

**PEMODELAN *GEOGRAPHICALLY WEIGHTED PANEL
REGRESSION* PADA DATA ANGKA HARAPAN HIDUP
DI PROVINSI SULAWESI SELATAN**

SKRIPSI



**MULIANA
H051191010**

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2023**

**PEMODELAN *GEOGRAPHICALLY WEIGHTED PANEL
REGRESSION* PADA DATA ANGKA HARAPAN HIDUP
DI PROVINSI SULAWESI SELATAN**

SKRIPSI

**Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Sains pada
Program Studi Statistika Departemen Statistika Fakultas Matematika dan
Ilmu Pengetahuan Alam Universitas Hasanuddin**

**MULIANA
H051191010**

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN**

**MAKASSAR
AGUSTUS 2023**

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

Pemodelan Geographically Weighted Panel Regression Pada Data Angka Harapan Hidup Di Provinsi Sulawesi Selatan

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun

Makassar, 10 Agustus 2023



Muliana


NIM H051191010

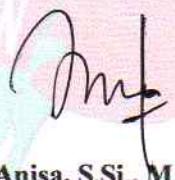
**PEMODELAN *GEOGRAPHICALLY WEIGHTED PANEL
REGRESSION* PADA DATA ANGKA HARAPAN HIDUP
DI PROVINSI SULAWESI SELATAN**

Disetujui Oleh:

Pembimbing Utama

Pembimbing Pertama


Drs. Raupong, M.Si.
NIP. 196210151988101001


Anisa, S.Si., M.Si.
NIP. 197302271998022001

Ketua Program Studi


Dr. Anna Islamiyati, S.Si., M.Si.
NIP. 197708082005012002

Pada 10 Agustus 2023

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Muliana
NIM : H051191010
Program Studi : Statistika
Judul Skripsi : *Pemodelan Geographically Weighted Panel Regression*
Pada Data Angka Harapan Hidup Di Provinsi Sulawesi Selatan

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1. Ketua : Drs. Raupong, M.Si. (.....)
2. Sekretaris : Anisa, S.Si., M.Si. (.....)
3. Anggota : Siswanto, S.Si., M.Si. (.....)
4. Anggota : Sitti Sahriman, S.Si., M.Si. (.....)

Ditetapkan di : Makassar

Tanggal : 10 Agustus 2023

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi ini. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa Sallam* beserta keluarga dan para sahabatnya. *Alhamdulillahirobbil'alamin*, berkat nikmat kemudahan dan pertolongan yang diberikan oleh Allah *Subhanahu Wa Ta'ala*, penulis dapat menyelesaikan skripsi yang berjudul “Pemodelan *Geographically Weighted Panel Regression* Pada Data Angka Harapan Hidup Di Provinsi Sulawesi Selatan” yang disusun sebagai salah satu syarat akademik untuk memperoleh gelar sarjana pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulis menyadari bahwa dalam penyelesaian skripsi ini tidak lepas dari bantuan dan dorongan dari berbagai pihak yang senantiasa turut membantu dalam bentuk moril maupun materil sehingga dengan segala keterbatasan kemampuan dan pengetahuan, penulis dapat menyelesaikan skripsi ini. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang setulus-tulusnya kepada kepada semua pihak yang terlibat. Dengan penuh kesadaran dan kerendahan hati, pada kesempatan ini perkenankanlah penulis menyampaikan ucapan terima kasih dan penghargaan setinggi-tingginya kepada yang terhormat:

1. Terima kasih dan penghargaan yang setinggi-tingginya kepada ayahanda Manye dan Ibunda Nuria atas kasih sayang, pengorbanan, kesabaran hati, dukungan penuh serta doa yang selalu dipanjatkan demi kelancaran setiap langkah dalam hidup penulis.
2. Terima kasih kepada kakak tersayang Nur Ida, Mustafa, Jufri, Aminah dan Ramlah serta keluarga besar yang selalu memberikan dukungan dan harapan besar kepada penulis yang membuat penulis selalu termotivasi untuk terus berusaha dan pantang menyerah.
3. Terima kasih kepada Prof. Dr. Ir. Jamaluddin Jompa, M.Sc., selaku Rektor Universitas Hasanuddin beserta seluruh staf jajarannya.

4. Terima kasih Bapak Dr. Eng. Amiruddin, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh staf jajarannya.
5. Terima kasih kepada Ibu Dr. Anna Islamiyati, S.Si., M.Si. selaku Ketua Departemen Statistika yang telah seperti orang tua sendiri. Segenap dosen pengajar dan staf Departemen Statistika yang telah membekali ilmu dan kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
6. Terima kasih kepada Ibu Dr. Nurtiti Sunusi S.Si., M.Si., selaku Ketua Departemen Statistika sekaligus Penasihat Akademik penulis pada masa-masa awal perkuliahan penulis yang banyak memberikan nasihat yang sangat berguna dalam kehidupan perkuliahan penulis.
7. Terima kasih kepada Bapak Drs. Raupong, M.Si. sebagai Pembimbing Utama dan Ibu Anisa, S.Si., M.Si. sebagai Pembimbing Pertama yang telah bersedia untuk meluangkan waktu untuk memberi masukan serta memberikan solusi atas semua permasalahan yang ada pada penelitian ini.
8. Terima kasih kepada Bapak Siswanto, S.Si., M.Si., sebagai Dosen Penguji sekaligus Penasihat Akademik penulis yang telah bersedia senantiasa meluangkan waktu memberikan saran, arahan serta solusi terbaik dalam proses perkuliahan hingga proses penyelesaian skripsi penulis.
9. Terima kasih kepada Ibu Sitti Sahruman, S.Si., M.Si. sebagai Dosen Penguji yang telah bersedia meluangkan waktunya untuk memberikan penilaian dan masukan terhadap skripsi ini.
10. Terima kasih yang setulus-tulusnya kepada Muhammad Syamsul Bahri yang selalu ada menemani dan memberikan asupan semangat bagi penulis sehingga dapat menuntaskan masa perkuliahan dan penyusunan skripsi dengan baik.
11. Terima kasih kepada sobat Axis Seli Lisnayati, Yasmin Pratiwi, Nurazizah, Nursyahfika dan Mayashari yang telah memberikan waktu dan pengalaman yang berharga dan menyenangkan di masa perkuliahan penulis. Kalian hebat, terimakasih atas cerita yang sudah diukir selama perkuliahan.
12. Terima kasih kepada teman-teman: Nur Aisyah, Nurul Hikmah, Andi Muhammad Rajab, Fadilah Amirul Adel, Muhammad Fathurrahman, Arief

Rahman Nur dan Sapriadi Rasyid. Terima kasih telah memberikan tempat berada di sekitar kalian, terima kasih atas waktu dan cerita yang menyenangkan di masa perkuliahan penulis.

