

**PEMODELAN *ROBUST GEOGRAPHICALLY WEIGHTED  
BINARY LOGISTIC REGRESSION* MENGGUNAKAN  
METODE *M-ESTIMATION* DENGAN PEMBOBOT *ANDREW*  
(Studi Kasus Persentase Penduduk Miskin di Sulawesi Selatan  
Tahun 2022)**

**SKRIPSI**



**KRISNA DWI KAYANA**

**H051201009**

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

**UNIVERSITAS HASANUDDIN**

**MAKASSAR**

**2024**



Optimization Software:  
[www.balesio.com](http://www.balesio.com)

**PEMODELAN *ROBUST GEOGRAPHICALLY WEIGHTED  
BINARY LOGISTIC REGRESSION* MENGGUNAKAN  
METODE *M-ESTIMATION* DENGAN PEMBOBOT *ANDREW*  
(Studi Kasus Persentase Penduduk Miskin di Sulawesi Selatan  
Tahun 2022)**

**SKRIPSI**

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

**KRISNA DWI KAYANA**

**H051201009**

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

**MAKASSAR**

**JUNI 2024**



## LEMBAR PERNYATAAN KEOTENTIKAN

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

**Pemodelan *Robust Geographically Weighted Binary Logistic Regression*  
Menggunakan Metode *M-Estimation* dengan pembobot *Andrew*  
(Studi Kasus Persentase Penduduk Miskin di Sulawesi Selatan Tahun 2022)**

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

Makassar, 10 Juni 2024



**Krisna Dwi Kayana**


**NIM H051201009**



**PEMODELAN ROBUST GEOGRAPHICALLY WEIGHTED  
BINARY LOGISTIC REGRESSION MENGGUNAKAN  
METODE M-ESTIMATION DENGAN PEMBOBOT ANDREW  
(Studi Kasus Persentase Penduduk Miskin di Sulawesi Selatan  
Tahun 2022)**

Disetujui Oleh:

Pembimbing

  
Andi Kresna Jaya, S.Si., M.Si.  
NIP. 19731228 200003 1 001

Ketua Program Studi

  
  
Dr. Anna Islamiyati, S.Si., M.Si.  
NIP. 19770808 200501 2 002



Pada 10 Juni 2024

## HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Krisna Dwi Kayana  
NIM : H051201009  
Program Studi : Statistika  
Judul Skripsi : *Pemodelan Robust Geographically Weighted Binary Logistic Regression Menggunakan Metode M-Estimation dengan pembobot Andrew* (Studi Kasus Persentase Penduduk Miskin di Sulawesi Selatan Tahun 2022)

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 : Andi Kresna Jaya, S.Si., M.Si. (.....)
2. Anggota : Prof. Dr. Dr. Georgina Maria Tinungki, M.Si. (.....)
3. Anggota : Drs Raupong, M.Si. (.....)

Ditetapkan di : Makassar

Tanggal : 10 Juni 2024



## KATA PENGANTAR

*Assalamu 'alaikum Warahmatullahi Wabaraktuh*

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'amin*, berkat nikmat kemudahan dan pertolongan yang diberikan oleh Allah *Subhanahu Wa Ta'ala*, penulis dapat menyelesaikan skripsi yang berjudul "**Pemodelan *Robust Geographically Weighted Binary Logistic Regression* Menggunakan Metode *M-Estimation* dengan pembobot *Andrew* (Studi Kasus Persentase Penduduk Miskin di Sulawesi Selatan Tahun 2022)**" 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 serta penghargaan yang setinggi-tingginya kepada bapak **La Nudi** dan Ibunda **Sarumina** yang telah memberikan dukungan penuh, limpahan cinta dan kasih sayang, kesabaran hati, serta dengan ikhlas telah menemani setiap langkah penulis dengan doa dan restu mulianya. Ucapan terima kasih juga penulis haturkan kepada Adik tersayang **Aldi Surya Putra** yang menjadi penyemangat dan motivasi terbesar untuk penulis agar segera menyelesaikan skripsi ini. Terima kasih juga kepada nenek **Wa Ido** dan nenek **Nuru Halida** serta kakek **La Ode Salih** dan kakek **La Daira** serta keluarga besar penulis **Usti, Muli, Siti**, ibunda **Ukasa**, Ibunda **Suriana**, Ibunda **Ayu**, ibunda **Muhuria**, bapak **La Kamu**, dan bapak **Dewa** yang selalu mendampingi dimasa sulit dan senang penulis serta selalu menghibur dan memberi semangat kepada penulis dalam menyelesaikan skripsi ini. Penulis kasih atas doa mulia dan dukungannya selama ini.



Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan dan ketulusan juga penulis ucapkan kepada:

1. **Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Anna Islamiyati, S.Si., M.Si.**, selaku Ketua Departemen Statistika yang dengan penuh kesabaran telah memberikan arahan, dorongan semangat dan motivasi kepada penulis selama menjadi mahasiswa di Departemen Statistika.
4. **Bapak Andi Kresna Jaya, S.Si., M.Si.**, selaku Pembimbing sekaligus Penasehat Akademik penulis yang dengan penuh kesabaran telah meluangkan waktu dan pemikirannya untuk senantiasa memberikan arahan, dorongan semangat, dan motivasi kepada penulis dari awal hingga selesainya penulisan tugas akhir ini.
5. **Ibu Prof. Dr. Dr. Georgina Maria Tinungki, M.Si.** dan **Bapak Drs Raupong, M.Si.**, selaku Tim Penguji yang telah meluangkan waktu dalam memberikan motivasi serta kritikan yang membangun kepada penulis dalam penyempurnaan tugas akhir ini.
6. Segenap **Dosen Pengajar** dan **Staf** yang telah memberikan ilmu dan kemudahan kepada penulis dalam berbagai hal selama menempuh pendidikan sarjana di Departemen Statistika.
7. Keluarga baru penulis, **Ibunda Neli** serta **ibunda Atmasari** dan anak-anaknya, **Rista** dan **Alan** yang senantiasa kebersamai, memberi semangat dan nasehatnya untuk penulis bisa melampaui masa sulit itu agar bisa kembali menyelesaikan skripsi ini.
8. Sahabat-sahabat **RANDOM, Laurine, Fadlan, Alif, Aisyah, Fahmi, Mukhlis, Rais, Dzaky,** dan **Bahar** yang menemani penulis dari awal perkuliahan yang bermula dari grup belajar bersama hingga terbentuk grup **Random** yang sekaligus menjadi tempat berbagi cerita serta bermain  
penulis merasa sangat bersyukur dipertemukan kalian, terima kasih  
kir kenangan manis bersama penulis selama kuliah.



9. Sahabat-sahabat tercinta, **Nanda** dan **Resti**, terimakasih selalu mendengarkan cerita random penulis, memberikan semangat, dan menghibur selama ini.
10. Sahabat **CIWI-CIWI AMONG**, **Rahmi**, **Parida**, **Ayu**, **Aulia**, **Aliah**, dan kawan-kawan. Terima kasih untuk kebersamaan dalam belajar selama kuliah.
11. Sahabat-sahabat diperantauan, **Sukma** dan **Dita**, terimakasih sudah berjuang sama-sama masuk di kampus ini dan selalu tolong penulis jika ada keperluan di luar kampus.
12. Teman seperjuangan di **STAT'20**. Terima kasih atas ilmu, kebersamaan, suka dan duka dalam menjalani perkuliahan di Departemen Statistika. Terima kasih sudah menerima kehadiran penulis. Kalian hebat dan luar biasa.
13. **Keluarga Besar Himastat FMIPA Unhas**, terima kasih atas ilmu, pengalaman, dan telah menjadi tempat belajar bagi penulis.
14. Kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih setinggi-tingginya untuk segala dukungan, partisipasi, dan apresiasi 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 Juni 2024



Krisna Dwi Kayana





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

---

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

Nama : Krisna Dwi Kayana  
NIM : H051201009  
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:

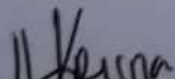
**“Pemodelan *Robust Geographically Weighted Binary Logistic Regression*  
Menggunakan Metode *M-Estimation* dengan pembobot *Andrew* (Studi Kasus  
Persentase Penduduk Miskin di Sulawesi Selatan Tahun 2022)”**

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 Juni 2024.

Yang menyatakan,





## ABSTRAK

Model *Geographically Weighted Binary Logistic Regression* (GWBLR) menggabungkan regresi logistik biner dengan GWR untuk analisis yang lebih mendalam berdasarkan lokasi pengamatan. Penerapan model ini memungkinkan identifikasi faktor-faktor kemiskinan dengan lebih akurat di setiap wilayah, sehingga kebijakan dapat disesuaikan untuk meningkatkan efektivitas pengentasan kemiskinan di Sulawesi Selatan. Dalam menganalisis data dengan menggunakan model GWBLR, terkadang ditemukan adanya *outlier* yang akan berdampak terhadap hasil estimasi parameter. Metode regresi *robust* adalah salah satu alternatif dalam mengatasi permasalahan regresi berupa data yang mengandung *outlier*. Salah satu metode regresi *robust* yang sering digunakan adalah *M-Estimation*. Penelitian ini bertujuan untuk mendapatkan model *Robust Geographically Weighted Binary Logistic Regression* (RGWBLR) menggunakan metode *M-Estimation* dengan pembobot *Andrew* yang diaplikasikan pada persentase penduduk miskin di Sulawesi Selatan tahun 2022. Variabel respon yang digunakan pada penelitian ini adalah persentase penduduk miskin dan variabel prediktor meliputi angka melek huruf, rumah tangga miskin yang menerima beras sejahtera (Rastra) atau BPNT, rata-rata lama sekolah, dan umur harapan hidup. Hasil yang diperoleh adalah nilai MSE dari penduga parameter RGWBLR (0.1180) lebih kecil dibandingkan dengan model GWBLR (9.1224). Faktor-faktor yang mempengaruhi persentase penduduk miskin di Provinsi Sulawesi Selatan adalah angka melek huruf dan rata-rata lama sekolah.

