

**PEMODELAN GEOGRAPHICALLY WEIGHTED BIVARIATE POISSON
INVERSE GAUSSIAN REGRESSION PADA JUMLAH
KASUS KUSTA DI PULAU SULAWESI**



**RIDWAN
H051201050**



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

**PEMODELAN *GEOGRAPHICALLY WEIGHTED BIVARIATE POISSON*
INVERSE GAUSSIAN REGRESSION PADA JUMLAH
KASUS KUSTA DI PULAU SULAWESI**

**RIDWAN
H051201050**



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

**PEMODELAN GEOGRAPHICALLY WEIGHTED BIVARIATE POISSON
INVERSE GAUSSIAN REGRESSION PADA JUMLAH
KASUS KUSTA DI PULAU SULAWESI**

RIDWAN
H051201050



Skripsi

Diajukan sebagai salah satu syarat untuk mencapai gelar Sarjana Statistika

Program Studi Statistika

pada

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

SKRIPSI

**PEMODELAN GEOGRAPHICALLY WEIGHTED BIVARIATE POISSON
INVERSE GAUSSIAN REGRESSION PADA JUMLAH
KASUS KUSTA DI PULAU SULAWESI**

**RIDWAN
H051201050**

Skripsi,

telah dipertahankan di hadapan Panitia Ujian Sarjana Statistika pada 14 Agustus
2024 dan dinyatakan telah memenuhi syarat kelulusan

pada

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

Mengesahkan:
Pembimbing tugas akhir,

Andi Kresna Jaya, S.Si., M.Si.
NIP. 19731228 200003 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 Bivariate Poisson Inverse Gaussian Regression* pada Jumlah Kasus Kusta di Pulau Sulawesi" adalah benar karya saya dengan arahan pembimbing skripsi saya (Andi Kresna Jaya, S.Si., M.Si.). 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 perbuatan tersebut berdasarkan aturan yang berlaku.

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

Makassar, 14 Agustus 2024



Ridwan
NIM. H051201050

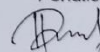
UCAPAN TERIMA KASIH

Bismillahirrahmanirrahim. Alhamdulillah Alhamdulillah'aalamiin puji syukur penulis panjatkan kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat, taufiq, hidayah, dan inayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul "**Pemodelan Geographically Weighted Bivariate Poisson Inverse Gaussian Regression pada Jumlah Kasus Kusta di Pulau Sulawesi**". Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa Sallam*, penerima wahyu al-Qur'anul karim yang senantiasa mengandung mukjizat di segala zaman. Semoga kita termasuk golongan yang mendapat syafaatnya *ila yaumil qiyamah, amin*.

Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada Bapak **Andi Kresna Jaya, S.Si., M.Si.**, selaku Pembimbing Utama yang senantiasa meluangkan waktu dalam kesibukannya untuk memberikan kritik-korektif sekaligus konstruktif terhadap proses pemikiran, penataan, dan penyelesaian skripsi ini. Terima kasih kepada Ibu **Dra. Nasrah Sirajang, M.Si.** dan Bapak **Drs. Raupong, M.Si.**, selaku Tim Penguji atas saran dan kritikan yang berharga serta waktu yang telah diberikan kepada penulis. Terima kasih juga kepada Pimpinan Universitas Hasanuddin beserta Jajarannya, Ketua Departemen Statistika, Jajaran Dosen dan Staf Departemen Statistika atas ilmu yang bermanfaat dan fasilitas selama penulis menempuh studi.

Penghargaan setinggi-tingginya dan ucapan terima kasih penulis haturkan kepada orang tua tercinta, Ayahanda **Syarifuddin S.** dan Ibunda **Maraddia** yang selalu berjuang dalam mengupayakan yang terbaik untuk penulis dan telah memberikan dukungan, pengorbanan, kasih sayang, serta doa yang senantiasa mengiringi langkah penulis. Terima kasih kepada saudara penulis beserta keluarga besar lainnya yang selalu memotivasi, memberikan semangat, dukungan, dan doa mulia kepada penulis. Terima kasih juga kepada teman-teman Statistika 2020 terkhususnya dela, afa dan ed serta semua pihak yang tidak dapat disebutkan satu-persatu. 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,



Ridwan

ABSTRAK

RIDWAN. **Pemodelan *Geographically Weighted Bivariate Poisson Inverse Gaussian Regression* pada Jumlah Kasus Kusta di Pulau Sulawesi** (dibimbing oleh Andi Kresna Jaya, S.Si., M.Si.).

Latar belakang. *Generalized Linear Model* (GLM) merupakan perluasan model regresi linier klasik untuk memodelkan data diskrit yang tidak berdistribusi normal, salah satunya regresi *Poisson*. Dalam regresi *Poisson* sering terjadi pelanggaran asumsi yaitu overdispersi yang dapat diatasi dengan model *mixed Poisson* seperti regresi *Poisson Inverse Gaussian* (PIG). Sedangkan untuk sepasang data diskrit *Poisson* teroverdispersi dapat dimodelkan dengan regresi *Bivariate Poisson Inverse Gaussian* (BPIG). Selain itu, jika data menunjukkan adanya heterogenitas spasial maka dapat dimodelkan dengan *Geographically Weighted Bivariate Poisson Inverse Gaussian Regression* (GWBPIGR). Salah satu data diskrit yang diasumsikan berdistribusi *Poisson* adalah jumlah kasus kusta yang diklasifikasikan oleh *World Health Organization* (WHO) menjadi dua tipe yaitu *Paucibacillary* (PB) dan *Multibacillary* (MB). Data jumlah kasus kusta di Pulau Sulawesi tahun 2021 mengalami peningkatan sebesar 8.32% dari tahun sebelumnya sehingga dianalisis menggunakan model GWBPIGR. **Tujuan.** Penelitian ini bertujuan untuk memperoleh estimasi parameter model GWBPIGR dan faktor-faktor yang memengaruhi jumlah kasus kusta di Pulau Sulawesi menggunakan model tersebut. **Metode.** Penelitian ini meliputi dua tahap yaitu mengestimasi parameter model GWBPIGR menggunakan metode *maximum likelihood estimation* (MLE) dengan algoritma *Berndt-Hall-Hall-Hausman* (BHHH) dan studi kasus model GWBPIGR pada data jumlah kasus kusta di Pulau Sulawesi. **Hasil.** Pemodelan GWBPIGR menggunakan pembobot *adaptive bisquare kernel* menghasilkan penduga parameter yang bersifat lokal dan berbeda-beda tiap kabupaten/kota. Pengujian signifikansi menghasilkan variabel signifikan yang berbeda-beda pula tiap kabupaten/kota. **Kesimpulan.** Faktor-faktor yang berpengaruh signifikan terhadap jumlah kasus kusta PB terbagi menjadi 3 kelompok sedangkan terhadap jumlah kasus kusta MB terbagi menjadi 2 kelompok.

Kata kunci: *Poisson*, Overdispersi, Heterogenitas Spasial, *Geographically Weighted Bivariate Poisson Inverse Gaussian Regression*, Jumlah Kasus Kusta.

ABSTRACT

RIDWAN. **Geographically Weighted Bivariate Poisson Inverse Gaussian Regression Modelling on the Number of Leprosy Cases in Sulawesi Island** (supervised by Andi Kresna Jaya, S.Si., M.Si.).

Introduction. Generalized Linear Model (GLM) is an extension of the classical linear regression model used to model discrete data that is non-normally distributed, such as Poisson regression. In Poisson regression, overdispersion often occurs, which can be overcome by mixed Poisson models such as Poisson Inverse Gaussian Regression (PIGR). For overdispersed Poisson discrete data pairs, it can be modeled with Bivariate Poisson Inverse Gaussian Regression (BPIGR). Additionally, if the data shows spatial heterogeneity, it can be modeled with Geographically Weighted Bivariate Poisson Inverse Gaussian Regression (GWBPIGR). One of the discrete data assumed to follow a Poisson distribution is the number of leprosy cases which has been classified by the World Health Organization (WHO) into two types, Paucibacillary (PB) and Multibacillary (MB). In 2021, the number of leprosy cases in Sulawesi Island increased by 8.32% from the previous year so it is analyzed using GWBPIGR model. **Objectives.** This research aims to estimate the parameter of GWBPIGR model and identify factors affecting the number of leprosy cases in Sulawesi Island using that model. **Methods.** This research includes two stages, estimating the parameters of GWBPIGR model using maximum likelihood estimation (MLE) method with Berndt-Hall-Hausman (BHHH) algorithm and a case study on the number of leprosy cases in Sulawesi Island. **Results.** GWBPIGR modeling using adaptive bisquare kernel weights produces parameter estimates that vary across regencies/cities. The significance testing results in different significant variables for each regencies/cities. **Conclusion.** The factors significantly influencing the PB type are divided into three groups, while those affecting the MB type are divided into two groups.

Keywords: Poisson, Overdispersion, Spatial Heterogeneity, Geographically Weighted Bivariate Poisson Inverse Gaussian Regression, The Number of Leprosy Cases.

DAFTAR ISTILAH

Istilah	Arti dan Penjelasan
Algoritma	Prosedur yang digunakan untuk menghitung data.
Apusan kulit	Pengambilan sampel dari permukaan kulit untuk pemeriksaan mikrobiologis.
<i>Bandwidth</i>	Parameter penghalus untuk mengetahui seberapa jauh pengaruh lokasi pengamatan saling berdekatan.
Basil	Bakteri berbentuk batang.
Cepat-asam	Sifat bakteri yang dapat menghasilkan asam dengan cepat.
Diskrit	Jenis data yang hanya dapat mengambil nilai bilangan bulat tertentu yang diperoleh dari hasil perhitungan.
Dispersi	Ukuran variabilitas dari suatu distribusi.
Distribusi	Penyebaran data.
Eksplisit	Kondisi suatu parameter atau variabel tidak dapat dipisahkan secara langsung dengan komponen lain dalam persamaan.
Equidispersi	Kondisi yang memperlihatkan ragam variabel respon dan rataannya bernilai sama pada regresi <i>Poisson</i> .
Fagosit	Sel dalam sistem kekebalan tubuh yang memiliki kemampuan untuk mencerna partikel asing, seperti bakteri, virus, sel-sel mati, dan sisa-sisa seluler lainnya.
Famili	Tingkatan dalam sistem klasifikasi makhluk hidup yang dikelompokkan berdasarkan kemiripan karakteristik yang lebih spesifik daripada ordo.
Fungsi kepadatan peluang	Fungsi probabilitas variabel kontinu.
Fungsi <i>likelihood</i>	Fungsi yang menggambarkan probabilitas data yang diobservasi berdasarkan parameter model.
Fungsi massa peluang	Fungsi probabilitas variabel diskrit.
Fungsi penghubung	fungsi yang menghubungkan nilai ekspektasi variabel respon dengan kombinasi linier dari variabel prediktor.
Gram-positif	Golongan bakteri yang mempertahankan zat warna kristal violet sewaktu proses pewarnaan Gram.
Heterogenitas spasial	Variasi atau keragaman antar lokasi pengamatan.
Implisit	Kondisi suatu parameter atau variabel dapat dipisahkan secara langsung dengan komponen lain dalam persamaan.
Independen	Saling bebas atau tidak saling bergantung.
Intraseluler obligat	Bakteri yang hanya dapat tumbuh dan bereproduksi di dalam sel organisme inang.