13. Terima kasih kepada semua teman-teman di angkatan Statistika 2019 yang telah mengisi masa perkuliahan penulis dengan banyak sekali cerita yang akan terus terkenang dalam ingatan penulis.
14. Terima kasih yang setinggi-tingginya kepada seluruh pihak yang mungkin tidak sempat penulis sebutkan satu persatu. Terima kasih atas segala dukungan, partisipasi, dan apresiasinya yang diberikan kepada penulis.

Penulis menyadari bahwa masih banyak kekurangan dalam penyusunan skripsi ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini dapat memberikan manfaat untuk berbagai pihak.

Makassar, 10 Agustus 2023



Muliana

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIK**

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Muliana
NIM : H051191010
Program Studi : Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty- Free Right*)** atas tugas akhir saya yang berjudul:

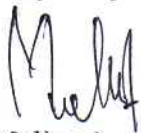
**“Pemodelan *Geographically Weighted Panel Regression* Pada Data Angka
Harapan Hidup Di Provinsi Sulawesi Selatan”**

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar tanggal 10 Agustus 2023.

Yang menyatakan,



(Muliana)

ABSTRAK

Data panel merupakan gabungan antara data *time series* dengan data *cross section*. Regresi data panel merupakan model yang digunakan untuk menentukan hubungan antara variabel dependen dengan independen pada data panel. Namun, perbedaan karakteristik lingkungan dan geografis antar lokasi pengamatan menyebabkan terdapat perbedaan pengaruh variabel independen terhadap variabel dependen pada setiap lokasi pengamatan yang disebut sebagai heterogenitas spasial. Model regresi spasial yang digunakan ketika terdapat heterogenitas spasial pada data panel adalah *geographically weighted panel regression* (GWPR). Penelitian ini bertujuan untuk memperoleh model pendugaan GWPR serta hasil pengelompokan wilayah kabupaten/kota berdasarkan variabel independen yang berpengaruh signifikan pada data angka harapan hidup di Provinsi Sulawesi Selatan. Fungsi pembobot spasial yang digunakan pada model GWPR adalah *fixed gaussian kernel*. Data yang digunakan dalam penelitian ini adalah data angka harapan hidup di Provinsi Sulawesi Selatan tahun 2017 - 2021 dengan faktor yang diduga mempengaruhi adalah harapan lama sekolah, pengeluaran perkapita, persentase rumah tangga bersanitasi layak, persentase rumah tangga yang memiliki sumber air minum layak, persentase penduduk miskin dan rata-rata lama sekolah. Hasil penelitian menunjukkan bahwa enam variabel independen yang digunakan mampu menjelaskan angka harapan hidup di Provinsi Sulawesi Selatan dengan nilai koefisien determinasi sebesar 97.86% serta memiliki nilai RMSE yaitu 0.05. Model pendugaan GWPR yang terbentuk berbeda setiap kabupaten/kota. Hasil uji parsial menunjukkan bahwa semua variabel independen berpengaruh signifikan secara parsial terhadap variabel dependen dikarenakan terdapat minimal satu wilayah kabupaten/kota memiliki hasil yang signifikan sehingga terbentuk sembilan kelompok wilayah berdasarkan variabel independen yang berpengaruh signifikan secara parsial.

Kata Kunci: Angka Harapan Hidup, Data Panel, Heterogenitas Spasial, Model GWPR

ABSTRACT

Panel data is combination of time series and cross section data. Panel regression is a model used to determine relationship between dependent and independent variables in panel data. Differences in environmental and geographical characteristics between observation locations cause differences in the effect of independent on dependent variable at each observation location which referred to as spatial heterogeneity. Spatial regression used when there is spatial heterogeneity in panel data is geographically weighted panel regression (GWPR). This study aims to obtain GWPR estimation and the results of grouping cities based on independent variables that have a significant effect on life expectancy data in South Sulawesi. Spatial weighting function used in GWPR model is a fixed gaussian kernel. Data used in this study are life expectancy in South Sulawesi in 2017 - 2021 with factors that influence the expected length of schooling, percapita spending, percentage of households with proper sanitation, percentage of households with proper drinking sources, the percentage of the population poor and average length of schooling. The results showed that six independent variables used were able to explain life expectancy in South Sulawesi Province with a coefficient determination of 97.86% an RMSE of 0.05. The GWPR estimation model formed is different for each city. The results of partial test show that all independent variables have partially significant effect on dependent variable because there is at least one city region that has significant results so that nine regional groups are formed based on independent variables that have a partial significant.

Keywords: *Life Expectancy, Panel Data, Spatial Heterogeneity, GWPR Model*

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
LEMBAR PERNYATAAN KEOTENTIKAN	iii
HALAMAN PERSETUJUAN PEMBIMBING	iv
HALAMAN PENGESAHAN	v
KATA PENGANTAR	vi
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	ix
ABSTRAK	x
ABSTRACT	xi
DAFTAR ISI	xii
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
BAB II TINJAUAN PUSTAKA	4
2.1 Regresi Data Panel.....	4
2.1.1 Estimasi Model Regresi Data Panel	5
2.1.2 Penentuan Model Terbaik Regresi Data Panel.....	7
2.2 Heterogenitas Spasial.....	8
2.3 Fungsi Pembobot Spasial.....	9
2.4 <i>Geographically Weighted Panel Regression</i>	10
2.4.1 Estimasi Parameter Model <i>Geographically Weighted Panel</i> <i>Regression</i>	11
2.4.2 Pengujian Hipotesis Model <i>Geographically Weighted Panel</i> <i>Regression</i>	11

2.5	Ukuran Keباikan Model.....	13
2.6	Angka Harapan Hidup	13
BAB III METODOLOGI PENELITIAN		17
3.1	Data.....	17
3.2	Metode Analisis	17
BAB IV HASIL DAN PEMBAHASAN		20
4.1	Analisis Deskriptif	20
4.2	Regresi Data Panel.....	23
	4.2.1 Estimasi Model Regresi Data Panel	23
	4.2.2 Penentuan Model Regresi Data Panel Terbaik	24
4.3	Pengujian Heterogenitas Spasial.....	25
4.4	Fungsi Pembobot Spasial.....	26
4.5	Pemodelan <i>Geographically Weighted Panel Regression</i>	29
4.6	Pengujian Hipotesis	30
	4.6.1 Uji Kecocokan Model	30
	4.6.2 Uji Parsial Signifikansi Parameter Model.....	31
4.7	Ukuran Keباikan Model.....	34
4.8	Interpretasi Model.....	35
BAB V KESIMPULAN DAN SARAN		37
5.1	Kesimpulan	37
5.2	Saran	37
DAFTAR PUSTAKA		38
LAMPIRAN.....		42