**Kata Kunci:** *Andrew*, *M-Estimation*, Persentase Penduduk Miskin, RGWBLR



**ABSTRACT**

*The Geographically Weighted Binary Logistic Regression (GWBLR) model combines binary logistic regression with GWR for a more in-depth analysis based on observation locations. The application of this model allows for more accurate identification of poverty factors in each region, enabling policies to be adjusted to improve the effectiveness of poverty alleviation efforts in South Sulawesi. In analyzing data using the GWBLR model, outliers are sometimes found that can affect parameter estimation results. Robust regression methods are an alternative to addressing regression issues involving outlier data. One commonly used robust regression method is M-Estimation. This research aims to develop a Robust Geographically Weighted Binary Logistic Regression (RGWBLR) model using the M-Estimation method with Andrew's weighting applied to the percentage of poor people in South Sulawesi in 2022. The response variable used in this research is the percentage of poor people, and the predictor variables include literacy rate, poor households receiving rice subsidies (Rastra) or BPNT, average years of schooling, and life expectancy. The results show that the MSE value of the RGWBLR parameter estimator (0.1180) is smaller than that of the GWBLR model (9.1224). This indicates that the RGWBLR model is better at explaining the percentage of poor people in South Sulawesi in 2022. Factors influencing the percentage of poor population in South Sulawesi Province are literacy rate and average years of schooling.*

**Keywords:** Andrew, M-Estimation, Percentage of Poor People, RGWBLR



DAFTAR ISI

**HALAMAN SAMPUL**..... i

**HALAMAN JUDUL** ..... ii

**LEMBAR PERNYATAAN KEOTENTIKAN** ..... iii

**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** ..... vi

**BAB I PENDAHULUAN**..... 1

1.1 Latar Belakang ..... 1

1.2 Rumusan Masalah ..... 2

1.3 Batasan Masalah..... 3

1.4 Tujuan Penelitian ..... 3

1.5 Manfaat Penelitian ..... 3

**BAB II TINJAUAN PUSTAKA**..... 4

2.1 Regresi Logistik Biner ..... 4

2.2 Pengujian Model Regresi ..... 5

2.2.1. Uji Simultan ..... 5

2.2.2. Uji Parsial ..... 5

2.3 Uji Multikolinearitas ..... 6

2.4 Uji Heterogenitas Spasial ..... 6

2.5 Model *Geographically Weighted Regression* ..... 7

2.6 Fungsi Pembobot Spasial ..... 7

Model *Geographically Weighted Binary Logistic Regression*..... 8

*lier* ..... 9

Regresi *Robust M-Estimation* ..... 9



2.10 Fungsi Pembobot *Andrew* ..... 9

2.11 Kemiskinan ..... 10

**BAB III METODOLOGI PENELITIAN ..... 11**

3.1 Sumber Data..... 11

3.2 Variabel Penelitian ..... 11

3.3 Metode Analisis ..... 12

**BAB IV HASIL DAN PEMBAHASAN ..... 14**

4.1 Estimasi Parameter..... 14

4.1.1 Estimasi Parameter *Geographically Weighted Regression*..... 14

4.1.2 Estimasi Parameter Model *Geographically Weighted Binary Logistic Regression* ..... 15

4.1.3 Estimasi Parameter Model *Robust Geographically Weighted Binary Logistic Regression* ..... 19

4.2 Eksplorasi Data ..... 23

4.3 Multikolinearitas ..... 27

4.4 Model Regresi Logistik Biner..... 28

4.5 Identifikasi *Outlier* ..... 28

4.6 Efek Heterogenitas Spasial ..... 29

4.7 Jarak Euclidean ..... 29

4.8 Pembobot Kernel *Gaussian* ..... 31

4.9 Estimasi Parameter Model *Geographically Weighted Regression* ..... 32

4.10 Estimasi Parameter Model *Geographically Weighted Binary Logistic Regression* ..... 32

4.11 Pembobot *Andrew* ..... 35

4.12 Estimasi Parameter *Robust* dengan *M-Estimation* ..... 35

**BAB V KESIMPULAN DAN SARAN ..... 39**

5.1. Kesimpulan ..... 39

5.2. Saran..... 39

**DAFTAR PUSTAKA ..... 40**

..... 43



## DAFTAR GAMBAR

<b>Gambar 4.1.</b> Grafik Fungsi Pembobot <i>Andrew</i> .....	22
<b>Gambar 4.2.</b> Peta Tematik Sebaran Persentase Penduduk Miskin di Sulawesi Selatan Tahun 2022 .....	23
<b>Gambar 4.3.</b> Peta Tematik Angka Melek Huruf.....	24
<b>Gambar 4.4.</b> Peta Tematik Persentase Rumah Tangga Miskin yang Menerima Beras Sejahtera (Rastra) atau BPNT .....	25
<b>Gambar 4.5.</b> Peta Tematik Rata-Rata Lama Sekolah .....	26
<b>Gambar 4.6.</b> Peta Tematik Umur Harapan Hidup .....	27
<b>Gambar 4.7.</b> Plot Nilai DFFITS.....	29



