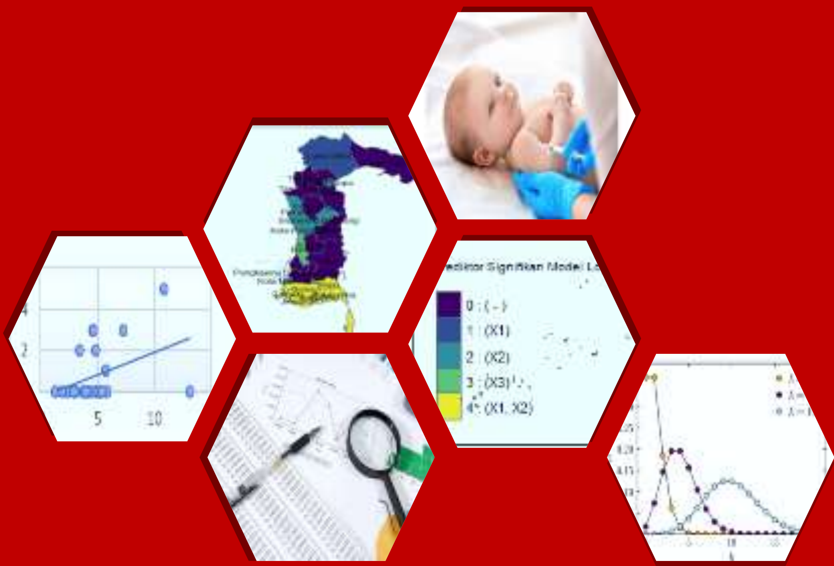
PEMODELAN *GEOGRAPHICALLY WEIGHTED ZERO INFLATED NEGATIVE BINOMIAL REGRESSION* PADA DATA JUMLAH KASUS KEMATIAN BAYI AKIBAT DEMAM DI PROVINSI SULAWESI SELATAN



AHMAD RYAN AL AQSHA

H051201075

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2024

PEMODELAN *GEOGRAPHICALLY WEIGHTED ZERO INFLATED NEGATIVE BINOMIAL REGRESSION* PADA DATA JUMLAH KASUS KEMATIAN BAYI AKIBAT DEMAM DI PROVINSI SULAWESI SELATAN

AHMAD RYAN AL AQSHA
H051201075

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024

PEMODELAN *GEOGRAPHICALLY WEIGHTED ZERO INFLATED NEGATIVE BINOMIAL REGRESSION* PADA DATA JUMLAH KASUS KEMATIAN BAYI AKIBAT DEMAM DI PROVINSI SULAWESI SELATAN

AHMAD RYAN AL AQSHA
H051201075

Skripsi

sebagai salah satu syarat untuk mencapai gelar sarjana

Program Studi Statistika

pada

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

2024

SKRIPSI

PEMODELAN *GEOGRAPHICALLY WEIGHTED ZERO INFLATED NEGATIVE BINOMIAL REGRESSION* PADA DATA JUMLAH KASUS KEMATIAN BAYI AKIBAT DEMAM DI PROVINSI SULAWESI SELATAN

AHMAD RYAN AL AQSHA
H051201075

Skripsi,

telah dipertahankan di depan Panitia Ujian Sarjana Statistika pada tanggal 20 Juni 2024 dan dinyatakan telah memenuhi syarat kelulusan pada

Program Studi Statistika
Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Hasanuddin
Makassar

Mengesahkan:
Pembimbing Tugas Akhir,

Drs. Raupong, M.Si.
NIP. 19621015 198810 1 001

Mengetahui:
Ketua Program Studi,

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

PERNYATAAN KEASLIAN SKRIPSI
DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, skripsi berjudul "Pemodelan *Geographically Weighted Zero Inflated Negative Binomial Regression* pada Data Jumlah Kasus Kematian Bayi Akibat Demam di Provinsi Sulawesi Selatan" adalah benar karya saya dengan arahan dari pembimbing Drs. Raupong, M.Si. sebagai Pembimbing Utama. Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka skripsi ini. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini adalah karya orang lain, maka saya bersedia menerima sanksi atas pembuatan tersebut berdasarkan aturan yang berlaku.

Dengan ini saya melimpahkan hak cipta (hak ekonomis) dari karya tulis saya berupa skripsi ini kepada Universitas Hasanuddin.

Makassar, 20 Juni 2024

Ahmad Ryan Al Aqsha
NIM. H051201075

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul "**Pemodelan *Geographically Weighted Zero Inflated Negative Binomial Regression* pada Data Jumlah Kasus Kematian Bayi Akibat Demam di Provinsi Sulawesi Selatan**". Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa Sallam* beserta keluarga dan para sahabatnya.

Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada **Bapak Drs. Raupong, M.Si.** selaku Pembimbing Utama yang dengan penuh kesabaran telah meluangkan waktu dan pemikirannya, memberikan motivasi, dorongan semangat, dan bimbingan kepada penulis selama proses penulisan tugas akhir ini. Terima kasih kepada **Bapak Dr. Nirwan, M.Si.** dan **Ibu Sitti Sahriman, S.Si., M.Si.**, selaku Tim Penguji yang senantiasa memberikan saran dan kritikan yang membangun dalam penyempurnaan penulisan tugas akhir ini. Terima kasih kepada **Pimpinan Universitas Hasanuddin, Departemen Statistika, Jajaran Dosen,** dan **Staf Departemen Statistika** yang telah memfasilitasi, memberikan ilmu bermanfaat, dan membantu penulis selama menempuh studi.

Penghargaan dan ucapan terima kasih yang tulus juga penulis ucapkan kepada kedua orang tua penulis yang tercinta, Ibunda **Marlinda** dan Ayahanda **Susanto Sulle** yang telah memberikan kasih sayang, dukungan, semangat, dan doa yang senantiasa mengiringi langkah penulis. Terima kasih kepada keluarga besar **Himastat FMIPA Unhas** khususnya untuk teman-teman **POIS20N** dan teman-teman seperjuangan **STATISTIKA 2020** untuk segala ilmu, cerita dan pengalamannya. Terima kasih kepada **Sri Rahayu Yusri** yang telah menjadi teman, sahabat dan pasangan hidup penulis dengan segala sikap dan sifat yang memenuhi segala aspek kekurangan dalam diri penulis. Terima kasih juga kepada sahabat-sahabat **Fadlan, Mukhlis, Fahmi, Kurr, Izzul, Razy, Azhar, Engkal, Hakam, Ridwan, Yoel, Sabila, Nisa, Uci, Liza, Nahda, Linda, Irma dan Azal** yang setia menemani dalam menyelesaikan skripsi serta momen-momen kebersamaan belajar, diskusi yang terus membangun diri penulis.

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

Penulis,

Ahmad Ryan Al Aqsha

ABSTRAK

AHMAD RYAN AL AQSHA. **Pemodelan *Geographically Weighted Zero Inflated Negative Binomial Regression* pada Data Jumlah Kasus Kematian Bayi Akibat Demam di Provinsi Sulawesi Selatan** (dibimbing oleh Raupong).

Latar belakang. Distribusi Poisson terjadi pada data diskrit dengan variabel respon berupa data *count* dengan asumsi equidispersi. Namun, sering ditemukan kejadian overdispersi pada data. Nilai nol yang berlebih pada data (*excess zeros*) juga dapat menyebabkan overdispersi. Regresi *Zero Inflated Negative Binomial* (ZINB) dapat menjadi alternatif yang dapat digunakan dalam proses analisis untuk mengatasi hal tersebut. Selain itu, faktor geografis juga sering muncul pada data yang dapat menyebabkan heterogenitas spasial sehingga digunakan *Geographically Weighted Regression* (GWR) dalam proses analisis. ZINB dan GWR dapat digabungkan menjadi *Geographically Weighted Zero Inflated Negative Binomial Regression* (GWZINBR) yang menjadi pengembangan model regresi untuk memodelkan data yang mengalami overdispersi pada regresi Poisson, *excess zeros*, dan heterogenitas spasial. Data yang digunakan dalam penelitian ini adalah jumlah kasus kematian bayi akibat demam di Provinsi Sulawesi Selatan Tahun 2020 yang mengandung 70,83% nilai nol. **Tujuan.** Penelitian ini bertujuan untuk memperoleh estimator parameter dan membentuk model GWZINBR pada data. **Metode.** Penelitian ini dibagi menjadi dua tahap, yaitu 1) estimasi parameter model GWZINBR menggunakan *Maximum Likelihood Estimation* (MLE) dengan algoritma *Expectation Maximization* (EM); dan 2) penerapan model GWZINBR pada data. **Hasil.** Pengujian secara parsial pada parameter model GWZINBR dengan pembobot *adaptive bisquare kernel* menghasilkan variabel prediktor signifikan yang beragam antar kabupaten/kota baik itu model log ataupun model logit. Beberapa kabupaten/kota memiliki variabel prediktor signifikan yang sama, dan juga terdapat daerah yang sama sekali tidak memilik variabel prediktor signifikan. **Kesimpulan.** Model GWZINBR dapat diterapkan pada data dan menghasilkan model bersifat lokal untuk setiap kabupaten/kota dalam pengamatan.

Kata kunci: Distribusi Poisson, Overdispersi, *Excess Zeros*, Pembobot, Heterogenitas Spasial, Kematian Bayi.

ABSTRACT

AHMAD RYAN AL AQSHA. **Modeling Geographically Weighted Zero Inflated Negative Binomial Regression on Infant Mortality Due to Fever Cases in South Sulawesi Province** (Supervised by Raupong)

Background: Poisson distribution occurs in discrete data with a response variable in the form of count data under the assumption of equidispersion. However, overdispersion is often encountered in data. Excess zeros in the data can also cause overdispersion. Zero Inflated Negative Binomial (ZINB) regression can be an alternative for addressing this issue in the analysis process. Additionally, geographic factors often appear in the data, leading to spatial heterogeneity, thus, Geographically Weighted Regression (GWR) is used in the analysis process. ZINB and GWR can be combined into Geographically Weighted Zero Inflated Negative Binomial Regression (GWZINBR), which is a model development to handle overdispersion in Poisson regression, excess zeros, and spatial heterogeneity. The data used in this study is the number of infant deaths due to fever in South Sulawesi Province in 2020, which contains 70.83% zero values. **Objective**: This study aims to obtain parameter estimators and form the GWZINBR model for the data. **Method**: This research is divided into two stages: 1) parameter estimation of the GWZINBR model using Maximum Likelihood Estimation (MLE) with the Expectation Maximization (EM) algorithm; and 2) application of the GWZINBR model to the data. **Results**: Partial testing of the GWZINBR model parameters with adaptive bisquare kernel weighting produced various significant predictor variables across different districts/cities in both the log and logit models. Some districts/cities had the same significant predictor variables, while some areas had no significant predictor variables at all. **Conclusion**: The GWZINBR model can be applied to the data and produce a local model for each district/city in the observation.