DAFTAR GAMBAR

Gambar 4.1. Peta Sebaran AHH di Sulawesi Selatan Tahun 2017-2021 23

Gambar 4.2. Peta Tematik Pengelompokan Wilayah Berdasarkan Uji Parsial. 34

DAFTAR TABEL

Tabel 2.1. Struktur Data panel.....	4
Tabel 3.1. Variabel Dependen dan Independen.....	17
Tabel 4.1. Statistika Deskriptif.....	20
Tabel 4.2. Uji <i>Chow</i>	24
Tabel 4.3. Uji <i>Hausman</i>	25
Tabel 4.4. Pengujian Heterogenitas Spasial	25
Tabel 4.5. <i>Bandwidth</i> Optimum dan Nilai CV	26
Tabel 4.6. Jarak <i>Euclidean</i> dan Pembobot Spasial di Kabupaten Enrekang	28
Tabel 4.7. Uji Kecocokan Model	30
Tabel 4.8. Uji Signifikansi Parameter Kabupaten Enrekang.....	31
Tabel 4.9. Pengelompokkan Wilayah Berdasarkan Variabel yang Signifikan	32
Tabel 4.10. Ukuran Kebaikan Model	34

DAFTAR LAMPIRAN

Lampiran 1. Data Angka Harapan Hidup di Provinsi Sulawesi Selatan dan Faktor yang Memengaruhinya.....	43
Lampiran 2. Perhitungan Model Regresi Data Panel	44
Lampiran 3. Perhitungan Pengujian Penentuan Model Regresi Panel Terbaik ..	45
Lampiran 4. Pengujian Heterogenitas Spasial.....	46
Lampiran 5. Jarak <i>Euclidean</i> Antar Lokasi Pengamatan	47
Lampiran 6. Fungsi Pembobot <i>Fixed Gaussian Kernel</i>	48
Lampiran 7. Hasil Transformasi Data dengan <i>Within Group</i>	50
Lampiran 8. Estimasi Parameter Model GWPR dengan Pendekatan FEM	51
Lampiran 9. Pengujian Kesesuaian Model	53
Lampiran 10. Nilai T_{hitung} Berdasarkan Uji Parsial Signifikansi Parameter	54
Lampiran 11. Nilai Koefisien Determinasi Secara Lokal	56

BAB I PENDAHULUAN

1.1 Latar Belakang

Analisis regresi linier adalah metode yang digunakan untuk menentukan hubungan linier antara satu variabel dependen dengan satu atau lebih variabel independen (Amaliah dkk., 2020). Jenis data yang biasanya digunakan pada analisis regresi linier umumnya adalah data *cross section* sehingga penelitian hanya dilakukan dalam satu kurun waktu. Namun, dalam suatu penelitian tidak cukup hanya melakukan pengamatan terhadap unit amatan dalam satu kurun waktu, tetapi perlu juga mengamati unit tersebut pada berbagai periode waktu agar informasi yang diperoleh lebih akurat (Wati dkk., 2021). Metode regresi yang digunakan untuk memodelkan data *cross section* dan data *time series* adalah regresi data panel. Keuntungan menggunakan regresi data panel adalah mampu menyediakan data yang lebih banyak. Banyaknya data yang digunakan bertujuan untuk meningkatkan nilai derajat kebebasan sehingga memperoleh hasil estimasi yang lebih efisien (Widarjono, 2005).

Model regresi data panel secara umum terdapat tiga macam yaitu *common effect model* (CEM), *fixed effect model* (FEM) dan *random effect model* (REM). CEM merupakan pendekatan model data panel yang tidak memperhatikan perbedaan karakteristik pada dimensi waktu maupun unit *cross section* (lokasi) sehingga untuk mengetahui adanya perbedaan dapat digunakan FEM dengan mengasumsikan bahwa intersep berbeda antar lokasi (Nengsih & Martaliah, 2021). Pada REM, mengasumsikan nilai *residual* tidak berkorelasi dengan variabel independen sehingga penggunaan REM jarang digunakan pada bidang ekonomi karena sulitnya memenuhi asumsi tersebut (Bell & Jones, 2015). Penerapan regresi pada beberapa kasus, sering terjadi bahwa variabel dependen yang diamati memiliki keterkaitan dengan hasil pengamatan antar satu wilayah dengan wilayah lain yang berbeda. Hal ini disebabkan adanya efek spasial yang dapat mempengaruhi model (Wati dkk., 2021).

Metode regresi yang digunakan untuk mengatasi adanya efek spasial disebut sebagai regresi spasial. Model regresi spasial secara umum terbagi atas dependensi spasial dan heterogenitas spasial. Dependensi spasial menunjukkan

adanya keterkaitan antar lokasi objek penelitian sedangkan heterogenitas spasial menunjukkan adanya keragaman antar lokasi (Mahmud & Pasaribu, 2021). Perbedaan karakteristik lingkungan dan geografis antar lokasi pengamatan memungkinkan munculnya heterogenitas spasial (Astuti dkk., 2018). Pada kasus heterogenitas spasial dikembangkanlah analisis regresi spasial yaitu *geographically weighted regression* (GWR) yang dapat mengakomodasi efek spasial sehingga parameter regresi yang dihasilkan bersifat lokal. Berdasarkan uraian tersebut, berkembanglah metode *geographically weighted panel regression* (GWPR) yang merupakan gabungan antara model regresi panel dengan GWR.

Penelitian sebelumnya terkait pemodelan GWPR yaitu penelitian yang dilakukan oleh Utami dan Yanti (2021) mengenai pemodelan GWPR pada kasus pneumonia balita di Kota Bandung menggunakan fungsi pembobot *adaptive bisquare*. Selain itu, GWPR diterapkan pada pemodelan persentase penduduk miskin di Provinsi Kalimantan Timur menggunakan fungsi pembobot *adaptive bisquare* (Pratama dkk., 2021). Hasil penelitian tersebut menunjukkan bahwa model GWPR lebih baik dibandingkan model regresi panel global. Berdasarkan penelitian yang dilakukan oleh Maulana dkk (2019), fungsi pembobot *fixed gaussian* memberikan hasil yang lebih baik dibandingkan dengan *fixed bisquare*. Selain itu, pada pemodelan regresi spasial juga menunjukkan bahwa fungsi pembobot *fixed gaussian* merupakan model terbaik (Pratiwi dkk., 2019).

