

BAB I PENDAHULUAN

1.1 Latar Belakang

Regresi linier merupakan salah satu metode analisis statistik yang umum digunakan untuk memodelkan hubungan antara variabel respon dan variabel prediktor, dengan asumsi bahwa koefisien regresi berlaku sama di seluruh pengamatan. Metode ini banyak digunakan karena menawarkan pendekatan yang sederhana namun efektif dalam memprediksi atau memahami pola hubungan antar variabel dalam berbagai bidang, seperti ekonomi dan kesehatan masyarakat (James dkk., 2021). Namun, dalam menghadapi kompleksitas data yang semakin meningkat, terutama yang melibatkan dimensi waktu dan spasial model regresi linier sederhana sering kali tidak memadai, karena tidak mampu menangkap dinamika serta heterogenitas yang bersifat lokal antar unit pengamatan. Jika heterogenitas ini diabaikan, maka model regresi dapat menghasilkan estimasi yang bias dan tidak efisien, sehingga mengurangi akurasi prediksi serta validitas inferensi yang dihasilkan. Hal ini terjadi karena hubungan antara variabel respon dan variabel prediktor bisa berbeda-beda di setiap lokasi atau waktu, yang tidak bisa dijelaskan hanya dengan satu koefisien global.

Seiring dengan kebutuhan untuk mengatasi keterbatasan tersebut, regresi data panel menjadi salah satu pendekatan analisis yang relevan, khususnya untuk data yang mencakup dimensi waktu (*time-series*) dan lintas individu (*cross-section*). Regresi data panel mampu mengakomodasi perubahan antar waktu sekaligus perbedaan antar unit pengamatan, sehingga dapat mengurangi bias akibat variabel yang tidak teramati dan meningkatkan efisiensi estimasi (Bergh dkk., 2021). Regresi data panel terdiri dari beberapa pendekatan, di antaranya *Common Effect Model* (CEM), *Fixed Effect Model* (FEM), dan *Random Effect Model* (REM). Dalam penelitian ini, REM dipilih karena variasi antar unit atau lokasi diperlakukan sebagai efek acak yang diasumsikan tidak berkorelasi dengan variabel independen. Berdasarkan penelitian yang dilakukan oleh Salsabila (2021) mengenai ketimpangan pendapatan daerah di Kalimantan Timur, regresi data panel dapat menjelaskan variasi antar unit pengamatan, tetapi belum mampu menangkap secara spesifik pengaruh karakteristik lokal masing-masing daerah. Keterbatasan ini menunjukkan perlunya pendekatan yang mampu menangkap variasi spasial secara eksplisit, salah satunya melalui penggunaan model regresi spasial (Cai dkk., 2021).

Data dengan dimensi spasial memiliki karakteristik unik berupa heterogenitas spasial dan dependensi spasial. Heterogenitas spasial merujuk pada perbedaan karakteristik lokal antar wilayah yang berpengaruh terhadap hubungan antar variabel, sementara dependensi spasial mengindikasikan adanya pengaruh dari wilayah sekitar terhadap suatu unit pengamatan (Liu dkk., 2023). Untuk mengakomodasi heterogenitas spasial dalam analisis data panel, metode *Geographically Weighted Panel Regression* (GWPR) merupakan pendekatan yang lebih relevan dibandingkan model panel global seperti FEM dan REM, yang mengasumsikan homogenitas hubungan antar variabel di seluruh lokasi. GWPR

dirancang untuk menangkap variasi lokal melalui pemberian bobot geografis pada masing-masing wilayah, sehingga estimasi parameter menjadi lebih representatif terhadap kondisi setempat.

Penerapan GWPR pada data panel dengan REM menghadapi tantangan karena dalam REM, efek individu atau wilayah dimodelkan sebagai bagian dari galat dan diasumsikan tidak berkorelasi dengan variabel independen. Pendekatan ini dapat menyebabkan bias dalam analisis spasial karena tidak mempertimbangkan variasi lokal secara eksplisit. Oleh karena itu, dibutuhkan transformasi *quasi demeaning* untuk menyesuaikan struktur data panel dengan pendekatan spasial. Transformasi ini dilakukan dengan menghitung rata-rata setiap variabel untuk tiap unit pengamatan sepanjang waktu, kemudian mengurangkan nilai individual dari rata-rata tersebut dengan penyesuaian menggunakan faktor theta (θ), yang mencerminkan proporsi varians efek acak terhadap total varians (Jordan & Philips, 2023). θ digunakan dalam proses transformasi untuk mengalikan rata-rata variabel unit pengamatan sebelum dikurangkan dari nilai individualnya, sehingga mempertahankan sebagian variabilitas antar unit dan memastikan bahwa galat hasil transformasi tidak lagi berkorelasi antarwaktu. Dengan demikian, GWPR dapat mengakomodasi heterogenitas spasial secara lebih tepat tanpa kehilangan keakuratan estimasi akibat efek acak global.

Metode GWPR telah digunakan secara luas dalam analisis data di berbagai bidang, termasuk kesehatan, sosial, dan ekonomi. Salah satu variabel penting yang dapat dianalisis dengan GWPR adalah Angka Harapan Hidup (AHH). AHH merupakan indikator utama dalam menilai kesejahteraan dan kualitas hidup masyarakat, serta dipengaruhi oleh berbagai faktor lokal seperti distribusi infrastruktur kesehatan, akses terhadap layanan kesehatan, dan kondisi sosial ekonomi wilayah (Ratih Nurfitri & Teti Sofia Yanti, 2023). Dengan pendekatan GWPR, pengaruh faktor-faktor tersebut dapat dianalisis secara spasial dan temporal, sehingga membantu pemerintah dan pemangku kepentingan dalam merumuskan kebijakan berbasis bukti yang lebih tepat sasaran (Federova dkk., 2024).

Beberapa penelitian sebelumnya telah menunjukkan efektivitas GWPR dalam memodelkan data spasial panel. Tambunan (2024) menggunakan model GWPR untuk memodelkan pengangguran terbuka di Provinsi Sumatera Utara pada periode 2018–2022 dan menemukan bahwa GWPR memberikan hasil estimasi yang lebih baik dibandingkan model panel global. Penelitian serupa dilakukan oleh Nasri (2023) pada data kemiskinan di Provinsi Jawa Barat tahun 2019–2021, yang menunjukkan keunggulan GWPR dalam menjelaskan persentase penduduk miskin dibandingkan dengan model FEM. Selain itu, Muliana (2023) menemukan bahwa fungsi pembobot *gaussian* memberikan hasil model terbaik dibandingkan dengan pembobot *exponential* dan *bisquare*.

Berdasarkan uraian di atas, maka penelitian ini bertujuan untuk mengkaji penerapan metode *Geographically Weighted Panel Regression* dengan pendekatan *Random Effect* yang telah ditransformasi menggunakan *quasi demeaning*. Penelitian ini diberi judul: **"Pemodelan *Geographically Weighted Panel Regression* dengan**

Random Effect Menggunakan Quasi Demeaning (Studi Kasus: Angka Harapan Hidup di Indonesia Tahun 2018–2022)

1.2 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS), yaitu data angka harapan hidup di Indonesia tahun 2018-2022
2. Model yang digunakan dalam penelitian ini adalah *geographically weighted panel regression* dengan *random effect* menggunakan fungsi pembobot terpilih

1.3 Tujuan dan Manfaat Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Memodelkan data angka harapan hidup di Indonesia menggunakan *geographically weighted panel regression* dengan *random effect*.
2. Menganalisis pengaruh variabel prediktor terhadap angka harapan hidup di berbagai provinsi menggunakan model *geographically weighted panel regression* dengan *random effect*.
3. Mengklasifikasikan wilayah provinsi di Indonesia berdasarkan variabel prediktor yang berpengaruh signifikan terhadap angka harapan hidup menggunakan model *geographically weighted panel regression* dengan *random effect*.

