data persentase penduduk miskin di Provinsi Sulawesi Selatan pada tahun 2023. Sehingga penulis mengangkat judul “**Pemodelan *Mixed Geographically Weighted Regression* Menggunakan Metode *MM-estimator* Pada Data Kemiskinan Di Provinsi Sulawesi Selatan**”.

## 1.2 Batasan Penelitian

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Pemodelan MGWR menggunakan fungsi pembobot *tricube fixed* kernel.
2. Pemodelan MGWR dengan metode *MM-estimator* menggunakan fungsi pembobot *tukey bisquare*.
3. Data yang digunakan dalam penelitian ini merupakan data yang diperoleh dari Publikasi Badan Pusat Statistika (BPS) Provinsi Sulawesi Selatan tahun 2023.

## 1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Memperoleh estimasi parameter model MGWR menggunakan metode *MM-estimator*.
2. Memperoleh faktor yang berpengaruh pada data persentase penduduk miskin di Provinsi Sulawesi Selatan berdasarkan model MGWR menggunakan metode *MM-estimator*.

## 1.4 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Memperluas wawasan dan pengetahuan dalam mengaplikasikan model MGWR menggunakan metode *MM-estimator*.
2. Memberikan informasi tambahan pada pemerintah agar menetapkan kebijakan dalam mengatasi kemiskinan di Provinsi Sulawesi Selatan.
3. Sebagai bahan pertimbangan dan bahan referensi untuk melakukan penelitian selanjutnya tentang permasalahan kemiskinan di Provinsi Sulawesi Selatan.

## 1.5 Teori

### 1.5.1 Analisis Regresi

Analisis regresi merupakan suatu metode yang digunakan untuk menganalisis hubungan antar peubah, dimana terdapat peubah prediktor dan peubah respon. Selain itu, metode estimasi biasanya digunakan untuk memperkirakan nilai peubah respon yang dipengaruhi oleh peubah prediktor, dan ini sering dilakukan dengan menggunakan metode *Ordinary Least Squares* atau kuadrat terkecil (Rahayu dkk., 2023). Model regresi biasanya digunakan untuk menganalisis data dari berbagai bidang (Abonazel dan Rabie, 2019). Model regresi yang memuat satu peubah respon dan satu peubah prediktor disebut dengan regresi linier sederhana, sedangkan jika

menggunakan beberapa peubah prediktor maka dinamakan regresi linier berganda. Bentuk umum dari regresi linier berganda dapat ditulis pada persamaan (1) sebagai berikut (Draper dan Smith, 1998):

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i \quad (1)$$

dengan

$y_i$  : nilai pengamatan ke- $i$  peubah respon,  $i = 1, 2, \dots, n$

$x_{ij}$  : nilai pengamatan ke- $i$  peubah prediktor  $j$

$\beta_0$  : nilai intersep model regresi

$\beta_j$  : koefisien parameter regresi untuk peubah prediktor ke- $j$

$\varepsilon_i$  : eror pada pengamatan ke- $i$  dengan asumsi  $\varepsilon_i \sim NIID(0, \sigma^2)$

Estimator atau penduga parameter model dapat diperoleh dengan jumlah kuadrat eror atau yang dikenal dengan *Ordinary Least Square* (OLS) (Rencher, 2000 dalam Apriyani dkk., 2018). Berdasarkan dari persamaan (1) maka dapat diuraikan menjadi persamaan-persamaan berikut:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_k x_{1k} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_k x_{2k} + \varepsilon_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_k x_{nk} + \varepsilon_n \end{aligned}$$

Dapat dibuat dalam bentuk matriks seperti berikut:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

yang dapat ditulis menjadi:

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

Persamaan (2) dapat ditulis dalam bentuk matriks sebagai berikut:

$$\begin{aligned} S &= \varepsilon^t \varepsilon = (Y - X\beta)^t (Y - X\beta) \\ &= Y^t Y - \beta^t X^t Y - Y^t X \beta + \beta^t X^t X \beta \\ &= Y^t Y - 2\beta^t X^t Y + \beta^t X^t X \beta \end{aligned} \quad (3)$$

Untuk memperoleh penduga parameter  $\hat{\beta}$ , maka persamaan (3) dapat diturunkan terhadap  $\beta^t$  dan disamakan dengan nol maka diperoleh sebagai berikut:

$$\begin{aligned} \left. \frac{\partial S}{\partial \beta} \right|_{\beta=\hat{\beta}} &= \frac{\partial (Y^t Y - 2\beta^t X^t Y + \beta^t X^t X \beta)}{\partial \beta} = 0 \\ -2X^t Y + 2X^t X \beta &= 0 \\ X^t X \beta &= X^t Y \\ (X^t X)^{-1} X^t X \beta &= (X^t X)^{-1} X^t Y \\ \hat{\beta} &= (X^t X)^{-1} X^t Y \end{aligned} \quad (4)$$

dengan

$Y$ : vektor variabel respon yang berukuran  $n \times 1$

$X$ : vektor variabel respon yang berukuran  $n \times (k + 1)$

$\hat{\beta}$ : vektor koefisien variabel prediktor yang berukuran  $(k + 1) \times 1$

### 1.5.2 Pengujian Heterogenitas Spasial

Heterogenitas spasial adalah efek yang menunjukkan adanya keragaman antar lokasi. Jadi setiap lokasi mempunyai struktur dan parameter hubungan yang berbeda pada setiap lokasi. Pengujian efek spasial dilakukan dengan uji heteogenitas yaitu uji *Breusch-Pagan test* (Rati, 2013 dalam Nurul, 2017). Heterogenitas spasial terjadi ketika variabel prediktor yang sama mempunyai pengaruh yang tidak sama pada lokasi yang berbeda dalam suatu wilayah (Tyas dan Puspitasari, 2023). Menurut Anselin (1988) hipotesis yang digunakan dalam analisis ini adalah sebagai berikut:

- 1) Hipotesis Pengujian

$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$  (tidak terjadi heterogenitas spasial)

$H_1: \text{minimal ada satu } \sigma_i^2 \neq \sigma_{i^*}^2, \text{ dengan } i = 1, 2, \dots, n \text{ (terjadi heterogenitas spasial)}$

- 2) Statistik Uji

$$BP = \left(\frac{1}{2}\right) \mathbf{f}^t \mathbf{Z} (\mathbf{Z}^t \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{f} \sim \chi_{(\alpha, k)}^2 \quad (5)$$

dengan elemen vektor  $\mathbf{f}$  adalah:

$$f_i = \left( \frac{e_i^2}{\sigma^2} - 1 \right)$$

dan

$e_i$ : eror untuk observasi ke- $i$

$\mathbf{Z}$ : matriks berukuran  $n \times (k + 1)$  yang berisi vektor konstan

- 3) Kriteria Keputusan

Tolak  $H_0$  dan terima  $H_1$  jika  $BP > \chi_{(\alpha, k)}^2$  atau jika  $p\text{-value} < \alpha = 5\%$  dengan  $k$  adalah banyaknya prediktor, sehingga dapat disimpulkan bahwa terjadi heterogenitas spasial.