Salah satu penerapan GWPR yaitu pada kasus angka harapan hidup (AHH). Berdasarkan data Badan Pusat Statistik, nilai AHH di Indonesia mengalami peningkatan, tetapi terdapat 25 Provinsi di Indonesia yang memiliki nilai AHH di bawah rata-rata nasional. Salah satunya adalah Provinsi Sulawesi Selatan. Nilai AHH di Sulawesi Selatan dapat dimodelkan menggunakan model GWPR karena merupakan data panel spasial yang mengandung unsur lokasi dan waktu yang berbeda. Berdasarkan uraian tersebut, maka penelitian ini difokuskan dalam pembentukan model pendugaan GWPR dan diharapkan hasil dari pengelompokan wilayah diperoleh informasi terkait faktor-faktor yang dapat memberikan pengaruh signifikan terhadap angka harapan hidup agar pemerintah dapat membuat suatu kebijakan yang tepat dalam meningkatkan angka harapan hidup tiap kabupaten/kota di Provinsi Sulawesi Selatan.

1.2 Rumusan Masalah

Rumusan masalah yang dibahas pada penelitian ini adalah sebagai berikut:

1. Bagaimana memperoleh model pendugaan *Geographically Weighted Panel Regression* yang terbentuk pada data angka harapan hidup di Provinsi Sulawesi Selatan?
2. Bagaimana pengelompokkan wilayah kabupaten/kota berdasarkan variabel independen yang berpengaruh signifikan pada data angka harapan hidup di Provinsi Sulawesi Selatan menggunakan model *Geographically Weighted Panel Regression*?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data panel berupa data *time series* dari tahun 2017-2021 dan data *cross section* meliputi 24 kabupaten/kota di Sulawesi Selatan.
2. Pemilihan model *Geographically Weighted Panel Regression* menggunakan pendekatan FEM dengan fungsi pembobot *fixed gaussian kernel*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Memperoleh model pendugaan *Geographically* yang terbentuk pada data angka harapan hidup di Provinsi Sulawesi Selatan.
2. Memperoleh hasil pengelompokkan wilayah kabupaten/kota berdasarkan variabel yang berpengaruh signifikan pada data angka harapan hidup di Provinsi Sulawesi Selatan menggunakan model *Geographically Weighted Panel Regression*.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari hasil penelitian ini adalah sebagai berikut:

1. Memperoleh wawasan baru mengenai model GWPR yang diterapkan pada data angka harapan hidup di Provinsi Sulawesi Selatan.
2. Menjadi sumber acuan dalam menetapkan kebijakan pemerintah untuk meningkatkan angka harapan hidup setiap kabupaten/kota di Provinsi Sulawesi Selatan.

BAB II
TINJAUAN PUSTAKA

2.1 Regresi Data Panel

Data panel adalah gabungan antara data silang (*cross section*) dengan data runtun waktu (*time series*). Unit *cross section* berupa individu, perusahaan dan wilayah atau lokasi, sedangkan unit *time series* berupa harian, bulanan dan tahunan. Pada data panel, setiap unit *cross section* diamati secara berulang-ulang selama beberapa waktu. Struktur data panel disajikan pada Tabel 2.1.

Tabel 2.1. Struktur Data panel

Wilayah	Tahun	y_{it}	x_{1it}	x_{2it}	...	x_{pit}
1	1	$y_{1,1}$	$x_{1,1,1}$	$x_{2,1,1}$...	$x_{p,1,1}$
2	1	$y_{2,1}$	$x_{1,2,1}$	$x_{2,2,1}$...	$x_{p,2,1}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	1	$y_{n,1}$	$x_{1,n,1}$	$x_{2,n,1}$...	$x_{p,n,1}$
1	2	$y_{1,2}$	$x_{1,1,2}$	$x_{2,1,2}$...	$x_{p,1,2}$
2	2	$y_{2,2}$	$x_{1,2,2}$	$x_{2,2,2}$...	$x_{p,2,2}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	2	$y_{n,2}$	$x_{1,n,2}$	$x_{2,n,2}$...	$x_{p,n,2}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	T	$y_{1,T}$	$x_{1,1,T}$	$x_{2,1,T}$...	$x_{p,1,T}$
2	T	$y_{2,T}$	$x_{1,2,T}$	$x_{2,2,T}$...	$x_{p,2,T}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	T	$y_{n,T}$	$x_{1,n,T}$	$x_{2,n,T}$...	$x_{p,n,T}$

Sumber: Rusgiyono & Prahutama (2021)

Regresi data panel merupakan regresi yang menggabungkan data *cross section* dan data *time series* (Amaliah dkk., 2020). Model regresi data panel secara umum menurut Baltagi (2005) dituliskan seperti pada Persamaan (2.1).

$$y_{it} = \beta_{oit} + \sum_{k=1}^p \beta_k x_{kit} + \varepsilon_{it} \tag{2.1}$$

Keterangan:

- i : lokasi pengamatan, dengan $i = 1, 2, \dots, n$
- t : waktu pengamatan, dengan $t = 1, 2, \dots, T$
- y_{it} : variabel dependen pada lokasi ke- i dan waktu ke- t
- β_{oit} : koefisien intersep pada lokasi ke- i dan waktu ke- t
- β_k : koefisien regresi variabel independen ke- k
- x_{kit} : variabel independen ke- k pada lokasi ke- i dan waktu ke- t
- ε_{it} : *residual* lokasi ke- i dan waktu ke- t

2.1.1 Estimasi Model Regresi Data Panel

Terdapat 3 pendekatan dalam mengestimasi model regresi dengan data panel yaitu *common effect model*, *fixed effect model* dan *random effect model*.

1. *Common effect model*

Common effect model (CEM) merupakan pendekatan estimasi data panel yang menggabungkan seluruh data tanpa memperhatikan lokasi maupun waktu pengamatan. Menurut Baltagi (2005) CEM menggunakan pendekatan *ordinary least square* (OLS) untuk menduga parameternya. Asumsi yang digunakan pada metode ini adalah nilai intersep bersifat konstan yaitu sama untuk semua lokasi dan waktu. Bentuk umum model CEM adalah seperti pada Persamaan (2.2).

$$y_{it} = \beta_0 + \sum_{k=1}^p \beta_k x_{kit} + \varepsilon_{it} \quad (2.2)$$

Keterangan:

y_{it} : variabel dependen pada lokasi ke- i dan waktu ke- t

β_0 : koefisien intersep

β_k : koefisien regresi variabel independen ke- k

x_{kit} : variabel independen ke- k pada lokasi ke- i dan waktu ke- t

ε_{it} : *residual* lokasi ke- i dan waktu ke- t

2. *Fixed effect model*

Salah satu cara memperhatikan keberagaman lokasi pada model regresi data panel adalah dengan membedakan nilai intersep pada setiap lokasi dengan mengasumsikan koefisien regresi konstan. Model ini dikenal dengan *fixed effect model* (FEM). Bentuk umum model FEM adalah seperti pada Persamaan (2.3).