Berdasarkan tujuan penelitian di atas maka setelah melakukan penelitian diharapkan dapat memberikan manfaat sebagai berikut:

1. Menambah literatur mengenai penerapan model *geographically weighted panel regression* dengan *random effect* dalam analisis data kesehatan, khususnya Angka Harapan Hidup
2. Menyediakan model berbasis data untuk mendukung perencanaan kebijakan kesehatan yang lebih tepat sasaran sesuai dengan karakteristik regional.
3. Membantu dalam klasifikasi wilayah berdasarkan karakteristik kesehatan dan sosial-ekonomi, sehingga intervensi dapat dilakukan secara lebih efektif dan efisien.

1.4 Landasan Teori

1.4.1 Regresi Data Panel

Data panel adalah jenis data gabungan dari data *cross-section* dan *time series*, yang berarti data mencakup pengamatan dari beberapa objek yang berbeda pada beberapa titik waktu yang berurutan (Alviani dkk., 2021). Dalam tabel data panel, setiap baris mewakili kombinasi spesifik dari objek dan waktu tertentu, sehingga memungkinkan analisis yang lebih mendalam terhadap perubahan dari waktu ke waktu dan perbedaan antar objek. Sehingga dapat dikatakan data panel memberikan

keunggulan dalam menganalisis dinamika dan variasi dalam data karena menggabungkan dimensi waktu dan objek sekaligus. Struktur data panel disajikan pada Tabel 1.

Tabel 1. Struktur Data Panel

Individu (i)	Tahun (t)	Variabel				
		(Y _{it})	(X _{1it})	(X _{2it})	...	(X _{kit})
1	1	y _{1,1}	x _{1,1,1}	x _{2,1,1}		x _{k,1,1}
	2	y _{1,2}	x _{1,1,2}	x _{2,1,2}	...	x _{k,1,2}
	⋮	⋮	⋮	⋮		
	T	y _{1,T}	x _{1,1,T}	x _{2,1,T}		x _{k,1,T}
2	1	y _{2,1}	x _{1,2,1}	x _{2,2,1}		x _{k,2,1}
	2	y _{2,2}	x _{1,2,2}	x _{2,2,2}	...	x _{k,2,2}
	⋮	⋮	⋮	⋮		
	T	y _{2,T}	x _{1,2,T}	x _{2,2,T}		x _{k,2,T}
⋮	⋮	⋮	⋮	⋮	...	
n	1	y _{n,1}	x _{1,n,1}	x _{2,n,1}		x _{k,n,1}
	2	y _{n,2}	x _{1,n,2}	x _{2,n,2}	...	x _{k,n,2}
	⋮	⋮	⋮	⋮		
	T	y _{n,T}	x _{1,n,T}	x _{2,n,T}		x _{k,n,T}

Analisis regresi data panel adalah salah satu metode statistik yang digunakan untuk melihat pengaruh beberapa peubah prediktor terhadap satu peubah respon dengan struktur data berupa data panel (Alamsyah dkk., 2022). Persamaan umum analisis data panel ditunjukkan pada Persamaan (1).

$$y_{it} = \beta_{0it} + \sum_{j=1}^k \beta_j x_{jit} + e_{it} \quad i = 1, 2, \dots, n; t = 1, 2, \dots, T; j = 1, 2, \dots, k \quad (1)$$

Keterangan:

- y_{it} : variabel respon pada unit amatan ke-i dan waktu ke-t
 β_{0it} : intersep model regresi pada unit lintas individu ke-i dalam periode ke-t
 β_j : koefisien regresi untuk variabel prediktor ke-j
 x_{jit} : variabel prediktor ke-j dari unit lintas individu ke-i dalam periode ke-t
 e_{it} : galat pada amatan ke-i dalam periode ke-t

1.4.2 Model Regresi Data Panel

Dalam regresi data panel, terdapat tiga pendekatan utama yang umum digunakan, yaitu *Common Effect Model* (CEM), *Fixed Effect Model* (FEM), dan *Random Effect Model* (REM). Masing-masing model ini memiliki asumsi yang berbeda terkait dengan hubungan antara variabel respon dan variabel prediktor, serta bagaimana variabel individu dan waktu diperhitungkan dalam model (Baltagi, 2005). Pemilihan

model yang tepat sangat bergantung pada karakteristik data dan tujuan analisis yang ingin dicapai. Berikut adalah penjelasan mengenai ketiga model tersebut:

1. Common Effect Model

Common Effect Model (CEM) adalah metode di mana seluruh data, baik *cross-section* maupun *time series*, digabungkan dengan mengabaikan perbedaan waktu dan lokasi penelitian (Salsabila dkk., 2022). Dalam model ini, diasumsikan bahwa nilai intersep dari setiap variabel adalah sama, begitu pula koefisien regresi untuk semua unit *cross-section* dan *time series*. Persamaan dari CEM ditunjukkan pada Persamaan (2).

$$y_{it} = \beta_0 + \sum_{j=1}^k \beta_j x_{jit} + e_{it} \quad i = 1, 2, \dots, n; t = 1, 2, \dots, T; j = 1, 2, \dots, k \quad (2)$$

Keterangan:

- y_{it} : variabel respon pada unit amatan ke-i dan waktu ke-t
- β_0 : intersep model regresi
- β_j : koefisien regresi untuk variabel prediktor ke-j
- x_{jit} : variabel prediktor ke-j untuk unit amatan ke-i dalam periode ke-t
- e_{it} : galat pada amatan ke-i dalam periode ke-t

2. Fixed Effect Model

Fixed Effect Model (FEM) adalah metode yang digunakan dalam analisis data panel untuk mengestimasi perbedaan intersep antara unit *cross-section* dan waktu, dengan asumsi bahwa intersep tetap (Gamayanti & Junaidi, 2021). Pada model ini, intersep antara unit dan waktu diasumsikan berbeda, sehingga variabel *dummy* digunakan untuk menjelaskan perbedaan tersebut. Variabel *dummy* ini memungkinkan model untuk mengakomodasi karakteristik spesifik dari setiap unit dan waktu yang tidak teramati namun tetap, sehingga intersep yang dihasilkan akan berbeda untuk masing-masing unit. Persamaan dari FEM ditunjukkan pada Persamaan (3).

$$y_{it} = \beta_{0i} + \sum_{j=1}^k \beta_j x_{jit} + e_{it} \quad i = 1, 2, \dots, n; t = 1, 2, \dots, T; j = 1, 2, \dots, k \quad (3)$$

dengan nilai $\hat{\beta}_{0i} = \bar{y} - (\hat{\beta}_{01}\bar{x}_{i1}) - (\hat{\beta}_{02}\bar{x}_{i2}) - \dots - (\hat{\beta}_{0k}\bar{x}_{ik})$

Keterangan:

- y_{it} : variabel respon pada unit amatan ke-i dan waktu ke-t
- β_{0i} : intersep model regresi pada unit amatan ke-i
- β_j : koefisien regresi untuk variabel prediktor ke-j
- x_{jit} : variabel prediktor ke-j untuk unit amatan ke-i dalam periode ke-t
- e_{it} : galat pada amatan ke-i dalam periode ke-t

3. Random Effect Model

Random Effect Model (REM) adalah metode estimasi dalam analisis data panel di mana variabel gangguan (*galat term*) mungkin saling berhubungan baik antar waktu maupun antar individu (Salsabila dkk., 2022). Berbeda dengan FEM, di mana setiap unit individu atau waktu memiliki intersep yang berbeda, REM mengasumsikan bahwa variasi antar unit atau waktu adalah acak dan tidak berkorelasi dengan variabel prediktor. Salah satu kelemahan FEM adalah menurunnya derajat kebebasan (*degree of freedom*) karena penambahan banyak variabel *dummy* untuk setiap unit atau waktu. Hal ini dapat mengurangi efisiensi estimasi parameter. Untuk mengatasi masalah ini, REM memperlakukan perbedaan antar unit atau waktu sebagai bagian dari komponen galat yang acak, bukan sebagai bagian dari intersep tetap. Persamaan dari REM ditunjukkan pada Persamaan (4).