## DAFTAR TABEL

<b>Tabel 3.1.</b> Variabel Respon dan Prediktor .....	11
<b>Tabel 4.1.</b> Pengelompokan Wilayah Miskin dan Tidak Miskin di 24 Kabupaten/Kota.....	24
<b>Tabel 4.2.</b> Nilai VIF Setiap Variabel Prediktor .....	27
<b>Tabel 4.3.</b> Pengujian Heterogenitas Spasial .....	29
<b>Tabel 4.4.</b> Jarak <i>Euclidean</i> Kabupaten Bantaeng dengan Lokasi Lain .....	30
<b>Tabel 4.5.</b> Pembobot Kernel <i>Gaussian</i> untuk Kabupaten Bantaeng .....	31
<b>Tabel 4.6.</b> Hasil Estimasi Parameter dengan metode WLS.....	32
<b>Tabel 4.7.</b> Hasil Estimasi Parameter GWBLR dengan Metode MLE.....	33
<b>Tabel 4.8.</b> Pengujian Simultan Model GWBLR.....	33
<b>Tabel 4.9.</b> Pengujian Model GWBLR .....	34
<b>Tabel 4.10.</b> Hasil Estimasi Parameter RGWBLR dengan Metode <i>M-Estimation</i> .....	36
<b>Tabel 4.11.</b> Pengujian Simultan Model RGWBLR.....	36
<b>Tabel 4.12.</b> Pengujian Model RGWBLR .....	37
<b>Tabel 4.13.</b> Nilai MSE GWBLR dan RGWBLR .....	38



## DAFTAR LAMPIRAN

<b>Lampiran 1.</b> Data.....	44
<b>Lampiran 2.</b> Nilai DFFITS .....	45
<b>Lampiran 3.</b> Uji Heterogenitas Spasial.....	46
<b>Lampiran 4.</b> Output <i>Bandwidth</i> Optimum Fungsi Pembobot Kernel <i>Gaussian</i> .	47
<b>Lampiran 5.</b> Jarak <i>Euclidean</i> .....	48
<b>Lampiran 6.</b> Pembobot Kernel <i>Gaussian</i> .....	50
<b>Lampiran 7.</b> Pembobot <i>Andrew</i> .....	52
<b>Lampiran 8.</b> Hasil Estimasi Parameter GWR.....	54
<b>Lampiran 9.</b> Hasil Estimasi Parameter GWBLR.....	55
<b>Lampiran 10.</b> Hasil Pengujian Pendugaan Parameter Secara Parsial GWBLR ..	56
<b>Lampiran 11.</b> Model <i>Geographically Weighted Binary Logistic Regression</i> .....	57
<b>Lampiran 12.</b> Hasil Estimasi Parameter RGWBLR.....	61
<b>Lampiran 13.</b> Hasil Pengujian Pendugaan Parameter Secara Parsial RGWBLR	62
<b>Lampiran 14.</b> Model <i>Robust Geographically Weighted Binary Logistic Regression</i> .....	63





# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Model regresi logistik biner merupakan metode analisis data yang digunakan untuk menentukan hubungan antara variabel respon dan variabel prediktor. Analisis regresi logistik biner mengasumsikan bahwa nilai parameter regresi selalu tepat, dalam artian bahwa respon tidak dipengaruhi oleh lokasi pengamatan. Tetapi kenyataannya, data yang diambil dari beberapa lokasi yang berbeda, pasti memiliki kondisi yang berbeda pula. Hal ini dapat dipengaruhi dari letak geografis, sosial budaya, hingga kondisi perekonomian lokasi pengamatan tersebut. Berdasarkan perkembangan zaman, telah dikembangkan model regresi yang dapat diterapkan pada data spasial yaitu *Geographically Weighted Regression (GWR)*.

GWR adalah teknik yang efektif dalam memodelkan data spasial nonstasioner (Fathurahman & Ratnasari, 2019). Hasil analisis ini adalah model regresi yang nilai-nilai parameternya berlaku hanya pada setiap lokasi pengamatan dan berbeda dengan lokasi lainnya. Salah satu perkembangan dari model GWR adalah menggabungkan antara regresi logistik biner dan GWR yang disebut dengan *Geographically Weighted Binary Logistic Regression (GWBLR)*.

GWBLR adalah sebuah metode analisis spasial yang mampu memberikan hasil regresi yang lebih akurat, sesuai dengan karakteristik dan tantangan yang terjadi di setiap wilayah (Leniati dkk., 2023). Dalam penerapan model GWBLR terkadang ditemukan adanya *outlier*. Dampak dari adanya *outlier* adalah membuat estimasi koefisien regresi yang diperoleh tidak tepat (Ihsan dkk, 2018). Hal ini berarti estimasi koefisien regresi dalam suatu model dapat dipengaruhi oleh salah satu data ekstrim yang merupakan *outlier*. Dengan demikian, diperlukan sebuah alternatif terhadap keberadaan dari *oulier* yaitu dengan menggunakan regresi *robust*.

Regresi *robust* diperkenalkan oleh Andrews pada tahun 1972. Regresi

digunakan untuk menganalisis data yang dipengaruhi *outlier* sehingga model yang *robust* atau resisten terhadap *outlier* tanpa menghilangkan tersebut. Suatu estimasi dikatakan resisten apabila estimasi yang relatif



tidak dipengaruhi oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data (Setiarini & Listyani, 2017). Salah satu metode regresi *robust* yang sering digunakan adalah *M-Estimation* dengan pembobot *Andrew*.

Penelitian ini akan diaplikasikan pada data persentase penduduk miskin di Sulawesi Selatan. Salah satu provinsi di Indonesia yang memiliki tantangan terkait kemiskinan. Data menunjukkan bahwa persentase penduduk miskin di provinsi ini mengalami fluktuasi selama periode tertentu. Pada Maret 2022, persentase penduduk miskin di wilayah pedesaan Sulawesi Selatan mencapai 11.63%, sementara pada bulan September 2022, angka tersebut meningkat menjadi 11.81%. Sementara itu, di wilayah perkotaan, pada bulan yang sama, persentase penduduk miskin mengalami penurunan dari 5.07% menjadi 4.98%. Fluktuasi ini mencerminkan kompleksitas masalah kemiskinan yang dipengaruhi oleh berbagai faktor, termasuk geografi, ekonomi, dan sosial.