Keywords: Poisson Distribution, Overdispersion, Excess Zeros, Weighting, Spatial Heterogeneity, Infant Mortality

DAFTAR ISTILAH

Istilah	**Arti dan Penjelasan**
Bobot	Nilai yang diberikan untuk menunjukkan besar pengaruh dari lokasi amatan, memiliki nilai dari 0 sampai 1
Data *count*	Jumlah observasi dalam data berbentuk bilangan bulat positif termasuk nol.
Distribusi	Pola penyebaran data atau probabilitas suatu kejadian dalam ruang sampel
Dispersi	Tingkat variasi data
Excess Zeros	Kondisi data memiliki lebih banyak nilai nol (>50%) yang bisanya terjadi pada data *count*
Geografis	Sesuatu yang berhubungan dengan Lokasi, posisi atau karakteristik suatu wilayah
Heterogenitas Spasial	Variasi dalam lokasi atau karakteristik yang berbeda di berbagai Lokasi pengamatan
Mean	Rata-rata dari sekumpulan nilai pada data
Model	Representasi statistik untuk menggambarkan hubungan antara variabel-variabel dalam suatu data
Non-Linear	Hubungan tidak proporsional antar variabel
Overdispersi	Kondisi nilai variansi lebih besar dari nilai rata-rata
Parameter	Nilai atau koefisien tetap dalam model atau fungsi
Prediktor	Variabel independen dalam penelitian yang dianggap mempengaruhi variabel respon
Regresi	Metode analisis hubungan antara variabel respon dan satu atau lebih variabel prediktor
Respon	Variabel dependen dalam penelitian
Varians	Ukuran penyebaran data dari nilai rata-rata
Variabel	Entitas yang nilainya dapat berubah

DAFTAR LAMBANG DAN SINGKATAN

Lambang/singkatan	Arti dan Penjelasan
y_i	Variabel respon pada lokasi pengamatan ke-i
λ_i	Rata-rata variabel respon pada pengamatan ke-i
ε_i	Galat ke i
$\boldsymbol{x}_i^T$	Vektor variabel prediktor pada lokasi pengamatan ke-i
$\boldsymbol{\beta}$	Vektor parameter regresi Poisson berukuran $(p+1)\times 1$
e	Bilangan euler yang bernilai 2,71828
exp	Eksponensial
ln	Logaritma natural
$E(Y)$	Ekspektasi dari variabel respon y
$Var(Y)$	Variansi dari variabel respon y
D	Nilai deviasi
ϕ	Nilai statistik uji variansi
db	Derajat Bebas
κ	Parameter disperse
π_i	Parameter nilai untuk y duga bernilai nol
β_p	Parameter model log yang diestimasi sebanyak p
γ_p	Parameter model logit yang diestimasi sebanyak p
z_i	Variabel indikator yang dianggap hilang atau nol
$\hat{I}$	Nilai Indeks Moran
w_{ij}	Pembobot antara lokasi ke-i dan lokasi ke-j
(u_i, v_i)	Koordinat geografis (lintang, bujur) pada lokasi ke-i
d_{ij}	Jarak antara lokasi (u_i, v_i) ke lokasi $\left(u_j, v_j\right)$
h	Parameter penghalus (*bandwidth*)
G	Nilai deviasi
$L\left(\hat{\Omega}\right)$	Parameter model penuh
$L(\hat{\omega})$	Parameter model intersep
W	Nilai Uji Wald
VIF	*Variance Inflation Factors*
UCI	*Universal Child Immunization*
BBLR	Berat Bayi Lahir Rendah
CV	*Cross Validation*
BP	*Breusch-Pagan*
NB	*Negative Binomial*
ZINB	*Zero Infalted Negative Binomial*
GWR	*Geographically Weighted Regression*
GWZINBR	*Geographically Weighted Zero Infalted Negative Binomial Regression*
MLE	*Maksimum Likelihood Estimation*
EM	*Expectation Maximization*

DAFTAR ISI

Halaman

DAFTAR TABEL

Tabel **Halaman**

DAFTAR GAMBAR

Gambar **Halaman**

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis regresi adalah penerapan ilmu statistika yang diperuntukan untuk memodelkan hubungan antara variabel respon dengan variabel prediktor. Persamaan regresi dapat berbentuk hubungan linear atau nonlinear. Ketika variabel respon berupa data kontinu dan mengikuti distribusi normal maka digunakan persamaan regresi linear. Namun, variabel respon seringnya ditemukan tidak berdistribusi normal sehingga hubungannya nonlinear. Oleh karena itu, dikembangkan *Generalized Linear Model* (GLM) untuk mengatasi hal tersebut. GLM adalah perluasan dari model regresi umum dan variabel responnya berdistribusi keluarga eksponensial (Myers dkk., 2010). Salah satu yang termasuk dalam keluarga eksponensial adalah distribusi Poisson. Data yang berbentuk *count* (jumlahan) dalam periode, ruang atau volume tertentu serta sifatnya jarang terjadi adalah ciri dari data berdistribusi Poisson.

Pembentukan model regresi ketika variabel respon berupa data diskrit dan berdistribusi Poisson dapat menggunakan model regresi Poisson. Terdapat asumsi yang harus dipenuhi dalam regresi Poisson yaitu nilai variansi dan nilai rata-rata dari variabel respon memiliki nilai yang sama atau equidispersi (Myers dkk., 2010). Namun, proses analisis pada regresi Poisson terkadang mengalami overdispersi, yaitu nilai variansi lebih besar dari nilai rata-ratanya sehingga menghasilkan galat baku yang lebih kecil dari nilai sesungguhnya (*underestimate*) yang dapat menyebabkan kesimpulan menjadi tidak valid (McCullagh, 1989).

Distribusi binomial negatif dapat menjadi alternatif yang cukup baik apabila terjadi overdispersi (Hilbe, 2011). Distribusi binomial negatif yang dimaksud adalah hasil dari campuran distribusi Poisson-gamma. Regresi binomial negatif memiliki karakteristik yang sama dengan regresi Poisson. Hanya saja regresi binomial negatif lebih fleksibel dibandingkan regresi Poisson karena regresi ini memiliki parameter dispersi yang membuat nilai variansi dapat bervariasi menjadi lebih besar dari rata-rata.

Variabel respon yang mengandung banyak nilai nol (*excess zeros*) dapat menyebabkan overdispersi pada regresi Poisson. *Excess zeros* adalah kondisi variabel respon memiliki nilai nol yang berlebih untuk data *count* yang artinya data tidak mengandung nilai negatif. Lambert (1992) mengatakan bahwa untuk kasus *excess zeros* pada data dapat dimodelkan dengan menggunakan regresi *Zero-Inflated Poisson* (ZIP) yang merupakan model campuran untuk data diskrit dengan banyak peristiwa bernilai nol. Tetapi regresi ini tidak mampu mengatasi overdispersi tinggi yang terjadi. Ismail & Zamani (2013) menyarankan bahwa data yang mengalami *excess zeros* dan overdispersi tinggi lebih sesuai menggunakan regresi *Zero-Inflated Negative Binomial* (ZINB) karena dalam distribusi *negative binomial* memperbolehkan variansi lebih besar dari rata-rata. Model regresi ZINB merupakan distribusi campuran Poisson-gamma dan menghasilkan estimasi parameter yang bersifat global atau sama untuk setiap pengamatan (Hilbe, 2011).

Ketika pengamatan yang dilakukan mengandung faktor geografis, maka data mengandung efek spasial. Ketergantungan spasial yang terjadi karena adanya korelasi antarlokasi yang melingkupi dependensi spasial dan heterogenitas spasial merupakan bentuk dari efek spasial. Heterogenitas spasial inilah yang memberikan perbedaan respon dari berbagai faktor antara lokasi yang satu dengan lokasi yang lain. Untuk mengatasi heterogenitas spasial, maka digunakan sebuah model regresi terboboti secara spasial, yaitu model *Geographically Weighted Regression* (GWR).

Model GWR adalah pengembangan dari model regresi untuk setiap parameter dihitung pada setiap titik lokasi pengamatan (Fotheringham dkk., 2002). Metode ini dapat digunakan untuk mengestimasi data yang memiliki keragaman spasial dengan hasil analisis yaitu parameter model bersifat lokal untuk setiap lokasi pengamatan. GWR dan ZINB adalah dua metode statistik yang dapat dikombinasikan untuk menganalisis data spasial dengan nol berlebih pada variabel respon serta mengalami overdispersi.

Penelitian terkait GWR telah dilakukan oleh Daulay & Simamora (2023) dengan hasil penelitian bahwa metode ini baik dalam menunjukkan faktor yang menyebabkan kemiskinan di Provinsi Sumatera Utara. Sedangkan penelitian terkait ZINB telah dilakukan oleh Fajri (2022) yang memperlihatkan hasil dari model regresi ZINB adalah model yang baik dalam mengatasi overdispersi pada data jumlah kematian neonatal di Kota Makassar tahun 2018. Aicha (2022) juga telah melakukan penelitian yang menggabungkan GWR dengan regresi ZIP yaitu *Geographically Weighted Zero Inflated Poisson Regression* (GWZIPR) pada kasus kematian demam berdarah *Dengue* di Sulawesi Selatan dan hasil penelitannya menunjukkan model untuk setiap lokasi pengmatan memiliki faktor signifikan yang berbeda dan model tersebut layak untuk digunakan.

Berdasarkan uraian tersebut, penulis tertarik untuk melakukan penelitian yang menggabungkan metode GWR dengan regresi ZINB yaitu *Geographically Weighted Zero Inflated Negative Binomial Regression* (GWZINBR). Proses estimasi parameter model GWZINBR dapat mengacuh pada proses estimasi parameter model ZINB yang kemudian diberikan unsur geeografis dan bobot dalam model sebagaimana proses estimasi pada model GWR. Metode estimasi dapat menggunakan *Maksimum Likelihood Estimation* (MLE) untuk memaksimalkan fungsi peluang dengan algoritma *Expectation Maximization* (EM) yang didalamnya terdapat proses pembentukan distribusi gabungan dan tahap ekspektasi yang dapat mengakomodasi data yang dianggap hilang akibat *excess zeros* pada variabel respon. Nilai estimasi parameter dari model ZINB kemudian menjadi parameter awal dalam proses iterasi *Newton Rapshon* untuk memperoleh nilai estimasi parameter model GWZINBR.

Penelitian ini diaplikasikan pada data jumlah kasus kematian bayi akibat demam di Provinsi Sulawesi Selatan tahun 2020. Data kematian bayi di setiap kabupaten/kota Sulawesi Selatan dipetakan berdasarkan penyakit yang dialami bayi dan demam menjadi salah satu penyakitnya. Tedapat berbagai faktor baik secara endogen maupun eksogen yang menyebabkan kematian bayi ini. Ketika mempertimbangkan kondisi wilayah untuk melihat perbandingan maka perlu dilakukan pengecekan terhadap faktor eksogen di setiap kabupaten/kota Sulawesi Selatan. Salah satu upaya dalam menentukan strategi penanggulangannya, dapat dilakukan adalah menganalisis faktor-faktor geografis yang diduga memiliki pengaruh terhadap tingginya kasus kematian bayi akibat demam di setiap kabupaten/kota Sulawesi Selatan. Penulis menyusun penelitian ini dengan judul "Pemodelan *Geographically Weighted Zero Inflated Negative Binomial Regression* pada Data Jumlah Kasus Kematian Bayi Akibat Demam di Provinsi Sulawesi Selatan".

1.2 Batasan Masalah

Batasan masalah yang dimasukkan dalam penelitian ini adalah sebagai berikut:

1. Data penelitian yang digunakan yaitu kasus kematian bayi akibat demam di Provinsi Sulawesi Selatan tahun 2020 berdasarkan data Dinas Kesehatan Provinsi Sulawesi Selatan: "Profil Kesehatan Tahun 2021".
2. Pendugaan parameter model menggunakan metode *Maximum Likelihood Estimation* dengan algoritma *Expectation Maximization*.
3. Jarak yang digunakan dalam penentuan pembobot yaitu jarak dalam satuan kilometer berdasarkan *Google Maps*.
   Pembobot yang digunakan dalam pendugaan parameter yaitu *Adaptive Bisquare Kernel*.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan penelitian adalah sebagai berikut:

1. Memperoleh estimator parameter model *Geographically Weighted Zero Inflated Negative Binomial Regression* yang bersesuaian dengan data menggunakan metode *Maximum Likelihood Estimation*.
2. Mendapatkan model yang terbentuk dari kasus kematian bayi akibat demam di Provinsi Sulawesi Selatan tahun 2020 berdasarkan *Geographically Weighted Zero Inflated Negative Binomial Regression*.