$$y_{it} = \beta_0 + \sum_{j=1}^k \beta_j x_{jit} + u_{it}; \quad u_{it} = \mu_i + e_{it} \quad (4)$$

Keterangan:

- y_{it} : variabel respon pada unit amatan ke-i dan waktu ke-t
- β_0 : intersep model regresi
- β_j : koefisien regresi untuk variabel prediktor ke-j
- x_{jit} : variabel prediktor ke-j untuk unit amatan ke-i dalam periode ke-t
- e_{it} : galat pada amatan ke-i dalam periode ke-t
- μ_i : galat pada amatan individu ke-i

Persamaan (4) dapat ditulis dalam bentuk matriks yang ditunjukkan pada Persamaan (5)

$$Y = X\beta + u \quad (5)$$

Keterangan:

- Y = vektor amatan dengan ukuran $nT \times 1$
- X = matriks variabel prediktor dengan ukuran $(nT \times (k + 1))$
- β = vektor koefisien dengan ukuran $((k + 1) \times 1)$
- u = vektor galat komponen dengan ukuran $nT \times 1$

Berdasarkan struktur galat pada Persamaan (4), maka varians dari galat total u_{it} dapat dihitung sebagai:

$$Var(u_{it}) = Var(\mu_i + e_{it}) = Var(\mu_i) + Var(e_{it}) + 2Cov(\mu_i, e_{it}) = \hat{\sigma}_\mu^2 + \hat{\sigma}_e^2 \quad (6)$$

Selain itu, karena efek individu μ_i bersifat tetap untuk setiap waktu t , maka terdapat kovarians antar waktu pada unit yang sama, yaitu:

$$Cov(u_{it}, u_{is}) = Cov(\mu_i + e_{it}, \mu_i + e_{is}) = Cov(\mu_i, \mu_i) = \hat{\sigma}_\mu^2 \text{ untuk } t \neq s \quad (7)$$

Berdasarkan Persamaan (6) dan (7), struktur matriks varians-kovarians untuk u_i dalam REM dapat ditulis sebagai:

$$\boldsymbol{\Omega}_i = \text{Var}(u_i) = \begin{bmatrix} \hat{\sigma}_\mu^2 + \hat{\sigma}_{e_1}^2 & \hat{\sigma}_\mu^2 & \cdots & \hat{\sigma}_\mu^2 \\ \hat{\sigma}_\mu^2 & \hat{\sigma}_\mu^2 + \hat{\sigma}_{e_2}^2 & \cdots & \hat{\sigma}_\mu^2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_\mu^2 & \hat{\sigma}_\mu^2 & \cdots & \hat{\sigma}_\mu^2 + \hat{\sigma}_{e_T}^2 \end{bmatrix} \quad (8)$$

Untuk keseluruhan individu, Persamaan (8) dituliskan dalam bentuk blok-diagonal, yaitu:

$$\boldsymbol{\Omega} = \text{Var}(u) = \begin{bmatrix} \boldsymbol{\Omega}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Omega}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\Omega}_n \end{bmatrix} \quad (9)$$

Namun, untuk menyederhanakan notasi, sering kali diasumsikan bahwa semua amatan memiliki struktur varian-kovarians yang sama untuk semua individu. Maka Persamaan (9) dapat ditulis ulang menjadi Persamaan (10)

$$\boldsymbol{\Omega} = \hat{\sigma}_\mu^2 (\mathbf{I}_n \otimes \mathbf{J}_T) + \hat{\sigma}_e^2 \mathbf{I}_{nT} \quad (10)$$

dengan

$\mathbf{I}_n$: matriks identitas dengan ukuran $(n \times n)$

$\mathbf{J}_T$: matriks dengan ukuran $(T \times T)$ yang semua elemennya bernilai 1

$\mathbf{I}_{nT}$: matriks identitas dengan ukuran $(nT \times nT)$

Perlu dilakukan transformasi untuk membuat galat menjadi homoskedastis dan tidak berkorelasi. Transformasi dilakukan dengan mengalikan kedua sisi Persamaan (5) dengan invers akar kuadrat dari Persamaan (10). Sehingga diperoleh:

$$\boldsymbol{\Omega}^{-\frac{1}{2}} \mathbf{Y} = \boldsymbol{\Omega}^{-\frac{1}{2}} \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\Omega}^{-\frac{1}{2}} \mathbf{u} \quad (11)$$

Persamaan (11) dapat ditulis ulang menjadi Persamaan (12)

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}} \boldsymbol{\beta} + \tilde{\mathbf{u}} \quad (12)$$

Sehingga diperoleh estimasi $\boldsymbol{\beta}$ pada Persamaan (13)

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}' \tilde{\mathbf{Y}} = (\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{Y} \quad (13)$$

Estimasi model *Random Effect* menggunakan *Generalized Least Squares* (GLS) yang merupakan pendekatan yang kuat dalam analisis data panel karena memiliki kemampuan untuk menghasilkan estimasi yang efisien, tidak bias, dan varians minimum (Rosita., 2021). Untuk mendapatkan estimasi parameter yang efisien, GLS melakukan transformasi variabel sehingga galat menjadi homoskedastik dan tidak berkorelasi.

1.4.3 Penentuan Model Terbaik Regresi Data Panel

Terdapat dua uji yang umumnya dilakukan untuk menentukan model terbaik dari ketiga model tersebut. Uji yang dilakukan yaitu:

1. Uji Chow

Uji Chow adalah metode statistik yang digunakan untuk menentukan model terbaik dalam analisis data panel antara FEM dan CEM. Uji ini memeriksa hipotesis bahwa semua unit *cross section* memiliki perilaku serupa, yang dalam banyak kasus tidak realistis karena adanya variasi perilaku antar unit (Abdussamad dkk., 2024). Hipotesis yang digunakan adalah sebagai berikut:

$H_0: \beta_{01} = \beta_{02} = \dots = \beta_{0n}$ (model terbaik adalah CEM)

H_1 : terdapat $\beta_{0i} \neq \beta_{0i^*}; i \neq i^*; i = i^* = 1, 2, \dots, n$ (model terbaik adalah FEM)

Statistik uji yang digunakan adalah (Alamsyah dkk., 2022):

$$F = \frac{\frac{[JKG_1 - JKG_2]}{n-1}}{\frac{JKG_2}{nT-n-k}} \sim F_{\alpha; (n-1), (nT-n-k)} \quad (14)$$

Keterangan:

n : jumlah Lokasi pengamatan

T : jumlah periode waktu pengamatan

k : jumlah variabel prediktor

JKG_1 : jumlah kuadrat galat dari estimasi model CEM

JKG_2 : jumlah kuadrat galat dari estimasi model FEM

Jika nilai $F \geq F_{\alpha; (n-1), (nT-n-k)}$ atau $p - value < \alpha$, maka H_0 ditolak, yang berarti *FEM* lebih baik. Setelah FEM dipilih, langkah selanjutnya adalah melakukan uji Hausman.

2. Uji Hausman

Uji Hausman digunakan untuk menguji keberadaan korelasi antara galat dengan satu atau lebih variabel prediktor dalam model. Jika tidak ada korelasi antara galat dan variabel prediktor, maka model terbaik yang diasumsikan antara FEM dan REM adalah model REM. Hipotesis yang digunakan dalam pengujian ini adalah sebagai berikut.:

$H_0: corr(x_{it}, e_{it}) = 0$ (tidak ada korelasi antara galat dengan variabel prediktor)

$H_1: corr(x_{it}, e_{it}) \neq 0$ (ada korelasi antara galat dengan variabel prediktor)

Statistik uji Hausman sebagai berikut (Alamsyah dkk., 2022):

$$\chi^2(k) = (b - \beta)' [var(b - \beta)]^{-1} (b - \beta) \sim \chi^2_{\alpha; k} \quad (15)$$

Keterangan:

b : koefisien *random effect*

β : koefisien *fixed effect*

Statistik Hausman mengikuti distribusi chi-square. Jika nilai $\chi^2(k) \geq \chi^2_{\alpha,k}$ atau $p - value < \alpha$, maka H_0 ditolak, yang berarti model terbaik yang digunakan adalah FEM.