Penelitian terdahulu oleh Deria dkk. (2019) yang membandingkan antara pembobot *Andrew*, pembobot *Ramsay*, dan pembobot *Welsch* dalam regresi *robust M-Estimation* pada tingkat kemiskinan di Jawa Tengah tahun 2018. Hasil dari penelitian tersebut adalah fungsi pembobot *Andrew* memiliki nilai MSE yang lebih rendah daripada pembobot *Ramsay* dan pembobot *Welsch*. Selanjutnya, Leniati dkk. (2023) membandingkan antara model regresi logistik biner dengan model GWBLR pada kasus kemiskinan di provinsi Sumbar pada tahun 2022. Berdasarkan penelitian tersebut model GWBLR memiliki nilai AIC yang lebih rendah dibandingkan dengan model regresi logistik biner. Oleh karena itu, penelitian ini menggunakan fungsi pembobot *Andrew* untuk menduga parameter *Robust Geographically Weighted Binary Logistic Regression (RGWBLR)* menggunakan metode *M-Estimation* pada data persentase penduduk miskin di Sulawesi Selatan tahun 2022.

## 1.2 Rumusan Masalah

Rumusan masalah yang dibahas pada penelitian ini adalah sebagai berikut:

1. Bagaimana model *Robust Geographically Weighted Binary Logistic Regression* pada tingkat persentase penduduk miskin di Sulawesi Selatan tahun 2022?



2. Bagaimana perbandingan hasil model *Geographically Weighted Binary Logistic Regression* dan *Robust Geographically Weighted Binary Logistic Regression* pada tingkat persentase penduduk miskin di Sulawesi Selatan tahun 2022?

### 1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Metode estimasi parameter model *Robust Geographically Weighted Binary Logistic Regression* menggunakan *M-Estimation* dengan pembobot *Andrew*.
2. Penelitian diaplikasikan pada data persentase penduduk miskin di Sulawesi Selatan tahun 2022 dengan variabel respon berbentuk biner.

### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mendapatkan model *Robust Geographically Weighted Binary Logistic Regression* pada tingkat persentase penduduk miskin di Sulawesi Selatan tahun 2022.
2. Untuk membandingkan hasil model *Geographically Weighted Binary Logistic Regression* dan *Robust Geographically Weighted Binary Logistic Regression* pada tingkat persentase penduduk miskin di Sulawesi Selatan tahun 2022?

### 1.5 Manfaat Penelitian

Manfaat yang diharapkan dalam penelitian ini adalah mampu menambah pemahaman teoritis dan praktis baik peneliti maupun pembaca tentang *Robust Geographically Weighted Binary Logistic Regression* dalam proses analisisnya menggunakan metode *M-Estimation* dengan pembobot *Andrew*, serta pemahaman menganalisis data persentase penduduk miskin di Sulawesi Selatan tahun 2022.



## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Regresi Logistik Biner

Model regresi logistik biner adalah suatu metode statistika yang menggambarkan hubungan antara variabel respon yang hanya terdiri dari dua kategori dengan variabel prediktor. Variabel respon mengikuti distribusi Bernoulli. Fungsi distribusi peluang untuk  $y$  dengan parameter  $\pi(x_i)$  adalah

$$f(y_i) = \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}, y_i = 0,1 \quad (2.1)$$

Persamaan regresi logistik biner sebagai berikut (Yusuf dkk., 2020):

$$\pi(x_i) = \frac{e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_m x_{im})}}{1 + e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_m x_{im})}}$$

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_m x_{im})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_m x_{im})} \quad (2.2)$$

dengan  $\pi(x_i)$  adalah peluang untuk pengamatan ke- $i$ ,  $\beta_0$  sebagai konstanta (*intercept*),  $\beta_j$  sebagai parameter untuk variabel prediktor ke- $j$ ,  $x_{ij}$  sebagai nilai pengamatan ke- $i$  pada variabel prediktor ke- $j$ , dan  $n$  adalah banyaknya variabel prediktor.

Bentuk transformasi logit dari Persamaan (2.2) dapat dituliskan sebagai berikut:

$$\{\pi(x_i)\} \{1 + e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_m x_{im})}\} = e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_m x_{im})}$$

$$\{\pi(x_i)\} + \{\pi(x_i) e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_m x_{im})}\} = e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_m x_{im})}$$

$$\pi(x_i) = e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_m x_{im})} - \pi(x_i) e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_m x_{im})}$$

$$\pi(x_i) = \{1 - \pi(x_i)\} e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_m x_{im})}$$

$$\frac{\pi(x_i)}{1 - \pi(x_i)} = e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_m x_{im})}$$

$$\ln \left[ \frac{\pi(x_i)}{1 - \pi(x_i)} \right] = \ln e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_m x_{im})}$$

$$\ln \left[ \frac{\pi(x_i)}{1 - \pi(x_i)} \right] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_m x_{im}$$

sehingga di peroleh Persamaan (2.3).