1.4 Manfaat Penelitian

Manfaat yang ingin diperoleh dari penelitian ini adalah sebagai berikut:

1. Sebagai bahan informasi, masukan dan pertimbangan masyarakat ataupun pemerintah dalam mengambil keputusan/kebijakan yang dapat menekan angka kasus kematian bayi akibat demam di Provinsi Sulawesi Selatan.
2. Sebagai bahan referensi bagi peneliti lain yang akan mengadakan penelitian dalam bidang yang sama maupun bahan referensi pembelajaran terkait metode *Geographically Weighted Zero Inflated Negative Binomial Regression*.

1.5 Teori

1.5.1 *Generalized Linier Model*

Dalam analisis regresi terdapat model regresi yang mengalami perluasan yang disebut *Generalized Linear Model* (GLM) dengan variabel respon mengikuti distribusi keluarga eksponensial. Distribusi yang termasuk dalam keluarga eksponensial adalah distribusi normal, binomial, binomial negatif, Poisson, geometrik, eksponensial, dan gamma. Terdapat tiga komponen utama yang harus ada dalam suatu GLM, yaitu (McCullagh, 1989):

1. Komponen acak, yaitu suatu komponen yang mengidentifikasi distribusi dari variabel respon (Y) berasal dari keluarga eksponensial. Bentuk keluarga distribusi eksponensial dapat dinyatakan dalam fungsi peluang berikut:
$$\begin{aligned} f(y) &= s(y)t(\kappa)e^{a(y)b(\kappa)} \\ &= \exp[a(y)b(\kappa) + c(\kappa) + d(y)] \end{aligned} \tag{1}$$
dengan $a(.)$, $b(.)$, $s(.)$, $t(.)$ adalah fungsi yang diketahui $s(y) = \exp[d(y)]\,, t(\kappa) = \exp[c(\kappa)]$ dan κ adalah parameter dispersi.
2. Komponen sistematik, meliputi variabel-variabel prediktor dari model untuk menjelaskan variabel-variabel yang berhubungan dalam sebuah model linear.
$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} \tag{2}$$
3. Fungsi penghubung (*link function*), yaitu fungsi yang menghubungkan rata-rata (λ) dari variabel respon (Y) dengan variabel-variabel prediktor melalui persamaan linear. Misalkan $\lambda_i = E(Y_i), i = 1,2, \dots, n,$ fungsi untuk menghubungkan λ_i dengan η_i adalah $g(.)$ Sehingga $g(\lambda_i) = \eta_i$. Fungsi penghubung $g(.)$ Menghubungkan $E(Y_i)$ dengan variabel prediktor, yaitu:
$$g(\lambda_i) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\,, i = 1,2, \dots, n \tag{3}$$

1.5.2 Regresi Poisson

Regresi Poisson merupakan model regresi nonliniear yang sering digunakan untuk mengatasi data *count* (jumlahan) dengan variabel respon mengikuti distribusi Poisson dan merupakan metode yang digunakan untuk memprediksi pola hubungan antara variabel respon (Y) dan prediktor (X) (Agresti, 2002). Distribusi Poisson merupakan anggota dari distribusi keluarga eksponensial dengan fungsi massa peluang dari distribusi ini memenuhi Persamaan (1).

Distribusi Poisson adalah pendekatan distribusi binomial dengan banyaknya *n* percobaan relatif besar. Pendekatan ini diperoleh dengan pertimbangan jika *n* percobaan besar, maka perhitungan peluang menggunakan distribusi binomial akan sulit dikerjakan dan membutuhkan waktu lama (Yulisti'anah, 2018). Fungsi massa peluang distribusi Poisson dapat dinyatakan sebagai berikut:
$$f(y|\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}, y = 0,1,2, \dots \tag{4}$$

dengan $\lambda > 0$ adalah rata-rata variabel respon (Y) yang berdistribusi Poisson untuk nilai rata-rata dan variansi dari acak Y mempunyai nilai lebih dari nol, $E(Y) = \lambda$ dan $Var(Y) = \lambda$. Regresi Poisson memenuhi tiga komponen utama dalam GLM sehingga termasuk dalam keluarga eksponensial yang dapat dilihat dalam rincian sebagai berikut (Agresti, 2002):

1. Komponen acak untuk variabel respon Y terdiri dari $(y_1, y_2, \dots y_n)$ dan fungsi massa peluang distribusi Poisson yang dapat dinyatakan sebagai berikut:
   $Y \sim$ Poisson (λ)
   $$\begin{aligned} f(y|\lambda) &= \frac{e^{-\lambda}\lambda^y}{y!} \\ &= \exp\left\{\ln \frac{e^{-\lambda}\lambda^y}{y!}\right\} \\ &= \exp\left\{\ln\left(e^{-\lambda}\lambda^y\right) - \ln(y!)\right\} \\ &= \exp\left\{\ln\left(\lambda^y e^{-\lambda}\right) - \ln(y!)\right\} \\ &= \exp\{y \ln \lambda - \lambda) - \ln\,(y!)\} \\ &= \exp[a(y)b(\kappa) + c(\kappa) + d(y)] \end{aligned}$$
   dengan fungsi peluang tersebut memenuhi Persamaan (1) sehingga diperoleh $a(y) = y$, $b(\kappa) = \ln \lambda$, $c(\kappa) = -\lambda$ dan $d(y) = -\ln y!$
2. Komponen sistematis adalah vektor $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)$ menghubungkan ke variabel prediktor. Masing-masing dari elemen $\boldsymbol{\eta}$ merupakan kombinasi linear dari variabel prediktor, dapat dinyatakan dengan $\eta_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ji}$ yang memenuhi Persamaan (2).
3. Fungsi penghubung yang digunakan adalah $g(.)$ yang kemudian berfungsi untuk menghubungkan komponen acak dan komponen sestematis sehingga dapat dinyatakan dengan $g(\lambda_i) = \eta_i$. Fungsi $g(.)$ menghubungkan λ_i dengan elemen $\boldsymbol{\eta}$ yaitu $g(\lambda_i) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ji}$ yang memenuhi Persamaan (3).

Myers dkk. (2010) menjelaskan model regresi Poisson adalah $y_i = \lambda_i + \varepsilon_i$ dengan y_i adalah variabel respon pada lokasi pengamatan ke-i, λ_i adalah rata-rata variabel respon pada pengamatan ke i dan ε_i adalah galat ke i. Model regresi Poisson merupakan salah satu model GLM dengan fungsi penghubung log dari nilai ekspektasi sehingga $\ln E(y_i) = \ln(\lambda_i)$ dan bergantung pada variabel prediktor $\boldsymbol{x}_i$, model regresi Poisson dapat ditulis sebagai berikut:

$$\begin{aligned} \ln(\lambda_i) &= \beta_0 + \sum_{j=1}^{p} \beta_j x_{ji}\,, i = 1,2,\dots,n \\ \ln(\lambda_i) &= \boldsymbol{x}_i^T \boldsymbol{\beta} \\ \exp(\ln(\lambda_i)) &= \exp \boldsymbol{x}_i^T \boldsymbol{\beta} \\ \lambda_i &= \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) \end{aligned} \tag{5}$$

dengan λ_i adalah rata-rata variabel respon pada pengamatan ke- i yang merupakan ekspektasi dari Y_i ; $\boldsymbol{x}_i^T = [1 \quad x_{1i} \quad x_{2i} \quad \cdots \quad x_{pi}]$ adalah vektor variabel prediktor pada lokasi pengamatan ke-i ; $\boldsymbol{\beta} = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \cdots \quad \beta_p]^T$ adalah vektor parameter regresi Poisson berukuran $(p+1) \times 1$.

Suatu data dapat dilakukan pengujian terhadap variabel responnya sehingga dapat dikatakan berdistribusi Poisson. Salah satu uji yang dapat digunakan adalah uji *Kolmogorov-Smirnov* (Ruliana dkk., 2016), hipotesisnya dinyatakan sebagai berikut:

H_0: variabel respon berdistribusi Poisson

H_1: variabel respon tidak berdistribusi Poisson

Statistika uji:

$$D_{hitung} = maks|F_n(Y) - F_0(Y)| \tag{6}$$

dengan

$F_0(Y)$: fungsi distribusi kumulatif yang dihipotesiskan

$F_n(Y)$: fungsi distribusi kumulatif yang diamati

Kriteria pengujiannya yaitu tolak H_0 apabila $D_{hitung} > D_{tabel}$ atau $p - value$ lebih kecil dari α yang ditentukan. Maka kesimpulan yang diperoleh adalah variabel respon tidak berdistribusi Poisson. D_{tabel} adalah tabel nilai kritis uji *Kolomogrov-Smirnov*.

1.5.3 Equidispersi

Asumsi penting pada model regresi Poisson adalah equidispersi yaitu kesamaan variansi dengan nilai rata-rata. Kondisi ini sering tidak terpenuhi pada data diskrit. Pada umumnya ditemukan data diskrit yang memiliki variansi lebih besar dari nilai rata-rata (overdispersi) atau variansi lebih kecil dari nilai rata-rata (underdispersi). Model regresi Poisson yang diterapkan pada data yang overdispersi mengakibatkan galat baku dari parameter dugaan menjadi berbias ke bawah (*underestimate*) dan signifikansi dari pengaruh variabel prediktor menjadi berbias ke atas (*overstate*).

Pemeriksaan overdispersi dapat dilakukan menggunakan nilai deviasi (*deviance*) dibagi dengan derajat bebas dan mempunyai nilai lebih besar dari 1, sedangkan underdispersi dideteksi menggunakan nilai deviasi dibagi dengan derajat bebas dan mempunyai nilai kurang dari 1. Menurut Agresti (2002), nilai deviasi dapat dinyatakan sebagai berikut:

$$\phi = \frac{D}{db} \tag{7}$$

$$D = 2\sum_{i=1}^{n} y_i \ln\left(\frac{y_i}{\hat{\lambda}}\right) \tag{8}$$

dengan

ϕ : Uji Variansi (parameter dispersi κ)

D : Nilai deviasi

y_i : Nilai variabel respon dari pengamatan ke-i

db : Derajat bebas

$\hat{\lambda}$: Penduga rataan

dengan, $db = n - c\ (n > c), n$ merupakan banyaknya pengamatan, $c = p + 1$ merupakan banyaknya parameter. Jika nilai statistik uji (ϕ) lebih dari 1 maka dapat disimpulkan data mengalami overdispersi.