1.4.4 Uji Asumsi Klasik

Pengujian asumsi perlu dilakukan agar hasil yang diperoleh merupakan pendugaan terbaik. Beberapa uji asumsi yang umum digunakan adalah uji normalitas, uji multikolinearitas, dan uji heterogenitas.

1. Uji Normalitas

Uji normalitas bertujuan untuk mengetahui distribusi galat dalam model regresi berdistribusi normal. Salah satu metode yang sering digunakan untuk menguji normalitas galat adalah uji Kolmogorov-Smirnov. Normalitas galat penting karena mempengaruhi keakuratan uji signifikansi dalam analisis regresi, seperti uji t dan uji F, yang digunakan untuk menguji hubungan antara variabel bebas dan variabel terikat (Wooldridge, 2018). Uji ini membandingkan distribusi empiris galat dengan distribusi normal teoretis. Hipotesis yang diuji dalam uji Kolmogorov-Smirnov adalah:

$H_0: F(x) = F_0(x)$ (galat berdistribusi normal)

$H_1: F(x) \neq F_0(x)$ (galat tidak berdistribusi normal)

Statistik uji Kolmogorov-Smirnov adalah sebagai berikut :

$$D = \max|F_n(x) - F_0(x)| \quad (16)$$

Keterangan:

$F_n(x)$: probabilitas distribusi kumulatif dari data

$F_0(x)$: probabilitas distribusi kumulatif yang dihipotesis berdistribusi normal

Jika nilai $D < D_{\alpha,n}$ atau $p - value \geq \alpha$, maka tidak ada cukup bukti untuk menolak H_0 . Hal ini berarti galat yang diuji berdistribusi normal.

2. Uji Multikolinearitas

Multikolinearitas terjadi ketika terdapat hubungan linier yang kuat antara dua atau lebih variabel prediktor dalam model regresi. Jika terdapat multikolinearitas, maka akan sulit untuk menentukan pengaruh masing-masing variabel prediktor terhadap variabel respon, sehingga mengurangi keakuratan estimasi parameter regresi (Baltagi, 2005). Hubungan tersebut dapat ditunjukkan dengan menggunakan matriks korelasi sesuai dengan Persamaan (17).

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix} \quad (17)$$

dengan

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Nilai r_{xy} yang mendekati 1 atau -1 mengindikasikan adanya hubungan linier yang kuat antara dua variabel. Jika nilai $r_{xy} \geq 0,7$ maka hubungan antar variabel dikategorikan sebagai kuat (Tichy, 1993).

Selain itu, terdapat metode lain yang umum digunakan adalah dengan menggunakan nilai *Variance Inflation Factor* (VIF). VIF mengukur seberapa besar varians dari suatu koefisien regresi meningkat karena adanya korelasi antar variabel predictor. Rumus VIF ditampilkan pada Persamaan (18) (Chatterjee & Hadi, 2012):

$$VIF = \frac{1}{(1 - R_k^2)} \quad (18)$$

dengan R_k^2 adalah koefisien determinasi regresi antar variabel prediktor. Jika nilai $VIF > 10$ maka dapat disimpulkan terjadi multikolinearitas pada variabel prediktor.

3. Uji Heterogenitas Spasial

Heterogenitas spasial adalah variasi karakteristik yang terjadi antar lokasi dalam unit spasial tertentu, kondisi data dari suatu lokasi cenderung tidak seragam dan memiliki perbedaan dengan data dari lokasi lain (Mujiarti dkk., 2024). Heterogenitas spasial mencerminkan perbedaan karakteristik atau keunikan spesifik di setiap wilayah, yang dapat menyebabkan perbedaan nilai parameter regresi di tiap lokasi pengamatan. Uji heterogenitas spasial bertujuan untuk mengidentifikasi keberadaan dan pengaruh perbedaan tersebut antar lokasi. Metode yang dapat digunakan untuk menguji keberadaan heterogenitas spasial adalah uji Breusch-Pagan (Supratman dkk., 2022) (Fitriatusakiah dkk., 2021). Hipotesis yang digunakan dalam uji Breusch-Pagan yaitu:

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

$H_1: \text{terdapat } \sigma_i^2 \neq \sigma_{i^*}^2; i \neq i^*; i = i^* = 1, 2, \dots, n$ (heterogenitas antar wilayah)

Statistik uji:

$$BP = \frac{1}{2} \mathbf{f}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{f} \sim \chi_{\alpha, k}^2 \quad (19)$$

dengan, elemen vektor $\mathbf{f}$ dirumuskan $f_i = \left(\frac{u_i^2}{\sigma^2} - 1 \right)$

Keterangan:

BP : nilai uji Breusch-Pagan

u_i : galat untuk pengamatan ke- i dengan asumsi $e \sim IIDN(0, \sigma^2)$

$\mathbf{Z}$: matriks $\mathbf{X}$ berukuran $nT \times (k + 1)$ yang sudah distandarisasi untuk setiap amatan

σ^2 : variansi dari u_i

Pengambilan keputusan pada uji Breusch-Pagan akan melakukan penolakan terhadap H_0 apabila nilai statistik $BP \geq \chi_{\alpha, k}^2$ atau nilai $p - value < \alpha$, yang mengindikasikan terjadinya heterogenitas spasial (Ikhsanudin & Pasaribu, 2023).

1.4.5 Fungsi Pembobot Spasial

Fungsi pembobot adalah alat penting dalam analisis spasial yang memberikan bobot berbeda pada elemen berdasarkan kedekatan geografis, sehingga pengamatan yang lebih dekat memiliki pengaruh lebih besar dalam model GWPR. Prinsip ini mendukung estimasi parameter yang bervariasi di lokasi berbeda, meningkatkan akurasi dan relevansi analisis. Pemilihan fungsi pembobot yang tepat dapat mengurangi bias, menangkap variasi signifikan dalam data, serta menentukan *bandwidth* optimal, yaitu radius pengaruh setiap lokasi. Dengan demikian, fungsi pembobot menjadi elemen esensial untuk validitas dan keandalan analisis spasial, memberikan pemahaman yang lebih baik tentang dinamika data.

Fungsi pembobot yang lazim digunakan dalam GWPR adalah fungsi kernel. Fungsi kernel merupakan komponen utama dalam pembobotan spasial yang menentukan cara pemberian bobot pada setiap pengamatan berdasarkan jarak geografisnya (Ramadayani dkk., 2022). Berdasarkan penentuan *bandwidth*, fungsi kernel terbagi menjadi dua jenis, yaitu *adaptive* dan *fixed*. Kernel *adaptive* menyesuaikan *bandwidth* untuk setiap lokasi berdasarkan kepadatan titik data di sekitarnya. Hal ini berbeda dengan kernel tetap yang menggunakan *bandwidth* yang sama untuk semua lokasi pengamatan (Arsyi dkk., 2024). Kedua jenis kernel ini dapat menggunakan bentuk fungsi seperti *gaussian*, *exponential*, *bisquare*, dan *tricube*, yang masing-masing memiliki pola penurunan bobot berbeda sesuai jarak. Rumus fungsi pembobot tersebut adalah sebagai berikut (Pamungkas dkk., 2016):

Rumus fungsi pembobot untuk masing-masing fungsi kernel adalah sebagai berikut:

1. Fungsi kernel *gaussian*

$$W_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right) \quad (20)$$

2. Fungsi kernel *bisquare*

$$W_{ij} = \begin{cases} \left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right)^2, & \text{untuk } d_{ij} \leq h \\ 0, & \text{untuk } d_{ij} > h \end{cases} \quad (21)$$