$$\ln \left[ \frac{\pi(x_i)}{1 - \pi(x_i)} \right] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_m x_{im} \quad (2.3)$$



## 2.2 Pengujian Model Regresi

### 2.2.1. Uji Simultan

Uji simultan atau disebut uji model *chi-square*, dilakukan untuk memeriksa peranan variabel prediktor dalam model secara bersama-sama dengan menggunakan *likelihood ratio test*, dengan hipotesis sebagai berikut:

$H_0: \beta_1 = \beta_2 = \dots = \beta_m = 0$  (tidak ada pengaruh antara variabel prediktor terhadap variabel respon)

$H_1: \beta_j \neq 0$  ( $j = 1, 2, \dots, m$ ) (ada variabel prediktor yang berpengaruh terhadap variabel respon)

Statistik uji:

$$G_2 = 2[\sum_{i=1}^n (y_i \ln(\pi(u_i, v_i)) + (1 - y_i) \ln(1 - \pi(u_i, v_i))) - \sum_{i=1}^n n_{1i} \ln(n_{1i}) + n_{0i} \ln(n_{0i}) + n \ln(n)] \sim \chi_{(m, \alpha)}^2 \quad (2.4)$$

dengan  $n_i = \sum_{i=1}^n y_i$ ,  $n_{0i} = \sum_{i=1}^n (1 - y_i)$ , dan  $n$  adalah jumlah pengamatan keseluruhan. Statistik uji akan mengikuti sebaran *chi-square* sehingga  $H_0$  ditolak jika  $G_2 > \chi_{(m, \alpha)}^2$  dengan  $m$  derajat bebas adalah banyaknya parameter dalam model tanpa  $\beta_0$ .

### 2.2.2. Uji Parsial

Pengujian secara parsial dilakukan untuk mengetahui signifikansi parameter terhadap variabel respon. Pengujian signifikansi parameter menggunakan uji Wald dengan hipotesis sebagai berikut (Caraka & Yasin, 2017):

$H_0: \beta_j = 0$  (tidak ada signifikansi parameter  $\beta$  terhadap variabel respon secara parsial)

$H_1: \beta_j \neq 0$  (ada signifikansi parameter  $\beta$  terhadap variabel respon secara parsial)

Statistik uji:

$$W = \frac{\hat{\beta}_j}{\sqrt{\text{Var}(\hat{\beta}_j)}} \sim N(0,1) \quad (2.5)$$

dengan  $\text{Var}(\hat{\beta}_j)$  merupakan elemen diagonal ke- $j$  dari invers matriks *fisher-*

$I(\hat{\beta})^{-1} = -[H(\hat{\beta})]^{-1}$  dan kriteria ujinya yaitu  $H_0$  ditolak jika  $|W| >$



### 2.3 Uji Multikolinearitas

Uji multikolinearitas adalah asumsi yang menunjukkan adanya hubungan linear yang kuat antara beberapa variabel prediktor dalam suatu model regresi. Multikolinearitas dapat dilihat dari nilai *Variance Inflation Factors* (VIF). Nilai VIF dirumuskan sebagai berikut:

$$VIF = \frac{1}{1 - R_j^2}; j = 1, 2, \dots, m \quad (2.6)$$

dengan  $R_j^2$  adalah koefisien determinasi antara  $x_j$  dengan variabel prediktor lain. Jika nilai  $VIF > 10$ , maka menunjukkan adanya multikolinearitas antar variabel prediktor (Leniati dkk., 2023).

### 2.4 Uji Heterogenitas Spasial

Masalah yang biasa ditemukan dalam data spasial adalah ragam yang tidak homogen pada setiap lokasi yang diamati atau disebut heterogenitas spasial. Uji yang digunakan untuk menguji adanya heterogenitas spasial adalah uji *Breusch-Pagan* dengan hipotesis sebagai berikut:

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

$H_1: \text{ada } \sigma_i^2 \neq \sigma_j^2 \text{ dengan } i \neq j$  (terdapat heterogenitas spasial)

Nilai *Breusch-Pagan test* :

$$BP = \frac{1}{2} \mathbf{f}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{f} \quad (2.7)$$

Nilai dari elemen vektor  $f$  dan residual  $\varepsilon_i$

$$f_i = \frac{\varepsilon_i^2}{\sigma^2} - 1, \varepsilon_i = y_i - \hat{y}_i, \sigma^2 = \frac{\varepsilon_i^t \varepsilon_i}{n}$$

dengan  $\varepsilon_i$  adalah residual untuk pengamatan ke- $i$  dengan matriks berukuran  $(n \times 1)$ ;  $f$  sebagai vektor berukuran  $(n \times 1)$ ;  $\sigma^2$  sebagai ragam residual  $\varepsilon_i$ ; dan  $Z$  adalah matriks berukuran  $n \times (m + 1)$  yang berisi vektor yang sudah distandarisasi ( $z$ ) untuk setiap pengamatan. Daerah penolakan  $H_0$  yaitu  $BP > \chi_{\alpha, (m)}^2$  dengan  $m$  adalah jumlah prediktor sehingga dapat disimpulkan bahwa terdapat heterogenitas spasial (Arif dkk., 2019).