### 1.5.3 Fungsi Pembobot

Fungsi pembobot bertujuan untuk memberikan hasil pendugaan parameter yang berbeda pada lokasi yang berbeda. Pada analisis spasial, pendugaan parameter di suatu lokasi ke- $i$  dipengaruhi oleh titik-titik yang lebih jauh (Leung dkk., 2000). Pendugaan parameter titik spasial dipengaruhi oleh titik-titik yang berdekatan lokasinya dibandingkan dengan lokasi yang tempat nya berjauhan, pembobot spasial berperan penting karena nilai pembobot yang mewakili letak data observasi satu dengan yang lainnya (Dewi, 2014). Matriks pembobot dapat diperoleh berdasarkan informasi jarak dari ketetanggaan atau dengan kata lain jarak antara satu lokasi pengamatan dengan lokasi pengamatan yang lain. Apabila lokasi ke- $j$  terletak pada koordinat  $(u_i, v_i)$  maka akan diperoleh jarak *euclidean* ( $d_{ij}$ ) antara lokasi pengamatan ke- $i$  dan lokasi pengamatan ke- $j$  dengan menggunakan persamaan sebagai berikut (Amaliasari, 2015):

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (6)$$

dengan  $d_{ij}$  merupakan jarak *euclidean* antara lokasi pengamatan ke- $i$  dengan lokasi pengamatan ke- $j$ ,  $u_i$  merupakan koordinat *longitude* lokasi pengamatan ke- $i$ ,  $u_j$

merupakan koordinat *longitude* lokasi pengamatan ke- $j$ ,  $v_i$  merupakan koordinat *latitude* lokasi pengamatan ke- $i$ ,  $v_j$  merupakan koordinat *latitude* lokasi pengamatan ke- $j$ .

Pembobotan pada model GWR dapat menggunakan fungsi kernel dan pada penelitian ini menggunakan fungsi pembobot *tricube fixed* kernel yang dapat dituliskan sebagai berikut (Pratomo dkk., 2024 ):

$$w_j(u_i, v_i) = \begin{cases} \left[ 1 - \left( \frac{d_{ij}}{h} \right)^3 \right]^3, & \text{untuk } d_{ij} < h \\ 0, & \text{lainnya} \end{cases} \quad (7)$$

dengan  $d_{ij}$  merupakan jarak *euclidean* antara lokasi pengamatan ke- $i$  dengan lokasi pengamatan ke- $j$ , dan  $h$  merupakan nilai *bandwidth*.

#### 1.5.4 Penentuan *Bandwidth*

*Bandwidth* adalah radius suatu lingkaran, sehingga sebuah titik lokasi pengamatan yang berada dalam radius lingkaran masih dianggap berpengaruh. Oleh karena itu, metode pemilihan *bandwidth* optimum sangat penting digunakan untuk pendugaan fungsi pembobot yang tepat. Penentuan *bandwidth* optimum pada proses pembentukan matriks pembobot merupakan hal yang penting dilakukan karena akan berpengaruh pada ketepatan model terhadap data, yaitu mengatur *varians* dan bias dari model (Wulandari, 2017). *Bandwidth* optimum adalah *bandwidth* yang menghasilkan *Cross Validation* (CV) minimum. Secara sistematis nilai *Cross Validation* (CV) dituliskan sebagai berikut (Fotheringham dkk., 2002):

$$CV(h) = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h))^2 \quad (8)$$

dengan

$y_i$  : pengamatan ke- $i$

$\hat{y}_{\neq i}(h)$  : penduga  $y_i$  dengan pengamatan lokasi ke- $i$  dihilangkan dari proses pendugaan.

#### 1.5.5 *Geographically Weighted Regression*

*Geographically Weighted Regression* (GWR) adalah metode yang menghasilkan koefisien regresi yang bervariasi secara sistematis dalam ruang. Model GWR merupakan pengembangan dari model regresi linear global dengan memperhatikan aspek spasial. Dengan memasukkan unsur pembobot geografis dalam pendugaan parameternya, model GWR mampu menghasilkan parameter pada tiap lokasi pengamatan. Model GWR dinyatakan sebagai berikut (Fotheringham dkk., 2002):

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^k \beta_j(u_i, v_i)x_{ij} + \varepsilon_i \quad i = 1, 2, \dots, n \quad (9)$$

dengan

$y_i$  : variabel respon pada lokasi pengamatan ke- $i$

$(u_i, v_i)$  : koordinat letak geografis (*latitude*, *longitude*) lokasi pengamatan ke- $i$

$\beta_0(u_i, v_i)$  : nilai intersep model GWR

$\beta_j(u_i, v_i)$  : koefisien regresi ke- $j$  pada lokasi pengamatan ke- $i$

$x_{ij}$  : variabel prediktor ke- $j$  pada lokasi pengamatan ke- $i$

$\varepsilon_i$  : eror yang diasumsikan berdistribusi normal dengan *mean* nol dan *varians*  $\sigma^2$

Dapat dibuat dalam bentuk matriks seperti berikut:

$$Y = X\beta(u_i, v_i) + \varepsilon \quad (10)$$

Pendugaan parameter pada model GWR menggunakan metode *Weighted Least Square* (WLS), yaitu dengan memberikan bobot yang berbeda untuk setiap wilayah pengamatan (Fotheringham dkk., 2002). Berdasarkan dengan metode WLS maka diperoleh penduga parameter GWR sebagai berikut:

$$\hat{\beta}(u_i, v_i) = [X^t W(u_i, v_i) X]^{-1} X^t W(u_i, v_i) Y \quad (11)$$

dengan  $W(u_i, v_i) = \text{diag}[W_1(u_i, v_i), \dots, W_n(u_i, v_i)]$  adalah matriks pembobot,  $X$  merupakan matriks data dari variabel prediktor, dan  $Y$  merupakan vektor variabel respon.

### 1.5.6 Uji Variabilitas Model GWR

Pengujian variabilitas model GWR dilakukan untuk mengetahui apakah terdapat perbedaan pengaruh yang signifikan dari variabel prediktor antar lokasi pengamatan (Leung dkk., 2000). Pengujian ini akan menghasilkan variabel yang bersifat global dan variabel yang bersifat lokal. Pengujian ini dapat dilakukan dengan hipotesis sebagai berikut:

1) Hipotesis pengujian

$H_0$  :  $\beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \dots = \beta_k(u_i, v_i)$  (Tidak terdapat perbedaan pengaruh yang signifikan dari variabel prediktor  $X_k$  antara satu lokasi dengan lokasi yang lainnya)

$H_1$  : minimal ada satu  $\beta_j(u_i, v_i) \neq \beta_{j^*}(u_i, v_i)$  untuk setiap  $i = 1, 2, \dots, n$  (Terdapat perbedaan pengaruh yang signifikan dari variabel prediktor  $X_k$  antara satu lokasi dengan lokasi yang lainnya)

2) Statistik Uji

$$F = \frac{V_j^2 / \text{tr} \left( \frac{1}{k} B_j^t \left[ I - \frac{1}{k} J \right] B_j \right)}{Y^t (I - S)^t (I - S) Y / b_1} \sim F_{(\alpha; df_1; df_2)} \quad (12)$$

dengan

$$V_j^2 = \frac{1}{n} \sum_{i=1}^n \left( \hat{\beta}_j(u_i, v_i) - \frac{1}{n} \sum_{i=1}^n \hat{\beta}_j(u_i, v_i) \right) = \frac{1}{n} \beta_j^t \left[ I - \frac{1}{n} J \right] \beta_j$$