Istilah	Arti dan Penjelasan
Iterasi	Proses pengulangan langkah-langkah dalam sebuah algoritma hingga mencapai hasil konvergen.
Kontinu	Jenis data yang mengambil nilai dalam suatu rentang yang diperoleh dari hasil pengukuran.
Koordinat geografis	Sistem yang digunakan untuk menentukan lokasi suatu titik di permukaan bumi berdasarkan garis lintang dan bujur.
Korelasi	Ukuran statistik yang menunjukkan seberapa kuat dan arah hubungan antara dua variabel.
<i>Least squares</i>	Metode untuk mengestimasi parameter dalam model regresi dengan meminimumkan jumlah kuadrat residual.
Lesi kulit	Perubahan struktur atau fungsi kulit.
Linier	Hubungan yang berbentuk garis lurus antara dua variabel.
<i>Mean</i>	Nilai rata-rata.
Mukosa	Jaringan lembab yang melapisi organ-organ tertentu dan rongga tubuh di seluruh tubuh.
Multikolinieritas	Kondisi dimana terdapat korelasi antar variabel prediktor.
Multivariat	Analisis statistik yang melibatkan lebih dari dua variabel.
Ordo taksonomi	Tingkatan dalam sistem klasifikasi makhluk hidup yang dikelompokkan berdasarkan kemiripan karakteristik morfologis dan filogenetik.
Overdispersi	Kondisi yang memperlihatkan ragam variabel respon yang lebih besar dari rataannya pada regresi <i>Poisson</i>
<i>Overestimate</i>	Nilai perkiraan lebih besar daripada nilai sebenarnya.
<i>Over-smoothing</i>	Kondisi nilai terlalu besar yang menyebabkan model yang dihasilkan terlalu halus.
Parameter	Nilai menggambarkan karakteristik dari suatu populasi.
Parsial	Hanya sebagian dari keseluruhan.
Pencilan	Pengamatan yang secara signifikan berbeda atau terpaut jauh dari sebagian besar data lainnya.
Rasio	Perbandingan antara dua kuantitas.
Regresi	Metode statistika untuk menentukan hubungan antara satu variabel respon dengan satu atau lebih variabel prediktor.
Saraf tepi	Sistem saraf yang terdiri dari serabut-serabut saraf yang menghubungkan sistem saraf pusat dengan organ tubuh lainnya.
<i>Satellites</i>	Keberadaan kepiting tapal kuda jantan di sarang betina.
Simpangan baku/ standar deviasi	Ukuran statistik yang menggambarkan seberapa besar penyebaran suatu data dari nilai rata-ratanya.
Simultan	Secara keseluruhan.
<i>Standard error</i>	Ukuran kesalahan standar dari suatu estimasi.
Standarisasi	Transformasi data sehingga memiliki rata-rata nol dan variansi satu.

Istilah	Arti dan Penjelasan
Underdispersi	Kondisi yang memperlihatkan ragam variabel respon yang lebih kecil dari rataannya pada regresi <i>Poisson</i> .
<i>Underestimate</i>	Nilai perkiraan lebih kecil daripada nilai sebenarnya.
<i>Under-smoothing</i>	Kondisi nilai terlalu kecil yang menyebabkan model yang dihasilkan terlalu kasar.
Variabel	Karakteristik yang dapat diukur atau diobservasi.
Variansi	Ukuran penyebaran data menyimpang dari nilai rata-ratanya.

DAFTAR SINGKATAN DAN LAMBANG

Lambang/singkatan	Arti dan Penjelasan
α	“alpha” taraf signifikansi
β	“beta” parameter regresi
ε	“epsilon” galat/ <i>error</i>
θ	“theta” parameter
λ, σ^2	“lambda”, “sigma kuadrat” parameter skala
μ	“mu” parameter lokasi
π	“pi” bilangan konstanta
ρ	“rho” koefisien korelasi
τ	“tau” parameter dispersi
ϕ	“phi” nilai dispersi
Ω, ω	“omega” himpunan parameter
U, V, X, Y, v, x, y	Variabel acak
$f(\cdot), h(\cdot)$	Fungsi probabilitas
$g(\cdot)$	Fungsi penghubung
$\log(\cdot)$	logaritma
$\ln(\cdot)$	logaritma natural
$e, \exp(\cdot)$	eksponen/basis dari logaritma natural
$E(\cdot)$	Ekspektasi
$Var(\cdot)$	Variansi
$Cov(\cdot)$	Kovariansi
$\min\{\cdot\}$	minimum
$K_s(z)$	Fungsi <i>Bessel</i>
se	<i>standard error</i>
χ^2	Distribusi <i>chi</i> -kuadrat
H_0	Hipotesis nol
H_1	Hipotesis alternatif
$p - value$	<i>probability-value</i>
iid	<i>Independent and identically distributed</i>
a, c, q, s, z	Komponen fungsi <i>Bessel</i>
b, d, h, w	Komponen matriks pembobot
k, p	Jumlah variabel prediktor
n	Jumlah pengamatan
r	Bilangan konstanta
t, Z	Nilai statistik uji
BP	<i>Breusch-Pagan</i>
$D(\cdot), G^2$	<i>Deviance</i>
∂	“do” turunan parsial
!	Faktorial
T	<i>Transpose</i> matriks

Lambang/singkatan	Arti dan Penjelasan
<i>tr</i>	<i>Trace matriks</i>
AIDS	<i>Acquired Immunodeficiency Syndrome</i>
BB	<i>Mid-Borderline leprosy</i>
BHHH	<i>Berndt-Hall-Hall-Hausman</i>
BL	<i>Borderline Lepromatous leprosy</i>
BPIG	<i>Bivariate Poisson Inverse Gaussian</i>
BT	<i>Borderline Tuberculoid leprosy</i>
COVID-19	<i>coronavirus disease 2019</i>
CV	<i>cross-validation</i>
GCV	<i>Generalized cross-validation</i>
GWBPIGR	<i>Geographically Weighted Bivariate Poisson Inverse Gaussian Regression</i>
GWPR	<i>Geographically Weighted Poisson Regression</i>
GWR	<i>Geographically Weighted Regression</i>
HIV	<i>Human immunodeficiency virus</i>
IG	<i>Inverse Gaussian</i>
LL	<i>Lepromatous leprosy</i>
MB	<i>Multibacillary</i>
MLE	<i>maximum likelihood estimation</i>
MLRT	<i>Maximum Likelihood Ratio Test</i>
PB	<i>Paucibacillary</i>
PIG	<i>Poisson Inverse Gaussian</i>
TT	<i>Tuberculoid leprosy</i>
VIF	<i>Variance Inflation Factor</i>
WHO	<i>World Health Organization</i>

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGANTAR	ii
HALAMAN PENGESAHAN	Error! Bookmark not defined.
PERNYATAAN KEASLIAN SKRIPSI DAN PELIMPAHAN HAK CIPTA	Error! Bookmark not defined.
UCAPAN TERIMA KASIH	Error! Bookmark not defined.
ABSTRAK	xi
ABSTRACT	xiii
DAFTAR ISTILAH	xv
DAFTAR SINGKATAN DAN LAMBANG	xix
DAFTAR ISI	xxi
DAFTAR TABEL	xxiii
DAFTAR GAMBAR	xxv
DAFTAR LAMPIRAN	xxvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Batasan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
1.5 Teori	4
BAB II METODOLOGI PENELITIAN	17
2.1 Sumber Data	17
2.2 Identifikasi Variabel	17
2.3 Metode Analisis	18
BAB III HASIL DAN PEMBAHASAN	21
3.1 Estimasi Parameter Model <i>Geographically Weighted Bivariate Poisson Inverse Gaussian Regression</i>	21
3.2 Penerapan Model <i>Geographically Weighted Bivariate Poisson Inverse Gaussian Regression</i> pada Jumlah Kasus Kusta di Pulau Sulawesi	27
BAB IV KESIMPULAN DAN SARAN	43
4.1 Kesimpulan	43
4.2 Saran	44
DAFTAR PUSTAKA	45
LAMPIRAN	49

DAFTAR TABEL

Tabel	Halaman
1. Definisi operasional variabel penelitian.....	17
2. Statistik deskriptif.....	27
3. Korelasi antara jumlah kasus kusta PB dan kusta MB.....	29
4. Uji overdispersi.....	30
5. Nilai VIF variabel prediktor.....	30
6. Estimasi dan uji signifikansi parameter model regresi BPIG.....	31
7. Uji heterogenitas spasial.....	32
8. Nilai GCV pembobot spasial fungsi <i>kernel</i>	33
9. Variabel prediktor yang berpengaruh signifikan pada model GWBPIGR tiap kabupaten/kota.....	35
10. Pengelompokan kabupaten/Kota berdasarkan variabel prediktor yang berpengaruh signifikan terhadap jumlah kasus kusta PB.....	37
11. Pengelompokan kabupaten/Kota berdasarkan variabel prediktor yang berpengaruh signifikan terhadap jumlah kasus kusta MB.....	39
12. Estimasi dan uji signifikansi parameter model GWBPIGR di Kabupaten Bone Provinsi Sulawesi Selatan.....	40
13. Estimasi dan uji signifikansi parameter model GWBPIGR di Kabupaten Konawe Utara Provinsi Sulawesi Tenggara.....	41

DAFTAR GAMBAR

Gambar	Halaman
1. <i>Framework for risk factors in Leprosy determinant</i>	16
2. Peta sebaran jumlah kasus kusta PB (a) dan kusta MB (b) di Pulau Sulawesi tahun 2021.....	28
3. Peta pengelompokan kabupaten/kota berdasarkan variabel prediktor yang berpengaruh signifikan pada jumlah kasus kusta PB.....	38
4. Peta pengelompokan kabupaten/kota berdasarkan variabel prediktor yang berpengaruh signifikan pada jumlah kasus kusta MB	39