$$y_{it} = \beta_{0i} + \sum_{k=1}^p \beta_k x_{kit} + \varepsilon_{it} \quad (2.3)$$

Menurut Wooldridge (2002), salah satu metode yang dapat digunakan untuk melihat perbedaan intersep pada FEM adalah *within group*. Pendekatan ini dilakukan dengan cara mengeliminasi efek unit lokasi (β_{0i}), kemudian nilai variabel dependen dan variabel independen dari setiap lokasi dirata-ratakan terhadap waktu. Persamaan rata-rata dinyatakan sebagai berikut (Wooldridge, 2002):

$$\bar{y}_i = \beta_{0i} + \sum_{k=1}^p \beta_k \bar{x}_{ki} + \bar{\varepsilon}_i \quad (2.4)$$

Model FEM *within group* diperoleh dengan cara mengurangi Persamaan (2.3) dan (2.4) sehingga diperoleh persamaan seperti berikut:

$$y_{it} - \bar{y}_i = \beta_{0i} - \beta_{0i} + \sum_{k=1}^p \beta_k (x_{kit} - \bar{x}_{ki}) + \varepsilon_{it} - \bar{\varepsilon}_i \quad (2.5)$$

Persamaan (2.5) dapat disederhanakan menjadi:

$$y_{it}^* = \sum_{k=1}^p \beta_k x_{kit}^* + \varepsilon_{it}^* \quad (2.6)$$

dengan $i = 1, 2, \dots, n$; $t = 1, 2, \dots, T$

Model FEM dengan *within group* jika dituliskan dalam bentuk matriks adalah sebagai berikut:

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \quad (2.7)$$

dengan:

$$\mathbf{y}^* = y_{it} - \bar{y}_i$$

$$\mathbf{X}^* = x_{kit} - \bar{x}_{ki}$$

$$\boldsymbol{\varepsilon}^* = \varepsilon_{it} - \bar{\varepsilon}_i$$

\mathbf{y}^* merupakan vektor variabel dependen rata-rata terkoreksi berukuran $nT \times 1$, \mathbf{X}^* merupakan matriks variabel independen rata-rata terkoreksi berukuran $nT \times p$ dan $\boldsymbol{\varepsilon}^*$ merupakan vektor *residual* rata-rata terkoreksi berukuran $nT \times 1$. Estimasi parameter $\boldsymbol{\beta}$ pada Persamaan (2.7) dapat digunakan metode OLS sehingga diperoleh:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{*'} \mathbf{X}^*)^{-1} (\mathbf{X}^{*'} \mathbf{y}^*) \quad (2.8)$$

dengan $\hat{\boldsymbol{\beta}}$ merupakan vektor koefisien regresi berukuran $p \times 1$.

Perbedaan intersep antar individu dapat dihitung menggunakan estimasi sebagai berikut (Hufaini dkk., 2020):

$$\hat{\beta}_{0i} = \bar{y}_i - (\hat{\beta}_1 \bar{x}_{1i}) - (\hat{\beta}_2 \bar{x}_{2i}) - \dots - (\hat{\beta}_p \bar{x}_{pi}) \quad (2.9)$$

3. *Random effect model*

Pendekatan *random effect model* (REM) mengasumsikan setiap lokasi mempunyai perbedaan intersep, yang mana intersep tersebut adalah variabel

random yang dimasukkan ke dalam model sebagai *residual* (Nengsih & Martaliah, 2021). Model REM dapat diestimasi dengan metode *generalized least square* (GLS) sebab jika diestimasi dengan OLS hasil estimasi tidak efisien (Gujarati, 2004). Bentuk umum model REM adalah seperti pada Persamaan (2.10).

$$y_{it} = \beta_0 + \sum_{k=1}^p \beta_k x_{kit} + w_{it} \quad (2.10)$$

dengan $w_{it} = \mu_i + \varepsilon_{it}$.

Keterangan:

y_{it} : variabel dependen pada lokasi ke- i dan unit waktu ke- t

β_0 : koefisien intersep

μ_i : *residual* lokasi ke- i

β_k : koefisien regresi variabel independen ke- k

x_{kit} : variabel independen ke- k pada lokasi ke- i dan waktu ke- t

ε_{it} : *residual* lokasi ke- i dan waktu ke- t

2.1.2 Penentuan Model Terbaik Regresi Data Panel

1. Uji *Chow*

Uji *Chow* menguji signifikansi intersep β_{0i} yang berbeda-beda pada masing-masing lokasi atau tidak berbeda. Jika intersep tiap lokasi adalah sama, maka model regresi data panel yang digunakan adalah model CEM. Jika intersep antar lokasi berbeda maka perlu dilanjutkan dengan uji *hausman* untuk menentukan model terbaik antara model FEM atau model REM. Hipotesis yang digunakan adalah sebagai berikut (Baltagi, 2005):

$H_0: \beta_{01} = \beta_{02} = \dots = \beta_{0n} = 0$ (intersep tiap lokasi adalah sama)

H_1 : minimal ada satu $\beta_{0i} \neq 0 ; i = 1, 2, \dots, n$ (terdapat lokasi yang memiliki intersep berbeda)

Statistik uji yang digunakan merupakan uji F seperti berikut:

$$F_{hitung} = \frac{(RSS_1 - RSS_2)/(n - 1)}{RSS_2/(nT - n - p)} \quad (2.11)$$

Keterangan:

n : jumlah lokasi pengamatan

T : jumlah periode waktu pengamatan

p : jumlah variabel independen

RSS_1 : *residual sums of squares* (jumlah kuadrat galat) dari model CEM

RSS_2 : *residual sums of squares* (jumlah kuadrat galat) dari model FEM

Penarikan kesimpulan yaitu jika $F_{hitung} > F_{n-1, nT-n-p}$ atau $p - value < \alpha$ (taraf signifikansi) maka tolak H_0 , yang berarti bahwa terdapat lokasi yang memiliki intersep yang berbeda sehingga perlu dilanjutkan dengan uji *hausman*.