$V_j^2$  adalah nilai skalar yang merepresentasikan variabilitas lokal parameter  $\beta_j(u_i, v_i)$  di seluruh lokasi  $i = 1, 2, \dots, n$ ,  $\beta_j$  berukuran  $n \times 1$ ,  $\beta_j^t$  berukuran  $1 \times n$ ,  $\left[ I - \frac{1}{n} J \right]$  berukuran  $n \times n$ , hasil dari operasi  $\beta_j^t \left[ I - \frac{1}{n} J \right] \beta_j$  adalah sebuah skalar berukuran  $1 \times 1$ , jadi ukuran akhir  $V_j^2$  adalah  $1 \times 1$ ,

$$B_j = \begin{bmatrix} \varepsilon_k^t (X^t W(u_1, v_1) X)^{-1} X^t W(u_1, v_1) \\ \varepsilon_k^t (X^t W(u_2, v_2) X)^{-1} X^t W(u_2, v_2) \\ \vdots \\ \varepsilon_k^t (X^t W(u_n, v_n) X)^{-1} X^t W(u_n, v_n) \end{bmatrix}$$

Matriks  $B_j$  menunjukkan seberapa besar pengaruh prediksi  $\varepsilon_k$  terhadap variabel prediktor pada setiap lokasi  $(u_i, v_i)$ ,  $\varepsilon_k$  vektor kolom berukuran  $k + 1$

yang bernilai 1 untuk elemen ke-  $j$  dan 0 untuk lainnya,  $(\mathbf{X}^t \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}(u_i, v_i)$  matriks berukuran  $(k+1) \times n$ ,  $\varepsilon_k^t \dots$  vektor berukuran  $1 \times (k+1)$ ,  $\mathbf{B}_j$  berukuran  $n \times (k+1)$ .

$$\boldsymbol{\beta}_j(u_i, v_i) = \begin{bmatrix} \boldsymbol{\beta}_0(u_i, v_i) \\ \boldsymbol{\beta}_1(u_i, v_i) \\ \vdots \\ \boldsymbol{\beta}_k(u_i, v_i) \end{bmatrix}$$

$$b_l = \text{tr}[(\mathbf{I} - \mathbf{S})^t (\mathbf{I} - \mathbf{S})]^l, l = 1, 2.$$

$b_l$  adalah nilai yang menunjukkan informasi terkait *leverage* lokal dari matriks  $\mathbf{S}$ ,  $b_l$  berukuran  $1 \times 1$ .

$$c_l = \text{tr} \left( \frac{1}{n} \mathbf{B}_j^t \left[ \mathbf{I} - \frac{1}{n} \mathbf{J} \right] \mathbf{B}_j \right)^l, l = 1, 2.$$

$c_l$  adalah ukuran seberapa besar pengaruh setiap pengamatan terhadap keseluruhan model atau parameter secara keseluruhan,  $c_l$  berukuran  $1 \times 1$ .

$$\mathbf{S}_l = \begin{bmatrix} \mathbf{x}_{l1}^t (\mathbf{X}_l^t \mathbf{W}(u_1, v_1) \mathbf{X}_l)^{-1} \mathbf{X}_l^t \mathbf{W}(u_1, v_1) \\ \mathbf{x}_{l2}^t (\mathbf{X}_l^t \mathbf{W}(u_2, v_2) \mathbf{X}_l)^{-1} \mathbf{X}_l^t \mathbf{W}(u_2, v_2) \\ \vdots \\ \mathbf{x}_{ln}^t (\mathbf{X}_l^t \mathbf{W}(u_n, v_n) \mathbf{X}_l)^{-1} \mathbf{X}_l^t \mathbf{W}(u_n, v_n) \end{bmatrix}$$

$$\mathbf{S} = \mathbf{S}_l + (\mathbf{I} - \mathbf{S}_l) [\mathbf{X}_g^t (\mathbf{I} - \mathbf{S}_l)^t (\mathbf{I} - \mathbf{S}_l)]^{-1} \mathbf{X}_g^t (\mathbf{I} - \mathbf{S}_l)^t (\mathbf{I} - \mathbf{S}_l)$$

Matriks  $\mathbf{J}$  adalah matriks berukuran  $n \times n$  dengan semua elemen adalah 1.

Nilai statistik uji  $F$  mengikuti distribusi  $F$  dengan derajat bebas  $df_1 = \frac{c_1^2}{c_2}$  dan

$$df_2 = \frac{b_1^2}{b_2}.$$

### 3) Kriteria keputusan

Jika diberikan taraf signifikansi  $\alpha = 5\%$  maka  $H_0$  ditolak apabila nilai statistik uji  $F \geq F_{\alpha, df_1, df_2}$  atau  $p - \text{value} < \alpha$  artinya variabel prediktor bersifat lokal. Sebaliknya apabila nilai statistik uji  $F < F_{\alpha, df_1, df_2}$  atau  $p - \text{value} > \alpha$  artinya variabel prediktor bersifat global (Suritman, 2020).

### 1.5.7 Mixed Geographically Weighted Regression

Metode MGWR adalah suatu metode pemodelan yang menggabungkan model regresi global dengan model regresi lokal. Berdasarkan model GWR bila ternyata variabel prediktor tidak semuanya berpengaruh secara lokal dan ada variabel prediktor yang bersifat global maka model inilah yang disebut model MGWR (Purhadi dan Yasin, 2012). Pada model MGWR beberapa koefisien diasumsikan konstan untuk seluruh titik pengamatan sedangkan yang lain bervariasi sesuai lokasi pengamatan data (Fotheringham dkk., 2002). Model MGWR dengan  $p$  variabel prediktor dan  $q$  variabel prediktor diantaranya bersifat lokal, dengan mengasumsikan bahwa intersep model bersifat lokal dapat dituliskan sebagai berikut:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^q \beta_j(u_i, v_i) x_{ij} + \sum_{j=q+1}^k \beta_j x_{ij} + \varepsilon_i \quad i = 1, 2, \dots, n \quad (13)$$

dengan

- $y_i$  : variabel respon pada lokasi pengamatan ke- $i$
- $(u_i, v_i)$  : koordinat letak geografis (*longitude, latitude*) lokasi pengamatan ke- $i$
- $\beta_0(u_i, v_i)$  : nilai intersep model MGWR
- $\beta_j(u_i, v_i)$  : koefisien regresi ke- $j$  pada lokasi pengamatan ke- $i$
- $\beta_j$  : koefisien regresi global ke- $j$  yang tidak bergantung pada lokasi
- $x_{ij}$  : variabel prediktor ke- $j$  pada lokasi pengamatan ke- $i$
- $\varepsilon_i$  : eror pengamatan ke- $i$  yang diasumsikan berdistribusi normal dengan *mean* nol dan *varians*  $\sigma^2$