3. Fungsi kernel *tricube*

$$W_{ij} = \begin{cases} \left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^3\right)^3, & \text{untuk } d_{ij} \leq h \\ 0, & \text{untuk } d_{ij} > h \end{cases} \quad (22)$$

4. Fungsi kernel *exponential*

$$W_{ij} = \exp\left(\frac{-d_{ij}}{h_i}\right) \quad (23)$$

Perhitungan jarak *euclidean* ditunjukkan pada Persamaan (24)

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (24)$$

Keterangan:

W_{ij} : bobot dari amatan pada lokasi ke-j untuk penaksiran koefisien lokasi ke-i

d_{ij} : jarak *euclidean* antara lokasi ke-i dan lokasi ke-j

h_i : *bandwidth* pada lokasi ke-i

Terdapat parameter *bandwidth* yang nilainya tidak diketahui pada fungsi pembobot kernel. Parameter ini dapat digambarkan sebagai radius atau jari-jari suatu lingkaran. Pemilihan *bandwidth* terbaik dalam GWPR sangat krusial untuk menentukan fungsi pembobot kernel yang paling sesuai. Hal ini dikarenakan fungsi pembobot kernel berpengaruh terhadap keakuratan model dalam memprediksi data. *Bandwidth* optimum ditentukan berdasarkan nilai *cross-validation* (CV) minimum. Nilai CV dapat diperoleh dengan menggunakan Persamaan (25) (Fotheringham dkk., 2002).

$$CV = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h))^2 \quad (25)$$

dengan y_i adalah nilai pengamatan ke-i dan $\hat{y}_{\neq i}(h)$ adalah nilai duga pengamatan ke-i yang diperoleh tanpa melibatkan ke-i itu sendiri

1.4.6 Geographically Weighted Panel Regression

Metode GWPR merupakan perluasan langsung dari metode *cross-sectional geographically weighted regression* untuk mengestimasi koefisien yang bervariasi secara spasial (nonstasioneritas spasial) dengan data panel (Yu dkk., 2021). Hal ini menjadikan GWPR sangat berguna dalam menganalisis data yang memiliki hubungan antar variabel yang tidak bersifat konstan, melainkan dapat bervariasi di antara wilayah yang berbeda. Dalam praktiknya, GWPR menerapkan model regresi untuk setiap lokasi secara terpisah. Pendekatan ini memungkinkan identifikasi perbedaan signifikan dalam hubungan tersebut di seluruh wilayah yang dianalisis. Dengan demikian, melalui penggabungan data lintas waktu dan lintas lokasi, GWPR memberikan wawasan yang lebih mendalam dan komprehensif tentang faktor-faktor yang mempengaruhi fenomena tertentu, seperti kemiskinan atau kesehatan masyarakat, di berbagai daerah (Wati & Utami, 2020). Persamaan GWPR dengan *random effect* diturunkan dari Persamaan (4), yang kemudian menghasilkan Persamaan (26).

$$y_{it} = \beta_0(u_i, v_i) + \sum_{j=1}^k \beta_j(u_i, v_i)x_{jit} + u_{it}; \quad u_{it} = \mu_i + e_{it} \quad (26)$$

Keterangan:

y_{it} : variabel respon pada unit amatan ke-i dan waktu ke-t

$\beta_0(u_i, v_i)$: intersep yang bervariasi secara spasial, dihitung pada koordinat geografis unit i (u_i, v_i).

x_{jit} : variabel prediktor ke-j untuk unit amatan ke-i dalam periode ke-t

$\beta_j(u_i, v_i)$: koefisien regresi untuk variabel prediktor ke-j, yang bervariasi secara spasial sesuai dengan lokasi geografis (u_i, v_i).

e_{it} : galat pada amatan ke-i dalam periode ke-t

μ_i : galat pada amatan ke-i

1.4.7 Quasi Demeaning

Transformasi *quasi demeaning* dalam GLS adalah teknik yang digunakan untuk menangani efek individu dalam model panel, terutama dalam model efek acak. Transformasi tersebut penting karena memungkinkan penghilangan pengaruh efek individu yang tidak diamati (μ_i) tanpa mengubah struktur spasial dan temporal data. *Quasi demeaning* memungkinkan GWPR menghasilkan estimasi parameter lokal yang lebih efisien dan tidak bias, karena mengurangi potensi bias dari heterogenitas individu yang tidak teramati dan tetap mempertahankan variasi spasial parameter regresi (Jordan & Philips, 2023). Proses transformasi dilakukan dengan mengurangi setiap pengamatan terhadap rata-rata individu yang dikalikan dengan parameter transformasi θ , sehingga diperoleh bentuk:

$$y_{it}^* = y_{it} - \theta \bar{y}_i; \quad x_{it}^* = x_{it} - \theta \bar{x}_i; \quad u_{it}^* = u_{it} - \theta \bar{u}_i \quad (27)$$

dengan

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it}, \quad \bar{x}_i = \frac{1}{T} \sum_{t=1}^T x_{it}, \quad \bar{u}_i = \frac{1}{T} \sum_{t=1}^T u_{it}$$

Tujuan dari transformasi ini adalah menghilangkan korelasi serial dalam galat, sehingga kovarians antara galat tertransformasi untuk waktu yang berbeda sama dengan nol, yaitu:

$$Cov(u_{it}^*, u_{is}^*) = 0 \text{ untuk } t \neq s \quad (28)$$

dengan menggunakan bentuk transformasi $u_{it}^* = u_{it} - \theta \bar{u}_i$, maka Persamaan (28) dapat dihitung menjadi Persamaan (29):

$$Cov(u_{it}^*, u_{is}^*) = Cov(u_{it}, u_{is}) - 2\theta Cov(u_{it}, \bar{u}_i) + \theta^2 Var(\bar{u}_i) \quad (29)$$

Berdasarkan pada struktur galat $u_{it} = \mu_i + e_{it}$, diperoleh bahwa:

$$Cov(\mu_i, \mu_i) = Var(\mu_i) = \sigma_\mu^2 \quad (30)$$

Selanjutnya, rata-rata galat individu dinyatakan sebagai $\bar{u}_i = \mu_i + \frac{1}{T} \sum_{t=1}^T e_{it}$, sehingga:

$$Cov(u_{it}, \bar{u}_i) = \sigma_\mu^2 + \frac{1}{T} \sigma_e^2 \quad (31)$$

$$Var(\bar{u}_i) = \sigma_\mu^2 + \frac{\sigma_e^2}{T} \quad (32)$$

Agar tidak terdapat korelasi antarwaktu, kovarians tersebut harus sama dengan nol, sehingga berdasarkan Persamaan (30), (31), dan (32) diperoleh:

$$\sigma_\mu^2 - 2\theta \left(\sigma_\mu^2 + \frac{\sigma_e^2}{T} \right) + \theta^2 \left(\sigma_\mu^2 + \frac{\sigma_e^2}{T} \right) = 0 \quad (33)$$

Membagi seluruh Persamaan (33) dengan $\sigma_\mu^2 + \frac{\sigma_e^2}{T}$ sehingga membentuk persamaan kuadrat seperti pada Persamaan (34):

$$\theta^2 - 2\theta + \frac{\sigma_\mu^2}{\sigma_\mu^2 + \frac{\sigma_e^2}{T}} = 0 \quad (34)$$

Penyelesaian Persamaan (34) dengan menggunakan rumus ABC diperoleh :

$$\theta = 1 \pm \sqrt{1 - \frac{\sigma_\mu^2}{\sigma_\mu^2 + \frac{\sigma_e^2}{T}}} \quad (35)$$

Namun, θ harus berada dalam rentang (0,1), maka solusi yang digunakan dari Persamaan (35) adalah:

$$\theta = 1 - \sqrt{1 - \frac{\sigma_\mu^2}{\sigma_\mu^2 + \frac{1}{T} \sigma_e^2}} \quad (36)$$