1.5.4 Regresi Binomial Negatif

Regresi binomial negatif merupakan salah satu model terapan dari GLM. Sebagai penerapan dari GLM maka distribusi binomial negatif memiliki tiga komponen yaitu:

1. Komponen Acak

Pada model binomial negatif variabel respon diasumsikan berdistribusi binomial negatif yang dihasilkan dari distribusi campuran Poisson-gamma (Simarmata & Ispriyanti, 2011). Misalkan:

$y|\lambda \sim$ Poisson (λ)

$\lambda \sim$ Gamma (α, β)

Distribusi peluang campuran Poisson-gamma dapat diperoleh dengan cara:

$$P(y|\alpha,\beta) = \int_0^\infty \text{Poisson } (y|\lambda) \text{ Gamma } (\lambda|\alpha,\beta)\, d\lambda$$

$$= \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha}\, \lambda^{\alpha-1} e^{\left(-\frac{\lambda}{\beta}\right)}\, d\lambda$$

$$= \frac{\beta^{-\alpha}}{\Gamma(\alpha)y!} \int_0^\infty e^{-\lambda\left(1+\frac{1}{\beta}\right)}\, \lambda^{y+\alpha-1} d\lambda$$

Misalkan $v = \left(1+\frac{1}{\beta}\right)\lambda$ maka $dv = \left(1+\frac{1}{\beta}\right)d\lambda$, dan untuk $\lambda = 0 \to v = 0$ sedangkan $\lambda = \infty \to v = \infty$.

$$P(y|\alpha,\beta) = \frac{\beta^{-\alpha}}{\Gamma(\alpha)y!} \int_0^\infty e^{-v} \left(\frac{\beta v}{1+\beta}\right)^{y+\alpha-1} \left(\frac{\beta v}{1+\beta}\right) dv$$

$$= \frac{\beta^{-\alpha}}{\Gamma(\alpha)y!} \left(\frac{\beta}{1+\beta}\right)^{y+\alpha} \int_0^\infty e^{-v}\, v^{y+\alpha-1} dv$$

$$= \frac{\beta^{-\alpha}}{\Gamma(\alpha)y!} \left(\frac{\beta}{1+\beta}\right)^{\alpha} \left(\frac{\beta}{1+\beta}\right)^{y_i} \Gamma(y+\alpha)$$

$$P(Y|\alpha,\beta) = \frac{\Gamma(y+\alpha)}{\Gamma(\alpha)y!} \left(\frac{1}{1+\beta}\right)^{\alpha} \left(\frac{\beta}{1+\beta}\right)^{y}, \text{ dengan } y = 0,1,2,\ldots$$

$P(Y|\alpha,\beta)$ merupakan fungsi massa peluang binomial negatif yang dihasilkan dari distribusi campuran Poisson-gamma. Nilai rata-rata dan variansi campuran Poisson-gamma adalah:

$$E[Y] = \alpha\beta \text{ dan } Var[Y] = \alpha\beta + \alpha\beta^2$$

untuk membentuk suatu model regresi pada distribusi binomial negatif, maka nilai parameter distribusi campuran Poisson-gamma dinyatakan dalam bentuk $\lambda = \alpha\beta$ dan $\kappa = 1/\alpha$ sehingga diperoleh rata-rata dan variansi dalam bentuk:

$$E[Y] = \lambda \text{ dan } Var[Y] = \lambda + \kappa\lambda^2$$

Kemudian fungsi massa peluang binomial negatif menjadi:

$$f(y|\lambda,\kappa) = \frac{\Gamma\left(y+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y!} \left(\frac{1}{1+\kappa\lambda}\right)^{\frac{1}{\kappa}} \left(\frac{\kappa\lambda}{1+\kappa\lambda}\right)^{y}, y = 0,1,2,\ldots,n \qquad (9)$$

Pada saat $\kappa \to 0$ maka distribusi binomial negatif memiliki variansi $Var(Y) \to \lambda$. Distribusi binomial negatif akan mendekati suatu distribusi Poisson yang mengasumsikan rata-rata dan variansi sama yaitu $E(Y) = Var(Y) = \lambda$. Fungsi distribusi keluarga eksponensial dari distribusi binomial negatif seperti pada Persamaan (2.9) (Greene, 2008) adalah sebagai berikut:

$$f(y|\lambda,\kappa) = \frac{\Gamma\left(y+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y!}\left(\frac{1}{1+\kappa\lambda}\right)^{\frac{1}{\kappa}}\left(\frac{\kappa\lambda}{1+\kappa\lambda}\right)^{y}$$

$$= \exp\left[\ln\left(\frac{\Gamma\left(y+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y!}\left(\frac{1}{1+\kappa\lambda}\right)^{\frac{1}{\kappa}}\left(\frac{\kappa\lambda}{1+\kappa\lambda}\right)^{y}\right)\right]$$

$$= \exp\left[\ln\left(\frac{\Gamma\left(y+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y!} + \ln\left(\frac{1}{1+\kappa\lambda}\right)^{\frac{1}{\kappa}} + \ln\left(\frac{\kappa\lambda}{1+\kappa\lambda}\right)^{y}\right)\right]$$

$$= \exp\left[y\ln\left(\left(\frac{\kappa\lambda}{1+\kappa\lambda}\right) + \frac{1}{\kappa}\ln\left(\frac{1}{1+\kappa\lambda}\right) + \ln\frac{\Gamma\left(y+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y!}\right)\right]$$

$$= \exp\left[\, a(y).b(\kappa) + c(\kappa) + d(y)\right]$$

dengan $a(y) = y$, $b(\kappa) = \ln\left(\frac{\kappa\lambda}{1+\kappa\lambda}\right)$, $c(\kappa) = -\frac{1}{\kappa}\ln\left(\frac{1}{1+\kappa\lambda}\right)$ dan $d(y) = \ln\left(\frac{\Gamma\left(y_i+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y_i!}\right)$.

2. Komponen Sistematis

Kontribusi variabel prediktor dalam model regresi binomial negatif dinyatakan dalam bentuk kombinasi linier antara parameter $\boldsymbol{\eta}$ dengan parameter regresi yang diestimasi yaitu:

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_j x_{pi} \tag{10}$$

atau dituliskan dalam bentuk matriks:

$$\boldsymbol{\eta} = \boldsymbol{X\beta}$$

dengan $\boldsymbol{\eta}$ adalah vektor berukuran $n \times 1$ dari pengamatan, $\boldsymbol{X}$ adalah matriks berukuran $n \times c$ yang elemennya terdiri dari variabel prediktor, $\boldsymbol{\beta}$ adalah vektor berukuran $c \times 1$ yang elemennya terdiri dari parameter regresi, dengan $c = p + 1$.

3. Fungsi Penghubung

Nilai rata-rata dari variabel respon adalah diskrit dan bernilai positif. Maka untuk mentransformasikan nilai η_i (bilangan riil) ke rentang yang sesuai dengan rentang pada variabel respon diperlukan suatu fungsi penghubung $g(.)$ yaitu:

$$g(\lambda_i) = \ln \lambda_i = \boldsymbol{x}_i^T\boldsymbol{\beta}$$

$$\lambda_i = \exp(\boldsymbol{x}_i^T\boldsymbol{\beta}) \tag{11}$$

Berdasarkan Persamaan (11), diperoleh model regresi binomial negatif adalah $\hat{y}_i = \lambda_i = \exp\left(\beta_0 + \sum_{j=1}^{p}\beta_j x_{ji}\right)$. Model regresi binomial negatif dapat digunakan untuk memodelkan data Poisson yang mengalami overdispersi karena distribusinya merupakan perluasan dari distribusi Poisson-gamma yang memuat parameter dispersi κ (Hilbe, 2011).

1.5.5 *Excess Zeros*

Salah satu permasalahan pada regresi Poisson adalah nilai nol yang berlebih (*excess zeros*). *Excess zeros* dapat dilihat pada proporsi variabel respon yang bernilai nol lebih besar dari data diskrit lainnya. Menurut Winkelmann (2008), banyaknya proporsi data yang bernilai nol dari proporsi data lainnya ($> 50\%$) dapat berakibat pada ketepatan (presisi) dalam pengambilan keputusan. Selain itu, regresi Poisson juga menjadi tidak tepat digunakan karena *excess zeros* merupakan salah satu penyebab terjadinya overdispersi (Hinde & Demétrio, 1998).

1.5.6 Regresi *Zero Inflated Negative Binomial*

Model regresi *Zero Inflated Negative Binomial* (ZINB) merupakan model yang dibentuk dari distribusi campuran Poisson-gamma. Jika y_i adalah variabel acak dengan $i = 1,2,\ldots,n$ nilai nol pada pengamatan diduga muncul dalam dua cara yang sesuai untuk keadaan (*state*) yang terpisah. Menurut Garay dkk. (2011), regresi ZINB dengan keadaan pertama disebut *zero state* terjadi dengan probabilitas π_i dan menghasilkan hanya pengamatan bernilai nol, sementara keadaan kedua disebut *negative binomial state* terjadi dengan probabilitas $(1-\pi_i)$ dan berdistribusi binomial negatif. Fungsi peluang model regresi ZINB dapat dinyatakan seperti pada Persamaan (12).

$$P(Y_i = y_i) = \begin{cases} \pi_i + (1-\pi_i)\left(\frac{1}{1+\kappa\lambda_i}\right)^{\frac{1}{\kappa}}, & \text{untuk } y_i = 0 \\ (1-\pi_i)\frac{\Gamma\left(y_i+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y_i!}\left(\frac{1}{1+\kappa\lambda_i}\right)^{\frac{1}{\kappa}}\left(\frac{\kappa\lambda_i}{1+\kappa\lambda_i}\right)^{y_i}, & \text{untuk } y_i > 0 \end{cases} \quad (12)$$

dengan $0 \le \pi_i \le 1, \lambda_i \ge 0$, κ adalah parameter dispersi untuk $\frac{1}{\kappa} > 0$ dan $\Gamma(.)$ adalah fungsi gamma. Ketika $\pi_i = 0$, variabel acak y_i memiliki distribusi binomial negatif dengan rata-rata λ_i dan parameter dispersi κ, sehingga $Y_i \sim NB(\lambda_i, \kappa)$.

Pada GLM, fungsi penghubung logit digunakan ketika parameter model regresi bernilai 0 dan 1 dan fungsi penghubung log digunakan jika parameter diharapkan bernilai positif. Oleh karena itu fungsi penghubung yang tepat digunakan pada regresi ZINB adalah fungsi penghubung logit untuk parameter π_i dan fungsi penghubung log untuk parameter λ_i.

Model regresi ZINB dapat dinyatakan seperti pada Persamaan (13) dan (14).

Model untuk *negative binomial state* $\hat{\lambda}_i$

$$\ln \hat{\lambda}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij}, i = 1,2,\ldots,n \text{ dan } j = 1,2,\ldots,p \quad (13)$$

Model untuk *zero state* π_i

$$\text{Logit } \hat{\pi}_i = \ln\frac{\pi}{1-\pi} = \hat{\gamma}_0 + \sum_{j=1}^{p} \hat{\gamma}_j x_{ij}, i = 1,2,\ldots,n \text{ dan } j = 1,2,\ldots,p \quad (14)$$

dengan:

$\beta_0, \beta_1, \beta_2, \ldots \beta_p$: parameter model regresi ZINB yang diestimasi

$\gamma_0, \gamma_1, \gamma_2, \ldots \gamma_p$: parameter model regresi ZINB yang diestimasi

1.5.7 Metode Estimasi Parameter

1.5.7.1 Metode *Maximum Likelihood Estimation*

Metode *Maximum Likelihood Estimation* (MLE) dapat digunakan untuk mengestimasi parameter suatu model yang diketahui fungsi peluangnya. Misalkan $Y_1, Y_2, \ldots, Y_n$ adalah sampel acak dari populasi dengan fungsi peluang $f(y|\theta)$ yang bergantung pada $\boldsymbol{\theta} = \{\lambda, \boldsymbol{\beta}, \boldsymbol{\gamma}, \kappa\}$ dengan $\boldsymbol{\theta}$ adalah parameter yang tidak diketahui. Karena $Y_1, Y_2, \ldots, Y_n$ saling bebas, maka fungsi peluang bersama dari $Y_1, Y_2, \ldots, Y_n$ dapat dinyatakan sebagai berikut:

$$f(y_1, y_2, \ldots, y_n|\boldsymbol{\theta}) = f(y_1|\boldsymbol{\theta}) \cdot f(y_2|\boldsymbol{\theta}) \cdot \ldots \cdot f(y_n|\boldsymbol{\theta})$$