Estimasi parameter untuk model MGWR menggunakan teknik yang sama dengan GWR yaitu dengan menggunakan metode *Weighted Least Square* (WLS). Model MGWR dapat dinyatakan dalam bentuk matriks (Nisa dkk., 2022):

$$Y = X_l \beta_l(u_i, v_i) + X_g \beta_g + \varepsilon \quad (14)$$

dengan

- $X_g$  : matriks variabel global
- $X_l$  : matriks variabel lokal
- $\beta_g$  : vektor dari parameter global
- $\beta_l(u_i, v_i)$  : vektor dari parameter lokal

Berdasarkan persamaan (13) misal  $\tilde{Y}$  mendefinisikan  $Y - X_g \beta_g$  maka

$$\tilde{Y} = Y - X_g \beta_g = X_l \beta_l(u_i, v_i) + \varepsilon \quad (15)$$

Jika diperhatikan persamaan tersebut menjadi seperti pada model GWR, sehingga seperti estimasi pada model GWR harus mendefinisikan eror terlebih dahulu, berdasarkan persamaan (15) maka eror dapat didefinisikan dengan

$$\varepsilon = \tilde{Y} - X_l \beta_l(u_i, v_i) \quad (16)$$

Dan fungsi kuadrat eror terbobotnya jika dinotasikan dengan  $R$  adalah sebagai berikut:

$$\begin{aligned} R &= \varepsilon^t W(u_i, v_i) \varepsilon \\ R &= (\tilde{Y} - X_l \beta_l(u_i, v_i))^t W(u_i, v_i) (\tilde{Y} - X_l \beta_l(u_i, v_i)) \\ R &= \tilde{Y}^t W(u_i, v_i) \tilde{Y} - \tilde{Y}^t W(u_i, v_i) X_l \beta_l(u_i, v_i) - \beta_l^t(u_i, v_i) X_l^t W(u_i, v_i) \tilde{Y} \\ &\quad + \beta_l^t(u_i, v_i) X_l^t W(u_i, v_i) X_l \beta_l(u_i, v_i) \\ R &= \tilde{Y}^t W(u_i, v_i) \tilde{Y} - 2 \beta_l^t(u_i, v_i) X_l^t W(u_i, v_i) \tilde{Y} + \beta_l^t(u_i, v_i) X_l^t W(u_i, v_i) X_l \beta_l(u_i, v_i) \quad (17) \end{aligned}$$

Untuk memperoleh penduga parameter  $\beta_l(u_i, v_i)$ , maka persamaan di atas diturunkan terhadap  $\beta_l^t(u_i, v_i)$  dan disamadengankan nol sehingga diperoleh:

$$\begin{aligned} \frac{\partial R}{\partial \beta_l^t(u_i, v_i)} &= 0 \\ -2 X_l^t W(u_i, v_i) \tilde{Y} + 2 X_l^t W(u_i, v_i) X_l \beta_l(u_i, v_i) &= 0 \\ X_l^t W(u_i, v_i) X_l \beta_l(u_i, v_i) &= 0 \\ \beta_l(u_i, v_i) &= [X_l^t W(u_i, v_i) X_l]^{-1} X_l^t W(u_i, v_i) \tilde{Y} \end{aligned}$$

Sehingga estimasi parameter model GWR pada lokasi pengamatan ke- $i$  adalah sebagai berikut (Fotheringham dkk., 2002):

$$\hat{\beta}_l(u_i, v_i) = [X_l^t W(u_i, v_i) X_l]^{-1} X_l^t W(u_i, v_i) \tilde{Y} \quad (18)$$

dengan

$$\widehat{\beta}_l(u_i, v_i) = [\widehat{\beta}_0(u_i, v_i), \widehat{\beta}_1(u_i, v_i), \dots, \widehat{\beta}_q(u_i, v_i)]^t$$

Jika  $x_{li}^t = (1, x_{1i}, x_{2i}, \dots, x_{qi})$  adalah elemen baris ke- $i$  dari matriks  $X_l$ , maka nilai dugaan untuk  $\tilde{y}$  pada lokasi pengamatan ke- $i$  yaitu (Purhadi dan Yasin, 2012):

$$\widehat{y}_i = x_{li}^t \widehat{\beta}_l(u_i, v_i)$$

$$\widehat{y}_i = x_{li}^t ((X_l^t W(u_i, v_i) \omega X_l)^{-1} X_l^t W(u_i, v_i) \omega \tilde{Y})$$

Sehingga secara umum dapat dituliskan dengan:

$$\widehat{Y} = (\widehat{y}_1, \widehat{y}_2, \dots, \widehat{y}_n) = S_l \tilde{Y} \text{ dengan nilai } S_l \text{ adalah}$$

$$S_l = \begin{bmatrix} x_{l1}^t (X_l^t W(u_1, v_1) X_l)^{-1} X_l^t W(u_1, v_1) \\ x_{l2}^t (X_l^t W(u_2, v_2) X_l)^{-1} X_l^t W(u_2, v_2) \\ \vdots \\ x_{ln}^t (X_l^t W(u_n, v_n) X_l)^{-1} X_l^t W(u_n, v_n) \end{bmatrix}$$

Kemudian, substitusikan elemen dari  $\widehat{\beta}_l(u_i, v_i)$  ke dalam model MGWR maka diperoleh sebagai berikut:

$$Y = X_l \beta_l(u_i, v_i) + X_g \beta_g + \varepsilon$$

$$Y = S_l \tilde{Y} + X_g \beta_g + \varepsilon$$

$$Y = S_l (Y - X_g \beta_g) + X_g \beta_g + \varepsilon$$

$$\varepsilon = (I - S_l) Y - (I - S_l) X_g \beta_g$$

dengan menggunakan metode kuadrat terkecil (OLS) maka diperoleh estimasi untuk parameter global yaitu:

$$\varepsilon^t \varepsilon = [(I - S_l) Y - (I - S_l) X_g \beta_g]^t [(I - S_l) Y - (I - S_l) X_g \beta_g]$$

$$\varepsilon^t \varepsilon = Y^t (I - S_l)^t (I - S_l) Y - 2 X_g^t \beta_g^t (I - S_l)^t (I - S_l) Y \\ + X_g^t \beta_g^t (I - S_l)^t (I - S_l) X_g \beta_g$$

Jika persamaan  $\varepsilon^t \varepsilon$  diturunkan terhadap  $\beta_g^t$  dan hasilnya disamadengankan nol maka diperoleh estimasi parameter model global yaitu sebagai berikut:

$$\widehat{\beta}_g = [X_g^t (I - S_l)^t (I - S_l) X_g]^{-1} X_g^t (I - S_l)^t (I - S_l) Y \quad (20)$$

$$\text{dengan } S_g = X_g (X_g^t X_g)^{-1} X_g^t$$

Dengan mensubstitusikan  $\widehat{\beta}_g$  ke dalam persamaan (18) maka akan menghasilkan estimasi koefisien lokal pada lokasi  $(u_i, v_i)$  yaitu:

$$\widehat{\beta}_l(u_i, v_i) = [X_l^t W(u_i, v_i) X_l]^{-1} X_l^t W(u_i, v_i) (Y - X_g \widehat{\beta}_g) \quad (21)$$

$\widehat{\beta}_g$  dan  $\widehat{\beta}_l(u_i, v_i)$  merupakan estimator tak bias untuk estimator  $\beta_g$  dan  $\beta_l(u_i, v_i)$ .