Apabila Persamaan (36) disederhanakan menjadi persamaan (37)

$$\theta = 1 - \sqrt{\frac{\hat{\sigma}_e^2}{T \hat{\sigma}_\mu^2 + \hat{\sigma}_e^2}} \quad (37)$$

1.4.8 Estimasi Parameter Model Geographically Weighted Panel Regression

Estimasi parameter dalam GWPR dilakukan dengan transformasi GLS yaitu dengan transformasi *quas demeaning*. Transformasi ini digunakan karena pada GLS biasa tidak dapat langsung mengurangi efek individu tanpa perlu menghitung kovarians lengkap. Transformasi ini juga menjaga struktur data panel dan memungkinkan estimasi parameter lokal yang lebih efisien dalam konteks spasial GWPR. Transformasi dilakukan berdasarkan Persamaan (27), dengan komponen utama berupa parameter transformasi θ yang diturunkan melalui Persamaan (37). Nilai θ

tersebut digunakan untuk menghilangkan autokorelasi galat antarwaktu dan mengurangi pengaruh efek individu yang tidak teramati (Jordan & Philips, 2023). Sehingga, diperoleh bentuk model yang telah ditransformasi sebagai berikut:

$$Y^* = X^* \beta + u^* \quad (38)$$

Keterangan:

Y^* = vektor amatan hasil transformasi dengan ukuran $nT \times 1$

X^* = matriks variabel prediktor hasil transformasi dengan ukuran $(nT \times (k + 1))$

β = vektor koefisien dengan ukuran $((k + 1) \times 1)$

u^* = vektor galat hasil transformasi dengan ukuran $nT \times 1$

Persamaan (38) dapat ditulis ulang menjadi Persamaan (39)

$$u^* = Y^* - X^* \beta \quad (39)$$

Pembobot yang digunakan untuk pengamatan (u_{it}, v_{it}) diperoleh dari Persamaan (18). Sehingga matriks pembobot spasial terbentuk adalah berbentuk matriks diagonal. Matriks pembobot tersebut ditunjukkan pada Persamaan (40).

$$W(u_{it}, v_{it}) = \begin{bmatrix} w(u_{1;1}, v_{1;1}) & 0 & \cdots & 0 \\ 0 & w(u_{2;1}, v_{2;1}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w(u_{n;T}, v_{n;T}) \end{bmatrix} \quad (40)$$

Pada Persamaan (41) dilakukan pendekatan *Weighted Least Square* (WLS). Pendekatan ini melibatkan penggunaan pembobot yang berbeda pada setiap lokasi dan waktu pengamatan, sehingga mampu menangkap variasi spasial dan temporal secara lokal (Karunia, 2021). Pendekatan WLS yang dilakukan yaitu:

$$WLS = u^{*'} W u^*$$

$$WLS = (Y^* - X^* \beta)' W(u_i, v_i) (Y^* - X^* \beta)$$

$$WLS = Y^{*'} W(u_i, v_i) Y^* - 2\beta' X^{*'} Y^* W(u_i, v_i) + \beta' X^{*'} X^* \beta \quad (41)$$

Koefisien regresi β diperoleh melalui differensiasi fungsi WLS pada Persamaan (41) terhadap β . Koefisien regresi tersebut ditunjukkan pada Persamaan (42).

$$\begin{aligned} \frac{\partial(WLS)}{\partial \beta} &= 0 \\ \frac{\partial(Y^{*'} W(u_i, v_i) Y^* - 2\beta' X^{*'} Y^* W(u_i, v_i) + \beta' X^{*'} X^* \beta)}{\partial \beta} &= 0 \\ -2X^{*'} Y^* W(u_i, v_i) + 2X^{*'} X^* \beta &= 0 \end{aligned}$$

$$\hat{\beta}(u_i, v_i) = (X^{*'}W(u_i, v_i)X^*)^{-1}X^{*'}W(u_i, v_i)Y^* \quad (42)$$

1.4.9 Pengujian Model *Geographically Weighted Panel Regression*

Pengujian model perlu dilakukan untuk mengevaluasi model yang diperoleh dan pengaruh variabel prediktor. Beberapa uji statistik yang sering digunakan dalam GWPR adalah uji *goodness of fit* dan uji parsial signifikansi parameter.

1. Uji *Goodness of Fit*

Uji *Goodness of Fit* dalam GWPR merupakan metode statistik yang digunakan untuk mengevaluasi sejauh mana model GWPR merepresentasikan data yang dianalisis. Uji *Goodness of Fit* penting untuk menentukan model yang digunakan sudah memadai atau memerlukan penyesuaian lebih lanjut (Raihani dkk., 2023). Selain itu, uji tersebut memberikan indikator kemampuan model GWPR dalam menangkap pola heterogenitas spasial dari data (Sifriyani dkk., 2024). Hal ini sangat relevan karena GWPR dirancang untuk mengakomodasi variabilitas spasial dan temporal dalam dataset panel. Berdasarkan hasil uji *Goodness of Fit*, dapat diketahui model GWPR lebih sesuai dibandingkan model global. Hipotesis yang digunakan dalam uji *Goodness of Fit* yaitu:

$H_0: \beta_j(u_i, v_i) = \beta_j; j = 1, 2, \dots, k$ (Model GWPR tidak lebih baik dibandingkan model global)

$H_1: \text{terdapat } \beta_j(u_i, v_i) \neq \beta_{j^*}(u_i, v_i); j \neq j^*; j = j^* = 1, 2, \dots, k$ (Model GWPR lebih baik dibandingkan model global)

Statistik uji:

$$F = \frac{\left(\frac{JKG_r - JKG_f}{df_r - df_f}\right)}{\left(\frac{JKG_f}{df_f}\right)} \sim F_{\alpha, df_1, df_2} \quad (43)$$

Keterangan:

JKG_r : jumlah kuadrat galat untuk model global.

JKG_f : jumlah kuadrat galat untuk model GWPR.

df_r : derajat bebas model global ($k_r - 1$).

df_f : derajat bebas model GWPR ($k_f - 1$).

Uji *Goodness of Fit* dengan nilai F memberikan cara untuk mengevaluasi apakah model GWPR secara signifikan lebih baik daripada model pembanding. Jika $F \geq F_{\alpha, df_1, df_2}$ atau $p - \text{value} < \alpha$, maka H_0 ditolak, yang berarti GWPR menangkap heterogenitas spasial dan temporal yang tidak ditangkap oleh model global.

2. Uji Parsial Signifikansi Prameter Model

Uji parsial signifikansi parameter dalam model GWPR bertujuan untuk menentukan parameter-parameter tertentu dalam model secara signifikan memengaruhi variabel respon pada wilayah atau lokasi tertentu. Sehingga model dapat dioptimalkan dengan mengecualikan variabel yang tidak signifikan (Nasri dkk., 2023). Model GWPR dirancang untuk menangkap variabilitas spasial, sehingga uji parsial membantu memastikan bahwa kontribusi parameter dapat berbeda secara signifikan antar wilayah (Yansi dkk., 2024). Hipotesis yang digunakan dalam uji parsial yaitu:

$H_0: \beta_j(u_i, v_i) = 0 ; j = 1, 2, \dots, k$ (tidak ada variabel prediktor yang berpengaruh secara parsial terhadap variabel respon)

$H_1: \text{terdapat } \beta_j(u_i, v_i) \neq 0$ (ada variabel prediktor yang berpengaruh secara parsial terhadap variabel respon)

Statistik uji:

$$t = \frac{\hat{\beta}_j(u_i, v_i)}{SE(\hat{\beta}_j(u_i, v_i))} \sim t_{\frac{\alpha}{2}, df} \quad (44)$$

Keterangan:

$\beta_j(u_i, v_i)$: koefisien regresi untuk variabel prediktor ke-j, yang bervariasi secara spasial sesuai dengan lokasi geografis (u_i, v_i) .