2. Uji *Hausman*

Uji *Hausman* digunakan untuk menguji adanya korelasi antara *residual* dengan satu atau lebih variabel independen dalam model (Alwi dkk., 2018). Jika terdapat korelasi antara *residual* dengan variabel independen, maka diasumsikan model terbaik antara FEM dan REM adalah model FEM. Hipotesis yang digunakan adalah sebagai berikut:

$H_0: corr(x_{it}, \varepsilon_{it}) = 0$ (tidak ada korelasi antara *residual* dengan variabel independen)

$H_1: corr(x_{it}, \varepsilon_{it}) \neq 0$ (ada korelasi antara *residual* dengan variabel independen)

Statistik uji yang digunakan mengikuti distribusi *chi-square* berdasarkan kriteria *Wald* sebagai berikut:

$$W = (\hat{\beta}_{FEM} - \hat{\beta}_{REM})' [var(\hat{\beta}_{FEM}) - var(\hat{\beta}_{REM})]^{-1} (\hat{\beta}_{FEM} - \hat{\beta}_{REM}) \quad (2.12)$$

$\hat{\beta}_{FEM}$ merupakan vektor dari estimasi parameter FEM dan $\hat{\beta}_{REM}$ merupakan vektor estimasi parameter REM. Penarikan kesimpulan yaitu jika $W > \chi^2_{\alpha, p}$ atau $p - value < \alpha$ maka tolak H_0 , yang berarti terdapat korelasi antara *residual* dengan variabel independen dalam model sehingga model terbaik antara REM dan FEM adalah model FEM.

2.2 Heterogenitas Spasial

Heterogenitas spasial terjadi akibat adanya perbedaan karakteristik satu wilayah dengan wilayah lainnya. Ketika ragam antar lokasi pengamatan tidak sama, maka mengindikasikan terdapat heterogenitas spasial. Menguji heterogenitas spasial dalam model regresi sangat penting karena mengabaikan hal tersebut akan menyebabkan estimasi tidak efisien. Pengujian heterogenitas spasial dilakukan dengan menggunakan statistik uji *Breusch-Pagan* (Rahayu, 2017). Hipotesis yang digunakan adalah seperti berikut:

$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$ (ragam *residual* sama untuk semua lokasi)

H_1 : minimal ada satu $\sigma_i^2 \neq \sigma^2$ (ragam *residual* tidak sama untuk semua lokasi)

Statistik uji yang digunakan pada uji *Breusch-Pagan* sebagai berikut:

$$BP = \left(\frac{1}{2}\right) \mathbf{f}' \mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' \mathbf{f} \quad (2.13)$$

dengan elemen vektor f adalah

$$f_{it} = \left(\frac{\varepsilon_{it}^2}{\sigma^2} - 1 \right)$$

ε_{it} adalah *residual* untuk lokasi pengamatan ke- i dan waktu ke- t , σ^2 adalah ragam *residual* dan \mathbf{Z} merupakan matriks variabel independen berukuran $nT \times (p + 1)$ yang sudah distandarisasi. Penarikan kesimpulan yaitu jika nilai $BP > \chi_{(\alpha, p)}^2$ atau jika $p - value < \alpha$ maka tolak H_0 , yang berarti bahwa ragam *residual* tidak sama untuk semua lokasi pengamatan sehingga dapat dikatakan terdapat heterogenitas spasial.

2.3 Fungsi Pembobot Spasial

Fungsi pembobot yang digunakan pada model GWPR sama saja dengan fungsi pembobot pada model GWR. Peran pembobot sangat penting karena nilai pembobot tersebut mewakili letak data pengamatan satu dengan lainnya. Terdapat beberapa metode yang dapat digunakan untuk menentukan besarnya pembobot untuk masing-masing lokasi yang berbeda pada model GWPR, salah satunya dengan menggunakan fungsi *kernel*. Fungsi *kernel* digunakan untuk memperkirakan parameter dalam model GWPR. Terdapat dua macam fungsi *kernel* yaitu fungsi *fixed kernel* dan fungsi *adaptive kernel* (Mahabbi, 2019).

Fungsi *fixed kernel* merupakan fungsi *kernel* yang memiliki *bandwidth* yang sama pada setiap lokasi pengamatan. Terdapat tiga macam fungsi *fixed kernel* yaitu *fixed gaussian kernel*, *fixed bisquare kernel* dan *fixed tricube kernel*. Adapun fungsi *adaptive kernel* merupakan fungsi *kernel* yang memiliki *bandwidth* yang berbeda pada masing-masing lokasi pengamatan. Terdapat tiga macam fungsi *adaptive kernel* yaitu *adaptive gaussian kernel*, *adaptive bisquare kernel* dan *adaptive tricube kernel*. Fungsi pembobot spasial yang digunakan pada penelitian ini adalah fungsi *fixed gaussian kernel* seperti pada Persamaan (2.14).

$$w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right) \quad (2.14)$$

Perhitungan jarak *euclidean* dengan Persamaan (2.15).

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (2.15)$$

Keterangan:

d_{ij} : jarak antar lokasi (u_i, v_i) ke lokasi (u_j, v_j) yang didapatkan dari jarak *euclidean*

h : *bandwidth*

Terdapat parameter *bandwidth* yang nilainya tidak diketahui pada fungsi pembobot *kernel*. Parameter ini dapat direpresentasikan sebagai radius/jari-jari suatu lingkaran. Penentuan *bandwidth* terbaik dalam GWPR sangat penting untuk memperkirakan fungsi *kernel* yang sesuai, karena akan mempengaruhi keakuratan model terhadap data. Penentuan *bandwidth* optimum dapat dilihat dari nilai *cross validation* (CV) minimum. Rumus untuk menghitung nilai CV adalah sebagai berikut (Caraka & Yasin, 2017):

$$CV = \sum_{i=1}^n [y_i - \hat{y}_{\neq i}(h)]^2 \quad (2.16)$$

dengan y_i adalah rata-rata dari waktu ke waktu variabel dependen di lokasi pengamatan ke- i dan $\hat{y}_{\neq i}(h)$ adalah nilai taksiran untuk y_i dengan menghilangkan pengamatan di lokasi (u_i, v_i) dari proses estimasi parameter.

2.4 Geographically Weighted Panel Regression

Geographically weighted panel regression (GWPR) adalah pengembangan dari regresi panel dengan memadukan faktor lokasi geografis dari data yang diperoleh, sehingga estimasi parameter yang diperoleh bersifat lokal (Fotheringham dkk., 2002). Dalam GWPR, diasumsikan bahwa proses mengikuti distribusi pengamatan terdekat lebih berhubungan daripada pengamatan yang jauh (Utami & Yanti, 2021). Pada penelitian ini diasumsikan bahwa kondisi tiap unit pengamatan saling berbeda, sehingga digunakan regresi panel dengan model FEM. Selain itu, penggunaan regresi panel model FEM juga berdasarkan hasil uji *chow* dan uji *hausman*.