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Tabel integral fungsi <i>Bessel</i>	51
2. Data penelitian dan koordinat geografis kabupaten/kota	52
3. Peta variabel prediktor	54
4. <i>Output</i> pengujian distribusi <i>Poisson</i>	57
5. <i>Output</i> pengujian korelasi variabel respon	58
6. <i>Output</i> pendeteksian overdispersi.....	59
7. <i>Output</i> pengujian multikolinieritas	60
8. <i>Output</i> estimasi parameter dan uji signifikansi parameter model regresi <i>Bivariate Poisson Inverse Gaussian</i>	61
9. <i>Output</i> pengujian heterogenitas spasial.....	62
10. Jarak <i>euclidean</i> antar lokasi.....	63
11. <i>Bandwidth optimum</i>	64
12. Matriks pembobot spasial <i>adaptive bisquare kernel</i>	66
13. Hasil estimasi parameter model <i>Geographically Weighted Bivariate Poisson Inverse Gaussian Regression</i>	67
14. Hasil uji signifikansi parameter model <i>Geographically Weighted Bivariate Poisson Inverse Gaussian Regression</i>	71

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan bentuk analisis yang digunakan untuk mengetahui hubungan antara variabel prediktor dengan variabel respon. Secara umum, analisis regresi yang sering digunakan adalah regresi linier klasik dengan variabel responnya berupa data kontinu yang mengikuti distribusi normal. Namun, beberapa kasus sering dijumpai variabel respon berupa data diskrit dan tidak berdistribusi normal sehingga regresi linier klasik kurang tepat dalam memodelkan data tersebut. Dalam situasi seperti ini, *Generalized Linear Model* (GLM) dapat digunakan (Sinharay, 2010). GLM merupakan perluasan dari model regresi linier klasik dimana variabel responnya dapat berdistribusi selain normal tetapi termasuk dalam keluarga eksponensial. GLM memiliki tiga komponen yaitu komponen acak, komponen sistematis, dan fungsi penghubung. Salah satu distribusi dalam keluarga eksponensial adalah distribusi *Poisson* yang merupakan distribusi yang menggambarkan peristiwa yang probabilitas kejadiannya kecil yang bergantung pada interval waktu tertentu atau di suatu daerah tertentu (Agresti, 2007).

Pendekatan dengan model regresi *Poisson* merupakan model regresi non-linier yang menggambarkan adanya hubungan antara variabel respon diskrit dan berdistribusi *Poisson* dengan variabel prediktor (Agresti, 2007). Sedangkan jika sepasang variabel respon berdistribusi *Poisson* yang saling berkorelasi dapat dimodelkan dengan regresi *Bivariate Poisson*. Dalam regresi *Poisson*, salah satu asumsi yang harus dipenuhi adalah equidispersi, yaitu nilai rata-rata dan variansi dari variabel respon harus bernilai sama. Namun dalam analisis data sering dijumpai data diskrit memperlihatkan variansi yang lebih besar dari rata-ratanya (*overdispersi*) atau variansi lebih kecil dari rata-ratanya (*underdispersi*). Kasus *overdispersi* lebih sering terjadi karena adanya keragaman dalam variabel respon serta adanya korelasi positif antar variabel respon yang terjadi pada analisis multivariat (Hilbe, 2011).

Menurut Cahyandari (2014), penyebab lain *overdispersi* yaitu banyaknya pengamatan bernilai nol, sumber keragaman antar individu tidak teramati, terdapat pengamatan yang hilang, adanya pencilan, atau kesalahan spesifikasi fungsi penghubung. *Overdispersi* dapat menyebabkan simpangan baku estimasi parameter *underestimate* dan uji signifikansi parameter *overestimate* sehingga kesimpulan yang diperoleh tidak valid (Ismail & Jemain, 2007). Untuk mengatasi masalah ini, para peneliti telah mengusulkan beberapa model dari perpaduan antara distribusi *Poisson* dengan distribusi kontinu yang disebut distribusi *mixed Poisson*. Distribusi ini digunakan untuk memodelkan data diskrit berdistribusi *Poisson* yang memiliki variansi yang sangat besar (*extra-Poisson variation*). Salah satu yang termasuk distribusi *mixed Poisson* adalah distribusi *Poisson Inverse Gaussian* (PIG) (Dean et al., 1989).

Distribusi PIG pertama kali diperkenalkan oleh Holla pada tahun 1966 yang merupakan campuran dari distribusi *Poisson* dan distribusi *inverse Gaussian* serta

mengandung parameter dispersi (Fogarty, 2015). Putri et al. (2020) telah melakukan penelitian dengan membandingkan performa model PIG dan model *Negative Binomial* pada data jumlah *satellites* belangkas. Selain itu Matonyo et al. (2020) juga melakukan penelitian menggunakan model PIG pada data *Salmonella Hadar* di Jerman. Penelitian-penelitian tersebut menghasilkan model PIG lebih baik dibandingkan dengan model *Negative Binomial* dalam mengatasi data teroverdispersi. Data diskrit berdistribusi *Bivariate Poisson* dan mengalami overdispersi dapat dimodelkan menggunakan regresi *Bivariate Poisson Inverse Gaussian* (BPIG). Wijaya (2017) telah melakukan penelitian menggunakan regresi BPIG pada data jumlah kasus baru HIV dan AIDS di Kabupaten Trenggalek dan Ponorogo. Namun pada penelitian tersebut kurang valid karena data yang digunakan merupakan data yang mengandung informasi geografis sehingga memungkinkan datanya dipengaruhi oleh efek spasial.

Menurut Anselin (1988), efek spasial berkaitan dengan data yang memiliki komponen lokasi atau efek lokasi yang berarti pengamatan pada suatu lokasi bergantung pada pengamatan di lokasi lain. Salah satu efek spasial adalah keragaman atau heterogenitas spasial yang menunjukkan adanya variasi atau keragaman antar lokasi. Metode statistika untuk memodelkan data yang mempertimbangkan heterogenitas spasial adalah *Geographically Weighted Regression* (GWR). Pemodelan GWR dilakukan pembobotan spasial sesuai dengan kedekatan antar lokasi pengamatan dan menghasilkan estimasi parameter yang berbeda-beda tiap lokasi pengamatan (Fotheringham et al., 2002). Jika data variabel respon merupakan data diskrit spasial berdistribusi *Poisson* maka dimodelkan dengan *Geographically Weighted Poisson Regression* (GWPR). Huang et al. (2020) telah melakukan penelitian pada data resiko COVID-19 di Hong Kong dengan hasil bahwa model GWPR lebih baik dibandingkan dengan model global regresi *Poisson*. Penelitian tersebut menghasilkan model yang baik namun tidak dapat mengatasi masalah overdispersi. Oleh karena itu, untuk mengatasi sepasang data diskrit berkorelasi dan berdistribusi *Poisson* yang mengalami overdispersi serta dipengaruhi heterogenitas spasial, digunakan model *Geographically Weighted Bivariate Poisson Inverse Gaussian Regression* (GWBPIGR).

Model GWBPIGR merupakan perluasan dari model regresi BPIG yang mempertimbangkan aspek spasial dengan menggunakan pembobot spasial dalam estimasi parameternya (Arniva et al., 2018). Untuk mengestimasi parameter, metode *maximum likelihood estimation* (MLE) dapat digunakan dengan memaksimumkan fungsi *likelihood* dari distribusi. Namun beberapa fungsi *likelihood* tidak dapat diselesaikan secara analitik menggunakan metode MLE karena diperoleh bentuk implisit sehingga harus diselesaikan secara iteratif, salah satunya dengan algoritma *Berndt-Hall-Hall-Hausman* (BHHH) yang merupakan pengembangan dari metode *Newton-Raphson*.

Salah satu data diskrit yang diasumsikan berdistribusi *Poisson* adalah data jumlah kasus kusta. *World Health Organization* (WHO) mengklasifikasikan penyakit kusta menjadi dua yaitu tipe *Paucibacillary* (PB) dan tipe *Multibacillary* (MB). Indonesia sendiri merupakan salah satu negara dengan jumlah kasus kusta tertinggi

di dunia terutama di tahun 2021 dilaporkan hampir mencapai 11 ribu kasus baru yang 89% diantaranya merupakan tipe MB. Walaupun angka tersebut menurun dari tahun sebelumnya, namun beberapa wilayah masih mengalami peningkatan seperti di Pulau Sulawesi dengan peningkatan sebesar 8.32% dari tahun sebelumnya (Kemenkes RI, 2022). Berdasarkan uraian tersebut, penulis tertarik untuk mengkaji model GWBPIGR yang diterapkan pada data jumlah kasus kusta di Pulau Sulawesi dengan menganalisis faktor-faktor geografisnya yang diduga berpengaruh. Penulis menyusun penelitian ini dengan judul “Pemodelan *Geographically Weighted Bivariate Poisson Inverse Gaussian Regression* pada Jumlah Kasus Kusta di Pulau Sulawesi”.

1.2 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Ruang lingkup penelitian adalah semua kabupaten/kota di Pulau Sulawesi.
2. Perhitungan jarak antar kabupaten/kota menggunakan jarak *euclidean* dan pembobotan spasial menggunakan pembobot dengan GCV minimum (*adaptive bisquare kernel*).
3. Metode estimasi parameter model *Geographically Weighted Bivariate Poisson Inverse Gaussian Regression* menggunakan metode *maximum likelihood estimation* dengan algoritma *Berndt-Hall-Hall-Hausman*.

1.3 Tujuan Penelitian

Adapaun tujuan dari penelitian adalah sebagai berikut.

1. Memperoleh estimasi parameter model *Geographically Weighted Bivariate Poisson Inverse Gaussian Regression* pada jumlah kasus kusta di Pulau Sulawesi, khususnya di kabupaten/kota dengan kasus kusta tertinggi dan terendah.
2. Memperoleh faktor-faktor yang memengaruhi jumlah kasus kusta di Pulau Sulawesi menggunakan model *Geographically Weighted Bivariate Poisson Inverse Gaussian Regression*.

1.4 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah dapat memperluas pemahaman teoritis dan praktis terkait model *Geographically Weighted Bivariate Poisson Inverse Gaussian Regression* terutama dalam estimasi parameter. Selain itu, penelitian ini dapat menjadi bahan informasi bagi masyarakat maupun instansi pemerintah terutama di Pulau Sulawesi dalam mengevaluasi upaya penurunan jumlah kasus kusta.

1.5 Teori

1.5.1 Generalized Linear Model

Generalized Linear Model (GLM) adalah perluasan dari model regresi linier klasik dimana variabel responnya dapat berdistribusi selain normal tetapi termasuk dalam keluarga eksponensial (McCullagh & Nelder, 1989). Menurut Agresti (2007), GLM memiliki tiga komponen yaitu:

1. Komponen Acak

Komponen acak dari GLM mengidentifikasi variabel respon Y beserta distribusi peluangnya. Misalkan suatu pengamatan pada Y dengan (Y_1, \dots, Y_n) , maka GLM menyatakan (Y_1, \dots, Y_n) saling bebas.

2. Komponen Sistematis

Komponen sistematis dari GLM menentukan variabel penjelas yang dimasukkan secara linier sebagai variabel prediktor di sisi kanan persamaan model, dengan kata lain menentukan variabel x_j pada persamaan berikut.

$$\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} \quad (1)$$

Kombinasi linier dari variabel penjelas ini disebut dengan *linear predictor*.