Fungsi *likelihood* didefinisikan sebagai fungsi peluang bersama dari $Y_1, Y_2, \dots, Y_n$ yang dianggap sebagai fungsi dari $\boldsymbol{\theta}$ yang dituliskan sebagai $L(\boldsymbol{\theta}|y)$, yaitu:

$$L(\boldsymbol{\theta}|y) = f(y_1, y_2, \dots, y_n|\boldsymbol{\theta}) = f(y_1|\boldsymbol{\theta}) \cdot f(y_2|\boldsymbol{\theta}) \cdot \dots \cdot f(y_n|\boldsymbol{\theta})$$

$$L(\boldsymbol{\theta}|y) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta}) \tag{15}$$

Estimator *maximum likelihood* $\widehat{\boldsymbol{\theta}}$ adalah nilai $\boldsymbol{\theta}$ yang memaksimumkan fungsi *likelihood* $L(\boldsymbol{\theta}|y)$. Namun, lebih mudah bekerja dengan log dari fungsi *likelihood*, yaitu $l(\boldsymbol{\theta}|y) = \ln L(\boldsymbol{\theta}|y)$. Karena fungsi log adalah fungsi yang monoton naik, maka nilai yang memaksimumkan fungsi $l(\boldsymbol{\theta}|y)$ sama dengan memaksimumkan fungsi $L(\boldsymbol{\theta}|y)$ sehingga dapat dituliskan sebagai berikut:

$$l(\boldsymbol{\theta}|y) = \ln L(\boldsymbol{\theta}|y) = \ln\left[\prod_{i=1}^{n} f(y_i|\boldsymbol{\theta})\right]$$

$$l(\boldsymbol{\theta}|y) = \sum_{i=1}^{n} \ln f(y_i|\boldsymbol{\theta}) \tag{16}$$

untuk memeroleh nilai $\boldsymbol{\theta}$ yang memaksimumkan fungsi $l(\boldsymbol{\theta}|y)$, maka $l(\boldsymbol{\theta}|y)$ diturunkan terhadap $\boldsymbol{\theta}$ dan kemudian menyamakannya dengan nol seperti pada Persamaan (17) berikut (Hogg dkk., 2012):

$$l'(\boldsymbol{\theta}|y) = \frac{\partial l(\boldsymbol{\theta}|y)}{\partial \boldsymbol{\theta}} = 0 \tag{17}$$

1.5.7.2 Algoritma *Expectation Maximization*

Algoritma *Expectation Maximization* (EM) digunakan untuk mengestimasi suatu parameter dengan memaksimumkan fungsi *likelihood* yang mengandung data hilang atau tidak lengkap (*incomplete data*). Algoritma ini terdiri dari dua tahap, yaitu tahap ekspektasi dan maksimalisasi. Misalnya, diasumsikan terdapat data pengamatan y berdistribusi tertentu yang mengandung data hilang z. Oleh karena itu, dibentuk distribusi antara y dan z, yaitu:

$$f(y, z|\boldsymbol{\theta}) = f(z) \cdot f(y|z) \tag{18}$$

Tahap ekspektasi (*E-Step*) pada algoritma EM dilakukan dengan menghitung ekspektasi dari fungsi log *likelihood* dari data hilang berdasarkan data pengamatan yang ada (tidak hilang), yang digunakan untuk mengganti keberadaan data yang dianggap hilang. Oleh karena itu, fungsi Q didefinisikan sebagai berikut:

$$Q\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(0)}; y\right) = E\left[\ln(\boldsymbol{\theta}^{(0)}|y, z)\right] \tag{19}$$

Pada tahap maksimalisasi (*M-Step*) dilakukan dengan mencari niali estimator yang dapat memaksimumkan fungsi Q yang telah didefinisikan pada tahap ekspektasi. Metode yang dapat digunakan untuk memaksimalkan fungsi Q adalah metode *Newton-Raphson* yang dilakukan secara numerik. Langkah Ekspektasi dan Maksimalisasi dilakukan secara iteratif sampai didapatkan estimator $\widehat{\boldsymbol{\theta}}$ yang konvergen, yaitu saat $\left\|\boldsymbol{\theta}^{(m+1)} - \boldsymbol{\theta}^{m}\right\| \leq \boldsymbol{\varepsilon}$, dengan $\boldsymbol{\varepsilon}$ adalah nilai galat yang sangat kecil (Purba, 2018).

1.5.7.3 Metode Iteratif Newton-Rapshon

Metode iteratif numerik *Newton Raphson* digunakan untuk proses estimasi parameter koefisien regresi sehingga diperoleh solusi dari fungsi log *likelihood*. Algoritma metode iteratif numerik *Newton Raphson* adalah sebagai berikut:

a. Menentukan taksiran awal untuk parameter $\widehat{\boldsymbol{\beta}}^{(0)}$ yang diperoleh dengan metode OLS, yaitu:
   $\widehat{\boldsymbol{\beta}}^{(0)} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$
b. Membentuk vektor gradien $\boldsymbol{g}$:
   $$\boldsymbol{g}^T\left(\widehat{\boldsymbol{\beta}}^{(m)}\right) = \left(\frac{\partial \ln L(\boldsymbol{\beta})}{\partial(\beta_0)}\ \frac{\partial \ln L(\boldsymbol{\beta})}{\partial(\beta_1)} \cdots \frac{\partial \ln L(\boldsymbol{\beta})}{\partial(\beta_p)}\right)_{\boldsymbol{\beta}=\boldsymbol{\beta}_{(m)}}$$
   dengan p adalah banyaknya parameter yang diestimasi.
c. Membentuk matriks Hessian ($\boldsymbol{H}$) yang simetris:
   $$\boldsymbol{H}\left(\widehat{\boldsymbol{\beta}}^{(m)}\right) = \begin{pmatrix} \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial^2(\beta_0)} & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial(\beta_0)\partial(\beta_1)} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial(\beta_0)\partial(\beta_0)_p} \\ & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial^2(\beta_1)} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial(\beta_1)\partial(\beta_p)} \\ & & \ddots & \vdots \\ & & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial^2(\beta_p)} \end{pmatrix}_{\boldsymbol{\beta}=\boldsymbol{\beta}_{(m)}}$$
d. Memasukkan nilai $\widehat{\boldsymbol{\beta}}^{(0)}$ ke dalam elemen-elemen vektor $\boldsymbol{g}$ dan matriks $\boldsymbol{H}$ sehingga diperoleh vektor $\boldsymbol{g}(\widehat{\boldsymbol{\beta}}^{(0)})$ dan matriks $\boldsymbol{H}(\widehat{\boldsymbol{\beta}}^{(0)})$.
e. Melakukan iterasi mulai dari $m = 0$ sebagaimana pada persamaan berikut:
   $\widehat{\boldsymbol{\beta}}^{(m+1)} = \widehat{\boldsymbol{\beta}}^{(m)} - \boldsymbol{H}^{-1}(\widehat{\boldsymbol{\beta}}^{(m)})\boldsymbol{g}(\widehat{\boldsymbol{\beta}}^{(m)})$

Proses iterasi akan berhenti jika telah diperoleh estimasi parameter yang konvergen dengan memenuhi $\left\|\widehat{\boldsymbol{\beta}}^{(m+1)} - \widehat{\boldsymbol{\beta}}^{(m)}\right\| \leq \boldsymbol{\varepsilon}$, untuk $\boldsymbol{\varepsilon}$ adalah nilai yang sangat kecil dan telah ditetapkan sebelumnya.

1.5.8 Estimasi Parameter Regresi *Zero Inflated Negative Binomial*

Estimasi parameter regresi *Zero Inflated Negative Binomial* (ZINB) menggunakan metode *Maximum Likelihood Estimation* (MLE). Metode ini biasanya digunakan untuk menduga parameter suatu model yang diketahui fungsi densitasnya. Fungsi peluang model regresi ZINB dapat dinyatakan seperti pada Persamaan (12) dan (Lambert, 1992) menyarankan model gabungan untuk λ dan π seperti pada Persamaan (13) dan Persamaan (14). Fajri, (2022) menyatakan model regresi ZINB berdasarkan model log dan logit diperoleh:

$$\lambda_i = \mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\beta}},\ \pi_i = \frac{\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}}{1+\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}},\ \text{dan } (1-\pi_i) = \frac{1}{1+\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\gamma}}} \tag{20}$$

Selanjutnya Persamaan (20) disubtitusikan ke Persamaan (12), diperoleh:

$$P(Y_i = y_i) = \begin{cases} \frac{\mathrm{e}^{x_i^T\gamma}}{1+\mathrm{e}^{x_i^T\gamma}} + \frac{1}{1+\mathrm{e}^{x_i^T\gamma}}\left(\frac{1}{1+\kappa\mathrm{e}^{x_i^T\beta}}\right)^{\frac{1}{\kappa}}, & \text{untuk } y_i = 0 \\ \frac{1}{1+\mathrm{e}^{x_i^T\gamma}}\frac{\Gamma\left(y_i+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y_i!}\left(\frac{1}{1+\kappa\mathrm{e}^{x_i^T\beta}}\right)^{\frac{1}{\kappa}}\left(\frac{1}{1+\kappa\mathrm{e}^{x_i^T\beta}}\right)^{y_i}, & \text{untuk } y_i > 0 \end{cases} \tag{21}$$

Misalkan diambil sampel $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ dari percobaan yang saling bebas. Fungsi *likelihood* dari Persamaan (21) adalah sebagai berikut:

$$L(\kappa, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \begin{cases} \prod_{i=1}^{n} \left\{ \frac{e^{x_i^T \gamma}}{1+e^{x_i^T \gamma}} + \frac{1}{1+e^{x_i^T \gamma}} \left(\frac{1}{1+\kappa e^{x_i^T \beta}} \right)^{\frac{1}{\kappa}} \right\}, & \text{untuk } y_i = 0 \\ \prod_{i=1}^{n} \left\{ \frac{1}{1+e^{x_i^T \gamma}} \frac{\Gamma\left(y_i + \frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right) y_i!} \left(\frac{1}{1+\kappa e^{x_i^T \beta}} \right)^{\frac{1}{\kappa}} \left(\frac{1}{1+\kappa e^{x_i^T \beta}} \right)^{y_i} \right\}, & \text{untuk } y_i > 0 \end{cases} \tag{22}$$

Sehingga fungsi log *likelihood* dari Persamaan (22) adalah:

$\ln L(\kappa, \boldsymbol{\beta}, \boldsymbol{\gamma}) =$

$$\sum_{i=1}^{n} \ln \left(e^{x_i^T \gamma} + \left(\frac{1}{1+\kappa e^{x_i^T \beta}} \right)^{\frac{1}{\kappa}} \right) - \sum_{i=1}^{n} \ln \left(1 + e^{x_i^T \gamma} \right), \quad \text{,untuk } y_i = 0$$

dan

$$- \sum_{i=1}^{n} \ln \left(1 + e^{x_i^T \gamma} \right) + \sum_{i=1}^{n} \ln \left(\Gamma \left(y_i + \frac{1}{\kappa} \right) \right) - \sum_{i=1}^{n} \ln \left(\Gamma \left(\frac{1}{\kappa} \right) \right) - \sum_{i=1}^{n} \ln(y_i!) + \frac{1}{\kappa} \sum_{i=1}^{n} \ln \left(\frac{1}{1+\kappa e^{x_i^T \beta}} \right) + y_i \sum_{i=1}^{n} \ln \left(\frac{\kappa e^{x_i^T \beta}}{1+\kappa e^{x_i^T \beta}} \right) \quad \text{,untuk } y_i > 0 \tag{23}$$

Penjumlahan fungsi log *likelihood* pada Persamaan (23) menghasilkan persamaan yang berbentuk implisit, sehingga fungsi log *likelihood* ini tidak dapat diselesaikan secara analitik. Persamaan (23) biasa disebut juga *incomplete* data *likelihood*. Oleh karena itu, digunakan algoritma *Expectation Maximization* (EM) untuk memaksimumkan fungsi log *likelihood* yang merupakan salah satu metode optimalisasi yang digunakan dalam memaksimumkan fungsi log *likelihood* yang mengandung data hilang.