### 1.5.8 Pengujian Hipotesis Model MGWR

Pengujian hipotesis model MGWR terdiri dari dua pengujian yaitu pengujian serentak dan pengujian individu (parsial satu-satu) parameter variabel global dan lokal model MGWR.

#### a. Pengujian Serentak Parameter Model MGWR

Pengujian serentak parameter variabel model MGWR terdiri dari dua yaitu pengujian pada parameter variabel global dan variabel lokal. Pengujian ini dilakukan untuk menguji secara serentak signifikansi dari parameter variabel model MGWR.

## 1) Pengujian Serentak Parameter Variabel Global

Hipotesis yang digunakan dalam pengujian serentak parameter variabel global model MGWR adalah sebagai berikut (Purhadi dan Yasmin, 2012):

$$H_0: \beta_{q+1} = \beta_{q+2} = \dots = \beta_k = 0$$

$H_1$ : minimal ada satu  $\beta_j \neq 0$  untuk suatu  $j = q + 1, q + 2, \dots, k$

Statistik uji:

$$F_1 = \frac{Y^t[(I - S_l)^t(I - S_l) - (I - S)^t(I - S)]Y/r_1}{Y^t(I - S)^t(I - S)Y/u_1} \sim F_{(\alpha; df_1; df_2)} \quad (22)$$

Dengan  $r_i = tr([(I - S_l)^t(I - S_l) - (I - S)^t(I - S)]^i)$ ,  $i = 1, 2, \dots, n$  dan  $u_t = tr([(I - S_l)^t(I - S_l)]^t)$ ,  $t = 1, 2$ . Nilai statistik uji  $F_1$  mengikuti distribusi  $F$  dengan derajat bebas  $df_1$  dan  $df_2$ .

Kriteria keputusan:

Jika diberikan taraf signifikansi  $\alpha = 5\%$  maka  $H_0$  ditolak apabila nilai statistik uji  $F_1 \geq F_{\alpha, df_1, df_2}$  dengan  $df_1 = \frac{r_1^2}{r_2}$  dan  $df_2 = \frac{u_1^2}{u_2}$ .

## 2) Pengujian Serentak Parameter Variabel Lokal

Hipotesis yang digunakan dalam pengujian serentak parameter variabel lokal model MGWR adalah sebagai berikut (Purhadi dan Yasmin, 2012):

$$H_0: \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \dots = \beta_q(u_i, v_i) = 0, \text{ untuk setiap } i = 1, 2, \dots, n$$

$H_1$ : minimal ada satu  $\beta_j(u_i, v_i) \neq 0$ , untuk setiap  $j = 1, 2, \dots, n$

Statistik uji:

$$F_2 = \frac{Y^t[(I - S_g)^t(I - S_g) - (I - S)^t(I - S)]Y/c_1}{Y^t(I - S)^t(I - S)Y/u_1} \sim F_{(\alpha; df_1; df_2)} \quad (23)$$

Dengan  $c_i = tr\left(\left[(I - S_g)^t(I - S_g) - (I - S)^t(I - S)\right]^i\right)$ ,  $i = 1, 2, \dots, n$  dan

$S_g = X_g(X_g^t X_g)^{-1} X_g^t$ . Nilai statistik uji  $F_2$  mengikuti distribusi  $F$  dengan derajat bebas  $df_1$  dan  $df_2$ .

Kriteria keputusan:

Jika diberikan taraf signifikansi  $\alpha = 5\%$  maka  $H_0$  ditolak apabila nilai statistik uji  $F_2 \geq F_{\alpha, df_1, df_2}$  dengan  $df_1 = \frac{c_1^2}{c_2}$  dan  $df_2 = \frac{u_1^2}{u_2}$ .

**b. Pengujian Individu (Parsial Satu-Satu) Parameter Model MGWR**

Pengujian individu (parsial satu-satu) parameter model MGWR terdiri dari dua yaitu pengujian individu parameter variabel global dan lokal. Pengujian ini dilakukan untuk mengetahui variabel global dan lokal yang berpengaruh signifikan secara individu terhadap variabel respon pada model MGWR.

## 1) Pengujian Individu Parameter Variabel Global

Hipotesis yang digunakan dalam pengujian individu parameter variabel global model MGWR adalah sebagai berikut (Purhadi dan Yasin, 2012):

$$H_0: \beta_j = 0, \text{ untuk setiap } j = q + 1, q + 2, \dots, k$$

$H_1$ : minimal ada satu  $\beta_j \neq 0$ , untuk suatu  $j = q + 1, q + 2, \dots, k$

Statistik uji:

$$t_{g_{hitung}} = \frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{g_{jj}}} \sim t_{(\alpha/2, df)} \quad (24)$$

Dengan  $g_{jj}$  adalah elemen diagonal ke- $j$  dari hasil perkalian matriks  $\mathbf{GG}^t$ .

Matriks  $\mathbf{G} = [\mathbf{X}_g^t(\mathbf{I} - \mathbf{S}_l)^t(\mathbf{I} - \mathbf{S}_l)\mathbf{X}_g]^{-1}\mathbf{X}_g^t(\mathbf{I} - \mathbf{S}_l)^t(\mathbf{I} - \mathbf{S}_l)$  dan  $\hat{\sigma}^2 = \frac{\mathbf{Y}^t(\mathbf{I} - \mathbf{S})^t(\mathbf{I} - \mathbf{S})\mathbf{Y}}{\text{tr}((\mathbf{I} - \mathbf{S})^t(\mathbf{I} - \mathbf{S}))}$ .

Kriteria Keputusan:

Pada taraf signifikansi  $\alpha = 5\%$ , maka dapat diambil keputusan tolak  $H_0$  jika

$|t_{g_{hitung}}| > t_{\alpha/2, df}$  dengan  $df = \frac{u_1^2}{u_2}$ .

## 2) Pengujian Individu Paramater Variabel Lokal

Hipotesis yang digunakan dalam pengujian individu parameter variabel lokal model MGWR adalah sebagai berikut (Purhadi dan Yasmin, 2012):

$H_0: \beta_j(u_i, v_i) = 0$ , untuk setiap  $j = 1, 2, \dots, q$  dan  $i = 1, 2, \dots, n$

$H_1$ : minimal ada satu  $\beta_j(u_i, v_i) \neq 0$

Statistik uji:

$$t_{l_{hitung}} = \frac{\hat{\beta}_j(u_i, v_i)}{\hat{\sigma}\sqrt{m_{jj}}} \sim t_{(\alpha/2, df)} \quad (25)$$

Dengan  $m_{jj}$  adalah elemen diagonal ke- $j$  dari hasil perkalian matriks  $\mathbf{MM}^t$ .

Matriks  $\mathbf{M} = [\mathbf{X}_l^t\mathbf{W}(u_i, v_i)\mathbf{X}_l]^{-1}\mathbf{X}_l^t\mathbf{W}(u_i, v_i)(\mathbf{I} - \mathbf{X}_g\mathbf{G})$ .

Kriteria Keputusan:

Pada taraf signifikansi  $\alpha = 5\%$ , maka dapat diambil keputusan tolak  $H_0$  jika

$|t_{l_{hitung}}| > t_{\alpha/2, df}$  dengan  $df = \frac{u_1^2}{u_2}$ .