$SE(\hat{\beta}_j(u_i, v_i))$: standar galat lokal untuk koefisien $\beta_j(u_i, v_i)$

Pengambilan keputusan pada uji parsial akan melakukan penolakan terhadap H_0 , apabila nilai statistik uji $|t| \geq t_{\frac{\alpha}{2}, df}$ atau nilai $p - \text{value} < \alpha$, yang berarti terdapat variabel prediktor yang berpengaruh secara parsial terhadap variabel respon.

1.4.10 Ukuran Kebaikan Model

Salah satu metode yang umum digunakan untuk menentukan ukuran kebaikan model adalah koefisien determinasi (R^2). Koefisien determinasi merupakan ukuran statistik yang menjelaskan seberapa baik model regresi menjelaskan variabilitas data amatan. Semakin tinggi nilai R^2 , semakin baik model tersebut dalam memprediksi variabel respon berdasarkan variabel prediktornya. Karena GWPR memperhitungkan heterogenitas spasial, R^2 lokal memberikan wawasan tentang variasi model di setiap lokasi, yang penting untuk kebijakan berbasis data lokal (Ananda dkk., 2023). Nilai R^2 dapat diperoleh dengan menggunakan Persamaan (45)

$$R^2 = 1 - \frac{JKG}{JKT} \quad (45)$$

dengan JKG adalah jumlah kuadrat galat dan JKT adalah jumlah kuadrat total.

Metode lain yang digunakan untuk menentukan ukuran kebaikan model adalah dengan menggunakan *Root Mean Square Error* (RMSE). Metode RMSE adalah metrik statistik yang digunakan untuk mengevaluasi kinerja model prediksi dengan mengukur rata-rata kesalahan antara nilai yang diprediksi dan nilai aktual dalam satuan yang sama dengan data asli. Dalam konteks GWPR, RMSE sering digunakan untuk menilai seberapa baik model dapat memprediksi data berdasarkan distribusi spasialnya. Penggunaan RMSE dilakukan karena metode ini dapat memberikan gambaran langsung tentang tingkat kesalahan model tanpa mempertimbangkan kompleksitasnya (Amansyah dkk., 2024). Nilai RMSE yang lebih rendah menunjukkan model yang lebih akurat dalam merepresentasikan data, karena memiliki kesalahan prediksi yang lebih kecil. Nilai RMSE dapat diperoleh dari Persamaan (35).

$$RMSE = \sqrt{\frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n (y_{it} - \hat{y}_{it})^2} \quad (44)$$

dengan y_{it} adalah nilai aktual dan $\hat{y}_{it}$ adalah nilai prediksi

1.4.11 Angka Harapan Hidup

Angka Harapan Hidup (AHH) adalah indikator demografi yang menunjukkan perkiraan rata-rata jumlah tahun yang dapat diharapkan untuk hidup oleh seseorang sejak lahir, dengan asumsi bahwa pola angka kematian saat ini tetap berlaku sepanjang hidupnya (Maryani & Kristiana, 2018). AHH di Indonesia sendiri terus menunjukkan peningkatan. Menurut data Badan Pusat Statistik (BPS), pada tahun 2024, rata-rata angka harapan hidup penduduk Indonesia mencapai 74,15 tahun, meningkat 0,22 tahun dibandingkan tahun sebelumnya yang sebesar 73,93 tahun. Peningkatan ini mencerminkan perbaikan dalam berbagai aspek, seperti akses terhadap layanan kesehatan, gizi, pendidikan, serta kondisi sosial dan ekonomi masyarakat. Namun, seiring dengan meningkatnya AHH, tantangan baru muncul, seperti peningkatan jumlah penduduk lanjut usia yang memerlukan perhatian khusus dalam hal kesehatan dan kesejahteraan. Oleh karena itu, penting bagi pemerintah dan masyarakat untuk terus bekerja sama dalam meningkatkan kualitas hidup, sehingga peningkatan angka harapan hidup sejalan dengan peningkatan kesejahteraan dan kualitas hidup seluruh penduduk Indonesia. AHH dipengaruhi oleh berbagai faktor yang saling terkait, termasuk faktor ekonomi, pendidikan, lingkungan, dan aspek sosial lainnya. Setiap faktor ini berkontribusi pada keputusan individu atau pasangan dalam memilih metode kontrasepsi yang sesuai. Variabel yang dapat menjelaskan faktor-faktor tersebut adalah sebagai berikut:

1. Persentase Penduduk yang Memiliki Keluhan Kesehatan dalam Sebulan Terakhir

Data penduduk yang memiliki keluhan kesehatan dalam sebulan terakhir menggambarkan persentase individu yang mengalami gangguan kesehatan dalam 30 hari terakhir sebelum survei dilakukan. Data ini mencakup berbagai jenis keluhan kesehatan, mulai dari penyakit ringan seperti flu dan batuk hingga penyakit kronis seperti diabetes dan hipertensi. Data ini dapat berguna memantau kondisi kesehatan masyarakat, mengevaluasi efektivitas sistem layanan kesehatan, serta membantu pemerintah dalam perencanaan kebijakan kesehatan yang lebih tepat sasaran (Novitasari dkk., 2021). Tingginya angka penduduk yang mengalami keluhan kesehatan dalam sebulan terakhir dapat menjadi indikator rendahnya kualitas kesehatan masyarakat, yang dalam jangka panjang dapat berdampak pada angka harapan hidup. Jika keluhan kesehatan yang dialami penduduk sering berulang dan tidak ditangani dengan baik, maka risiko penyakit kronis dan kematian dini dapat meningkat, sehingga menurunkan rata-rata angka harapan hidup di suatu daerah. Rumus perhitungan yang digunakan yaitu:

$$PPMKK = \frac{PMKK}{JP} \times 100\%$$

Keterangan:

PPMKK : persentase penduduk yang memiliki keluhan kesehatan

PMKK : jumlah penduduk yang memiliki keluhan kesehatan

JP : jumlah penduduk

2. Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak

Data ini menunjukkan sejauh mana infrastruktur air bersih tersedia dan dapat diakses oleh rumah tangga di suatu wilayah. Sumber air minum layak mencakup air yang berasal dari fasilitas yang memenuhi standar kebersihan, seperti air leding atau perpipaan, sumur bor, sumur terlindung, mata air terlindung, air hujan yang dikelola dengan baik, serta air minum dalam kemasan yang memenuhi persyaratan kesehatan (Permadi dkk., 2023). Akses terhadap air minum yang layak sangat berpengaruh terhadap kesehatan masyarakat, terutama dalam mencegah berbagai penyakit yang disebabkan oleh air yang terkontaminasi. Rumus yang digunakan untuk memperoleh data yaitu:

$$PJRTAML = \frac{JRTAML}{JRT} \times 100\%$$

Keterangan:

PJRTAML : persentase rumah tangga yang memiliki akses air minum layak

JRTAML : jumlah rumah tangga yang memiliki akses air minum layak

JRT : jumlah rumah tangga

3. Persentase Penduduk Miskin

Persentase penduduk miskin adalah indikator yang menggambarkan proporsi penduduk dalam suatu wilayah yang hidup di bawah garis kemiskinan. Penduduk miskin didefinisikan sebagai individu atau rumah tangga yang memiliki pengeluaran per kapita di bawah garis kemiskinan, yaitu standar minimum yang ditetapkan berdasarkan kebutuhan dasar makanan dan non-makanan, seperti sandang, papan, pendidikan, dan kesehatan. Data persentase penduduk miskin memiliki implikasi yang luas dalam perencanaan pembangunan, karena tingkat kemiskinan yang tinggi sering kali berhubungan dengan rendahnya kualitas hidup masyarakat. Salah satu dampak signifikan dari kemiskinan adalah rendahnya angka harapan hidup. Penduduk miskin cenderung memiliki akses terbatas terhadap layanan kesehatan, gizi yang kurang memadai, serta kondisi lingkungan yang tidak mendukung kesehatan yang baik. Akibatnya, semakin tinggi persentase penduduk miskin dalam suatu wilayah, semakin rendah angka harapan hidupnya (Melati dkk., 2024). Rumus yang digunakan untuk memperoleh data yaitu:

$$PPM = \frac{JPM}{JP} \times 100\%$$

Keterangan:

PPM : persentase penduduk miskin

JPM : jumlah penduduk miskin

JP : jumlah penduduk

4. Tingkat Pengangguran Terbuka

Tingkat Pengangguran Terbuka (TPT) adalah indikator yang mengukur persentase angkatan kerja yang tidak memiliki pekerjaan tetapi sedang aktif mencari pekerjaan dalam suatu periode waktu tertentu. Semakin tinggi angka pengangguran terbuka, semakin besar jumlah individu yang belum berhasil mendapatkan pekerjaan, yang dapat mengindikasikan adanya masalah dalam penciptaan lapangan kerja, ketidakseimbangan antara keterampilan tenaga kerja dan kebutuhan industri, atau perlambatan pertumbuhan ekonomi. Tingkat pengangguran terbuka yang tinggi dapat berdampak negatif terhadap berbagai aspek kehidupan, termasuk kesejahteraan sosial, daya beli masyarakat, dan stabilitas sosial. Pengangguran yang berkepanjangan dapat meningkatkan risiko kemiskinan, menurunkan kualitas hidup, dan berdampak pada kesehatan masyarakat, yang pada akhirnya dapat mempengaruhi angka harapan hidup (Cholily, 2024). Rumus yang digunakan untuk memperoleh data yaitu:

$$TPT = \frac{JPR}{JAK} \times 100\%$$

Keterangan:

TPT : tingkat pengangguran terbuka

JPR : jumlah pengangguran

JAK : jumlah angkatan kerja

5. Produk Domestik Bruto Atas Dasar Harga Konstan

Produk Domestik Regional Bruto Atas Dasar Harga Konstan (PDRB ADHK) adalah indikator ekonomi yang mengukur nilai total produksi barang dan jasa yang dihasilkan dalam suatu wilayah dalam satu periode tertentu dengan menggunakan harga yang tetap (konstan). PDRB ADHK digunakan untuk menganalisis pertumbuhan ekonomi suatu daerah dari waktu ke waktu secara riil. Jika PDRB ADHK meningkat, berarti ekonomi daerah tersebut mengalami pertumbuhan yang positif, sedangkan jika PDRB ADHK tetap atau menurun, maka pertumbuhan ekonominya melemah. PDRB ADHK memiliki keterkaitan dengan angka harapan hidup, karena pertumbuhan ekonomi yang positif cenderung meningkatkan kesejahteraan masyarakat. Wilayah dengan PDRB ADHK yang tinggi biasanya memiliki akses yang lebih baik terhadap layanan kesehatan, pendidikan, dan infrastruktur, yang berkontribusi terhadap peningkatan kualitas hidup dan umur panjang penduduknya (Wardhana, 2019).

6. Persentase Rumah Tangga yang Memiliki Akses Sanitasi Layak

Sanitasi layak mencakup fasilitas pembuangan limbah domestik yang aman dan tidak mencemari lingkungan, seperti jamban dengan septic tank, sistem saluran pembuangan air limbah yang terkelola dengan baik, serta sarana cuci tangan dengan air bersih dan sabun (Permadi dkk., 2023). Akses terhadap sanitasi yang baik berperan penting dalam mencegah penyebaran penyakit menular seperti diare, kolera, tifus, dan infeksi cacing. Persentase rumah tangga dengan sanitasi layak mencerminkan tingkat kesadaran masyarakat terhadap kebersihan serta efektivitas program pemerintah dalam menyediakan infrastruktur sanitasi yang memadai. Daerah dengan akses sanitasi yang rendah cenderung menghadapi permasalahan kesehatan masyarakat yang lebih serius, seperti peningkatan angka kematian akibat penyakit infeksi dan gizi buruk akibat gangguan pencernaan yang berulang. Rumus yang digunakan untuk memperoleh data yaitu:

$$PJRTSL = \frac{JRTSL}{JRT} \times 100\%$$

Keterangan:

PJRTSL : persentase rumah tangga yang memiliki akses sanitasi layak

JRTSL : jumlah rumah tangga yang memiliki akses sanitasi layak

JRT : jumlah rumah tangga

BAB II METODE PENELITIAN

2.1 Data

Penelitian ini menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS), yaitu publikasi Statistik Indonesia 2023 dan situs resmi BPS di <https://www.bps.go.id/id>. Data yang digunakan berupa data panel yang terdiri dari 170 amatan, mencakup data time series dari tahun 2018 hingga 2022 dan data *cross section* dari 34 provinsi di Indonesia. Data tersebut disajikan pada Lampiran1. Variabel respon dan prediktor yang digunakan dalam penelitian ini dapat dilihat pada Tabel 2.

Tabel 2. Variabel Respon dan Prediktor

Variabel	Keterangan	Satuan
Y	Angka harapan hidup	Tahun
X_1	Persentase penduduk yang memiliki keluhan kesehatan dalam sebulan terakhir	Persen
X_2	Persentase rumah tangga yang memiliki akses sumber air minum layak	Persen
X_3	Persentase penduduk miskin	Persen
X_4	Tingkat pengangguran terbuka	Persen
X_5	Produk domestik regional bruto atas dasar harga konstan	Triliun Rupiah
X_6	Persentase rumah tangga yang memiliki akses sanitasi layak	Persen

2.2 Metode Analisis

Langkah-langkah analisis data yang dilakukan pada penelitian adalah sebagai berikut:

1. Analisis deskriptif untuk melihat kondisi data menggunakan peta tematik
2. Estimasi model regresi data panel kedalam tiga pendekatan umum yaitu:

a) CEM

$$y_{it} = \beta_0 + \sum_{j=1}^k \beta_j x_{jit} + e_{it}$$

b) FEM

$$y_{it} = \beta_{0i} + \sum_{j=1}^k \beta_j x_{jit} + e_{it}$$

c) REM

$$y_{it} = \beta_0 + \sum_{j=1}^k \beta_j x_{jit} + u_{it}; \quad u_{it} = \mu_i + e_{it}$$

3. Penentuan model regresi data panel terbaik dengan menggunakan :
 - a) Uji Chow
Uji Chow dilakukan untuk membandingkan model CEM dengan FEM, menggunakan Persamaan (14).
 - b) Uji Hausman
Uji Hausman dilakukan untuk membandingkan model FEM dengan REM, menggunakan Persamaan (15).
4. Pengujian Asumsi klasik yaitu:
 - a) Uji normalitas
Uji Kolmogorov-Smirnov dilakukan untuk menguji kenormalan distribusi galat menggunakan Persamaan (16)
 - b) Uji multikolinearitas
Uji VIF dilakukan untuk menguji keberadaan multikolinearitas pada variabel prediktor dengan Persamaan (18)
 - c) Uji heterogenitas spasial
Breusch-Pagan dilakukan untuk menguji efek heterogenitas spasial menggunakan Persamaan (19).
5. Perhitungan jarak *eucledean* antar lokasi i terhadap lokasi j yang terletak pada koordinat (u_i, v_i) menggunakan Persamaan (24)
6. Penentuan nilai bandwit optimum menggunakan pendekatan *cross-validation* (CV) sesuai dengan Persamaan (25)
7. Perhitungan matriks pembobot spasial dengan fungsi *adaptive kernel Gaussian* sesuai dengan Persamaan (20)
8. Transformasi data *quasi demeaning* sesuai dengan Persamaan (27)
9. Estimasi nilai parameter pemodelan GWPR sesuai dengan Persamaan (42)
10. Pengujian kecocokan model GWPR sesuai dengan Persamaan (43)
11. Pengujian parsial pemodelan GWPR sesuai dengan Persamaan (44)
12. Pengujian kebaikan model GWPR menggunakan R^2 dan RMSE sesuai dengan Persamaan (45) dan (46)