Misalkan variabel $y_i (i = 1,2, \dots, n)$ berkaitan dengan variabel indikator z_i yaitu:

$$z_i = \begin{cases} 1 & \text{jika } y_i \text{ berasal dari } \textit{zero state} \\ 0 & \text{jika } y_i \text{ berasal dari } \textit{negative binomial state} \end{cases} \tag{24}$$

dengan $P(z_i = 1) = \pi_i$ dan $P(z_i = 0) = 1 - \pi_i$. Jika nilai variabel respon $y_i = 1,2, \dots$ maka nilai $z_i = 0$. Sedangkan jika variabel respon $y_i = 0$ maka nilai z_i belum dapat ditentukan atau dianggap hilang sebagian. Oleh karena itu, perlu dibentuk distribusi gabungan antara y dan z. Langkah-langkah estimasi parameter dengan algoritma EM sebagai berikut:

1. Menentukan distribusi dari variabel z_i:

$$P(z_i = 1) = \pi_i$$

$$P(z_i = 0) = P\left(Y \sim Negative\ Binomial\ (\lambda, \kappa)\right) = 1 - \pi_i$$

   Sehingga $z_i \sim Bernoulli\ (\pi_i), E(z_i) = \pi_i, Var\ (z_i) = \pi_i(1 - \pi_i)$

2. Membentuk distribusi gabungan antara y_i dan z_i yaitu

$$f(y, z | \pi, \lambda) = f(z) \cdot f\left(\frac{y}{z}\right) =$$

$$f(y, z | \pi, \lambda) = (\pi_i)^{z_i} (1 - \pi_i)^{(1-z_i)} \left(\frac{\Gamma\left(y_i + \frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right) y_i!} \left(\frac{1}{1+\kappa\lambda_i} \right)^{\frac{1}{\kappa}} \left(\frac{1}{1+\kappa\lambda_i} \right)^{y_i} \right)^{(1-z_i)} \tag{25}$$

   Kemudian dengan mensubtirusikan Persamaan (20) ke Persamaan (25) sehingga diperoleh persamaan log *likelihood*:

$\ln L(\kappa, \boldsymbol{\beta}, \boldsymbol{\gamma}|y, z) =$

$$\sum_{i=1}^{n} z_i x_i^T \gamma - \sum_{i=1}^{n} \ln\left(1 + \mathrm{e}^{x_i^T \gamma}\right) + \sum_{i=1}^{n} (1 - z_i) \ln\left(\frac{\prod_{b=0}^{y_i-1}\left(b + \frac{1}{k}\right)}{y_i!}\right)\left(\frac{1}{1 + \kappa \mathrm{e}^{x_i^T \beta}}\right)^{\frac{1}{k}}\left(\frac{\kappa \mathrm{e}^{x_i^T \beta}}{1 + \kappa \mathrm{e}^{x_i^T \beta}}\right)^{y_i} \quad (26)$$

Persamaan (26) biasa disebut *complete data likelihood*, Persamaan ini yang akan dimaksimumkan menggunakan algoritma EM, parameter $\kappa, \boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ dapat diestimasi secara terpisah menjadi:

$$\ln L((\kappa, \beta, \gamma|y, z) = \ln L((\kappa, \boldsymbol{\beta}|y, z) + \ln L((\boldsymbol{\gamma}|y, z)$$

dengan

$$\begin{aligned}\ln L((\kappa, \boldsymbol{\beta}|y, z) &= \sum_{i=1}^{n} (1 - z_i) \ln\left(\frac{\prod_{b=0}^{y_i-1}\left(b + \frac{1}{k}\right)}{y_i!}\right)\left(\frac{1}{1 + \kappa \mathrm{e}^{x_i^T \beta}}\right)^{\frac{1}{k}}\left(\frac{\kappa \mathrm{e}^{x_i^T \beta}}{1 + \kappa \mathrm{e}^{x_i^T \beta}}\right)^{y_i} \\ &= \sum_{i=1}^{n} (1 - z_i)\left(\sum_{b=0}^{y_i-1} \ln\left(b + \frac{1}{k}\right) - \ln(y_i!) + \right. \\ &\quad \left. y_i \ln\left(\kappa \mathrm{e}^{x_i^T \beta}\right) - \left(y_i + \frac{1}{k}\right) \ln\left(1 + \kappa \mathrm{e}^{x_i^T \beta}\right)\right) \qquad (27)\end{aligned}$$

$$\ln L((\boldsymbol{\gamma}|y, z) \quad = \sum_{i=1}^{n} z_i x_i^T \gamma - \sum_{i=1}^{n} \ln\left(1 + \mathrm{e}^{x_i^T \gamma}\right) \qquad (28)$$

Algortitma EM terdiri dari dua tahap, yaitu tahap ekspektasi (*E-Step*) dan tahap maksimalisasi (*M-Step*).

3. Tahap Ekspektasi (*E-Step*)

Mengganti variabel z_i dengan $z_i^{(m)}$ yang merupakan ekspektasi dari z_i dengan $m = 0,1,2, \ldots$

$$z_i^{(m)} = E(z_i|y_i, \boldsymbol{\beta}^{(m)}, \boldsymbol{\gamma}^{(m)} = \begin{cases} P(z_i = 1|y_i = 0, \boldsymbol{\beta}^{(m)}, \boldsymbol{\gamma}^{(m)} & \text{jika } y_i = 0 \\ 0 & \text{jika } y_i > 0 \end{cases}$$

dengan

$$P(z_i = 1|y_i = 0, \boldsymbol{\beta}^{(m)}, \boldsymbol{\gamma}^{(m)} = \frac{1}{1 + \mathrm{e}^{-x_i^T \gamma^{(m)}}\left(\frac{1}{1 + \kappa \mathrm{e}^{\boldsymbol{x}_i^T \boldsymbol{\beta}^{(m)}}}\right)^{\frac{1}{k}}}$$

Sehingga

$$Q\left(\boldsymbol{\beta}, \boldsymbol{\gamma}\middle|\boldsymbol{\beta}^{(m)}, \boldsymbol{\gamma}^{(m)}\right) = E\left[\ln L\left(\boldsymbol{\beta}^{(m)}, \boldsymbol{\gamma}^{(m)}\middle|y, z^{(m)}\right)\right] =$$

$$\sum_{i=1}^{n} \ln L\left(\boldsymbol{\beta}^{(m)}|y_i, z_i^{(m)}\right) + \sum_{i=1}^{n} \ln L\left(\boldsymbol{\gamma}^{(m)}|y_i, z_i^{(m)}\right)$$

Persamaan (27) dan (28) menjadi

$$\begin{aligned}\ln L\left(\boldsymbol{\beta}^{(m)}\middle|y_i, z_i^{(m)}\right) &= \sum_{i=1}^{n}\left(1 - z_i^{(m)}\right)\left(\sum_{b=0}^{y_i-1} \ln\left(b + \frac{1}{k}\right) - \ln(y_i!) + \right. \\ &\quad \left. y_i \ln\left(\kappa \mathrm{e}^{\boldsymbol{x}_i^T \boldsymbol{\beta}^{(m)}}\right) - \left(y_i + \frac{1}{k}\right) \ln\left(1 + \kappa \mathrm{e}^{\boldsymbol{x}_i^T \boldsymbol{\beta}^{(m)}}\right)\right) \qquad (29)\end{aligned}$$

$$\ln L(\left(\boldsymbol{\gamma}^{(m)}\middle|y_i, z_i^{(m)}\right) = \sum_{i=1}^{n} z_i^{(m)} \boldsymbol{x}_i^T \boldsymbol{\gamma}^{(m)} - \sum_{i=1}^{n} \ln\left(1 + \mathrm{e}^{\boldsymbol{x}_i^T \gamma^{(m)}}\right) \qquad (30)$$

4. Tahap Maksimalisasi (*M-Step*)

Tahap maksimalisasi meliputi maksimalisasi untuk parameter $\kappa, \boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$. Untuk proses maksimalisasi parameter digunakan metode iteratif numerik yaitu metode *Newton Rapshon.* Tahap maksimalisasi parameter $\kappa, \boldsymbol{\beta}$ dilakukan secara bersama-sama dengan κ (kappa) adalah parameter dispersi pada model regresi *Zero Inflated Negative Binomial* (ZINB). Algoritma metode iteratif numerik *Newton Rapshon* dengan taksiran awal untuk parameter dispersi κ dan parameter regresi $\boldsymbol{\beta}$ dinyatakan dalam parameter $\hat{\boldsymbol{\delta}}^{(0)}$ dengan $\boldsymbol{\delta} = (\kappa, \boldsymbol{\beta})$ dan taksiran awal untuk parameter $\boldsymbol{\gamma}$ dinyatakan dalam parameter $\hat{\boldsymbol{\gamma}}^{(0)}$.

Dengan menggunakan metode *Newton Rapshon* maka:

$$\hat{\boldsymbol{\delta}}^{(m+1)} = \hat{\boldsymbol{\delta}}^{(m)} - \boldsymbol{H}^{-1}\left(\hat{\boldsymbol{\delta}}^{(m)}\right)\boldsymbol{g}\left(\hat{\boldsymbol{\delta}}^{(m)}\right), m = 1,2, \dots, n$$

untuk $\boldsymbol{\delta}^{(0)}$ adalah nilai parameter awal yang diperoleh dari metode *Ordinary Least Square* (OLS), $\boldsymbol{H}$ adalah matriks Hessian yang berisikan turunan kedua dari $\ln L\left(\kappa, \boldsymbol{\beta}^{(m)} \middle| y_i, z_i^{(m)}\right)$ terhadap κ dan $\boldsymbol{\beta}$ serta turunan dari $\ln L\left(\kappa, \boldsymbol{\beta}^{(m)} \middle| y_i, z_i^{(m)}\right)$ terhadap $\kappa, \boldsymbol{\beta}$. Sedangkan $\boldsymbol{g}$ adalah vektor yang berisikan turunan pertama dari $\ln L\left(\kappa, \boldsymbol{\beta}^{(m)} \middle| y_i, z_i^{(m)}\right)$ terhadap κ dan $\boldsymbol{\beta}$.

$$\hat{\boldsymbol{\gamma}}^{(m+1)} = \hat{\boldsymbol{\gamma}}^{(m)} - \boldsymbol{H}^{-1}\left(\hat{\boldsymbol{\gamma}}^{(m)}\right)\boldsymbol{g}\left(\hat{\boldsymbol{\gamma}}^{(m)}\right), m = 1,2, \dots, n$$

untuk $\boldsymbol{\gamma}^{(0)}$ adalah nilai parameter awal yang diperoleh dari metode OLS, $\boldsymbol{H}$ adalah matriks Hessian yang berisikan turunan kedua dari $\ln L(\left(\boldsymbol{\gamma}^{(m)} \middle| y_i, z_i^{(m)}\right)$. Sedangkan $\boldsymbol{g}$ adalah vektor yang berisikan turunan pertama dari $\ln L(\left(\boldsymbol{\gamma}^{(m)} \middle| y_i, z_i^{(m)}\right)$.