3. Fungsi Penghubung (*link function*)

Fungsi penghubung dari GLM menentukan sebuah fungsi $g(\cdot)$ yang menghubungkan μ dengan prediktor linier sebagai berikut.

$$g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} = \beta_0 + \sum_{p=1}^k \beta_p x_{ip} \quad (2)$$

Fungsi penghubung $g(\cdot)$ menggabungkan komponen acak dan komponen sistematis. Fungsi penghubung yang paling sederhana adalah $g(\mu) = \mu$ yang memodelkan rata-rata secara langsung yang disebut *identity link*. Fungsi penghubung lainnya memungkinkan μ berhubungan secara non-linier dengan variabel prediktor seperti $g(\mu) = \ln(\mu)$ yang memodelkan logaritma natural dari rata-rata yang disebut *log link*.

1.5.2 Regresi Poisson

Regresi *Poisson* merupakan model regresi non-linier yang menggambarkan adanya hubungan antara variabel respon berupa data diskrit dan berdistribusi *Poisson* dengan variabel prediktor (Agresti, 2007). Distribusi *Poisson* adalah suatu distribusi yang menggambarkan peristiwa yang probabilitas kejadiannya kecil yang bergantung pada interval waktu tertentu atau di suatu daerah tertentu. Jika variabel acak diskrit Y berdistribusi *Poisson* dengan parameter $\mu > 0$ yang merupakan rata-rata suatu kejadian (Arkandi & Winahju, 2015). Sehingga fungsi massa peluang distribusi *Poisson* adalah sebagai berikut (Walpole et al., 2011).

$$f(y; \mu) = \frac{e^{-\mu} \mu^y}{y!}, y = 0, 1, 2, \dots \quad (3)$$

dengan nilai *mean* dan variansi distribusi *Poisson* sama yaitu $E(Y) = Var(Y) = \mu$.

Regresi *Poisson* menggunakan fungsi penghubung $\ln(\cdot)$ dari GLM agar modelnya dapat digunakan dalam data pengamatan tanpa mengharuskan variabel responnya berdistribusi normal. Oleh karena itu, model regresi *Poisson* dapat dituliskan sebagai berikut (Cahyandari, 2014).

$$\begin{aligned} g(\mu_i) &= \ln(\mu_i) \\ \ln(\mu_i) &= \mathbf{x}_i^T \boldsymbol{\beta} \\ \mu_i &= \exp(\mathbf{x}_i^T \boldsymbol{\beta}); i = 1, 2, \dots, n \end{aligned} \quad (4)$$

dengan

μ_i menunjukkan rata-rata variabel respon berdistribusi *Poisson* pengamatan ke- i .

$\mathbf{x}_i^T = [1 \ x_{i1} \ x_{i2} \ \dots \ x_{ik}]$ menunjukkan vektor dari variabel prediktor yang berukuran $1 \times (k + 1)$ pada pengamatan ke- i .

$\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_k]^T$ menunjukkan vektor dari parameter regresi yang berukuran $(k + 1) \times 1$.

1.5.3 Regresi *Bivariate Poisson*

Regresi *Bivariate Poisson* adalah metode yang digunakan untuk memodelkan sepasang data diskrit berdistribusi *Poisson* yang saling berkorelasi dan diduga secara bersama-sama dengan variabel prediktor. Sepasang variabel respon yang saling berkorelasi mengikuti distribusi *Bivariate Poisson*. Distribusi *Bivariate Poisson* terjadi apabila variabel acak U_0, U_1, U_2 saling bebas yang masing-masing berdistribusi *Poisson* dengan parameter μ_0, μ_1, μ_2 . Terdapat variabel acak Y_1 dan Y_2 seperti berikut (AlMuhayfith et al., 2016).

$$\begin{aligned} Y_1 &= U_1 + U_0 \\ Y_2 &= U_2 + U_0 \end{aligned}$$

Maka variabel acak Y_1 dan Y_2 secara bersama-sama berdistribusi *Bivariate Poisson* dengan fungsi massa peluang bersamanya sebagai berikut (Kawamura, 1973).

$$f(y_1, y_2) = e^{-(\mu_0 + \mu_1 + \mu_2)} \sum_{r=0}^{\min(y_1, y_2)} \frac{\mu_1^{y_1-r} \mu_2^{y_2-r} \mu_0^r}{(y_1-r)! (y_2-r)! r!}; y_1, y_2 \geq 0 \quad (5)$$

dengan r adalah bilangan yang membuat nilai non-negatif bagi $y_1 - r; y_2 - r; r$. Nilai *mean* dan variansi dari variabel acak Y_1 dan Y_2 bernilai sama yaitu $E(Y_1) = Var(Y_1) = \mu_1 + \mu_0$ dan $E(Y_2) = Var(Y_2) = \mu_2 + \mu_0$.

Model regresi *Bivariate Poisson* mensyaratkan kedua variabel respon berdistribusi *Poisson*. Salah satu pendekatan untuk menguji distribusi *Bivariate Poisson* pada variabel respon Y_1 dan Y_2 adalah dengan *index of dispersion test*. Hipotesis dan statistik uji yang digunakan sebagai berikut (Best, 1999).

$H_0: F(y) = F_0(y)$ untuk Y_1 dan Y_2 (variabel respon Y_1 dan Y_2 mengikuti distribusi *Bivariate Poisson*).

$H_1: F(y) \neq F_0(y)$ untuk Y_1 dan Y_2 (variabel respon Y_1 dan Y_2 tidak mengikuti distribusi *Bivariate Poisson*).

$$I_B = \frac{n(\bar{Y}_2 S_{Y_1}^2 - 2m_{11}^2 + \bar{Y}_1 S_{Y_2}^2)}{(\bar{Y}_1 \bar{Y}_2 - m_{11}^2)} \sim \chi^2(2n - 3) \quad (6)$$

dengan $S_{Y_1}^2 = \frac{\sum_{i=1}^n (Y_{i1} - \bar{Y}_1)^2}{n}$, $S_{Y_2}^2 = \frac{\sum_{i=1}^n (Y_{i2} - \bar{Y}_2)^2}{n}$, dan $m_{11} = \frac{\sum_{i=1}^n (Y_{i1} - \bar{Y}_1)(Y_{i2} - \bar{Y}_2)}{n}$. Kriteria pengambilan keputusan dalam *index of dispersion test* adalah menolak H_0 jika $|I_B| > \chi_{(\alpha; 2n-3)}^2$ pada taraf signifikansi α .

Nilai koefisien korelasi antara variabel respon Y_1 dan Y_2 berkisar antara -1 sampai 1. Nilai $\rho = \pm 1$ hanya terjadi ketika $\sigma^2 = 0$, dalam hal ini terjadi hubungan linier sempurna antara dua variabel, sedangkan nilai yang mendekati nol menunjukkan sedikit atau tidak ada korelasi keduanya. Korelasi antara Y_1 dan Y_2 dapat diuji menggunakan korelasi *Pearson* dengan hipotesis dan statistik uji sebagai berikut (Walpole et al., 2011).

$H_0: \rho_{Y_1 Y_2} = 0$ (tidak terdapat korelasi antara variabel respon Y_1 dan Y_2).

$H_1: \rho_{Y_1 Y_2} \neq 0$ (terdapat korelasi antara variabel respon Y_1 dan Y_2).

$$t = \frac{\rho_{Y_1 Y_2} \sqrt{n-2}}{\sqrt{1 - \rho_{Y_1 Y_2}^2}} \quad (7)$$

dengan $\rho_{Y_1 Y_2} = \frac{Cov(Y_1, Y_2)}{\sqrt{Var(Y_1)Var(Y_2)}} = \frac{\sum_{i=1}^n (y_{i1} - \bar{y}_1)(y_{i2} - \bar{y}_2)}{\sqrt{\sum_{i=1}^n (y_{i1} - \bar{y}_1)^2 \sum_{i=1}^n (y_{i2} - \bar{y}_2)^2}}$. Kriteria pengambilan

keputusan dalam uji korelasi *Pearson* adalah menolak H_0 jika $|t_{hitung}| > t_{(\alpha/2; n-2)}$ pada taraf signifikansi α .

Parameter μ_j dihubungkan dengan variabel prediktor menggunakan fungsi penghubung $\ln(\cdot)$. Sehingga model regresi *Bivariate Poisson* adalah sebagai berikut.

$$\begin{aligned} \ln(\mu_{ij} + \mu_0) &= \mathbf{x}_i^T \boldsymbol{\beta}_j \\ \mu_{ij} + \mu_0 &= \exp(\mathbf{x}_i^T \boldsymbol{\beta}_j); i = 1, 2, \dots, n; j = 1, 2 \end{aligned} \quad (8)$$

dengan

μ_{ij} menunjukkan rata-rata variabel respon ke- j pada pengamatan ke- i .

$\mathbf{x}_i^T = [1 \ x_{i1} \ x_{i2} \ \dots \ x_{ik}]$ menunjukkan vektor variabel prediktor yang berukuran $1 \times (k+1)$ pada pengamatan ke- i .

$\boldsymbol{\beta}_j = [\beta_{0j} \ \beta_{1j} \ \beta_{2j} \ \dots \ \beta_{kj}]^T$ menunjukkan vektor parameter regresi yang berukuran $(k+1) \times 1$ pada variabel respon ke- j .

1.5.4 Overdispersi

Salah satu asumsi yang harus dipenuhi dalam regresi *Poisson* adalah nilai rata-rata dan variansi dari variabel respon harus bernilai sama yang disebut dengan equidispersi. Namun, dalam analisis data sering dijumpai kasus dimana variansi dari variabel respon lebih besar dari rata-ratanya yang disebut overdispersi. Sebaliknya, jika variansinya lebih kecil dari rata-rata variabel respon disebut underdispersi (McCullagh & Nelder, 1989). Kasus overdispersi lebih sering terjadi karena adanya keragaman dalam variabel respon atau adanya korelasi positif antar variabel respon yang terjadi pada analisis multivariat (Hilbe, 2011).

Menurut Cahyandari (2014), penyebab lain overdispersi seperti banyaknya pengamatan yang bernilai nol (*excess zero*), adanya sumber keragaman antar

individu yang tidak teramati (*unobserved heterogeneity*), terdapat pengamatan yang hilang (*data missing*), adanya pencilan (*outlier*) pada data, atau kesalahan spesifikasi fungsi penghubung. Jika terjadi overdispersi, nilai simpangan baku dari estimasi parameter jauh lebih kecil dari nilai sebenarnya (*underestimate*) dan uji signifikansi dari parameter jauh lebih besar dari nilai sebenarnya (*overestimate*) sehingga model regresi *Poisson* tidak efisien dengan nilai *standard error* yang tinggi (Ismail & Jemain, 2007).