Model umum GWPR didapatkan dari gabungan antara model GWR dengan regresi panel. Berikut adalah kombinasi dari persamaan GWR dan persamaan regresi panel FEM dengan *within group*:

$$y_{it}^* = \sum_{k=1}^p \beta_k(u_{it}, v_{it})x_{kit}^* + \varepsilon_{it}^* \quad (2.17)$$

Keterangan:

- y_{it}^* : variabel dependen rata-rata terkoreksi (*demeaned*) pada lokasi ke- i dan waktu ke- t
- $\beta_k(u_{it}, v_{it})$: koefisien regresi variabel independen rata-rata terkoreksi ke- k pada lokasi pengamatan ke- i dan waktu ke- t
- (u_{it}, v_{it}) : titik koordinat lokasi pengamatan ke- i dan waktu ke- t
- x_{kit}^* : variabel independen rata-rata terkoreksi ke- k pada lokasi ke- i dan waktu ke- t
- ε_{it}^* : *residual* rata-rata terkoreksi pada lokasi ke- i dan waktu ke- t

2.4.1 Estimasi Parameter Model *Geographically Weighted Panel Regression*

Estimasi parameter dari model GWPR ini menggunakan pendekatan *weighted least square* (WLS) yaitu menggunakan unsur pembobot yang berbeda untuk setiap lokasi dan waktu pengamatan. Penduga bagi koefisien regresi pada model GWPR dalam bentuk matriks adalah sebagai berikut (Martha dkk., 2021):

$$\hat{\beta}(u_{it}, v_{it}) = (\mathbf{X}^* \mathbf{W}(u_{it}, v_{it}) \mathbf{X}^*)^{-1} \mathbf{X}^* \mathbf{W}(u_{it}, v_{it}) \mathbf{y}^* \quad (2.18)$$

Keterangan:

- $\hat{\beta}(u_{it}, v_{it})$: vektor dari koefisien regresi lokal, berukuran $p \times 1$
- $\mathbf{W}(u_{it}, v_{it})$: matriks pembobot spasial, berukuran $nT \times nT$

2.4.2 Pengujian Hipotesis Model *Geographically Weighted Panel Regression*

Pengujian hipotesis parameter model GWPR sama halnya seperti pengujian parameter pada GWR (Wati & Utami, 2020):

1. Uji Kecocokan Model (*Goodness of Fit*)

Pengujian kesesuaian model GWPR dilakukan dengan menguji kesesuaian parameter secara bersamaan. Tujuan dari pengujian ini adalah untuk mengetahui adanya perbedaan antara model regresi data panel (model regresi global) dan

model GWPR. Hipotesis yang digunakan adalah sebagai berikut :

H_0 : $\beta_k(u_{it}, v_{it}) = \beta_k ; k = 1, 2, \dots, p$ (tidak terdapat perbedaan yang signifikan antara model regresi panel dengan model GWPR)

H_1 : minimal ada satu $\beta_k(u_{it}, v_{it}) \neq \beta_k$ (terdapat perbedaan yang signifikan antara model regresi panel dengan model GWPR)

Statistik uji yang digunakan adalah uji F sebagai berikut:

$$F_{hitung} = \frac{RSS_{GWPR}/df_1}{RSS_{panel}/df_2} \quad (2.19)$$

dengan,

$$RSS_{GWPR} = \mathbf{y}'(\mathbf{I} - \mathbf{L})'(\mathbf{I} - \mathbf{L})\mathbf{y}$$

$$RSS_{panel} = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} \text{ dengan } \mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

$$df_1 = \frac{\delta_1^2}{\delta_2} \text{ dengan } \delta_l = tr[(\mathbf{I} - \mathbf{L})'(\mathbf{I} - \mathbf{L})^l] ; l = 1, 2$$

$$df_2 = nT - p - 1$$

\mathbf{I} adalah matriks identitas berukuran $NT \times NT$ dan \mathbf{L} matriks proyeksi dari model GWPR. Penarikan kesimpulan yaitu jika $F_{hitung} > F_{\alpha, df_1, df_2}$ atau $p - value < \alpha$ maka tolak H_0 , yang berarti bahwa terdapat perbedaan yang signifikan antara model regresi panel dengan model GWPR.

2. Uji Parsial Signifikansi Parameter Model

Uji parsial signifikansi parameter model GWPR diuji untuk mengetahui variabel independen x_k yang berpengaruh signifikan secara parsial terhadap variabel dependen pada setiap lokasi ke- i dan waktu ke- t . Hipotesis yang digunakan adalah sebagai berikut:

H_0 : $\beta_k(u_{it}, v_{it}) = 0 ; k = 1, 2, \dots, p$ (tidak terdapat variabel independen yang berpengaruh secara parsial terhadap variabel dependen)

H_1 : minimal ada satu $\beta_k(u_{it}, v_{it}) \neq 0$ (terdapat variabel independen yang berpengaruh secara parsial terhadap variabel dependen)

Statistik uji yang digunakan adalah uji t sebagai berikut:

$$T_{hitung} = \frac{\hat{\beta}_k(u_{it}, v_{it})}{\hat{\sigma} \sqrt{c_{kk}}} \quad (2.20)$$

dengan c_{kk} adalah elemen diagonal ke- k dari matriks $\mathbf{C}_{it}\mathbf{C}'_{it}$

$$\mathbf{C}_{it} = (\mathbf{X}^*\mathbf{W}(u_{it}, v_{it})\mathbf{X}^*)^{-1} \mathbf{X}^*\mathbf{W}(u_{it}, v_{it})$$

Adapun penarikan kesimpulan yaitu jika $|T_{hitung}| > t_{(\frac{\alpha}{2}, df)}$ dengan $df = \frac{\delta_1^2}{\delta_2}$ atau nilai $p - value < \alpha$ maka tolak H_0 , yang berarti terdapat variabel independen yang berpengaruh secara parsial terhadap variabel dependen pada setiap lokasi ke- i dan waktu ke- t .

2.5 Ukuran Kebaikan Model

Metode yang dapat digunakan untuk menentukan ukuran kebaikan model adalah koefisien determinasi (R^2). Koefisien determinasi merupakan suatu ukuran kebaikan model yang menggambarkan kebaikan model dalam menjelaskan keragaman variabel dependen. Nilai R^2 dapat ditentukan dengan menggunakan persamaan berikut (Fitri, 2019):

$$R^2 = \frac{TSS - RSS}{TSS} \quad (2.21)$$

dengan TSS adalah *total sum of squares* dan RSS adalah *residual sum of squares*.

Selain metode tersebut, untuk menentukan ukuran kebaikan model juga dapat menggunakan *Root Mean Square Error* (RMSE). Semakin kecil nilai RMSE, maka model yang dihasilkan semakin baik. Nilai ini dapat diperoleh dengan mengakarkan nilai MSE (*Mean Square Error*) atau dapat menggunakan rumus seperti berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (2.22)$$

dengan e_i adalah nilai *residual* pada pengamatan ke- i dan n adalah jumlah lokasi pengamatan.