Proses iterasi mulai dari $m = 0$ dan akan berhenti jika telah diperoleh estimasi parameter yang konvergen dengan memenuhi $\left\|\hat{\boldsymbol{\delta}}^{(m+1)} - \hat{\boldsymbol{\delta}}^{(m)}\right\| \leq \boldsymbol{\varepsilon}$ dan $\left\|\hat{\boldsymbol{\gamma}}^{(m+1)} - \hat{\boldsymbol{\gamma}}^{(m)}\right\| \leq \boldsymbol{\varepsilon}$, dengan $\boldsymbol{\varepsilon}$ adalah nilai yang sangat kecil dan telah ditetapkan sebelumya.

1.5.9 Efek Spasial

Efek spasial pada data dapat berupa *error* yang saling berkorelasi (dependensi spasial) maupun keragaman (heterogenitas) spasial antar lokasi. Pemodelan data spasial selalu melibatkan matriks bobot spasial. Maka dari itu diuraikan lebih lanjut mengenai efek spasial pada data dan penentuan matriks pembobot.

1.5.9.1 Dependensi Spasial

Dependensi spasial terjadi karena adanya dependensi atau ketergantungan nilai antar suatu wilayah (Mustika & Sulistyawan, 2019). Kondisi ini menunjukkan adanya kemiripan sifat untuk lokasi yang saling berdekatan atau terjadi autokorelasi antar residual di lokasi pengamatan (Anselin, 1988). Pengujian efek dependensi spasial adalah komponen penting dalam analisis spasial dan tidak dapat diabaikan karena dapat menimbulkan estimator yang bias dan kesimpulan yang tidak akurat. Dalam mengidentifikasi adanya efek ketergantungan spasial pada model dapat dilakukan dengan melakukan uji signifikansi *Morans'I* (Usali dkk., 2021). Hipotesis yang digunakan adalah sebagai berikut:

$H_0: I = 0$ (tidak ada dependensi spasial)

$H_1: I \neq 0$ (ada dependensi spasial)

Statistik uji *Morans'I* dirumuskan sebagai berikut:

$$Z_{hit} = \frac{\hat{I} - E(\hat{I})}{\sqrt{Var(\hat{I})}} \tag{31}$$

dengan

$\hat{I} = \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$; $E(\hat{I}) = -\frac{1}{n-1}$ dan $Var(\hat{I}) = \frac{n^2 S_1 - nS_2 + 3S_0^2}{(n^2-1)S_0^2} - \left[E(\hat{I})\right]^2$

$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$;

$S_1 = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(w_{ij} + w_{ji}\right)^2$;

$S_2 = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji}\right)^2$

Keterangan:

$\hat{I}$: nilai Indeks Moran

n : banyaknya lokasi pengamatan

x_i : nilai pada lokasi ke-i $(i = 1,2, \ldots, n)$

x_j : nilai pada lokasi ke-j $(j = 1,2, \ldots, n)$

$\bar{x}$: rata-rata nilai pengamatan pada semua wilayah

w_{ij} : bobot terstandarisasi antara lokasi ke-i dan lokasi ke-j

$E(\hat{I})$: nilai ekspektasi dari nilai Indeks Moran

$Var(\hat{I})$: nilai variansi dari nilai Indeks Moran

Dasar pengambilan keputusan adalah tolak H_0 jika $\left|Z(\hat{I})\right| > Z_{\alpha/2}$ atau jika p-value $< \alpha$ pada taraf signifikansi α yang berarti terdapat dependensi spasial dalam antar lokasi pengamatan (Wong & Lee, 2001).

1.5.9.2 Heterogenitas Spasial

Heterogenitas spasial disebabkan karena adanya perbedaan karakteristik dari pengamatan di lokasi yang satu dengan pengamatan di lokasi yang lain. Menurut Anselin (1988), efek spasial yang berupa heterogenitas spasial ini dapat diidentifikasi menggunakan pengujian *Breusch-Pagan* (*BP-Test*). Digambarkan dengan hipotesis sebagai berikut:

H_0: $\sigma_1^2(u_i, v_i) = \sigma_2^2(u_i, v_i) = \cdots = \sigma_n^2(u_i, v_i) = \sigma^2$ (variansi antar lokasi sama)

H_1: minimal ada satu $\sigma_i^2(u_i, v_i) \neq \sigma^2$ (variansi antar lokasi berbeda)

Statistik Uji *Breusch-Pagan*:

$$BP = \left(\frac{1}{2}\right) \boldsymbol{f}^T \boldsymbol{Z}(\boldsymbol{Z}^T \boldsymbol{Z})^{-1} \boldsymbol{Z}^T \boldsymbol{f} \tag{32}$$

dengan

$\boldsymbol{f}$: $(f_1, f_2, \ldots, f_n)^T$ untuk $f_i = \frac{\varepsilon_i^2}{\sigma^2} - 1$, ε_i^2 adalah kuadrat sisaan untuk pengamatan ke-i $(\varepsilon_i = y_i - \hat{y}_i)$ dan σ^2 adalah variansi dari y

$\boldsymbol{Z}$: matriks berukuran $n \times (p+1)$ yang berisi vektor yang telah distandarisasi untuk setiap pengamatan.

Tolak H_0 jika $BP > \mathcal{X}^2_{(\alpha,p)}$ atau *p-value* < α yang berarti variansi antar lokasi berbeda.

1.5.10 *Geographically Weighted Regression*

Model *Geographically Weighted Regression* (GWR) merupakan pengembangan dari model regresi. Model ini adalah model regresi linier lokal yang menghasilkan parameter model yang bersifat lokal untuk setiap titik atau lokasi data tersebut dikumpulkan. Asumsi yang digunakan pada model GWR adalah *error* berdistribusi normal dengan rata-rata sama dengan σ^2. Dalam model GWR, persamaan regresi digambarkan dengan y_i sebagai variabel respon dan $x_{i1}, x_{i2}, \dots, x_{ip}$ sebagai variabel prediktor. Model GWR menghasilkan penaksir parameter model yang bersifat lokal untuk setiap titik atau lokasi data tersebut yang diamati. Sehingga model GWR dapat ditulis sebagai berikut:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{p} \beta_j(u_i, v_i)x_{ij} + \varepsilon_i, \text{ dengan } i = 1,2, \dots, n \tag{33}$$

Keterangan:

y_i	: nilai respon untuk lokasi ke-i
x_{ij}	: nilai prediktor ke-j pada lokasi ke-i, $j = 1,2, \dots, p$
$\beta_0(u_i, v_i)$	: konstanta regresi untuk model GWR
$\beta_j(u_i, v_i)$	: koefisien regresi untuk prediktor ke-j
(u_i, v_i)	: koordinat geografis (lintang, bujur) pada lokasi ke-i
ε_i	: residual pada pengamatan ke-i. $\varepsilon \sim (0, \sigma^2)$.

Berdasarkan hukum pertama geografi, "Semua hal selalu berkaitan dengan sesuatu yang lain. Namun sesuatu yang lebih dekat lebih berpengaruh dibanding sesuatu yang jauh". Sehingga model GWR dalam menggambarkan hubungan nonstasioner menjadi:

$$y_i = w_0(u_i, v_i)\beta_{0_{OLS}} + \sum_{j=1}^{p} w_j(u_i, v_i)\beta_{j_{OLS}}x_{ij} + \varepsilon_i$$

Keterangan:

$w_0(u_i, v_i)$	: bobot dari konstanta $\beta_{0_{OLS}}$
$w_j(u_i, v_i)$	: bobot dari koefisien $\beta_{j_{OLS}}$

Dengan mensubstitusikan persamaan OLS dari pendugaa β_j, maka

$$\hat{y}_i = \sum_{j=1}^{p} \hat{\beta}_j(u_i, v_i)x_{ij} = \sum_{j=1}^{p} w_j(u_i, v_i)\hat{\beta}_j(OLS)x_{ij}$$

Apabila direpresentasikan ke dalam bentuk matriks menjadi:

$$\hat{y}_i = x_i^T\hat{\beta}_j(u_i, v_i) = x_i^T\boldsymbol{W}(u_i, v_i)(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

$\boldsymbol{W}(u_i, v_i)$ adalah matriks pembobot spasial sebagai berikut (Ismah dkk., 2020):

$$\boldsymbol{W}(u_i, v_i) = \begin{bmatrix} w_0(u_i, v_i) & 0 & \dots & 0 \\ 0 & w_0(u_i, v_i) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & w_0(u_i, v_i) \end{bmatrix}$$

1.5.10.1 Fungsi Pembobot Model *Geographically Weighted Regression*

Pemilihan pembobot pada model GWR sangat penting dilakukan karena besarnya nilai pembobot tersebut mewakili masing-masing lokasi data pengamatan yang berbeda. Untuk menentukan besarnya pembobot pada model GWR dapat dilakukan dengan menggunakan fungsi kernel (*kernel function*). Fungsi kernel digunakan untuk mengestimasi parameter dalam model GWR jika fungsi jarak (w_{ij}) adalah fungsi yang kontinu dan monoton turun (Nakaya dkk., 2005). Menurut Fotheringham dkk. (2002), pembobot yang terbentuk dengan menggunakan fungsi kernel adalah sebgai berikut:

1. Fungsi *Fixed*
   a. Fungsi *Fixed Gaussian Kernel*

$$w_{ij} = \exp\,[-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2 \qquad (34)$$

   Fungsi kernel gauss akan memberikan bobot yang semakin menurun mengikuti fungsi Gaussian ketika d_{ij}semakin besar.

   b. Fungsi *Fixed Bisquare Kernel*

$$w_{ij} = \begin{cases} \left[1-\left(\frac{d_{ij}}{h}\right)^2\right]^2, \text{untuk } d_{ij} \leq h \\ 0, \qquad\qquad\qquad \text{untuk } d_{ij} > h \end{cases} \qquad (35)$$

   Fungsi *fixed bisquare kernel* akan memberi bobot nol ketika j berada pada atau di luar radius h dari lokasi i, sedangkan apabila lokasi j berada di dalam radius h maka akan mengikuti fungsi *fixed bisquare kernel*.

2. Fungsi Adaptif
   a. Fungsi *Adaptive Gaussian Kernel*

$$w_{ij} = \exp\left[-\left(\frac{d_{ij}}{h_{i(q)}}\right)^2\right] \qquad (36)$$

   b. Fungsi *Adaptive Bisquare Kernel*

$$w_{ij} = \begin{cases} \left[1-\left(\frac{d_{ij}}{h_{i(q)}}\right)^2\right]^2, \text{untuk } d_{ij} < h_i \\ 0, \qquad\qquad\qquad \text{untuk } d_{ij} > h_i \end{cases} \qquad (37)$$

dengan $d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2}$ adalah jarak *euclide* antara lokasi (u_i, v_i) ke lokasi $\left(u_j, v_j\right)$, *h* adalah parameter penghalus (*bandwidth*), dan $h_{i(q)}$ adalah *bandwidth adaptive* yang menyatakan q sebagai jarak terdekat dari titik lokasi pengamatan.

1.5.10.2 Bandwidth

Bandwidth merupakan lingkaran dengan radius h dari titik pusat lokasi yang digunakan sebagai dasar menentukan bobot setiap pengamatan terhadap model regresi pada lokasi tersebut (Yulisti'anah, 2018). Setiap pengamatan yang berada di dalam lokasi ke-i masih dianggap berpengaruh terhadap model regresi yang terbentuk. Menurut Fotheringham dkk. (2002), pemilihan *bandwidth* optimum dapat dilakukan dengan menghitung nilai *Cross Validation* (CV) dengan rumus sebagi berikut:

$$CV(h) = \sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(h)]^2 \tag{38}$$

dengan $\hat{y}_{\neq i}(h)$ adalah nilai dugaan y_i dengan pengamatan di lokasi ke-i dihilangkan dari proses estimasi. Untuk mendapatkan h yang optimal, maka diperoleh dari h yang menghasilkan nilai CV minimum.