### 1.5.9 Pencilan

Pencilan adalah pengamatan yang jauh dari pusat data yang mungkin berpengaruh besar terhadap koefisien regresi. Pencilan tidak dapat dihilangkan begitu saja dari pengamatan karena pencilan juga memberikan informasi yang tidak dapat diberikan oleh titik data yang lainnya. Pencilan dapat diabaikan apabila diakibatkan dari kesalahan dalam memasukkan data, kesalahan pengukuran, analisis, atau kesalahan yang lainnya (Draper dan Smith, 1992).

Pencilan berpengaruh terhadap proses analisis data, misalnya terhadap nilai *mean* dan standar deviasi. Oleh karena itu, pencilan dalam suatu pola data harus dihindari. Pencilan menyebabkan *varians* pada data menjadi lebih besar, interval dan *range* menjadi lebar, *mean* tidak dapat menunjukkan nilai yang sebenarnya (bias) dan pada beberapa analisis inferensi, pencilan dapat menyebabkan kesalahan dalam pengambilan keputusan dan kesimpulan (Shodiqin dkk., 2018). Adapun metode yang dapat digunakan untuk mengidentifikasi adanya pencilan yang berpengaruh dalam koefisien regresi adalah sebagai berikut:

### a. Metode Nilai Leverage

Nilai *leverage* dari suatu observasi adalah ukuran untuk mengidentifikasi pencilan dalam variabel prediktor (Fitrianto dan Xin, 2022). Metode ini mengukur pengaruh suatu observasi terhadap besarnya estimasi parameter. Hal ini dapat dilihat dari jarak nilai  $X$  semua observasi. Jika suatu kasus terdiri dari beberapa variabel prediktor maka untuk perhitungan nilai *leverage* dapat dilakukan dengan menggunakan persamaan matriks berikut (Wijaya, 2009 dalam Perihatini, 2018):

$$H = X(X'X)^{-1}X' \quad (26)$$

dengan  $X$  merupakan matriks  $n \times (k + 1)$  dengan  $n$  merupakan banyaknya data pengamatan dan  $k$  merupakan banyaknya variabel prediktor, diagonal dari  $H$  berisi nilai-nilai leverage ( $h_{ii}$ ). *Leverage* untuk kasus ke- $i$  merupakan nilai dari baris ke- $i$  dan kolom ke- $i$  dari  $H$ .

$$h_{ii} = x_i'(X'X)^{-1}x_i \quad (27)$$

dengan  $x_i'$  adalah baris ke- $i$  dari  $X$ , jika nilai  $h_{ii}$  lebih besar dari  $\frac{2(k+1)}{n}$  maka pengamatan ke- $i$  dikatakan pencilan terhadap  $x$  (Junus, 2021).

### b. Metode DFFITS

Metode DFFITS (*Difference in fit Standardized*) merupakan metode yang menampilkan nilai perubahan dalam harga yang diprediksi dimana kasus tertentu dikeluarkan dan yang sudah distandarkan (Dewi dkk., 2016). Nilai DFFITS dapat didefinisikan dalam bentuk persamaan berikut (Montgomery dan Peck, 1982):

$$DFFITS = t_i \left( \frac{h_{ii}}{1-h_{ii}} \right)^{\frac{1}{2}} \quad (28)$$

dengan

$$t_i = \varepsilon_i \sqrt{\frac{n-k-1}{JKG(1-h_{ii}) - \varepsilon_i^2}}$$

dengan  $\varepsilon_i$  merupakan eror pengaman ke- $i$  dari  $JKG$  merupakan jumlah kuadrat eror.

Suatu pengamatan ke- $i$  dikatakan sebagai pencilan apabila nilai  $|DFFITS_i| > 2\sqrt{\frac{k}{n}}$ ,

dengan  $p$  adalah banyaknya parameter dalam model dan  $n$  adalah banyaknya observasi (Mardiana dkk., 2021).

#### 1.5.10 Fungsi Objektif

Fungsi objektif adalah fungsi yang digunakan untuk mencari fungsi pembobot pada regresi *robust* (Alma, 2011). Fungsi objektif diturunkan terhadap  $u_i$  akan menghasilkan suatu fungsi yang disebut sebagai fungsi pengaruh ( $\psi(u_i)$ ). Fungsi pengaruh yang dibagi dengan  $u_i$  akan menghasilkan fungsi pembobot yang digunakan dalam perhitungan regresi *robust*. Adapun fungsi pembobot yang akan digunakan yaitu fungsi pembobot *tukey bisquare*, persamaan fungsi objektif untuk *tukey bisquare* sebagai berikut (Fox, 2002):

$$\rho'(u_i) = \begin{cases} \frac{c^2}{6} \left\{ 1 - \left[ 1 - \left( \frac{u_i}{c} \right)^2 \right]^3 \right\}, & |u_i| \leq c \\ \frac{c^2}{6}, & |u_i| > c \end{cases} \quad (29)$$

Dengan fungsi pengaruh yaitu:

$$\rho'(u_i) = \frac{\partial(\rho(u_i))}{\partial u_i} = \psi(u_i) = \begin{cases} \varepsilon_i \left[ 1 - \left( \frac{u_i}{c} \right)^2 \right]^3, & |u_i| \leq c \\ 0, & |u_i| > c \end{cases} \quad (30)$$

Sehingga diperoleh fungsi pembobot

$$\omega_i = \omega(u_i) = \frac{\psi(u_i)}{u_i} = \begin{cases} \left[ 1 - \left( \frac{u_i}{c} \right)^2 \right]^3, & |u_i| \leq c \\ 0, & |u_i| > c \end{cases} \quad (31)$$

dengan  $u_i = \frac{e_i}{\hat{\sigma}}$  dan konstanta  $c$  adalah konstanta yang menghasilkan efisiensi tinggi dengan eror berdistribusi normal yang dapat memberikan perlindungan terhadap pencilan, diketahui nilai konstanta  $c$  pada metode *tukey bisquare* untuk *S-estimator* yaitu 1.547 dan untuk *M-estimator* yaitu 4.685.

### 1.5.11 Regresi Robust

Regresi Robust adalah metode regresi yang digunakan ketika distribusi eror tidak normal atau terdapat beberapa pencilan yang mempengaruhi model (Begashaw dan Yohanes, 2020). Metode ini merupakan alat penting untuk menganalisa data yang dipengaruhi oleh pencilan sehingga dihasilkan model yang robust atau tahan terhadap pencilan (Sari dkk., 2020). Model regresi linier membutuhkan estimasi parameter yang kuat, jika data yang digunakan terdapat pencilan (Kalina dan Tichavsky, 2020). Adapun salah satu jenis estimasi parameter yang dapat digunakan untuk mengestimasi regresi *robust* pada penelitian ini yaitu estimasi *Method of Moment (MM-estimator)* (Shodiqin dkk., 2018).