### 2.5 Model Geographically Weighted Regression

GWR merupakan pengembangan dari metode regresi. Model ini merupakan model regresi linier lokal (*locally linier regression*) yang menghasilkan penaksir parameter model yang bersifat lokal untuk setiap titik atau lokasi dimana data tersebut dikumpulkan. Dalam model GWR, variabel respon diprediksi dengan variabel prediktor yang masing-masing koefisien regresinya bergantung pada lokasi dimana data tersebut diamati (Fadli dkk., 2018). Model GWR dapat ditulis sebagai berikut:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^m \beta_j(u_i, v_i)x_{ij} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.8)$$

dengan  $y_i$  adalah nilai pengamatan variabel respon ke- $i$ ;  $x_{ij}$  sebagai nilai pengamatan variabel prediktor ke- $j$  pada lokasi pengamatan ke- $i$ ;  $\beta_0(u_i, v_i)$  sebagai konstanta/*intercept* pada pengamatan ke- $i$ ;  $\beta_j(u_i, v_i)$  sebagai nilai pengamatan variabel prediktor ke- $j$  pada lokasi pengamatan ke- $i$ ; dan  $\varepsilon_i$  adalah galat pengamatan ke- $i$ .

### 2.6 Fungsi Pembobot Spasial

Pemilihan pembobot sangat penting, karena nilai pembobot dapat mewakili letak data pengamatan satu dengan lainnya. Lokasi yang berdekatan akan menunjukkan hubungan kemiripan dan lokasi yang berjauhan akan menunjukkan keragaman spasial. Keragaman spasial antara lokasi pengamatan satu dengan lainnya ditunjukkan dengan matriks pembobot  $\omega_i(u_i, v_i)$  berukuran  $n \times n$  dengan diagonal utama adalah bobot dari tiap lokasi untuk menduga parameter lokasi ke- $i$  (Lutfiani dkk., 2019).

$$\omega_i(u_i, v_i) = \begin{bmatrix} \omega_{i1} & 0 & \dots & 0 \\ 0 & \omega_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \omega_{in} \end{bmatrix}$$

Pembobot yang digunakan dalam penelitian ini adalah Kernel *Gaussian*. Bentuk fungsi Kernel *Gaussian* adalah :

$$\omega_{ij} = \exp \left[ -\frac{1}{2} \left( \frac{d_{ij}}{b} \right)^2 \right] \quad (2.9)$$

adalah *bandwidth* dan  $d_{ij}$  adalah jarak *euclidean* :

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (2.10)$$



Pemilihan *bandwidth* optimum dapat diperoleh dengan menghitung *Cross Validation (CV)*. *Bandwidth* yang optimum akan menghasilkan nilai CV yang paling kecil (Amalah dkk., 2023). Bentuk dari persamaan CV adalah sebagai berikut :

$$CV(h) = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h))^2 \tag{2.11}$$

dengan  $\hat{y}_{\neq i}(h)$  adalah nilai penduga  $y_i$  dimana pengamatan di lokasi  $(u_i, v_i)$  dihilangkan dari proses estimasi. Selanjutnya, pemilihan model terbaik dapat dilakukan dengan beberapa metode salah satunya dengan nilai *Mean Square Error (MSE)*. Model terbaik adalah model dengan nilai MSE yang paling kecil, dapat dihitung dengan persamaan berikut :

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \tag{2.12}$$

dengan  $y_i$  adalah nilai pengamatan variabel respon ke- $i$ ;  $\hat{y}_i$  sebagai nilai penduga; dan  $n$  adalah banyak lokasi pengamatan.

### 2.7 Model Geographically Weighted Binary Logistic Regression

GWBLR adalah salah satu metode yang digunakan untuk mengestimasi parameter regresi yang mempertimbangkan faktor spasial dengan menerapkan model GWR yang menggabungkan parameter non-stasioner dan data kategorikal untuk membentuk GWBLR (Muzdalifah dkk., 2020). Dalam model GWBLR, variabel respon diprediksi dengan variabel prediktor yang masing-masing koefisien regresinya bergantung pada lokasi pengamatan. Secara umum, bentuk dari model GWBLR adalah sebagai berikut (Mei, 2018):

$$\begin{aligned} \pi(x_i) &= \frac{\exp(\beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \dots + \beta_j(u_i, v_i)x_{ij})}{1 + \exp(\beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \dots + \beta_j(u_i, v_i)x_{ij})} \\ \pi(x_i) &= \frac{\exp(\sum_{j=0}^m \beta_j(u_i, v_i)x_{ij})}{1 + \exp(\sum_{j=0}^m \beta_j(u_i, v_i)x_{ij})} \end{aligned} \tag{2.13}$$

dengan  $i = 1, 2, \dots, n; j = 0, 1, 2, \dots, m$ . Bentuk logit untuk Persamaan (2.13) sebagai berikut:

$$\ln \left( \frac{\pi(x_i)}{1 - \pi(x_i)} \right) = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \dots + \beta_j(u_i, v_i)x_{ij} \tag{2.14}$$





**2.8 Outlier**

*Outlier* merupakan pengamatan yang jauh dari pusat data yang mungkin berpengaruh terhadap koefisien regresi. Salah satu metode yang dapat digunakan untuk mendeteksi outlier yaitu *Difference Fitted value of FITS* (DFFITS). Metode digunakan untuk mengetahui pengaruh suatu pengamatan ke-*i* terhadap model regresi yang ditinjau dari nilai fitnya. Besarnya nilai DFFITS adalah:

$$DFFITS = t_i \left( \frac{h_{ii}}{1 - h_{ii}} \right)^{\frac{1}{2}} \tag{2.15}$$