Overdispersi pada variabel respon Y_1 dan Y_2 dapat dideteksi menggunakan uji *Deviance* dengan hipotesis dan statistik uji sebagai berikut (McCullagh & Nelder, 1989).

$H_0: \phi = 1$ (variabel respon Y_1 dan Y_2 tidak mengalami overdispersi).

$H_1: \phi > 1$ (variabel respon Y_1 dan Y_2 mengalami overdispersi).

$$\hat{\phi} = \frac{D}{(n - k)} \quad (9)$$

dengan $D(y; \mu) = 2 \sum_{i=1}^n \left(y_i \ln \left(\frac{y_i}{\mu_i} \right) - (y_i - \mu_i) \right)$. Kriteria pengambilan keputusan dalam uji *Deviance* adalah menolak H_0 jika $\hat{\phi} > 1$.

1.5.5 Regresi *Poisson Inverse Gaussian*

Regresi *Poisson Inverse Gaussian* (PIG) adalah metode dari salah satu alternatif distribusi *mixed Poisson* yang digunakan untuk memodelkan data diskrit berdistribusi *Poisson* yang memiliki variansi yang sangat besar (*extra-Poisson variation*). Distribusi PIG merupakan campuran dari distribusi *Poisson* dan distribusi *Inverse Gaussian* yang mengandung parameter dispersi (Fogarty, 2015). Distribusi *Inverse Gaussian* dengan parameter μ sebagai parameter lokasi dan λ sebagai parameter bentuk atau dinotasikan dengan $v \sim IG(\mu, \lambda)$ memiliki fungsi kepadatan peluang sebagai berikut (Folks & Chhikara, 1978).

$$f_{IG}(v; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi v^3}} \exp\left(-\frac{\lambda(v - \mu)^2}{2\mu^2 v}\right); 0 < v < \infty; \mu > 0; \lambda > 0 \quad (10)$$

dengan nilai *mean* dan variansinya adalah $E(V) = \mu$ dan $Var(V) = \frac{\mu^3}{\lambda}$.

Misalkan Y adalah variabel respon yang berdistribusi PIG dengan parameter *mean* μ dan parameter dispersi τ atau dinotasikan dengan $Y \sim PIG(\mu, \tau)$, maka fungsi kepadatan peluang bagi Y dapat dihitung sebagai berikut (Zha et al., 2016).

$$f_{PIG}(y; \mu, \tau) = \int_0^{\infty} f(y; \mu, v) h(v; \tau) dv \quad (11)$$

dengan

$$f(y; \mu, v) = \frac{e^{-\mu v} (v\mu)^y}{y!} \quad (12)$$

dan $h(v; \tau)$ merupakan fungsi kepadatan peluang dari V yang bersifat independen dan mengikuti distribusi *Inverse Gaussian* dengan $\mu = 1$ dan $\lambda = 1/\tau$ atau dinotasikan dengan $V \sim IG(1, 1/\tau)$ dan fungsi kepadatan peluangnya sebagai berikut.

$$h_{IG}(v; \tau) = (2\pi\tau v^3)^{-1/2} \exp\left(-\frac{(v-1)^2}{2\tau v}\right); v \geq 0 \quad (13)$$

dengan nilai *mean* dan variansinya adalah $E(V) = 1$ dan $Var(V) = \tau$. Sehingga fungsi kepadatan peluang bagi Y dapat dihitung sebagai berikut.

$$\begin{aligned} f_{PIG}(y; \mu, \tau) &= \int_0^{\infty} f(y; \mu, v)h(v; \tau)dv \\ &= \int_0^{\infty} \frac{e^{-\mu v}(\mu v)^y}{y!} (2\pi\tau v^3)^{-1/2} \exp\left(-\frac{(v-1)^2}{2\tau v}\right) dv \\ &= \frac{\mu^y (2\pi\tau)^{-1/2} e^{1/\tau}}{y!} \int_0^{\infty} \exp\left[-\left(\mu + \frac{1}{2\tau}\right)v - \left(\frac{1}{2\tau}\right)v^{-1}\right] v^{y-3/2} dv \end{aligned}$$

Berdasarkan tabel integral oleh Gardshteyn and Ryzhik (1980) dalam (Shoukri et al., 2004) pada Lampiran 1 bahwa:

$$\int_0^{\infty} \exp(-ax^h - cx^{-q}) x^{s-1} dx = \left(\frac{2}{q}\right) \left(\frac{c}{a}\right)^{\frac{s}{2q}} K_{\frac{s}{q}}(2\sqrt{ac}) \quad (14)$$

maka,

$$\int_0^{\infty} \exp\left[-v\left(\mu + \frac{1}{2\tau}\right) - \frac{1}{2\tau v}\right] v^{y-3/2} dv = 2 \left(\frac{1}{2\mu\tau + 1}\right)^{\frac{y-1/2}{2}} K_{y-\frac{1}{2}}\left(2\sqrt{\frac{2\mu\tau + 1}{4\tau^2}}\right)$$

Sehingga diperoleh

$$\begin{aligned} f_{PIG}(y; \mu, \tau) &= \frac{\mu^y (2\pi\tau)^{-1/2} e^{1/\tau}}{y!} 2 \left(\frac{1}{2\mu\tau + 1}\right)^{\frac{y-1/2}{2}} K_{y-\frac{1}{2}}\left(2\sqrt{\frac{2\mu\tau + 1}{4\tau^2}}\right) \\ &= \frac{\mu^y e^{1/\tau}}{y!} \left(\frac{2}{\pi\tau}\right)^{1/2} \frac{\left(\frac{1}{\tau}\sqrt{1+2\mu\tau}\right)^{1/2}}{\left(\frac{\tau}{\tau}\sqrt{1+2\mu\tau}\right)^y} K_{y-\frac{1}{2}}\left(\frac{1}{\tau}\sqrt{1+2\mu\tau}\right) \\ &= \left(\frac{2z}{\pi}\right)^{1/2} \frac{\mu^y e^{1/\tau} K_s(z)}{(z\tau)^y y!}; y \geq 0 \end{aligned} \quad (15)$$

dengan $s = y - \frac{1}{2}$ dan $z = \sqrt{\frac{1}{\tau^2} + \frac{2\mu}{\tau}}$. Sehingga, $K_s(z) = K_{y-\frac{1}{2}}\left(\frac{1}{\tau}\sqrt{1+2\mu\tau}\right)$ yang merupakan jenis ketiga dari modifikasi fungsi *Bessel*. Nilai *mean* dan variansi dari distribusi PIG adalah $E(Y) = \mu$ dan $Var(Y) = \mu + \tau\mu^2$. Dalam Shoukri et al. (2004),

fungsi *Bessel* memiliki sifat-sifat diantaranya $K_{\frac{1}{2}}(a) = K_{-\frac{1}{2}}(a) = \left(\frac{\pi}{2a}\right)^{1/2} e^{-a}$ dan

$$K_{\frac{3}{2}}(a) = \left(1 + \frac{1}{a}\right) K_{\frac{1}{2}}(a). \text{ Sehingga, } \frac{K_{y+\frac{1}{2}}(a)}{K_{y-\frac{1}{2}}(a)} = (1 + 2\lambda\mu)^{1/2} M(y) \text{ dengan } M(y) = (y + 1) \frac{p(y+1)}{p(y)}.$$

Variabel Y pada model PIG merupakan bilangan bulat non-negatif $(0, \infty)$, sehingga parameter μ dihubungkan dengan variabel prediktor menggunakan fungsi

penghubung $\ln(\cdot)$. Sehingga model regresi PIG adalah sebagai berikut (Herindrawati et al., 2017).

$$\begin{aligned}\ln(\mu_i) &= \mathbf{x}_i^T \boldsymbol{\beta} \\ \mu_i &= \exp(\mathbf{x}_i^T \boldsymbol{\beta}); i = 1, 2, \dots, n\end{aligned}\quad (16)$$

dengan

$\mathbf{x}_i^T = [1 \ x_{i1} \ x_{i2} \ \dots \ x_{ik}]$ menunjukkan vektor variabel prediktor berukuran $1 \times (k + 1)$ pada pengamatan ke- i .

$\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_k]^T$ menunjukkan vektor dari parameter regresi yang berukuran $(k + 1) \times 1$.

1.5.6 Regresi *Bivariate Poisson Inverse Gaussian*

Regresi *Bivariate Poisson Inverse Gaussian* (BPIG) merupakan metode yang digunakan untuk memodelkan dua variabel respon berkorelasi yang berdistribusi *Poisson* dan mengalami overdispersi. Misalkan Y_1 dan Y_2 adalah dua variabel acak berdistribusi *Poisson* yang saling bebas dengan *mean* $E(Y_1) = v\mu_1$ dan $E(Y_2) = v\mu_2$ serta variansinya $Var(Y_1) = \mu_1 + \mu_1^2\tau$ dan $Var(Y_2) = \mu_2 + \mu_2^2\tau$. Variabel acak V mengikuti distribusi *Inverse Gaussian* yang bersifat independen (Mardalena et al., 2021). Distribusi BPIG didasarkan pada distribusi campuran *Bivariate Poisson* dengan $r = 0$ dan *Inverse Gaussian* yang memiliki fungsi kepadatan peluang bersama sebagai berikut.

$$f_{BPIG}(y_j; \mu_j, \tau; j = 1, 2) = \int_0^\infty \prod_{j=1}^2 f(y_j; \mu_j, v) h(v; \tau) dv \quad (17)$$

dengan

$$f(y_j; \mu_j, v; j = 1, 2) = \frac{e^{-\mu_j v} (\mu_j v)^{y_j}}{y_j!} \quad (18)$$

dan $h(v; \tau)$ merupakan fungsi kepadatan peluang pada Persamaan (13). Maka fungsi kepadatan peluang bersamanya dapat dihitung sebagai berikut.

$$\begin{aligned}f_{BPIG}(y_j; \mu_j, \tau; j = 1, 2) &= \int_0^\infty \prod_{j=1}^2 f(y_j; \mu_j, v) h(v; \tau) dv \\ &= \int_0^\infty \prod_{j=1}^2 \frac{e^{-\mu_j v} (\mu_j v)^{y_j}}{y_j!} (2\pi\tau v^3)^{-1/2} \exp\left(-\frac{(v-1)^2}{2\tau v}\right) dv \\ &= \frac{\mu_1^{y_1} \mu_2^{y_2} (2\pi\tau)^{-1/2} e^{1/\tau}}{y_1! y_2!} \int_0^\infty \exp\left[-\left(\mu_1 + \mu_2 + \frac{1}{2\tau}\right)v - \left(\frac{1}{2\tau}\right)v^{-1}\right] v^{y_1+y_2-3/2} dv\end{aligned}$$