2.6 Angka Harapan Hidup

Angka harapan hidup (AHH) didefinisikan sebagai rata-rata jumlah tahun yang dijalani oleh seseorang setelah orang tersebut mencapai ulang tahun yang ke- x . Ukuran umum yang digunakan adalah AHH saat lahir (0 tahun) yang mencerminkan kondisi kesehatan saat itu. Sehingga pada umumnya ketika membicarakan AHH, yang dimaksud adalah rata-rata tahun yang akan dijalani seseorang sejak orang tersebut lahir (BPS, 2021). Idealnya nilai AHH dihitung berdasarkan angka kematian menurut umur yang datanya diperoleh dari catatan

registrasi kematian. Namun, karena sistem registrasi penduduk di Indonesia belum berjalan dengan baik maka perhitungan nilai AHH digunakan cara tidak langsung yaitu menggunakan program *Mortpak Lite*. Data yang digunakan adalah rata-rata jumlah anak lahir hidup dan rata-rata anak masih hidup, yang bersumber dari hasil Survei Sosial Ekonomi Nasional (BPS, 2011).

Menurut Irevanie (2017), terdapat beberapa faktor yang berpengaruh terhadap angka harapan hidup seperti faktor ekonomi, faktor pendidikan, faktor lingkungan dan sebagainya. Variabel yang menjelaskan faktor-faktor tersebut berdasarkan beberapa penelitian sebelumnya adalah sebagai berikut:

1. Harapan lama sekolah

Harapan lama sekolah didefinisikan sebagai lamanya sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak usia 7 tahun di masa mendatang (BPS, 2021). Semakin tinggi nilai harapan lama sekolah di suatu kabupaten/kota maka kualitas pendidikan akan meningkat yang berakibat pada perilaku kesehatan penduduk yang semakin baik sehingga AHH dapat bertambah (Irevanie, 2017). Menurut *united nation development programme* (UNDP), HLS dihitung dengan cara sebagai berikut:

$$HLS_a^t = \sum_{i=a}^t \frac{E_i^t}{P_i^t}$$

Keterangan:

HLS_a^t : harapan lama sekolah pada usia a dan pada tahun t

E_i^t : partisipasi sekolah penduduk usia i pada tahun t

P_i^t : populasi penduduk usia i yang bersekolah pada tahun t

i : usia ($a, a + 1, \dots, n$)

2. Pengeluaran perkapita

Pengeluaran perkapita biaya yang dikeluarkan untuk konsumsi rumah tangga yang secara umum dibedakan atas konsumsi makanan dan konsumsi bukan makanan. Peningkatan pengeluaran perkapita mengindikasikan bahwa adanya pertumbuhan ekonomi. Dalam memilih barang dan jasa yang kualitasnya lebih tinggi, maka harga yang ditawarkan pasar pastinya lebih tinggi sehingga jika pengeluaran perkapita meningkat akan menjelaskan bahwa orang memilih produk yang lebih baik dari tahun ke tahunnya. Hal ini akan secara langsung

meningkatkan angka harapan hidup (Aris, 2018). BPS menggunakan pengeluaran perkapita dalam menghitung standar hidup layak melalui paritas daya beli berbasis formula *Rao* seperti berikut (BPS, 2021):

$$PPP_j = \prod_{i=1}^m \left(\frac{p_{ij}}{p_{ik}} \right)^{1/m}$$

Keterangan:

PPP_j : paritas daya beli di wilayah j

p_{ij} : harga komoditas i di kabupaten/kota j

p_{ik} : harga komoditas i di Jakarta Selatan

m : jumlah komoditas

3. Persentase rumah tangga yang memiliki sanitasi layak

Persentase rumah tangga yang memiliki sanitasi layak merupakan persentase perilaku rumah tangga yang mempunyai budaya hidup bersih. Keadaan lingkungan dengan sanitasi yang bersih menjadi pendorong kebersihan yang akan memberikan dampak kesehatan yang lebih baik untuk rumah tangga yang ditempati sehingga dapat meningkatkan nilai AHH (Tanadjaja, 2017). Rumus Perhitungan yang digunakan adalah sebagai berikut (Bappenas, 2019):

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

Keterangan:

$PLSL$: persentase rumah tangga yang memiliki sanitasi layak

$JRTSL$: jumlah rumah tangga yang memiliki sanitasi layak

JRT : jumlah rumah tangga seluruhnya

4. Persentase rumah tangga yang memiliki sumber air minum layak

Lingkungan berpengaruh besar terhadap derajat kesehatan. Salah satu indikator yang digunakan untuk melihat lingkungan yang baik adalah persentase rumah tangga yang memiliki sumber air minum layak. Air merupakan kebutuhan dasar penduduk untuk dikonsumsi. Jika air yang dikonsumsi kurang bersih dapat memicu berbagai penyakit yang menyebabkan kualitas kesehatan menurun sehingga nilai AHH menurun (Irevanie, 2017). Rumus Perhitungan yang digunakan adalah sebagai berikut (Bappenas, 2019):

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

dengan *PAML* merupakan persentase rumah tangga yang memiliki sumber air minum layak, *JRTAML* merupakan jumlah rumah tangga yang memiliki sumber air minum layak dan *JRT* merupakan jumlah rumah tangga seluruhnya.

5. Persentase penduduk miskin

Penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran perkapita per bulan di bawah garis kemiskinan (BPS, 2021). Semakin tingginya persentase penduduk miskin berdampak semakin rendahnya nilai AHH. Sebab penduduk yang miskin tidak memiliki kemampuan untuk memenuhi kebutuhan dengan baik termasuk menjaga kesehatan atau melakukan pengobatan ketika sakit (Setiawati, 2020). Persentase penduduk miskin dihitung dengan rumus seperti berikut:

$$\text{Persentase penduduk miskin} = \frac{\text{jumlah penduduk miskin}}{\text{jumlah penduduk}} \times 100\%$$

6. Rata-rata lama sekolah

Rata-rata lama sekolah adalah jumlah tahun yang dijalani oleh penduduk usia 25 tahun ke atas dalam menempuh pendidikan formal (BPS, 2021). Tingginya angka rata-rata lama sekolah menunjukkan semakin tinggi tingkat pendidikan suatu daerah maka semakin mengerti akan pentingnya kesehatan sehingga dapat meningkatkan angka harapan hidup (Nandawati, 2018). Secara matematis rata-rata lama sekolah dihitung dengan rumus seperti berikut:

$$RLS = \frac{1}{P_{15+}} \sum_{i=1}^{P_{15+}} (\text{lama sekolah penduduk ke } - i)$$

dengan P_{15+} adalah jumlah penduduk berusia 15 tahun ke atas.