Besar nilai dari DFFITS akan ikut naik jika nilai  $t_{ii}$  dan  $h_{ii}$  keduanya naik. Hal ini menunjukkan pengamatan tersebut mempunyai pengaruh yang besar pada hasil analisis regresi. Penentuan pengamatan *i* sebagai *outlier* berdasarkan *cut off* masing-masing. Untuk data ukuran kecil ( $n \leq 15$ ) dikatakan sebagai *outlier* jika nilai  $DFFITS > 1$ . Sedangkan untuk data yang ukuran besar, nilai  $|DFFITS| > 2 \sqrt{\frac{m+1}{n}}$  merupakan data *outlier* (Situmorang & Susanti, 2020). Dengan *m* adalah banyaknya variabel prediktor.

**2.9 Regresi Robust M-Estimation**

Regresi *robust* adalah metode yang digunakan untuk mengatasi *outlier* tanpa menghapus data *outlier* tersebut. Regresi *robust* terdiri dari beberapa metode estimasi yang salah satunya adalah *M-Estimation*. Pada prinsipnya, metode *M-Estimation* digunakan untuk meminimumkan suatu fungsi objektif.

$$\sum_{i=1}^n \rho(u) = \sum_{i=1}^n \rho \left( \frac{\varepsilon}{\hat{\sigma}} \right) = \sum_{i=1}^n \rho \left( \frac{\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(u_i, v_i)}{\hat{\sigma}} \right) \tag{2.16}$$

$\rho(u)$  adalah fungsi simetris dari residual atau fungsi yang memberikan kontribusi pada masing-masing residual pada fungsi objektif. Sedangkan,  $\hat{\sigma}$  adalah skala estimasi *robust* (Dewayanti & Widodo, 2016).

$$\hat{\sigma} = \frac{MAD}{0.6745} = \frac{\text{median}|\varepsilon - \text{median}(\varepsilon)|}{0.6745} \tag{2.17}$$

**2.10 Fungsi Pembobot Andrew**

Fungsi objektif adalah fungsi yang digunakan untuk mencari fungsi objektif pada regresi *robust*. Pembobot *Andrew* mempunyai fungsi objektif sebagai berikut:



$$\rho(u) = \begin{cases} c \left(1 - \cos\left(\frac{u}{c}\right)\right) & , |u| \leq c\pi \\ 2c & , |u| > c\pi \end{cases} \quad (2.18)$$

dengan fungsi *influence* ( $\psi(u)$ ) sebagai berikut:

$$\psi(u) = \rho'(u) = \frac{\partial(\rho(u))}{\partial u} = \begin{cases} \sin\left(\frac{u}{c}\right) & , |u| \leq c\pi \\ 0 & , |u| > c\pi \end{cases} \quad (2.19)$$

Sedangkan, fungsi pembobot *Andrew* adalah :

$$w(u) = \frac{\psi(u)}{u} = \begin{cases} \frac{\sin\left(\frac{u}{c}\right)}{\frac{u}{c}} & , |u| \leq c\pi \\ 0 & , |u| > c\pi \end{cases} \quad (2.20)$$

dengan  $\pi$  adalah 3.14 dan nilai  $c$  pada pembobot disebut *tunning constant*.

*Tunning constant* menentukan efisiensi dari teknik regresi *robust* yang digunakan.

*Tunning constant* untuk pembobot *Andrew* adalah  $c = 1.339$  (Deria dkk., 2019).

### 2.11 Kemiskinan

Menurut Badan Pusat Statistik (2023), kemiskinan adalah ketidakmampuan ekonomi untuk memenuhi kebutuhan dasar makanan dan non-makanan yang diukur dari sisi pengeluaran. Data kemiskinan yang akurat dan tepat sasaran sangat penting untuk strategi penanggulangan kemiskinan. Pengukuran kemiskinan yang andal dapat membantu pemerintah dalam memfokuskan perhatian pada kondisi hidup orang miskin. Data ini dapat digunakan untuk mengevaluasi kebijakan pemerintah, membandingkan kemiskinan antar waktu dan daerah, serta menentukan target penduduk miskin untuk memperbaiki kondisi mereka.

Nilai *Head Count Index* (HCI) provinsi Sulawesi Selatan pada tahun 2022 yaitu sebesar 8.63% (Risantika dkk., 2022). Rumus yang digunakan untuk mencari persentase penduduk miskin menggunakan *Head Count Index-P<sub>0</sub>*, yaitu persentase penduduk yang berada dibawah Garis Kemiskinan (GK).

$$P_0 = \frac{1}{n} \sum_{i=1}^q \left[ \frac{z - y_i}{z} \right]^0 \quad (2.21)$$

$P_0$  adalah persentase Penduduk Miskin (*Head Count Index*);  $z$  sebagai garis kemiskinan;  $y_i$  sebagai rata-rata pengeluaran per kapita sebulan penduduk yang berada dibawah garis kemiskinan,  $y_i < z$ ;  $q$  sebagai banyaknya penduduk yang berada dibawah garis kemiskinan; dan  $n$  adalah jumlah penduduk.