Berdasarkan tabel integral oleh Gardshteyn and Ryzhik (1980) dalam (Shoukri et al., 2004) pada Persamaan (14) pada Lampiran 1 maka,

$$\int_0^\infty \exp\left[-\left(\mu_1 + \mu_2 + \frac{1}{2\tau}\right)v - \left(\frac{1}{2\tau}\right)v^{-1}\right] v^{y_1+y_2-\frac{3}{2}} dv$$

$$= 2 \left(\frac{1}{2\tau(\mu_1 + \mu_2) + 1} \right)^{\frac{y_1 + y_2 - 1/2}{2}} K_{y_1 + y_2 - \frac{1}{2}} \left(2 \sqrt{\frac{2\tau(\mu_1 + \mu_2) + 1}{4\tau^2}} \right)$$

Sehingga diperoleh

$$\begin{aligned} f_{BPIG}(y_j; \mu_j, \tau; j = 1, 2) &= \frac{\mu_1^{y_1} \mu_2^{y_2} (2\pi\tau)^{-1/2} e^{1/\tau}}{y_1! y_2!} 2 \left(\frac{1}{2\tau(\mu_1 + \mu_2) + 1} \right)^{\frac{y_1 + y_2 - 1/2}{2}} K_{y_1 + y_2 - \frac{1}{2}} \left(2 \sqrt{\frac{2\tau(\mu_1 + \mu_2) + 1}{4\tau^2}} \right) \\ &= \frac{\mu_1^{y_1} \mu_2^{y_2} e^{1/\tau}}{y_1! y_2!} \left(\frac{2}{\pi\tau} \right)^{1/2} \left(\frac{1}{\tau} \sqrt{1 + 2\tau(\mu_1 + \mu_2)} \right)^{1/2} \left(\frac{1}{\tau} \sqrt{1 + 2\tau(\mu_1 + \mu_2)} \right)^{y_1 + y_2} K_{y_1 + y_2 - \frac{1}{2}} \left(\frac{1}{\tau} \sqrt{1 + 2\tau(\mu_1 + \mu_2)} \right) \\ f_{BPIG}(y_j; \mu_j, \tau; j = 1, 2) &= \left(\frac{2Z}{\pi} \right)^{\frac{1}{2}} \frac{\mu_1^{y_1} \mu_2^{y_2} e^{1/\tau} K_s(z)}{(z\tau)^{y_1 + y_2} y_1! y_2!}; y_1, y_2 \geq 0 \end{aligned} \quad (19)$$

dengan $s = y_1 + y_2 - \frac{1}{2}$ dan $z = \sqrt{\frac{1}{\tau^2} + \frac{2(\mu_1 + \mu_2)}{\tau}}$. Sehingga, $K_s(z) = K_{y_1 + y_2 - \frac{1}{2}} \left(\frac{1}{\tau} \sqrt{1 + 2\tau(\mu_1 + \mu_2)} \right)$ yang merupakan jenis ketiga dari modifikasi fungsi *Bessel*.

Parameter μ_j dihubungkan dengan variabel prediktor menggunakan fungsi penghubung $\ln(\cdot)$. Sehingga model regresi BPIG adalah sebagai berikut.

$$\begin{aligned} \ln(\mu_{ij}) &= \mathbf{x}_i^T \boldsymbol{\beta}_j \\ \mu_{ij} &= \exp(\mathbf{x}_i^T \boldsymbol{\beta}_j); i = 1, 2, \dots, n; j = 1, 2 \end{aligned} \quad (20)$$

dengan

$\mathbf{x}_i^T = [1 \ x_{i1} \ x_{i2} \ \dots \ x_{ik}]$ menunjukkan vektor variabel prediktor berukuran $1 \times (k + 1)$ pada pengamatan ke- i .

$\boldsymbol{\beta}_j = [\beta_{j0} \ \beta_{j1} \ \beta_{j2} \ \dots \ \beta_{jk}]^T$ menunjukkan vektor dari parameter regresi yang berukuran $(k + 1) \times 1$ pada variabel respon ke- j .

Oleh karena itu, variabel acak Y_1 dan Y_2 yang berdistribusi BPIG dapat dituliskan fungsi kepadatan peluang bersamanya sebagai berikut (Amalia et al., 2017).

$$\begin{aligned} f(y_j; \boldsymbol{\beta}_j, \tau; j = 1, 2) &= e^{1/\tau} K_s(z) \left(\frac{2}{\pi\tau} \right)^{\frac{1}{2}} \left(1 + 2\tau \sum_{j=1}^2 e^{\mathbf{x}_i^T \boldsymbol{\beta}_j} \right)^{-\frac{2\sum_{j=1}^2 y_j - 1}{4}} \times \\ &\quad \prod_{j=1}^2 \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}_j y_j}}{y_j!}; y_j \geq 0 \end{aligned} \quad (21)$$

1.5.7 Heterogenitas Spasial

Efek spasial berkaitan dengan data yang memiliki komponen lokasi atau efek lokasi. Efek spasial terbagi dua yaitu autokorelasi spasial dan heterogenitas spasial.

Autokorelasi spasial mengacu pada korelasi antara pengamatan pada lokasi tertentu dengan pengamatan di lokasi lain. Sedangkan heterogenitas spasial menunjukkan adanya variasi atau keragaman antar lokasi. Setiap lokasi memiliki struktur dan parameter hubungan yang berbeda. Heterogenitas spasial dapat diuji dengan menggunakan statistik *Breusch-Pagan* dengan hipotesis dan statistik uji sebagai berikut (Anselin, 1988).

$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2; i = 1, 2, \dots, n$ (tidak terdapat heterogenitas spasial).

$H_1: \text{minimal terdapat } \sigma_i^2 \neq \sigma_{i^*}^2$ (terdapat heterogenitas spasial).

$$BP = \frac{1}{2} \mathbf{f}' \mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' \mathbf{f} \sim \chi^2(k) \quad (22)$$

dengan $\mathbf{f} = [f_1 \ f_2 \ \dots \ f_n]^T$ dimana $f_i = (\sigma^{-1} \varepsilon_i)^2 - 1$ untuk $i = 1, 2, \dots, n$ dan \mathbf{Z} menunjukkan matriks \mathbf{X} terstandarisasi berukuran $n \times (k + 1)$. Kriteria pengambilan keputusan dalam uji *Breusch-Pagan* adalah menolak H_0 jika $BP > \chi_{(\alpha/2, k)}^2$ pada taraf signifikansi α .

1.5.8 Matriks Pembobot Spasial

Matriks pembobot spasial merupakan skema pembobotan berdasarkan kedekatan titik regresi i dengan titik-titik di sekitar i tanpa adanya hubungan eksplisit yang dinyatakan. Matriks pembobot spasial dapat ditunjukkan sebagai berikut (Fotheringham et al., 2002).

$$\mathbf{W} = \begin{bmatrix} 1 & w_{12} & \dots & w_{1n} \\ w_{21} & 1 & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & 1 \end{bmatrix}$$

Pembentukan fungsi pembobot berdasarkan jarak *euclidean* dapat menggunakan fungsi kernel (*kernel function*) sebagai berikut (Gollini et al., 2015).

$$\text{gaussian kernel} \quad : \quad w_{i^*i} = \exp \left[-\frac{1}{2} \left(\frac{d_{i^*i}}{b} \right)^2 \right] \quad (23)$$

$$\text{bisquare kernel} \quad : \quad w_{i^*i} = \begin{cases} \left[1 - \left(\frac{d_{i^*i}}{b} \right)^2 \right]^2 & , \ |d_{i^*i}| < b \\ 0 & , \ |d_{i^*i}| > b \end{cases} \quad (24)$$

$$\text{tricube kernel} \quad : \quad w_{i^*i} = \begin{cases} \left[1 - \left(\frac{d_{i^*i}}{b} \right)^3 \right]^3 & , \ |d_{i^*i}| < b \\ 0 & , \ |d_{i^*i}| > b \end{cases} \quad (25)$$

$$\text{exponential kernel} \quad : \quad w_{i^*i} = \exp \left(-\frac{|d_{i^*i}|}{b} \right) \quad (26)$$

$$\text{boxcar kernel} \quad : \quad w_{i^*i} = \begin{cases} 1 & , \ |d_{i^*i}| \leq b \\ 0 & , \ |d_{i^*i}| > b \end{cases} \quad (27)$$

dengan $d_{i^*i} = \sqrt{(u_{i^*} - u_i)^2 + (v_{i^*} - v_i)^2}$ merupakan jarak *euclidean* dan $b = \begin{cases} h & : \text{fixed kernel} \\ h_i & : \text{adaptive kernel} \end{cases}$ menunjukkan nilai *bandwidth* yang merupakan parameter

penghalus untuk mengetahui seberapa jauh pengaruh lokasi pengamatan saling berdekatan.

1.5.9 Bandwidth Selector

Pemilihan *bandwidth* pada matriks pembobot spasial sangat penting. Salah satu kriteria yang dapat digunakan dalam pemilihan *bandwidth* adalah dengan kriteria *least squares*. Namun dalam prosedurnya terdapat masalah yaitu ketika nilai *bandwidth* sangat besar yang menyebabkan bias parameter karena model yang dihasilkan terlalu mulus (*over-smoothing*) sehingga model lokal semakin mendekati model global. Sebaliknya, ketika nilai *bandwidth* sangat kecil akan menyebabkan model yang dihasilkan terlalu kasar (*under-smoothing*). Solusi untuk masalah ini adalah dengan pendekatan *cross-validation* (CV) untuk regresi lokal dan estimasi densitas kernel. Pendekatan statistik CV yang lebih mudah dalam perhitungannya dikenal sebagai *Generalized cross-validation* (GCV) dengan persamaannya sebagai berikut (Fotheringham et al., 2002).

$$GCV = \frac{n \sum_{i=1}^n [y_i - \hat{y}_i(b)]^2}{(n - v_1)^2} \quad (28)$$

dengan $v_1 = \text{tr}(\mathbf{S})$ dimana $\mathbf{S} = \mathbf{X}(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}$ adalah matriks duga (*hat matrix*).

1.5.10 Geographically Weighted Regression

Geographically Weighted Regression (GWR) adalah metode yang digunakan untuk memodelkan hubungan yang bervariasi secara spasial. Dibandingkan dengan regresi global, koefisien dalam GWR merupakan fungsi dari lokasi spasial (Lu et al., 2014). Model dari GWR adalah sebagai berikut (Fotheringham et al., 2002).

$$y_i = \beta_0(u_i, v_i) + \sum_{p=1}^k \beta_p(u_i, v_i) x_{ip} + \varepsilon_i \quad (29)$$

dengan

y_i menunjukkan variabel respon untuk lokasi pengamatan ke- i dimana $i = 1, 2, \dots, n$.

$\beta_0(u_i, v_i)$ menunjukkan intersep model GWR berdasarkan koordinat geografis lokasi pengamatan ke- i dengan u koordinat lintang dan v koordinat bujur.