1.5.11 *Geographically Weighted Zero Inflated Negative Binomial Regression*

Analisis regresi yang digunakan pada data *count* yang variabel responnya mengandung nilai nol dalam proporsi besar (*excess zeros*) dan mengalami overdispersi yaitu regresi *Zero Inflated Negative Binomial* (ZINB). Namun, penerapan data *count* yang melibatkan faktor lokasi membutuhkan alternatif solusi. *Geographycally Weighted Zero Inflated Negative Binomial Regression* (GWZINBR) menjadi metode yang dapat digunakan dalam kasus ini. Model GWZINBR adalah suatu bentuk lokal dari model regresi ZINB yang akan menghasilkan penaksiran parameter model bersifat lokal untuk setiap lokasi.

Setiap pengamatan pada variabel respon diambil dari lokasi u_i, v_i yang berbeda yaitu $y_1, y_2, \dots, y_n$. Pada model GWZINBR variabel respon Y_i memiliki peluang berbeda untuk $y_i = 0$ dan $y_i > 0$, yaitu pada Persamaan (12). Sehingga didapatkan model regresi GWZINBR sebagai berikut:

$$\lambda_i = \exp\left(\boldsymbol{x}_i^T \boldsymbol{\beta}(u, \boldsymbol{v})\right) \tag{39}$$

$$\pi_i = \frac{\exp\left(\boldsymbol{x}_i^T \boldsymbol{\gamma}(u, \boldsymbol{v})\right)}{1 + \exp\left(\boldsymbol{x}_i^T \boldsymbol{\gamma}(u, \boldsymbol{v})\right)} \tag{40}$$

dan fungsi *likelihood* $L(\kappa, \boldsymbol{\beta}(u_i, v_i), \boldsymbol{\gamma}(u_i, v_i); (\boldsymbol{x}, \boldsymbol{y})) =$

$$\begin{cases} \prod_{i=1}^{n} \left(\frac{\exp\left(\boldsymbol{x}_i^T \boldsymbol{\gamma}(u_i, v_i)\right)}{1+\exp\left(\boldsymbol{x}_i^T \boldsymbol{\gamma}(u_i, v_i)\right)} + \frac{1}{1+\exp\left(\boldsymbol{x}_i^T \boldsymbol{\gamma}(u_i, v_i)\right)} \left(\frac{1}{1+\kappa \exp \boldsymbol{x}_i^T \boldsymbol{\beta}(u_i, v_i)} \right)^{\frac{1}{\kappa}} \right) & , y_i = 0 \\ \prod_{i=1}^{n} \left(\frac{1}{1+\exp\left(\boldsymbol{x}_i^T \boldsymbol{\gamma}(u_i, v_i)\right)} \frac{\Gamma\left(y_i + \frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right) y_i!} \left(\frac{1}{1+\kappa \exp \boldsymbol{x}_i^T \boldsymbol{\beta}(u_i, v_i)} \right)^{\frac{1}{\kappa}} \left(\frac{\kappa \exp \boldsymbol{x}_i^T \boldsymbol{\beta}(u_i, v_i)}{1+\kappa \exp \boldsymbol{x}_i^T \boldsymbol{\beta}(u_i, v_i)} \right)^{y_i} \right) & , y_i > 0 \end{cases}$$

κ sebagai parameter dispersi, $\boldsymbol{\beta}(u_i, v_i)$ dan $\boldsymbol{\gamma}(u_i, v_i)$ adalah vektor dari parameter regresi akan ditaksir terletak pada lokasi (u_i, v_i), berukuran $(p + 1) \times 1$. $\boldsymbol{X}$ adalah matriks yang berisi variabel prediktor berbeda serta berukuran $n \times (p + 1)$.

1.5.12 Uji Serentak Parameter Model *Geographically Weighted Zero Inflated Negative Binomial Regression*

Uji hipotesis yang pertama dilakukan adalah pengujian model secara serentak untuk menguji signifikansi pada seluruh parameter $\boldsymbol{\beta}(u_i, v_i)$ dan $\boldsymbol{\gamma}(u_i, v_i)$. Bentuk hipotesisnya:

$H_0: \beta_1(u_i, v_i) = \cdots = \beta_p(u_i, v_i) = \gamma_1(u_i, v_i) = \cdots = \gamma_p(u_i, v_i) = 0$

H_1: paling sedikit terdapat satu $\beta_j(u_i, v_i) \neq 0$ atau $\gamma_j(u_i, v_i) \neq 0$

$i = 1,2, \ldots, n; j = 1,2, \ldots, p$

Statistik uji yang digunakan yaitu uji rasio *likelihood* (deviasi). Nilai deviasi dibentuk dari parameter model penuh $\left(\Omega = \left(\boldsymbol{\beta}(u_i, v_i), \boldsymbol{\gamma}(u_i, v_i)\right)\right.$ dan model intersep $\left(\omega =, \gamma(u_i, v_i)\right)$. Deviasi terbentuk dari dua log *likelihood* dari $L(\widehat{\Omega})$ dan $L(\widehat{\omega})$ yakni:

$$G = -2 \ln \left[L \left(\frac{\widehat{\omega}}{\widehat{\Omega}} \right) \right] = 2 \left[\ln L(\widehat{\Omega}) - \ln L(\widehat{\omega}) \right] \sim \chi^2_{\alpha,p} \qquad (41)$$

G berdistribusi *Chi-Square* dengan p adalah jumlah parameter. Tolak H_0 jika nilai $G > \chi^2_{\alpha,p}$. Kesimpulan yang diperoleh adalah variabel prediktor secara bersama-sama memiliki pengaruh terhadap variabel respon (Kurniawan, 2017).

1.5.13 Uji Parsial Parameter Model *Geographically Weighted Zero Inflated Negative Binomial Regression*

Uji hipotesis yang kedua adalah pengujian parameter model secara parsial. Pengujian ini untuk mengetahui signifikansi masing-masing parameter $\boldsymbol{\beta}(u_i, v_i)$ dan $\boldsymbol{\gamma}(u_i, v_i)$. Bentuk hipotesis pengujian secara parsial pada kedua parameter model adalah sebagai berikut:

Untuk parameter $\boldsymbol{\beta}(u_i, v_i)$

$H_0: \beta_j(u_i, v_i) = 0 \qquad H_1: \beta_j(u_i, v_i) \neq 0 \; ; j = 1,2, \ldots, p$

Untuk parameter $\boldsymbol{\gamma}(u_i, v_i)$

$H_0: \gamma_j(u_i, v_i) = 0 \qquad H_1: \gamma_j(u_i, v_i) \neq 0 \; ; j = 1,2, \ldots, p$

Statistik uji yang digunakan yaitu uji Wald (W):

$$W = \left| \frac{\widehat{\beta}(u_i, v_i)}{se\left(\widehat{\beta}(u_i, v_i)\right)} \right| \text{ dan } W = \left| \frac{\widehat{\gamma}(u_i, v_i)}{se\left(\widehat{\gamma}(u_i, v_i)\right)} \right| \qquad (42)$$

Kriteria keputusan tolak H_0 jika $W > Z_{\alpha/2}$ artinya variabel prediktor memiliki pengaruh yang signifikan terhadap variabel respon (Fericha, 2016).

1.5.14 Multikolinearitas

Adanya korelasi antar variabel prediktor dalam model regresi linear atau yang biasa disebut dengan multikolenearitas, akan menyebabkan error yang besar pada pendugaan parameter regresi, untuk itu perlu dilakukan uji multikolinearitas. Menurut Hocking (1996) dapat diketahui melalui nilai koefisien korelasi *Pearson* (r_{il}) antar variabel prediktor yang lebih besar dari 0,95. Selain itu, adanya kasus multikolinearitas juga dapat diketahui melalui *Variance Inflation Factors* (VIF) yang bernilai lebih besar dari 10, nilai VIF yang dinyatakan dalam Persamaan (43).

$$VIF_i = \frac{1}{1 - R_i^2} \tag{43}$$

dengan R_i^2 adalah koefisien determinasi antara x_i dengan variabel prediktor lainnya. Jika terdapat kasus multikolinearitas maka untuk mengatasinya menggunakan metode tertentu seperti *Principal Component Analysis* (PCA), yaitu dengan membentuk komponen-komponen utama sebagai variabel prediktor baru yang merupakan kombinasi linear dari variabel-variabel prediktor sebelumnya.

1.5.15 Jumlah Kematian Bayi Akibat Demam

Kematian bayi adalah kematian yang terjadi antara saat setelah bayi lahir sampai bayi belum berusia tepat satu tahun. Banyak faktor yang dikaitkan dengan kematian bayi. Secara garis besar dari sisi penyebabnya, kematian bayi ada dua macam yaitu endogen dan eksogen. Kematian bayi endogen atau yang umum disebut kematian neonatal. Kematian bayi yang terjadi pada bulan pertama setelah dilahirkan, dan umumnya disebabkan oleh faktor-faktor yang dibawa anak sejak lahir, yang diperoleh dari orang tuanya pada saat konsepsi atau didapat selama kehamilan. Kematian bayi eksogen atau kematian bayi post neonatal, adalah kematian bayi yang terjadi setelah usia satu bulan sampai menjelang usia satu tahun yang disebabkan oleh faktor-faktor yang bertalian dengan pengaruh lingkungan luar.

Kematian bayi akibat demam termasuk dalam kematian bayi eksogen dengan berbagai faktor luar yang dapat mengakibatkannya. Beberapa yang menjadi perhatian adalah terkait persentase kejadian BBLR di Sulawesi Selatan Tahun 2020 sebesar 4,24% dari 154.733 kelahiran hidup. Menurut hasil Riskesdas tahun 2013 Persentase BBLR pada perempuan (14,5%) lebih tinggi daripada laki-laki (10,3%), namun persentase berat lahir ≥ 4000 gr pada laki-laki (6,1%) lebih tinggi dibandingkan perempuan (4,3%). Menurut pendidikan dan kuintil indeks kepemilikan terlihat adanya kecenderungan semakin tinggi pendidikan dan kuintil indeks kepemilikan, semakin rendah prevalensi BBLR

Pelayanan yang diberikan saat kunjungan neonatal adalah pemeriksaan sesuai standar Manajemen Terpadu Bayi Muda (MTBM) dan konseling perawatan bayi baru lahir termasuk ASI Eksklusif dan perawatan tali pusat. Pelayanan kesehatan bayi ditujukan pada bayi usia 29 hari-11 bulan yang memperoleh pelayanan kesehatan sesuai dengan standar oleh tenaga kesehatan yang memiliki kompetensi klinis kesehatan (dokter, bidan, dan perawat) minimal empat kali termasuk didalamnya penyuluhan ASI ekslusif.

Pimunisasi bayi mencakup vaksinasi BCG, DPT (3 kali), polio (4 kali), hepatitis-B (3 kali) dan imunisasi campak (1 kali), yang dilakukan melalui pelayanan rutin di posyandu dan fasilitas pelayanan kesehatan lainnya. Pencapaian *Universal Child Immunization* (UCI) pada dasarnya merupakan proyeksi terhadap cakupan sasaran bayi yang telah mendapatkan imunisasi secara lengkap. Suatu desa/ kelurahan telah mencapai target UCI apabila >80 % bayi di desa/ kelurahan tersebut mendapat imunisasi lengkap (Dinas Kesehatan Sulawesi Selatan, 2021).