*MM-estimator* merupakan gabungan dari metode estimasi yang mempunyai nilai *breakdown* tinggi (*S-estimator*) dan metode *M-estimator* (Yohai, 1987). Metode *MM-estimator* dapat secara bersamaan mencapai nilai *breakdown* tinggi dan efisiensi yang tinggi (Alanamu dan Oyeyemi, 2018). Prosedur *MM-estimator* adalah mengestimasi parameter regresi menggunakan *S-estimator* yang meminimumkan skala eror dari *M-estimator* dan kemudian dilanjutkan dengan menghitung estimasi parameter akhir dengan *M-estimator* (Farouk dkk., 2022). *MM-estimator* didefinisikan sebagai berikut:

$$\hat{\beta}_{MM} = \min \sum_{i=1}^n \rho \left( \frac{e_i}{\hat{\sigma}} \right) = \min \sum_{i=1}^n \rho \left( \frac{y_i - \sum_{j=0}^k X_i \beta_j}{\hat{\sigma}} \right) \quad (32)$$

Langkah pertama dari *MM-estimator* dimulai dengan mencari *S-estimator* yang meminimumkan suatu skala residual, selanjutnya skala residual tetap konstan dan di akhiri dengan menetapkan parameter-parameter regresi menggunakan *M-estimator*. *S-estimator* sebagai permulaan dengan nilai *breakdown* yang tinggi dan di akhiri dengan *M-estimator* membuat estimasi mempunyai efisiensi yang tinggi.

Fungsi pembobot pada regresi *robust MM-estimator* yang sering digunakan adalah fungsi pembobot *tukey bisquare*.

### 1.5.12 Kemiskinan

Kemiskinan adalah sebuah kondisi dimana kemampuan kehidupan seseorang maupun berkelompok masyarakat yang hidup dibawah garis yang di tetapkan oleh pemerintah dalam hal ini mengenai ekonomi (Alwi dan Nurfadilah 2019). Kemiskinan di Indonesia mencapai angka 25,9 juta jiwa yang hidup dalam kemiskinan pada Maret 2023. Persentase penduduk miskin di provinsi Sulawesi Selatan sejak tahun 2015 tercatat sebesar 9,39% mengalami penurunan hingga tahun 2021 sebesar 8,53% . Namun, mengalami sedikit kenaikan dalam periode tahun 2022-2023, pada September 2022 tercatat sebesar 8,66% dan pada Maret 2023 tercatat sebesar 8,70%. Adapun faktor-faktor penyebab kemiskinan yang digunakan pada penelitian ini adalah sebagai berikut (BPS, 2020):

#### 1. Persentase Penduduk

Persentase penduduk adalah tingkat pertambahan penduduk disetiap tahun nya dalam jangka waktu tertentu yang dinyatakan dengan satuan persentase.

$$P_0 = \frac{Q}{N} \times 100\%$$

dengan  $P_0$  adalah persentase penduduk,  $Q$  adalah jumlah penduduk miskin, dan  $N$  adalah jumlah penduduk.

#### 2. Tingkat Pengangguran Terbuka

Tingkat pengangguran terbuka adalah persentase jumlah pengangguran terhadap jumlah angkatan kerja. Angkatan kerja merupakan penduduk usia kerja (15 tahun ke atas) yang bekerja atau punya pekerjaan namun sementara tidak bekerja, dan pengangguran.

$$TPT = \left( \frac{\text{Jumlah Pengangguran}}{\text{Angkatan Kerja}} \right) \times 100\%$$

3. Persentase penduduk miskin usia 15 tahun ke atas tidak tamat Sekolah Dasar  
Persentase penduduk miskin usia 15 tahun ke atas yang tidak tamat sekolah dasar adalah proporsi penduduk miskin berusia 15 tahun ke atas yang belum menyelesaikan pendidikan tingkat Sekolah Dasar atau sederajat, dibandingkan dengan total penduduk miskin dalam kelompok usia tersebut di suatu wilayah tertentu.

$$PPM_{<SD} = \left( \frac{P_{<SD}}{P_M} \right) \times 100\%$$

dengan  $PPM_{<SD}$  adalah persentase penduduk miskin usia 15 tahun ke atas yang tidak tamat sekolah dasar,  $P_{<SD}$  adalah jumlah penduduk miskin usia 15 tahun ke atas yang tidak tamat sekolah dasar, dan  $P_M$  adalah jumlah penduduk miskin usia 15 tahun ke atas.

#### 4. Harapan Lama Sekolah

Harapan Lama Sekolah (HLS) didefinisikan sebagai lamanya sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak pada umur tertentu di masa mendatang.

$$HLS = \sum_{i=1}^n (A_i \times S_i)$$

dengan  $n$  adalah jumlah kelompok umur yang digunakan dalam perhitungan HLS,  $A_i$  adalah proporsi penduduk pada kelompok umur ke- $i$  yang masih sekolah dan  $S_i$  adalah jumlah tahun sekolah yang diharapkan pada kelompok umur ke- $i$ .

5. Persentase Rumah Tangga Dengan Sumber Penerangan Listrik Non-PLN  
Persentase rumah tangga dengan sumber penerangan listrik non-PLN adalah proporsi rumah tangga yang menggunakan sumber listrik selain dari Perusahaan Listrik Negara (PLN) sebagai penerangan utama, dibandingkan dengan total rumah tangga di suatu wilayah tertentu.

$$PRT_{NonPLN} = \left( \frac{RT_{NonPLN}}{RT_{Total}} \right) \times 100\%$$

dengan  $PRT_{NonPLN}$  adalah persentase rumah tangga dengan sumber penerangan listrik non-PLN,  $RT_{NonPLN}$  adalah jumlah rumah tangga yang menggunakan sumber penerangan listrik selain dari PLN, dan  $RT_{Total}$  adalah jumlah rumah tangga.

6. Persentase Pengeluaran Per Kapita Sebulan Makanan  
Persentase pengeluaran per kapita sebulan makanan adalah proporsi pengeluaran yang dialokasikan untuk kebutuhan makanan dibandingkan dengan total pengeluaran per kapita dalam satu bulan.

$$PPEM = \left( \frac{E_M}{E_T} \right) \times 100\%$$

dengan  $PPEM$  adalah persentase pengeluaran per kapita sebulan makanan,  $E_M$  adalah total pengeluaran rumah tangga untuk makanan dalam satu bulan, dan  $E_T$  adalah total pengeluaran rumah tangga (makanan dan non-makanan) dalam satu bulan.

7. Indeks Kedalaman Kemiskinan  
Indeks kedalaman kemiskinan adalah ukuran rata-rata kesenjangan pengeluaran masing-masing penduduk miskin terhadap garis kemiskinan.

$$P_1 = \frac{1}{N} \sum_{i=1}^q \left( \frac{Z - Y_i}{Z} \right)$$

dengan  $P_1$  adalah indeks kedalaman kemiskinan,  $N$  adalah jumlah penduduk,  $q$  adalah jumlah penduduk miskin (yang memiliki pengeluaran per kapita di bawah garis kemiskinan),  $Z$  adalah garis kemiskinan, dan  $Y_i$  adalah pengeluaran per kapita penduduk miskin ke- $i$ .

8. Tingkat Setengah Pengangguran  
Tingkat setengah pengangguran adalah persentase penduduk usia kerja yang bekerja dengan jumlah jam kerja kurang dari 35 jam per minggu, tetapi tidak sedang mencari pekerjaan tambahan (Artinya, mereka puas atau tidak sedang berusaha meningkatkan jam kerja atau mendapatkan pekerjaan tambahan).

$$TSP = \left( \frac{\text{Jumlah Setengah Pengangguran}}{\text{Angkatan Kerja}} \right) \times 100\%$$

dengan jumlah setengah pengangguran adalah penduduk yang bekerja kurang dari 35 jam per minggu dan angkatan kerja adalah jumlah penduduk yang bekerja dan yang menganggur.