$\beta_p(u_i, v_i)$ menunjukkan koefisien regresi variabel prediktor ke- p dimana $p = 1, 2, \dots, k$ berdasarkan koordinat geografis lokasi pengamatan ke- i .

x_{ip} menunjukkan variabel prediktor ke- p pada lokasi pengamatan ke- i .

$\varepsilon_i \sim iid N(0, \sigma_\varepsilon^2)$ menunjukkan residual pengamatan ke- i .

1.5.11 Geographically Weighted Bivariate Poisson Inverse Gaussian Regression

Model *Geographically Weighted Bivariate Poisson Inverse Gaussian regression* (GWBPIGR) merupakan perluasan dari regresi BPIG dimana estimasi parameternya mempertimbangkan aspek spasial dengan menggunakan matriks pembobot spasial (Arniva et al., 2018). Misalkan Y_1 dan Y_2 berdistribusi GWBPIG atau dinotasikan dengan $Y_j \sim \text{GWBPIG}(\mu_j(u_i, v_i), \tau); j = 1, 2$, maka fungsi kepadatan peluang bagi Y_j adalah sebagai berikut (Amalia et al., 2017).

$$f(y_j; \boldsymbol{\beta}_j(u_i, v_i); \tau; j = 1, 2) = e^{\frac{1}{\tau} K_s(z(u_i, v_i))} \left(\frac{2}{\pi \tau} \right)^{\frac{1}{2}} \left(1 + 2\tau \sum_{j=1}^2 e^{\mathbf{x}_i^T \boldsymbol{\beta}_j(u_i, v_i)} \right)^{-\frac{2 \sum_{j=1}^2 y_j - 1}{4}} \times \prod_{j=1}^2 \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}_j(u_i, v_i) y_j}}{y_j!}; y_j \geq 0 \quad (30)$$

dengan $s = y_1 + y_2 - \frac{1}{2}$, $z(u_i, v_i) = \frac{1}{\tau} \left(1 + 2\tau \sum_{j=1}^2 e^{\mathbf{x}_i^T \boldsymbol{\beta}_j(u_i, v_i)} \right)^{1/2}$. Sehingga, $K_s(z(u_i, v_i)) = K_{y_1 + y_2 - \frac{1}{2}} \left[\frac{1}{\tau} \left(1 + 2\tau \sum_{j=1}^2 e^{\mathbf{x}_i^T \boldsymbol{\beta}_j(u_i, v_i)} \right)^{1/2} \right]$ yang merupakan jenis ketiga dari modifikasi fungsi *Bessel*.

Pada model GWBPIGR, variabel respon y_j diprediksi oleh variabel prediktor dan bergantung pada lintang dan bujur (u_i, v_i) lokasi dan menghasilkan estimasi parameter untuk setiap lokasi. Sehingga, model GWBPIGR adalah sebagai berikut (Arniva et al., 2018).

$$\ln(\mu_{ij}) = \mathbf{x}_i^T \boldsymbol{\beta}_j(u_i, v_i) \\ \mu_{ij} = \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}_j(u_i, v_i)\right); i = 1, 2, \dots, n; j = 1, 2 \quad (31)$$

dengan

$\mathbf{x}_i^T = [1 \ x_{i1} \ x_{i2} \ \dots \ x_{ik}]$ menunjukkan vektor variabel prediktor berukuran $n \times (k + 1)$ untuk lokasi pengamatan ke- i .

$\boldsymbol{\beta}_j(u_i, v_i) = [\beta_{j0}(u_i, v_i) \ \beta_{j1}(u_i, v_i) \ \beta_{j2}(u_i, v_i) \ \dots \ \beta_{jk}(u_i, v_i)]^T$ menunjukkan vektor dari parameter regresi yang berukuran $(k + 1) \times 1$ pada variabel respon ke- j .

1.5.12 Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) merupakan metode yang digunakan untuk mengestimasi parameter suatu model yang diketahui fungsi probabilitasnya. Misalkan Y_1, Y_2, \dots, Y_n adalah variabel acak yang bersifat identik dan independen dengan fungsi kepadatan peluangnya $f(y; \boldsymbol{\theta})$ dengan $\boldsymbol{\theta} \in \Omega$ merupakan parameter yang tidak diketahui. Maka fungsi *likelihood* didefinisikan sebagai fungsi kepadatan peluang bersama dari Y_1, Y_2, \dots, Y_n yang dianggap sebagai fungsi dari $\boldsymbol{\theta}$ yaitu sebagai berikut (Hogg et al., 2012).

$$L(\boldsymbol{\theta}; \mathbf{y}) = f(y_1, y_2, \dots, y_n; \boldsymbol{\theta}) \\ = f(y_1; \boldsymbol{\theta}) \cdot f(y_2; \boldsymbol{\theta}) \cdot \dots \cdot f(y_n; \boldsymbol{\theta}) \\ = \prod_{i=1}^n f(y_i; \boldsymbol{\theta}); \boldsymbol{\theta} \in \Omega \quad (32)$$

Estimator maksimum *likelihood* $\hat{\theta}$ adalah nilai θ yang memaksimalkan fungsi *likelihood* $L(\theta; y)$. Namun, lebih mudah bekerja dengan logaritma natural dari fungsi *likelihood*-nya yang dapat dituliskan sebagai berikut.

$$\begin{aligned}\ell(\theta; y) &= \ln L(\theta; y) \\ &= \ln \left(\prod_{i=1}^n f(y_i; \theta) \right) \\ &= \sum_{i=1}^n \ln f(y_i; \theta); \theta \in \Omega\end{aligned}\quad (33)$$

Untuk memperoleh nilai $\hat{\theta}$ maka fungsi $\ell(\theta; y)$ diturunkan terhadap θ kemudian disamakan dengan nol.

$$\frac{\partial \ell(\theta; y)}{\partial \theta} = \mathbf{0} \quad (34)$$

1.5.13 Algoritma *Berndt-Hall-Hall-Hausman*

Sebagian besar metode iteratif menggunakan metode gradien yang mengubah $\hat{\theta}_m$ ke arah yang ditentukan oleh gradien. Persamaan yang digunakan dengan metode gradien adalah sebagai berikut.

$$\hat{\theta}_{m+1} = \hat{\theta}_m + \mathbf{A}_m \mathbf{g}_m; m = 1, 2, \dots, M \quad (35)$$

dengan

\mathbf{A}_m menunjukkan matriks berukuran $(k+1) \times (k+1)$ bergantung pada vektor $\hat{\theta}_m$.

$\mathbf{g}_m = \left. \frac{\partial \ell(\theta; y)}{\partial \theta} \right|_{\hat{\theta}_m}$ menunjukkan vektor gradien berukuran $(k+1) \times 1$ yang dievaluasi pada vektor $\hat{\theta}_m$.

Algoritma *Berndt-Hall-Hall-Hausman* (BHHH) merupakan pengembangan dari metode *Newton-Raphson* yang menggunakan Persamaan (35) dengan $\mathbf{A}_m = -\mathbf{H}_{BHHH,m}^{-1}$ dimana matriks *Hessian* didefinisikan sebagai berikut.

$$\mathbf{H}_{BHHH,m} = - \left. \sum_{p=1}^{k+1} \mathbf{g}_p(\hat{\theta}_m) \mathbf{g}_p(\hat{\theta}_m)^T \right|_{\hat{\theta}_m} \quad (36)$$

Dibandingkan dengan metode *Newton-Raphson*, algoritma BHHH hanya memerlukan evaluasi turunan pertama dalam komputasinya. Algoritma ini dapat diterapkan pada metode MLE maupun selain MLE dan dapat digunakan pada banyak *cross-section m-estimator* karena dapat bekerja dengan baik dan hanya membutuhkan turunan pertama (Cameron & Trivedi, 2005).

1.5.14 Pengujian Signifikansi Model *Geographically Weighted Bivariate Poisson Inverse Gaussian Regression*

Pengujian signifikansi secara simultan terhadap parameter model GWBPIGR menggunakan metode *Maximum Likelihood Ratio Test* (MLRT). Metode ini melibatkan $L(\hat{\Omega})$ yang merupakan fungsi maksimum *likelihood* untuk model yang

melibatkan variabel penjelas dan $L(\hat{\omega})$ yang merupakan fungsi maksimum *likelihood* untuk model tanpa melibatkan variabel penjelas. Hipotesis dan statistik uji MLRT adalah sebagai berikut (Amalia et al., 2017).

$$H_0: \beta_{j1}(u_i, v_i) = \beta_{j2}(u_i, v_i) = \dots = \beta_{jp}(u_i, v_i) = 0; j = 1, 2; p = 1, 2, \dots, k.$$

$$H_1: \text{minimal terdapat } \beta_{jp}(u_i, v_i) \neq 0.$$

$$G^2 = -2 \ln \left(\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right) \quad (37)$$

dengan $L(\hat{\Omega})$ dibentuk dari himpunan parameter di bawah populasi yaitu $\Omega = \{\beta_j(u_i, v_i); \tau; i = 1, 2, \dots, n; j = 1, 2\}$ dan $L(\hat{\omega})$ dibentuk dari himpunan parameter di bawah H_0 yaitu $\omega = \{\beta_{j0}(u_i, v_i); \tau; i = 1, 2, \dots, n; j = 1, 2\}$. Kriteria pengambilan keputusan uji MLRT adalah menolak H_0 jika $G^2 > \chi^2_{(\alpha, 2np)}$ pada taraf signifikansi α .

Pengujian signifikansi secara parsial terhadap parameter model GWBPIGR menggunakan uji *Wald* dengan hipotesis dan statistik uji sebagai berikut (Mardalena et al., 2021).

$$H_0: \beta_{jp}(u_i, v_i) = 0; j = 1, 2; p = 1, 2, \dots, k.$$

$$H_1: \beta_{jp}(u_i, v_i) \neq 0.$$

$$Z = \frac{\hat{\beta}_{jp}(u_i, v_i)}{se(\hat{\beta}_{jp}(u_i, v_i))} = \frac{\hat{\beta}_{jp}(u_i, v_i)}{\sqrt{Var(\hat{\beta}_{jp}(u_i, v_i))}} \quad (38)$$

dengan $Var(\hat{\beta}_{jp})$ diperoleh dari diagonal utama matriks varian-kovarian $Cov(\hat{\theta}) = -E(\mathbf{H}^{-1}(\hat{\theta})) = -\mathbf{H}^{-1}(\hat{\theta})$. Kriteria pengambilan keputusan dalam uji *Wald* adalah menolak H_0 jika $|Z_{hitung}| > Z_{(\alpha/2)}$ pada taraf signifikansi α .