9. Persentase Rumah Tangga Yang Memiliki Akses Terhadap Sumber Air Minum Layak

Rumah tangga dikatakan memiliki akses air minum layak yaitu jika sumber air minum utama yang digunakan adalah leding, air terlindungi, dan air hujan. Air terlindungi mencakup sumur bor/pompa, sumur terlindung dan mata air terlindung.

$$PRT_{AirLayak} = \left( \frac{RT_{AirLayak}}{RT_{Total}} \right) \times 100\%$$

dengan  $PRT_{AirLayak}$  adalah persentase rumah tangga dengan akses terhadap sumber air minum layak,  $RT_{AirLayak}$  adalah jumlah rumah tangga yang memiliki akses ke sumber air minum layak, dan  $RT_{Total}$  adalah jumlah rumah tangga.

## BAB II METODE PENELITIAN

### 2.1 Sumber Data

Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh melalui publikasi Badan Pusat Statistik Provinsi Sulawesi Selatan pada tahun 2023. Penelitian ini juga menggunakan data letak astronomi yang meliputi garis lintang (*latitude*) dan garis bujur (*longitude*) tiap kabupaten/kota di Provinsi Sulawesi Selatan sebagai faktor pembobot geografis. Adapun lokasi pengamatan pada penelitian ini terdiri dari 24 kabupaten/kota di Provinsi Sulawesi Selatan.

### 2.2 Variabel Penelitian

Variabel dalam penelitian ini terdiri dari variabel respon ( $Y$ ) dan variabel prediktor ( $X$ ). Adapun variabel yang digunakan dalam penelitian ini ditampilkan pada tabel 1 sebagai berikut.

**Tabel 1.** Variabel Penelitian

Kode	Variabel	Skala
Y	Persentase penduduk miskin	Rasio
X1	Persentase penduduk	Rasio
X2	Tingkat pengangguran terbuka	Rasio
X3	Persentase penduduk miskin usia 15 tahun ke atas tidak tamat Sekolah Dasar	Rasio
X4	Harapan lama sekolah	Rasio
X5	Persentase rumah tangga dengan sumber penerangan listrik non-PLN	Rasio
X6	Persentase pengeluaran per kapita sebulan makanan	Rasio
X7	Indeks kedalaman kemiskinan	Rasio
X8	Tingkat setengah pengangguran	Rasio
X9	Persentase rumah tangga yang memiliki akses terhadap sumber air minum layak	Rasio

### 2.3 Metode Analisis

Adapun langkah-langkah analisis yang akan dilakukan dalam penelitian ini adalah sebagai berikut:

1. Analisis awal:
  - a) Melakukan pemodelan regresi linier berganda dengan metode OLS.
  - b) Melakukan uji heterogenitas spasial menggunakan uji *Breusch-Pagan*.

Hipotesis yang digunakan pada uji ini yaitu:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 \text{ (Homoskedastisitas)}$$

$$H_1: \text{minimal ada satu } \sigma_i^2 \neq \sigma_j^2 \text{ dengan } i \neq j \text{ (Heteroskedastisitas)}$$

Statistik uji yang digunakan pada uji *Breusch-Pagan* dapat dihitung menggunakan persamaan (5). Uji ini akan menolak  $H_0$  jika nilai  $BP > \chi^2_{(\alpha, k)}$  atau jika  $p\text{-value} < \alpha$  dengan  $k$  adalah banyaknya variabel prediktor.

- c) Mengidentifikasi pencilan dengan melihat nilai *leverage* dengan persamaan (27) dan nilai DFFITS dengan persamaan (28).
2. Melakukan analisis deskriptif data sebagai gambaran awal untuk mengetahui besarnya persentase penduduk miskin di Provinsi Sulawesi Selatan tahun 2023.
  3. Menghitung jarak *euclidean* ( $d_{ij}$ ) antara lokasi ke- $i$  terhadap lokasi ke- $j$  yang terletak pada koordinat  $(u_i, v_i)$  dan  $(u_j, v_j)$  dengan persamaan (6).
  4. Menentukan nilai bandwidth optimum menggunakan metode *Cross Validation* dengan persamaan (8).
  5. Menentukan matriks pembobot untuk mengidentifikasi kedekatan antar wilayah. Matriks pembobot yang digunakan pada penelitian ini yaitu fungsi *tricube fixed* kernel dengan persamaan (7).
  6. Melakukan uji variabilitas model GWR untuk menentukan variabel global dan lokal pada model MGWR dengan persamaan (12).
  7. Menentukan estimasi parameter model MGWR menggunakan metode *MM-estimator*. Adapun prosedur MGWR menggunakan metode *MM-estimator* dapat diuraikan sebagai berikut:

a) Menghitung nilai error  $e_i = y_i - \hat{y}$  dari OLS

b) Menghitung nilai  $\hat{\sigma}_S = \sqrt{\frac{n \sum_{i=0}^n (e_i^2) - (\sum_{i=0}^n e_i)^2}{n(n-1)}}$

c) Menghitung nilai  $u_i = \frac{e_i}{\hat{\sigma}_S}$

d) Menghitung pembobot untuk *S-estimator* sebagai berikut:

$$w_i = \begin{cases} \left[1 - \left(\frac{u_i}{1.547}\right)^2\right]^2 & \text{untuk } |u_i| \leq 1.547 \\ 0 & \text{untuk } |u_i| > 1.547 \end{cases}$$

e) Mengestimasi koefisien  $\beta_g^{(1)}$  dan  $\beta_l(u_i, v_i)^{(1)}$  dengan metode WLS menggunakan pembobot  $\omega_i$  sehingga diperoleh  $e_i = y_i - \hat{y}$

f) Error  $e_i^{(1)}$  pada langkah pertama digunakan untuk menghitung skala error *M-estimator*,  $\hat{\sigma}_M = \frac{\text{median}|e_i - \text{median}(e_i)|}{0.6745}$

g) Menghitung nilai  $u_i = \frac{e_i}{\hat{\sigma}_M}$  kemudian menghitung pembobot awal  $\omega_i^{(1)}$

$$\omega_i^{(1)} = \begin{cases} \left[1 - \left(\frac{u_i}{4.685}\right)^2\right]^2 & \text{untuk } |u_i| \leq 4.685 \\ 0 & \text{untuk } |u_i| > 4.685 \end{cases}$$

h) Error  $e_i^{(1)}$  dan skala estimasi  $\hat{\sigma}_M$  dari langkah f digunakan dalam iterasi awal dengan metode WLS (*Weighted Least Square*) untuk menghitung koefisien regresi  $\hat{\beta}_{MM}$  menggunakan  $\omega_i^{(1)}$ .

i) Mengulangi langkah f, g, dan h sampai diperoleh penaksir yang konvergen.

8. Menghitung estimasi parameter global dan lokal model MGWR menggunakan metode *MM-estimator*.
9. Melakukan pengujian hipotesis model MGWR yaitu uji serentak dan uji individu (parsial satu-satu).
10. Membuat interpretasi dan kesimpulan.