1.5.15 Multikolinieritas

Multikolinieritas merupakan kondisi dimana terdapat korelasi antar variabel prediktor dalam hal ini antar variabel prediktor tidak saling bebas. Multikolinieritas dapat menyebabkan nilai variansi estimasi parameter *overestimate*, kesalahan tanda koefisien parameter, dan masalah pada selang kepercayaan estimasi parameter sehingga perlu dideteksi. Multikolinieritas dapat dideteksi menggunakan *Variance Inflation Factor* (VIF) dengan hipotesis dan statistik uji sebagai berikut (Ryan, 2009).

$$H_0: VIF_p < 10 \text{ (tidak terdapat gejala multikolinieritas pada variabel prediktor).}$$

$$H_1: VIF_p \geq 10 \text{ (terdapat gejala multikolinieritas pada variabel prediktor).}$$

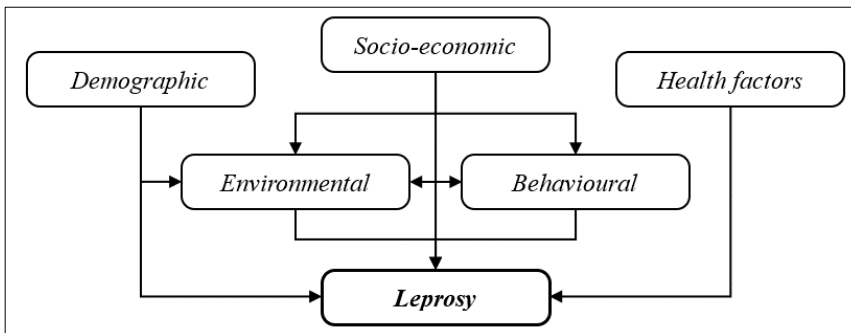
$$VIF_p = \frac{1}{1 - R_p^2}; p = 1, 2, \dots, k \quad (39)$$

dengan $R_p^2 = \frac{\sum_{i=1}^n (\hat{x}_{ip} - \bar{x}_p)^2}{\sum_{i=1}^n (x_{ip} - \bar{x}_p)^2}$. Kriteria pengambilan keputusan dalam uji multikolinieritas adalah menolak H_0 jika $VIF_p \geq 10$.

1.5.16 Kusta

Kusta (*Hansen's disease*) merupakan penyakit infeksi kronis yang disebabkan oleh sejenis penyakit bakteri yang disebut *Mycobacterium leprae* (*M. leprae*). Bakteri *M. leprae*, ordo taksonomi *Actinomycetales*, famili *Mycobacteriaceae* adalah basil intraseluler obligat gram-positif bersifat cepat-asam yang menunjukkan ketertarikan terhadap fagosit pada kulit dan sel *Schwann*. Penyakit kusta menyerang kulit, saraf tepi, mukosa saluran pernapasan bagian atas, dan mata (WHO, 2023). Penyakit kusta masih tersebar di lebih dari 120 negara, dengan jumlah kasus baru lebih dari 200,000 setiap tahunnya. Berdasarkan data dari *World Health Organization* (WHO) tahun 2019, Negara Brazil, India, dan Indonesia melaporkan lebih dari 10.000 kasus baru, sementara di 13 negara lainnya (Bangladesh, Republik Demokratik Kongo, Ethiopia, Madagaskar, Mozambik, Nepal, Nigeria, Filipina, Somalia, Sudan Selatan, Sri Lanka, dan Republik Demokratik Tanzania) masing-masing melaporkan 1,000-10,000 kasus baru. Sedangkan 99 negara melaporkan kurang dari 1,000 kasus baru, serta 45 negara lainnya melaporkan 0 kasus.

Kusta pertama kali diklasifikasikan oleh Rabello dan karakteristik polaritas penyakit ini telah ditetapkan. Pada tahun 1966, Ridley Jopling memperkenalkan metode klasifikasi kusta berdasarkan klinis pasien dan status kekebalan tubuh, antara lain *Tuberculoid leprosy* (TT), *Borderline Tuberculoid leprosy* (BT), *Mid-Borderline leprosy* (BB), *Borderline Lepromatous leprosy* (BL), dan *Lepromatous leprosy* (LL). Sementara WHO mengklasifikasikan kusta menjadi dua yaitu tipe *Paucibacillary* (PB) dan tipe *Multibacillary* (MB). Tipe PB menandakan kasus kusta dengan 1-5 lesi kulit tanpa menunjukkan adanya basil pada apusan kulit (*negative smear*), sedangkan tipe MB menandakan kasus kusta dengan lebih dari 5 lesi kulit dan menunjukkan adanya basil pada apusan kulit (*positive smear*) (Chen et al., 2022). Menurut Kerr-Pontes et al. (2006), risiko penyakit kusta dipengaruhi oleh beberapa aspek antara lain *demographic* (demografi), *socio-economic* (sosial-ekonomi), *environmental* (lingkungan), *health factors* (fasilitas dan pelayanan kesehatan), dan *behavioural* (perilaku). Berikut merupakan kerangka determinan faktor-faktor yang memengaruhi resiko kusta.



Gambar 1. Framework for risk factors in Leprosy determinant

BAB II METODOLOGI PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari laporan validasi kusta Kementerian Kesehatan Republik Indonesia (Kemenkes RI) dan publikasi Badan Pusat Statistik (BPS) Provinsi Sulawesi Utara, Sulawesi Barat, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Selatan, dan Gorontalo pada tahun 2021. Unit pengamatan data berupa kabupaten/kota di Pulau Sulawesi sebanyak 81 dan menggunakan koordinat geografis lintang dan bujur pada masing-masing lokasi pengamatan.

2.2 Identifikasi Variabel

Variabel yang digunakan dalam penelitian ini terdiri dari variabel respon dan variabel prediktor yang diuraikan pada Tabel 1 sebagai berikut.

Tabel 1. Definisi operasional variabel penelitian

Jenis Variabel	Nama Variabel	Definisi Operasional	Satuan
Respon	Kusta PB (Y_1)	Jumlah kasus kusta baru tipe PB tahun 2021.	Kasus
Respon	Kusta MB (Y_2)	Jumlah kasus kusta baru tipe MB tahun 2021.	Kasus
Prediktor	Kepadatan Penduduk (X_1)	Rasio banyaknya penduduk per kilometer persegi.	Jiwa/km ²
Prediktor	Persentase Penduduk Miskin (X_2)	Jumlah penduduk yang memiliki rata-rata pengeluaran per kapita per bulan di bawah garis kemiskinan dibagi dengan jumlah penduduk dikali 100%.	Persen (%)
Prediktor	Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (X_3)	Jumlah rumah tangga yang dapat mengakses fasilitas sanitasi layak (fasilitas buang air besar sendiri/bersama, jenis kloset leher angsa, tempat pembuangan berupa septik tank/SPAL) dibagi dengan jumlah rumah tangga keseluruhan dikali 100%.	Persen (%)
Prediktor	Persentase Rumah Tangga yang Memiliki Akses Terhadap	Jumlah rumah tangga yang memiliki akses terhadap air minum layak (air minum dengan jarak ke tempat pembuangan limbah mini-	Persen (%)

Lanjutan Tabel 1

Jenis Variabel	Nama Variabel	Definisi Operasional	Satuan
	Layanan Sumber Air Minum Layak (X_4)	mal 10 meter yang bersumber dari leding, sumur bor/pompa, sumur terlindung, mata air terlindung, dan termasuk air hujan) dibagi dengan jumlah rumah tangga keseluruhan dikali 100%.	
Prediktor	Jumlah Fasilitas Kesehatan (X_5)	Jumlah unit fasilitas kesehatan berupa rumah sakit umum, rumah sakit khusus, puskesmas rawat inap, puskesmas non rawat inap, dan klinik pratama	Unit
Prediktor	Jumlah Tenaga Medis (X_6)	Jumlah dokter termasuk dokter spesialis (selain spesialis gigi) dan dokter umum.	Individu

2.3 Metode Analisis

2.3.1 Estimasi Parameter Model *Geographically Weighted Bivariate Poisson Inverse Gaussian Regression*

1. Menentukan fungsi *likelihood* untuk lokasi ke- i berdasarkan fungsi kepadatan peluang $Y_j \sim \text{GWBPIG}(\mu_j(u_i, v_i), \tau); j = 1, 2$.

$$L(\boldsymbol{\beta}_j(u_i, v_i); \tau; i = 1, 2, \dots, n; j = 1, 2) = \prod_{i=1}^n f(y_j; \boldsymbol{\beta}_j(u_i, v_i); \tau; j = 1, 2) \quad (40)$$

2. Menentukan fungsi \ln *likelihood* untuk lokasi ke- i .

$$\ell = \ln L(\boldsymbol{\beta}_j(u_i, v_i); \tau; i = 1, 2, \dots, n; j = 1, 2) \quad (41)$$

3. Menentukan fungsi \ln *likelihood* dengan pembobot geografis w_{i^*i} untuk setiap lokasi dimana $i^* \neq i$.

$$\ell^* = \sum_{i=1}^n \ln f(y_j; \boldsymbol{\beta}_j(u_i, v_i); \tau; j = 1, 2) w_{i^*i} \quad (42)$$

4. Mencari turunan pertama dari fungsi ℓ^* terhadap parameter $\boldsymbol{\beta}_1(u_i, v_i)$, $\boldsymbol{\beta}_2(u_i, v_i)$, dan τ kemudian disamakan dengan nol.
5. Jika hasil turunan pertama terhadap masing-masing parameter menghasilkan bentuk implisit maka estimasi parameter diselesaikan menggunakan algoritma BHHH.

2.3.2 Penerapan Model *Geographically Weighted Bivariate Poisson Inverse Gaussian Regression*

1. Menguji distribusi *Bivariate Poisson* pada variabel respon Y_1 dan Y_2 menggunakan *index of dispersion test* pada Persamaan (6).
2. Menguji korelasi antara variabel respon Y_1 dan Y_2 menggunakan uji korelasi *Pearson* pada Persamaan (7).
3. Mendeteksi overdispersi pada variabel respon Y_1 dan Y_2 menggunakan nilai *Deviance* pada Persamaan (9).
4. Menguji asumsi non-multikolinieritas pada variabel prediktor menggunakan nilai *Variance Inflation Factor* (VIF) pada Persamaan (39).
5. Memodelkan data dengan regresi BPIG.
6. Menguji heterogenitas spasial dengan uji *Breusch-Pagan* pada Persamaan (22).
7. Menghitung jarak *euclidean* antar lokasi pengamatan berdasarkan letak geografis lintang dan bujur.
8. Membentuk matriks pembobot spasial dengan fungsi kernel dan menentukan *bandwidth optimum* menggunakan nilai GCV pada Persamaan (28).
9. Mengestimasi parameter model GWBPIGR menggunakan metode MLE dengan algoritma BHHH.
10. Melakukan uji signifikansi secara simultan terhadap parameter model GWBPIGR menggunakan MLRT pada Persamaan (37).
11. Melakukan uji signifikansi secara parsial terhadap parameter $\beta_j(u_i, v_i)$ menggunakan uji *Wald* pada Persamaan (38).
12. Menginterpretasikan model yang diperoleh dan mengelompokkan variabel yang signifikan.